Acquisition of grounded models of adjectival modifiers supporting semantic composition and transfer to a physical interactive robot

N. Mavridis
Informatics and Telematics Institute
NCSR Demokritos
Athens, Greece
nmav@alum.mit.edu

S.B. Kundig
Informatics and Telematics Institute
NCSR Demokritos
Athens, Greece
stefkundig@yahoo.gr

N. Kapellas
Informatics and Telematics Institute
NCSR Demokritos
Athens, Greece
kapelnick.mud@gmail.com

***Abstract*—Compositionality is a property of natural language which is of prime importance: It enables humans to form and conceptualize potentially novel and complex ideas, by combining words. On the other hand, the symbol grounding problem examines the way meaning is anchored to entities external to language, such as sensory percepts and sensory-motor routines. In this paper we aim towards the exploration of the intersection of compositionality and symbol grounding. We thus propose a methodology for constructing empirically derived models of grounded meaning, which afford composition of grounded semantics. We illustrate our methodology for the case of adjectival modifiers. Grounded models of adjectively modified and unmodified colors are acquired through a specially designed procedure with 134 participants, and then computational models of the modifiers "dark" and "light" are derived. The generalization ability of these learnt models is quantitatively evaluated, and their usage is demonstrated in a real-world physical humanoid robot. We regard this as an important step towards extending empirical approaches for symbol grounding so that they can accommodate compositionality: a necessary step towards the deep understanding of natural language for situated embodied agents, such as sensor-enabled ambient intelligence and interactive robots.**

***Keywords*—compositionality; symbol grounding; adjectival modifiers; interactive robots**

I. INTRODUCTION

Two of the three general questions motivating our research in this paper are the following: First, "How do words connect to entities which are outside language, such as sensory percepts, actions and goals?" and second, "How can we empirically (i.e. by experiment and / or observation) learn such grounded models?", i.e. without prescribing our own pre-conceptions regarding such models.

If one attempts to build situated embodied intelligent agents, such as robots, which perceive the world through their sensors, and communicate with humans using natural language about the shared situation that they are in, as is the case for example in [1] and [2], then theoretical and computational answers to the above two questions become of primary importance. Following these two fundamental questions, the third question that is central to our research is "How can we learn computational models of grounded semantic composition empirically?". Of course, all of the three above questions cannot be given trivial answers; indeed, multi-year research agendas can be created towards the examination and derivation of adequate answers and appropriate models.

Thus, in this paper, we will concentrate on a more specific endeavor: we will provide a methodology for deriving models of grounded meaning of modifiers that are computationally composable, and we will illustrate their utilization for the case of adjectival modifiers. The operational example that will be utilized is the following: Models of single-word color descriptors, such as "red" and "blue", will be derived from human subjects through an appropriately designed experimental procedure. Then, we will empirically learn models of compound two-word color descriptors, such as "dark red" or "light red". We will then provide a computational procedure, through which by observing in what way the grounded model of "red" is related to the grounded model of "dark red", we will abstract a model for the modifier "dark". Most importantly, this abstracted computational model of the modifier "dark", will be composable with other colors, beyond "red". Thus, our system, having been given examples of "red", "blue" and "dark red", will be able to predict the model of "dark blue", without ever having seen an instance of this.

We will then quantitatively evaluate how much the derived model of "dark blue", as created by the composition of our model of "dark" and the given model of "blue", indeed matches the human model. We will then validate that the deviation from the human model is of a magnitude that falls within the human inter-annotator agreement; and thus, the learnt model is certainly adequate for communication. Our composable models of such modifiers will be proven to be applicable to a range of various colors, thus proving their generalization abilities, and most importantly fall well within the human inter-annotator agreement bands, thus proving to be at least as good as human models for communication purposes.

Then, we will also ask another important question: how much would we gain, if we learnt the model of "dark" not only from a single color (i.e. from "dark red") but from larger training sets consisting of more than one colors? For

978-1-4673-7509-2/15/$31.00 © 2015 IEEE

example, one could train the model from "dark red" and "dark green", and then test it on "dark blue" and "dark yellow". Finally, and most importantly, we will illustrate that the learnt models transfer to the physical world, for the case of a humanoid robot equipped with a vision system and a conversational interface.

II. BACKGROUND

Although arguably aspects of the symbol grounding problem are first touched upon in earlier philosophical texts, the term is usually attributed to [3]. Since then, a sizeable corpus of research has examined multiple aspects of the symbol grounding problem. For example, one such aspect is concerned with the grounding of spatial relations; work on such models include the classic [4], and extend all the way to more recent unsupervised approaches such as [5]. Real-world computational approaches with applications to robotics include [1] as well more recent projects such as the one by Tellex [2]. Regarding models of the evolution of grounded meaning, a classic approach is that of semiotic dynamics [6]. Furthermore, noteworthy large-scale projects exist towards the acquisition of grounded lexica [7]. Last but not least, companion robots such as [8] would greatly benefit from such grounded models.

However, grounding lexical semantics is just the beginning of any large-scale approach towards grounded understanding of language. Any such approach should also cover the very important aspect of semantic composition, enabling the derivation of grounded semantics of larger expressions through the computational composition of the models of their constituents. As argued in [9], no such applicable computational theory of grounded semantic composition exists yet. This state of affairs provided strong motivation for us, and thus the work presented here forms one of our first steps towards a more general theory.

Extending the discussion beyond symbol grounding, one could ask: what are the forces that shape perceptual and linguistic categories, and how knowledge of these forces could come to our aid when learning such categories computationally? It is noteworthy that neither the linguistic relativity that the Sapir-Whorf [10][11] hypothesis proposes, nor its counterpart theories supporting perceptual determinism [12], can alone sufficiently explain this issue, and the empirical data support that in practice both forces are at play. Furthermore, the second of the above questions remains yet not sufficiently explored.

Let us now return to our main task, and describe the methodology that was utilized.

III. METHODS

The general methodology followed in this project can be summarized in 5 steps:

Step 1: Acquire the meaning of concepts <*A*>, <*B*> in an empirical learning process where the training set is extracted from human-derived data. (e.g. <*A*>=''red'', <*B*>=''blue")

Step 2: Acquire the meaning for {<*modifier* + *A*>} through an empirical process as in Step 1. (e.g. extract the grounded model of <*modifier* + *A*>=''dark red" from human-derived data)

Step 3: Given the above, deduct an approximate model of <*modifier*> (e.g. a composable computational model of dark" that can be composed not only with "red" but also with "blue", "yellow" etc., i.e. a model that can demonstrate generalization)

Step 4: Apply model of <*modifier*> as an operator of a compositionality function, which will produce higher level perceptual concepts (e.g. compose the learnt model of "dark" with several colors other than "red").

Step 5: Evaluate the generalization power of the learnt modifier, by comparing it to human models (e.g. see how well the computational "dark"+''blue" model that arose of the composition fits the human "dark blue" model as it is approximated by empirical data gathered by humans)

Formally, the composition operator * accomplishes:

$$\text{Meaning}(modifier + A) = \text{Meaning}(modifier) * \text{Meaning}(A)$$

The meaning space that was chosen in this study in order to demonstrate our methods was that of colors, and the modifiers whose composable grounded models were empirically learnt were the basic color modifiers <*light*> and <*dark*>. The basic workflow followed, as prescribed from the above general steps, is summarized in fig.1. The procedure can be generalized for other meaning spaces that can be modeled as an n-dimensional continuous feature space.

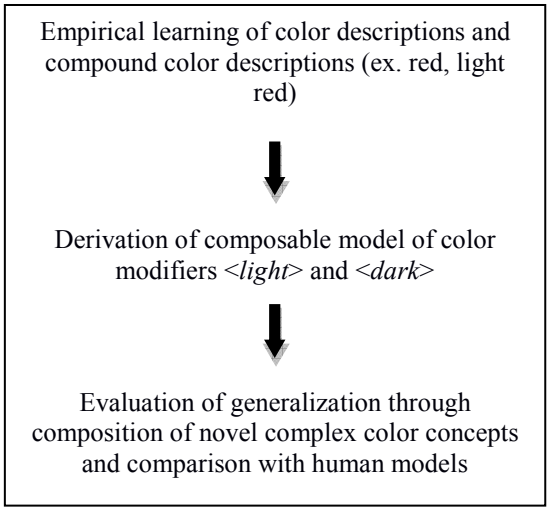

Fig.1. Methodology workflow

A. Survey Design

A total of 44 colors from the Hue-Saturation-Lightness (HSL) space were selected for our survey, which aimed towards acquiring the grounded meaning models from human subjects. The 44 points corresponded to four fixed hue and saturation values, the ones defined as Red, Green, Blue and Yellow, and 11 gradations of lightness equally spaced from 20% to 80%. Five surveys were distributed to a total of 134 non color-blind people, 73 male and 61 female, with an

average age of 28 years, and a range of ages of 21-40 years. Each survey consisted of four discrete parts:

1. Open description
2. Restricted description
3. Referent resolution
4. Exploration of similarity space

All questions were randomized for each respondent in order to minimize bias.

1. Open description

In the first part, one of the 44 colors was displayed in each question, and the subject was asked to give an unrestricted verbal description. The main objective here was to explore the richness of human descriptive space of words regarding colors, the ultimate goal being to delimit an appropriate subset that still permits sufficient interaction. Each questionnaire included 15 open description questions, in order to comprise all of the colors in a balanced way, as shown in table I.

Colors:	**Red**	**Green**	**Blue**	**Yellow**
Survey 1	4	4	4	3
Survey 2	4	4	3	4
Survey 3	4	4	3	4

TABLE I. Questions per survey distribution in part 1 and 2.

2. Restricted description

In this part, subjects were asked to evaluate their agreement of a given description for a color, in a scale from 1 to 7. Three descriptions were provided: 1) *Dark color* (for example, "dark red") 2) *Light color* and 3) *color*. The same population method was applied into part 2 as in part 1, with a total of 15 questions in each questionnaire.

3. Resolution of referring expressions to color

The questions in this section contained figures containing 4 colors corresponding to successive points of the total 11 for each color. The respondent was given a word description and was asked to pick the colors that fitted the expression in his opinion. Four shading clusters were created for each color, which were equally distributed in the questionnaires. The design pattern is illustrated in table II.

Shading scale	**Dark 1-4**	**Medium 4-7**	**Medium 5-8**	**Light 8-11**
Red	1	3	1	2
Green	2	2	3	3
Blue	1	2	3	2
Yellow	3	3	2	1

TABLE II. Distribution of questions in part 3.

4. Exploration of the similarity space of colors

Part 4 demanded the respondents to evaluate how similar two demonstrated colors appeared in a scale from 1 to 7. The exploration of the color similarity space was separated into two sub-parts, one were pairs of the same color were demonstrated (Part A), and one for color-points of different colors (Part B).

Due to economy considerations, not all of the possible (44×43)/2 combinations were covered in the questionnaires. The missing data were completed through interpolation and extrapolation, using linear methods.

B. Data Analysis

All data responses were gathered and then imported and processed on Matlab V7.12 from Mathworks. An average ranking (in the 1 to 7 likert scale), corresponding to each color point being described as light, dark or unmodified was obtained from restricted description responses and a curve fitting the 11 points for each description was calculated using a 4thorder polynomial approximation. Thus we obtained grounded models for each color, both in modified and in unmodified forms, as demonstrated in fig. 2. In total, twelve such models were derived: four colors (red, green, blue, yellow) x three versions (unmodified, light, dark).

The next step was to infer the mathematical transformation G representing the modifier that takes us from <*color*> to <*modified color*> (i.e. the composable computational model of the modifiers). G was assumed to be an affine transformation with the general form:

$$\begin{bmatrix}\vec{y}\\1\end{bmatrix} = \begin{bmatrix}A & \vec{b}\\0,\dots,0 & 1\end{bmatrix}\begin{bmatrix}\vec{x}\\1\end{bmatrix} \quad (1)$$

Affine transformations can represent not only translations, rotations, reflections, and scalings, but also shear transformations. However, an important point to note is that such a transformation cannot be applied directly to our color models, because their domain is restricted, to [1…11]. Thus, to evaluate the derived models, the domain of the curves was first mapped to infinity with a specially designed transformation T:

$$T(x) = \begin{cases} 1/\ln\left(1/\frac{x-6}{5}\right), & x > 6, \\ 0, & x = 6 \\ -1/\ln\left(1/-\frac{x-6}{5}\right), & x < 6 \end{cases} \quad (2)$$

which performs the mapping $[1\dots11] \rightarrow [-\infty, +\infty]$.

After applying T, the affine transform G was applied, and finally the curves were mapped back with T^{-1}, in order to be compared with the human acquired model. Consequently, the testing process can be summarized in the expression:

How close is $(T^{-1} * G_{Light} * T)(blue)$ to $(lightblue)$?

I.e. how well does the result of our derived model following composition match the human model? In section IV, quantitative as well as qualitative answers to this most important question will be presented, for a variety of settings.

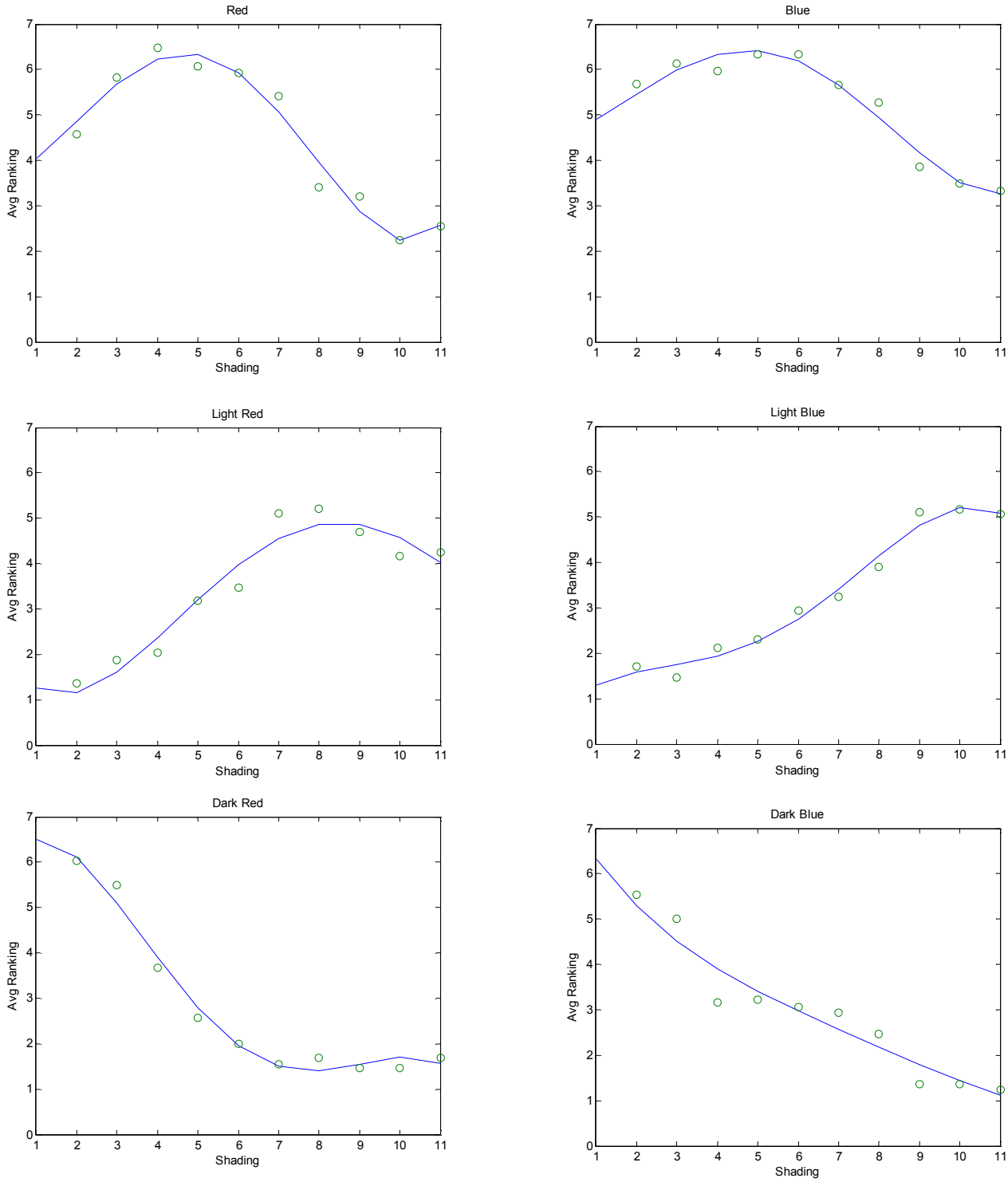

Fig. 2. The empirically acquired grounded models of “Red”, “Light Red”, “Dark Red” (left column, top to bottom) and “Blue”, “Light Blue”, “Dark Blue” (right column, top to bottom).

C. *Robotic embodiment*

To illustrate the applicability of the acquired models in the real world, an appropriate robotic embodiment was implemented. The highly-realistic equipped humanoid robot "IbnSina" was utilized [13][14] in an interactive testing process of color recognition (fig. 3). Seven colored paper folders were used in the experiment, four of which are shown in figure 4. A see-through-eye pinhole camera with resolution 640 x 480 pixels provided a feed to our matlab code using VFM framegrabber, and the robotic voice was created using voicebox.m. The possible answers, whose set was determined as described above, were blended with lip syncing and appropriate facial movements using Visual Show Automation by Brookshire software. After calibration, the camera-derived HSL color value to be evaluated was calculated using the median of the central 1/9 of the area of the picture, which was occupied with the paper folders used in the experiment. Therefore, a paper folder is placed in front of the robot and it is asked "what color is this?", then the robot answers by selecting the color which is closest in hue to the median of the central region, and selecting the one of the three modifiers {"light", "dark", (unmodified)} that has the highest average ranking in its learnt model for the shading (lightness) of the median of the central region. Then, matlab calls the VSA software through a matlabdos() call invoking VSA with the appropriate animation-audio file for the required answer as an argument. Thus, appropriate spoken descriptions such as "It is light green" are generated upon placement of the color paper folder in front of the robot and verbal inquiry from the human. The above procedure with the humanoid was repeated for multiple training and testing sets combinations.

In order to further validate the appropriateness of the real-world descriptions generated by the robot, which was trained by using the color and modifier models that were derived through the responses of the human subjects, we performed yet one more human survey with the actual physical paper folders that were shown to the robot. A total of 10 people, 6 male and 4 female, completed the survey with an average age of 34.2 years and a range of ages of 25-40 years.

IV. RESULTS

Results from open description questions indicated a preference for modifiers like 'light', 'dark' 'bright', 'deep' etc. as expected, although the remaining answers were highly scattered as shown in the example of fig. 5, which depicts the distribution of verbal descriptions attributed to the brick-like dark red color on the left.

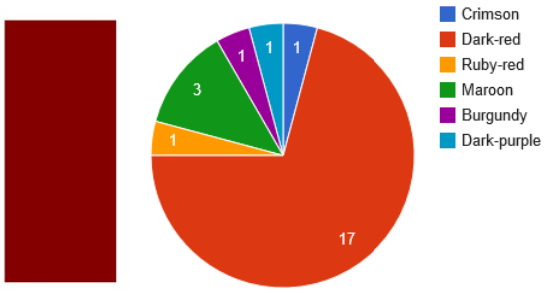

Fig 5. Left: Color to be described, Right: Open description responses.

Fig.3. Real-world transfer testing using the IbnSina Humanoid Robot

Fig. 4. Four of the color folders used in our robot transfer experiment.

To evaluate the derived models presented back in figure 2, the unmodified colors' curves underwent the inferred G transform and the output was compared with the corresponding human acquired model. In order to explore adequacy of training set and power of generalization, we decided to investigate all four colors as potential training sets, and see how well the learned modifiers generalize to all the remaining three colors of each combination. Thus, initially training occurred each time on a single color, and testing was performed on the remaining three. The average Root Mean Square Error (RMSE) was calculated in each case. Later, as we shall see, we extended also to larger training sets.

The results are presented in figure 6 and table III. In figure 6, you can see as an example the learnt model for "light" as trained from "red" and as applied to "blue", and the learnt model for "dark" trained from "blue" and applied to "red". The blue line is the actual human model for "light red" (top figure) or "dark red" (bottom figure), while the green line is the estimated model arising from the composition of the computationally learnt modifiers ("light" and "dark")

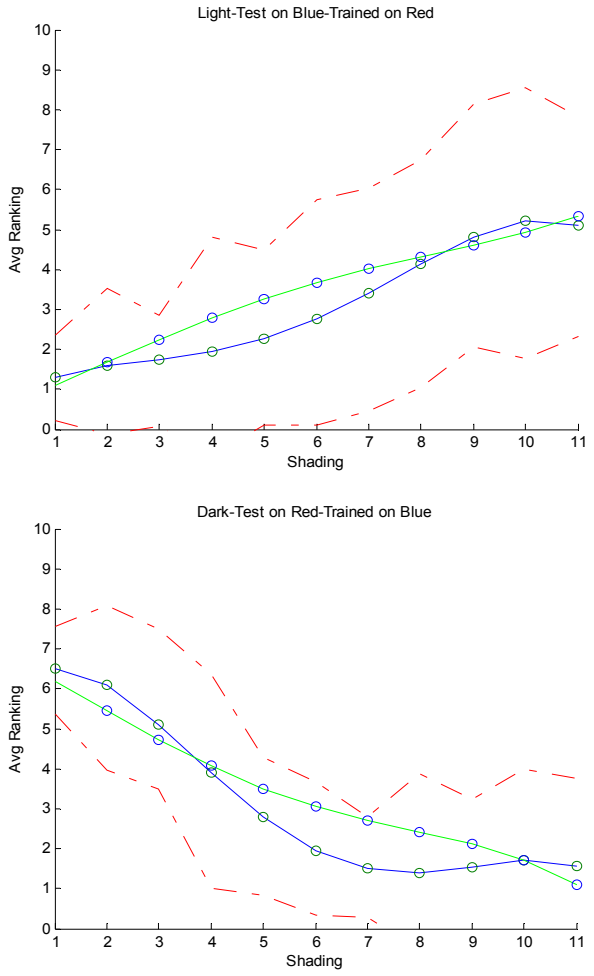

Fig. 6. Two of the learnt models, as derived from our computational model. Top: Learn composable model of modifier "light" from human model of "light red" and "red", and apply it to generalize to "light blue". Bottom: Similarly, for "dark". Green line: derived learnt model. Blue line: actual human model. Red lines: 1.5 sigma bands (μ-1.5σ,μ+1.5σ) where μ is mean value and σ is standard deviation.

with the human model of "red". The root mean square error for the top figure is 0.5904, and for the bottom figure is 0.7325, which given the 0…7 rating scale, corresponds to 8.4% and 10.4% error. The important point to notice though is that this is not only small, but is actually a very strong result: if one observed the two red lines in figure 6, which correspond to μ + 1.5σ and μ-1.5σ (where μ=mean and σ=standard deviation), then it is clear that the learnt curves fall completely within these lines. Now, taking into account that there is considerable variation across human annotators, which is reflected in the σ(standard deviation) values, the fact that the learnt curves (green lines) fall completely well within the two red curves means that they are well within the Φ(1.5) – Φ(-1.5) = 0.86 most typical sample of the human annotators. Thus, the learnt models are well within the most typical 86% of all human annotators, and thus are at least as good as most humans would be, given the level of inter-annotator agreement that exists (i.e. the natural variation that exists in human rankings). Moving beyond the specific cases of learning "light" (train from red, test to blue) and learning "dark" (train from red, test to blue), to all the possible combinations, we derived the results tabulated in table III.

Trained on	**Tested on**			
	Modifier <Light>			
	Red	Blue	Yellow	Green
Red	-	0.5904	0.7192	1.4553
Blue	0.8950	-	1.1332	1.7406
Yellow	0.9667	0.7893	-	1.2918
Green	1.5302	1.6214	1.1435	-
Trained on	**Modifier <Dark>**			
	Red	Blue	Yellow	Green
Red	-	0.7413	1.913	1.1341
Blue	0.7325	-	1.4216	1.3202
Yellow	1.2710	1.1268	-	0.7758
Green	1.4846	1.5673	2.8874	-

TABLE III. Mean Square Error for various training-testing combinations across the four colors, for the case of the "light" modifier (top) and for the "dark" modifier (bottom). All training sets consist of a single color.

The overall average RMSE was 1.1562 for "light", and 1.3661 for "dark". Regarding participation in the [μ-2σ, μ+2σ] variance bands, we had 95.5% points within the variance bands, which means that the modifier models we have learnt with our method are well within acceptable human variation.

After exploring all combinations of train-test pairs with training sets comprised of a single modified color, we began to train the affine transformation G on models of multiple modified colors (with the same modifier), and as always testing took place on the remaining colors. The results can be seen in detail in table IV. As the training set becomes larger, we notice a slight decrease of the average RMSE (fig.7).

Finally, detailed results regarding the robotic embodiment demonstration are presented in table V. We see that 94% of the participant's answers agreed with the folders' color description derived by our method. Furthermore, and quite importantly, the most frequent human description for each folder agreed with the description produced by the robot. Thus, the transfer of the models obtained through the survey to the physical real-world robot produced results that were well within the human inter-annotator agreement, and furthermore represented the most frequent human descriptions of the folders used in our experiments. Thus, not only we were able to derive empirically composable grounded models for adjectival modifiers of colors, but also to demonstrate their utility when transferred to a physical robot in the real world.

Trained on	Tested on			
	Modifier <Light>			
	Red	Blue	Yellow	Green
Red, Blue	-	-	0.8445	1.6532
Red, Green	-	1.2396	0.8289	-
Red, Yellow	-	0.6724	-	1.3036
Blue, Green	1.1690	-	1.0805	-
Blue, Yellow	0.8066	-	-	1.5012
Yellow, Green	1.0659	1.2470	-	-
Trained on	**Modifier <Dark>**			
	Red	Blue	Yellow	Green
Red, Blue, Yellow	-	-	-	1.4809
Red, Blue, Green	-	-	0.5821	-
Red, Yellow, Green	-	1.0411	-	-
Blue, Yellow, Green	0.8733	-	-	-

TABLE IV. Mean Square Error for various training-testing combinations across the four colors, for the case of the "light" modifier (top) and for the "dark" modifier (bottom). Training sets consist of multiple colors.

Age	Sex	B folder	L-B folder	D-B folder	G folder	L-G folder	D-G folder	L-Y folder
24	M	Blue	Light Blue	Dark Blue	Green	Light Green	Dark Green	Yellow
25	M	Blue	Light Blue	Dark Blue	Green	Light Green	Dark Green	Light Yellow
30	F	Blue	Light Blue	Dark Blue	Dark Green	Light Green	Dark Green	Light Yellow
23	F	Blue	Light Blue	Dark Blue	Green	Light Green	Dark Green	Light Yellow
58	F	Blue	Light Blue	Dark Blue	Green	Light Green	Dark Green	Yellow
59	M	Blue	Light Blue	Dark Blue	Green	Light Green	Dark Green	Yellow
27	M	Blue	Light Blue	Dark Blue	Green	Light Green	Dark Green	Light Yellow
24	M	Blue	Light Blue	Dark Blue	Green	Light Green	Dark Green	Light Yellow
33	F	Blue	Light Blue	Dark Blue	Green	Light Green	Dark Green	Light Yellow
29	M	Blue	Light Blue	Dark Blue	Green	Light Green	Dark Green	Light Yellow

TABLE V. Human descriptions of the color of the folders used in the robotic illustration.

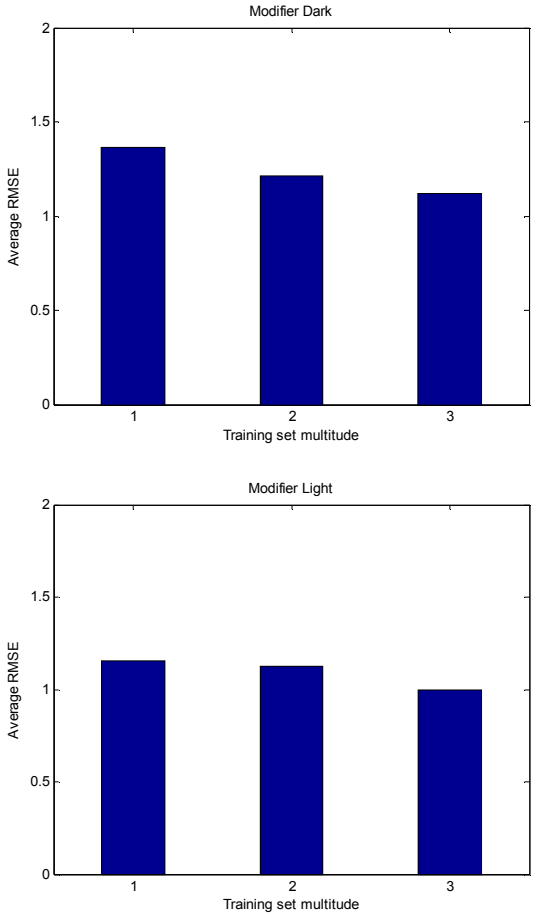

Fig. 7. Average RMSE of the computed modifiers for different training set multitude. Top: Modifier "dark". Bottom: Modifier "light".

V. DISCUSSION

Given the ever-increasing applications of situated intelligent agents, including physically embodied agents such as robots, it is important to equip them with computational models enabling situated language understanding, and thus addressing the symbol grounding problem. A prerequisite for the wider application of symbol grounding approaches is creating learnable grounded models that go beyond isolated lexical semantics and afford compositionality, as we have done for the case of adjectival color modifiers in this paper.

Thus, first of all, as a future work in continuation to this project we would like to illustrate the extension of our model to other not only adjectival but also adverbial modifiers and moreover to complex chains of modifiers as well, consisting of multi-word descriptions. Such adjectival and adverbial modifiers can furthermore extend to various sensory modalities, including audition and taste. It is also worth noting, that the models provided here are not only useful towards building intelligent agents, but could also be used as computational models representing potential behavioral-level theories of human performance, which could furthermore be augmented with neuroscientific considerations. Mathematical category theory [15] and also conceptual spaces [16] play an important role in this wider theory.

Second, we are currently devising a wider theory of grounded semantic composition, which can deal with a richer set of heterogeneous types and aims towards extending to utterance-level constructions, and can also handle appropriate ordering, thus discriminating "dark red" from "red dark". Last but not least, we envision that such grounded models which afford compositionality can help create more fluid and effective human-robot interaction leading to widespread and beneficial introduction of robots and other such intelligent entities in our everyday life.

VI. CONCLUSION

Given the important property of compositionality of natural language, humans are able to form and conceptualize potentially novel and complex ideas, by combining words. On the other hand, the symbol grounding problem examines the way meaning is anchored to entities external to language, such as sensory percepts and sensory-motor routines. In this paper we explored of the intersection of compositionality and symbol grounding, with our ultimate purpose of devising learnable models of grounded meaning that are composable, and demonstrating their utility in real-world interactive robots. We thus proposed a methodology for constructing empirically derived models of grounded meaning, which afford composition of grounded semantics. We illustrated our methodology for the case of adjectival modifiers. Grounded models of adjectively modified and unmodified colors were acquired through a specially designed procedure with 134 participants, and then computational models of the modifiers "dark" and "light" were derived. The generalization ability of these learnt models was quantitatively evaluated, and their usage was demonstrated in a real-world physical humanoid robot.

The results were highly positive: the quantitative deviation between the models produced by learnt composable modifiers when they were generalizing to new colors was very small, and furthermore was decreasing with training set size. Even for the case of a single-color training set, qualitatively we were well within human inter-annotator agreement, and thus we were effectively as good as the average human. Furthermore, and quite importantly, we demonstrated how the models can be transferred to a physical robot and be used in order to demonstrate their applicability to the real world. Again, we cross-verified with the descriptions of human subjects, and our results always agreed with the most frequent human descriptions.

We regard this as an important step towards extending empirical approaches for symbol grounding so that they can accommodate compositionality: a necessary step towards the deep understanding of natural language for situated embodied agents, such as sensor-enabled ambient intelligence and interactive robots. Thus, future intelligent machines will enjoy the benefits of the compositionality of natural language, which have enabled humans to imagine and plan about, and bring forth complex situations that were never seen before, through the power of composing and communicating imaginary worlds by putting together words.

REFERENCES

[1] Mavridis, Nikolaos, "Grounded situation models for situated conversational assistants", PhD Thesis, Massachusetts Institute of Technology, 2007

[2] Tellex, Stefanie, et al. "Approaching the symbol grounding problem with probabilistic graphical models." AI magazine 32.4 (2011): 64-76.

[3] Harnad, S. (1990) The Symbol Grounding Problem. Physica D 42: 335-346.

[4] T. Regier, L. A. Carlson, "Grounding Spatial Language in Perception: An Empirical and Computational Investigation," in Journal of Experimental Psychology: General 2001, vol. 130, No. 2,273-298, 2001.

[5] Vavrecka, Michal, and Igor Farkaš. "Unsupervised grounding of spatial relations." Proceedings of European Conference on Cognitive Science, Sofia, Bulgaria. 2011.

[6] Steels, Luc. "Semiotic dynamics for embodied agents." Intelligent Systems, IEEE 21.3 (2006): 32-38.

[7] Pastra, Katerina, et al. "The POETICON Corpus: Capturing Language Use and Sensorimotor Experience in Everyday Interaction." LREC. 2010.

[8] N. Mavridis, W. Kazmi at al, "On the synergies between online social networking, face recognition and interactive robotics", in proceedings of the IEEE International Conference of Computational Aspects of Social networks (CASON), 2009

[9] M. Daoutis and N. Mavridis, "Towards a model for Grounding Semantic Composition," in AISB - Artificial Intelligence and the Simulation of Behavior, Goldsmiths, University of London, United Kingdom, April 2014.

[10] P. Kay, W. Kempton, "What Is the Sapir-Whorf Hypothesis?"

[11] P. Kay and C. K. McDaniel, "The linguistic significance of the meaning of basic color terms," in Language, Vol. 54, No. 3 (Sep., 1978), pp. 610-646.

[12] L. Steels, T. Belpaeme, "Coordinating perceptually grounded categories through language: A case study for color," in Behavioral and Brain sciences, 28, pp.469-529, 2005.

[13] Riek, L., et al. "IbnSina steps out: exploring Arabic attitudes toward humanoid robots." Proceedings of the 2nd international symposium on new frontiers in human–robot interaction, AISB, Leicester. Vol. 1. 2010.

[14] Mavridis, Nikolaos, et al. "Transforming IbnSina into an advanced multilingual interactive android robot." GCC Conference and Exhibition (GCC), 2011 IEEE. IEEE, 2011.

[15] Asperti, Andrea, and Giuseppe Longo. Categories, types, and structures: an introduction to category theory for the working computer scientist. MIT press, 1991.

[16] Gärdenfors, Peter. Conceptual spaces: The geometry of thought. MIT press, 2004.