Co-learning nouns and adjectives

Güner Orhan, Sertaç Olgunsoylu, Erol Şahin, Sinan Kalkan
KOVAN Research Lab, Computer Engineering
Middle East Technical University, Turkey
Email: {guner.orhan, sertac.olgunsoylu, erol, skalkan}@ceng.metu.edu.tr

***Abstract*—In cognitive robotics community, categories belonging to adjectives and nouns have been learned separately and independently. In this article, we propose a prototype-based framework that conceptualize adjectives and nouns as separate categories that are, however, linked to and interact with each other. We demonstrate how this co-learned concepts might be useful for a cognitive robot, especially using a game called "What object is it?" that involves finding an object based on a set of adjectives.**

I. Introduction

Conceptualization, i.e., extracting relevant information from a set of exemplars in a category via an abstraction process, is an important step in building developing cognitive systems. Concepts that are formed after such abstraction processes are important for especially (i) recognizing an event or perceptual entity, (ii) comparing different categories, and (iii) reasoning in general.

The literature has extensively studied how nouns and adjectives can be learned or conceptualized from the sensorimotor experiences of a robot (e.g., [1], [2], [3], [4], [5]). However, learning these categories have been addressed separately without any interactions between the categories, as illustrated in Fig. 1a.

In this article, we go beyond the literature and our previous studies on noun and adjective learning [5] by co-learning nouns and adjectives. Co-learning is achieved by taking into account co-occurrences between adjectives and nouns (Fig. 1b). Conceptualizing nouns, adjectives and the interaction between adjectives and nouns are performed by a prototype-based approach that we have previously applied to only nouns and adjectives. We demonstrate that co-conceptualization helps in correcting wrong categorization. Moreover, it allows a robot to reason about the properties (i.e., adjectives) of an object as well as determine an object from its properties.

A. Related Studies

There has been many attempts to linking nouns to sensorimotor experiences of robots in the robotics community. For example, Yu and Ballard [6] proposed a system mapping words in speech to co-occurring features in images using a generative correspondence model. Sinapov et al. [2] increased the challenge by predicting object categories for 100 objects from the sensorimotor interactions, even using water [7]. Carbonetto and de Freitas [8] presented a system that splits a given image into regions and finds a proper mapping between regions and nouns inside the given dictionary using a probabilistic translation mode similar to a machine translation problem. Similar studies [9], [10] propose using neural networks to link words with objects and behaviors of robots to the extracted visual features.

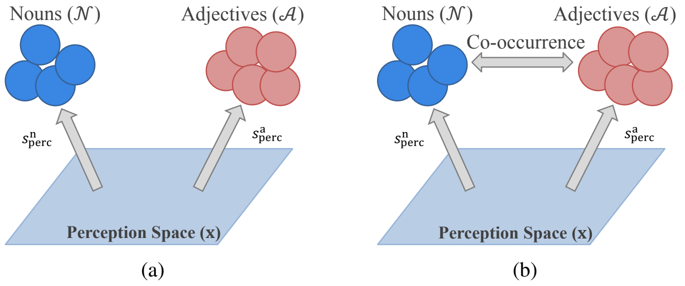

Nouns ($\mathcal{N}$)
Adjectives ($\mathcal{A}$)
$s^{\mathrm{n}}_{\mathrm{perc}}$
$s^{\mathrm{a}}_{\mathrm{perc}}$
Perception Space (x)
(a)
Nouns ($\mathcal{N}$)
Co-occurrence
Adjectives ($\mathcal{A}$)
$s^{\mathrm{n}}_{\mathrm{perc}}$
$s^{\mathrm{a}}_{\mathrm{perc}}$
Perception Space (x)
(b)

Fig. 1: (a) Existing methods to learning adjectives and nouns. (b) Our proposal, which involves using also co-occurrence between nouns and adjectives in learning them.

An important tool in linking language and sensori-motor data is artificial neural networks due to its biological plausibility and easy adaptability. Cangelosi [1] presents a review of their earlier work (all using multi-layer neural networks) on (i) the multi-agent modeling of grounding and language development, using simulated agents that discover labels, or words, for edible and non-edible food while navigating in a limited environment [11], (ii) the transfer of symbol grounding, using one simulated teacher (agent) and one simulated learner (agent) that learn new behaviors based on the symbolic representations of previously learned behaviors [12] and (iii) language comprehension in a humanoid robot, where the robot learns to associate words with its behaviors and the objects in the environment. Similarly, in an earlier work, Cangelosi and Parisi [10] use a neural network for linking nouns to two different objects (a vertical bar and a horizontal bar) and verbs to two different behaviors (pushing and pulling).

It is very well known that cross-situational co-occurrences of words and objects are very important for learning meanings of words [13] and there are already many computational models that incorporate this for word learning [4], [14] especially in a Bayesian framework [3], [4] to learn a mapping from features of objects to words.

As for learning adjectives from sensorimotor interactions of the robot, there are also many studies. McMahon et al. [15] developed a method for learning haptic adjectives from interactions whereas others [16], [17], [18] studied learning color, size and distance related adjectives based on visual features. Similar studies [9], [19], [20], [21], [22], [23] have been performed for learning object categories; however, co-learning of nouns and adjectives has not been studied previously.

II. Methods

A. Experimental Setup

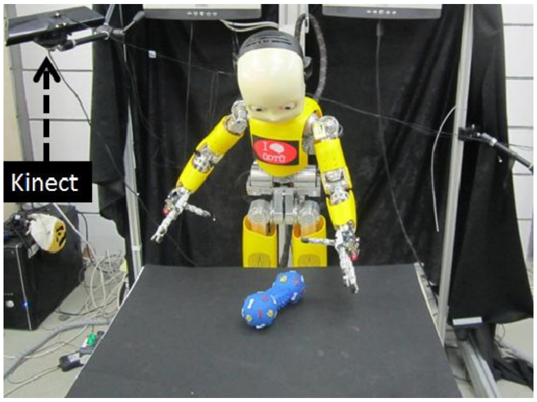
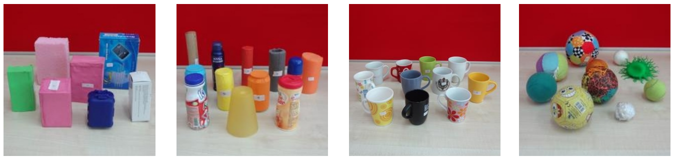


Fig. 2: The experimental setup. iCub perceives the environment with a Kinect camera, and collects tactile, audio, proprioceptive information in addition to visual information.

In order to collect the visual, haptic, audio and proprioceptive data from the objects, we used the iCub humanoid robot platform (*Fig.* 2). On the perception side, we employed a Kinect sensor[1]. Moreover, the setup included a motion capture system (VisualeyezTM VZ 4000[2]) to transform the tabletop object position and other related features to the coordinate frame of iCub. Visual processing of the RGB-D data from the Kinect sensor was performed using Point Cloud Library (PCL - [24]). For haptic sensing, we use the tactile sensor of iCub, placed on each fingertip. For audio sensing, we use a standard microphone. Lastly, we take the proprioceptive information (the joint values of fingers of iCub) into consideration.

B. Objects - Nouns and Adjectives

We have 40 objects, which are arbitrarily divided into two equal groups: 20 objects for conceptualizing nouns and adjectives, and 20 for testing them. The 40 objects have been labeled with nouns $\mathcal{N} = \{box, cylinder, cup, ball\}$. (Fig. 3) and with adjective pairs $\mathcal{A} = \{hard - soft; noisy - silent; round - edgy; tall - short; thin - thick\}$ (Fig. 4).



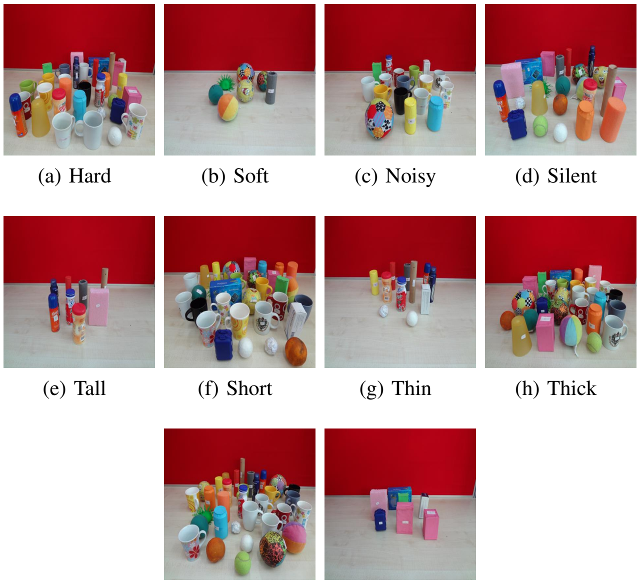
(a) Boxes (b) Cylinders (c) Cups (d) Balls

Fig. 3: The noun categories used in the experiments.

Table I shows the co-occurrences of the noun and adjective labels for the objects. We see that some adjectives are unique to certain nouns (e.g., *box* is *edgy*) whereas others are shared by different nouns categories. Note also that all cups in our dataset are noisy.

[1] http://www.xbox.com/en-US/kinect

[2] http://www.ptiphoenix.com/VZmodels.php



(a) Hard (b) Soft (c) Noisy (d) Silent

(e) Tall (f) Short (g) Thin (h) Thick

(i) Round (j) Edgy

Fig. 4: The adjective categories used in the experiments.

TABLE I: Co-occurence table of noun and adjective category pairs. It includes only the distributions of *training* objects of a specific noun category to the corresponding adjective

Noun	Hard	Soft	Noisy	Silent	Tall	Short	Thin	Thick	Round	Edgy
Box (4)	4	0	2	2	0	4	1	3	0	4
Cylinder (6)	5	1	2	4	3	3	4	2	6	0
Cup (4)	4	0	4	0	0	4	0	4	4	0
Ball (6)	2	4	1	5	0	6	1	5	6	0

C. Perception

We extract the following 100-dimensional features from each object and use $\mathbf{x}$ to denote them:

- Physical dimensions of the object (width, height, depth).
- Surface normal features (40 features: two 20-bin of histograms for azimuth and zenith of surface normals).
- Shape index features (20 features: a histogram of shape index values).
- Audio features (13 features: differences of maximum and minimum values of Mel-Frequency Cepstrum Coefficients (MFCC) coefficients [25]).
- Tactile features (12 features: range, minimum, maximum, mean, variance, standard deviation values of tactile sensor values on index and middle finger tips).
- Proprioceptive features (12 features: range, minimum, maximum, mean, variance, standard deviation values of the encoder values of two joints on iCub's arm).

For visual information, we have used the same orientations for each object in the same noun category. However, the position of an object is changed randomly on the table inside the area where iCub can reach the object. The surface normal and shape index features are extracted from the table-top segmented RGB-D data by using the PCL library [24]. Shape index is a combination of curvatures in orthogonal directions, which captures the local characteristics of a surface [26].

D. Data Collection

For collecting sensorimotor data from the objects, iCub first perceives the environment via the Kinect sensor, extracts the visual features and grasps the object based on the data from the visual features. The tactile and the proprioceptive data are extracted after the object has been grasped. After grasping, iCub shakes the object during which it collects audio information.

E. Conceptualization of Nouns and Adjectives

We adopt a prototype-based approach for conceptualizing each noun and adjective category (see [27] for other views on conceptualization). The method looks at the distribution of each feature dimension in a category to assign them into three categories: *consistently negative (-)*, *consistently positive (+)* and *inconsistent (*)*. Then, the distribution of features in a category are summarized in a string of '+', '-' and '*' symbols along with the corresponding mean and variance values for each feature dimension. The method is detailed in Algorithm 1.

Algorithm 1 Derivation of Prototypes

```
for all $l$ in the set of adjective categories $\mathcal{A}$ or noun categories $\mathcal{N}$ do
    - Compute the mean ${}_i\mu_l$ for each feature $i$:
        $${}_i\mu_l = \frac{1}{N}\sum_{e\in l} {}_i e,$$   (1)
      where $N$ is the cardinality of the set $\{e|e \in l\}$; and ${}_i\mathbf{e}$ is the $i^{th}$ value of vector $\mathbf{e}$.
    - Compute the variance ${}_i\sigma_l$ of each feature dimension $i$:
        $${}_i\sigma_l = \frac{1}{N}\sum_{e\in l}({}_i e -_i \mu_l)^2.$$   (2)
end for
- Apply Robust Neural Growing Gas (RGNG) algorithm [28] in the space of $\mu \times \sigma$.
- Manually assign the labels '+', '-', and '*' to the three clusters that emerge in the previous step.
```

The prototypes extracted for the adjective and nouns categories from the training set are listed in Tables IV and III respectively. We see that the relevant and irrelevant features are captured nicely. For example, since all the cups in our dataset are noisy, we see consistent dependence of cups to audio features. This allows us to predict the adjective and noun categories of an object using only relevant(consistent) features.

The distance between the perceptual features **x** of an object and the prototype f_l of a category l is calculated without using the irrelevant features which are marked as '*' in the prototype:

$$d(\mathbf{x}, f_l) = \sqrt{\sum_{i\in R(f_l)\setminus R^*(f_l)} ({}_i\mathbf{x} - \ {}_i\mu_l)^2}, \qquad (3)$$

where $R(f_l) \setminus R^*(f_l)$ is the set of *relevant* feature dimensions in prototype f_l; ${}_i\mathbf{v}$ is the i^{th} value of a vector **v**; and μ_l is the mean of the features in category l.

TABLE II: Prototypes of noun and adjective co-occurrences (Sect. II-E). '*', '+' and '-' respectively represent inconsistent co-occurrence, consistent co-occurrence and consistent non-co-occurrences.

Noun	Hard	Soft	Noisy	Silent	Tall	Short	Thin	Thick	Round	Edgy
Box	+	-	*	*	-	+	-	+	-	+
Cylinder	+	-	*	*	*	*	*	*	+	-
Cup	+	-	+	-	-	+	-	+	+	-
Ball	*	*	-	+	-	+	-	+	+	-

F. Prediction of Noun Categories

As we mentioned before, prediction of a noun (and adjective) category is based on perception as well as the interaction between nouns and adjectives. Therefore, the similarity between features **x** of an object and the prototype f of a noun $n \in \mathcal{N}$ is defined as follows (see also Fig. 1b):

$$s^{\text{n}}_{\text{comb}}(\mathbf{x}, f_n) = (1 - w_{\text{an}}) \times s^{\text{n}}_{\text{perc}}(\mathbf{x}, f_n) + w_{\text{an}} \times c_{\text{n}}(f_n, \hat{\mathcal{A}}_{\mathbf{x}}), \qquad (4)$$

where $w_{\text{an}} \in [0, 1]$ is a weight controlling the contribution of the prediction from the adjectives; $\hat{\mathcal{A}}_{\mathbf{x}} \subset \mathcal{A}$ is the set of adjectives that are predicted from the features **x** of the object; $s^{\text{n}}_{\text{perc}}(\mathbf{x}, f_n)$ is the similarity between the perceptual features **x** and the prototype f_n for noun n:

$$s^{\text{n}}_{\text{perc}}(\mathbf{x}, f_n) = \frac{\prod_{n_1\in\mathcal{N}\setminus\{n\}} d(\mathbf{x}, f_{n_1})}{\sum_{n_1\in\mathcal{N}}(\prod_{n_2\in\mathcal{N}\setminus\{n_1\}} d(\mathbf{x}, f_{n_2}))}, \qquad (5)$$

where $d(.,.)$ is the distance function in Eq. 3. $c_{\text{n}}(f_n, \hat{\mathcal{A}}_{\mathbf{x}}) \in [0, 1]$ in Eq. 4 is defined as:

$$c_{\text{n}}(f_n, \hat{\mathcal{A}}_{\mathbf{x}}) = \sum_{a\in\hat{\mathcal{A}}_{\mathbf{x}}} \frac{c(n, a)}{|\hat{\mathcal{A}}_{\mathbf{x}}|}, \qquad (6)$$

where A_p is the set of predicted adjectives based on the perceptual features of the object; $|\mathcal{S}|$ is the cardinality of set $\mathcal{S}$. $c(n, a)$ is the co-occurence value of noun n with adjective a taking into consideration only consistent dependencies (i.e., non-'*' entries in Table II). In other words, if a noun is not consistently co-occurring with an adjective, that adjective does not contribute any weight to c_{n}. The functions $s^{\text{n}}_{\text{comb}}(.,.)$, $s^{\text{n}}_{\text{perc}}$ and c_{n} take values in the range $[0, 1]$.

G. Prediction of Adjective Categories

Similar to the nouns, the similarity between features **x** of an object and the prototype f of an adjective $a \in \mathcal{A}$ is defined as (see also Fig. 1b):

$$s^{\text{a}}_{\text{comb}}(\mathbf{x}, f_a) = (1 - w_{\text{na}}) \times s^{\text{a}}_{\text{perc}}(\mathbf{x}, f_a) + w_{\text{na}} \times c(a, n_{\mathbf{x}}), \qquad (7)$$

where $w_{\text{na}} \in [0, 1]$ is a weight controlling the contribution of the prediction from the nouns; $n_{\mathbf{x}} \in \mathcal{N}$ is the noun that is predicted from the features of the object; $s^{\text{a}}_{\text{perc}}(\mathbf{x}, f_a)$ is the perceptual similarity between the object and the prototype f_a for adjective a:

$$s^{\text{a}}_{\text{perc}}(\mathbf{x}, f_a) = \frac{d(\mathbf{x}, \mu_{\overline{a}})}{d(\mathbf{x}, \mu_{\overline{a}}) + d(\mathbf{x}, \mu_a)}, \qquad (8)$$

where $\overline{a}$ is the pair of adjective a; and, $d(.,.)$ is the distance function in Eq. 3. As in Eq. 6, $c(n, a)$ is the co-occurrence value of noun n with adjective a taking into consideration only consistent dependencies (i.e., non-'*' entries in Table II).

TABLE III: Prototypes for noun categories.

Noun Categories	Visual Features	Audio Features	Haptic Features	Proprioceptive Features
Box	+-+++-++---++--------------++--+--------**--------------------+----	-**+**-++-*+*	------------	--++*+*+----
Cylinder	--+-+-+----++---------------+++++-----------------------------+----	*******_**+-+	**__********	--++++++----
Cup	--++-++----++---------------+++-----***------++---------------+----	++-++++++++++	------------	---+--------
Ball	+-+--++-+++--+--------------+-+++***---------------------------++++	_____*_______	__******____	++++++++--++

IEEE ICDL - Epigenetic Robotics 2013

TABLE IV: Prototypes for adjective categories.

Adjective Categories	Visual Features	Audio Features	Haptic Features	Proprioceptive Features
Hard	+-+*+++++++++--------------+*++*+--------*--------------------+++++	******+**+++*	**++****++**	--++*+*+----
Soft	+-+++*+-++*++*------------+*++++***---------------------------*+***	+---+*-++---+	------------	++******--++
Noisy	--+++*+---*++*------------++++*+*-****------------------------++---	+++++++++++++	*___****__*_	--++*+**----
Silent	+-+-++++*+*++*-------------+*++++**-----**--------------------*+***	_____________	************	++******--+-
Tall	--+-+-+----++---------------+**+*-----------------------------+----	________*___*	**__********	*-++++++----
Short	****_*+*******_____________*******_*__________________________+****	*************	*_******____	********__*_
Thin	--+-+-+-*-*++---------------+++++-----------------------------+-**-	******_+*_*_+	**__********	--++++++----
Thick	--+++++-++*++*-------------+++++-***--*-**--------------------+++*+	**+******++-*	*_******____	--++*+*+----
Round	--+-+*+-++*++*--------------+++++*-*--------------------------*+**+	**+***+**++-+	**+******_**	+-++*+*+----
Edgy	+-++++++---++--------------++--+--------**--------------------+----	-**+**-+*+*+*	------------	--++*+**----

In other words, if an adjective is not consistently co-occurring with a noun, that noun does not contribute any weight to c.

The functions $s^{a}_{\text{comb}}(.,.)$ and $s^{a}_{\text{perc}}(.,.)$ take values in the range $[0,1]$. To avoid the cyclic computation, $\hat{\mathcal{A}}_{\mathbf{x}}$ in Sect. II-F is determined by setting w_{na} to zero, and $n_{\mathbf{x}}$ is found by setting w_{an} to zero.

H. Prediction of Noun and Adjective Categories Using SVM

SVM (Support Vector Machines) [29] is a widely-used supervised method for learning a maximum-margin separation between labeled data. We apply SVM to learn two mappings from the perceptual features: $\mathbf{x} \rightarrow \mathcal{A}$ and $\mathbf{x} \rightarrow \mathcal{N}$. We use 5-fold cross-validation when training a SVM.

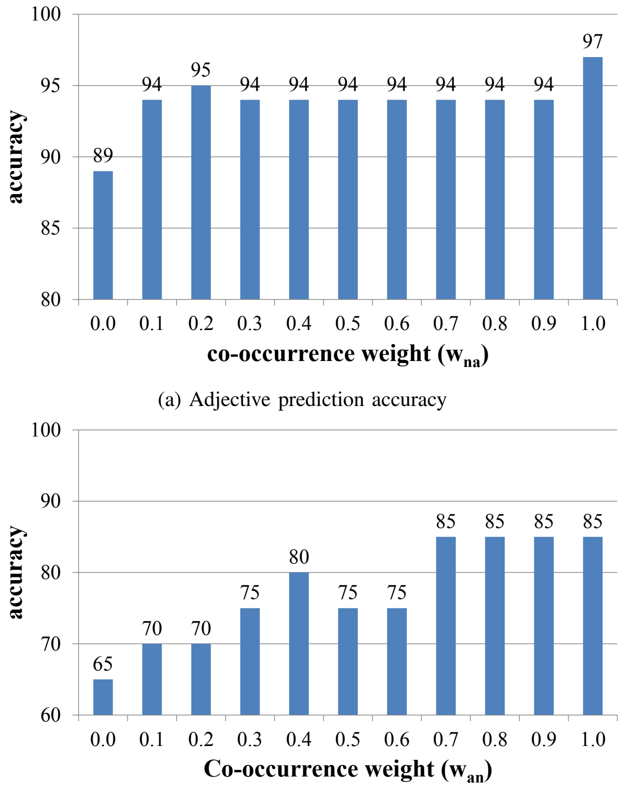
Table V lists the average prediction accuracies from the perceptual features for s_{perc} and SVM. We can see that nouns are learned better than adjectives. Moreover, for both noun and adjective predictions, using prototypes give better accuracies than using SVM.

TABLE V: Average noun and adjective prediction accuracy results on the **training** set.

	Perceptual Similarity (s_{perc})	SVM
Nouns	100%	90%
Adjectives	94%	88%

III. Results

Using the 20 objects (the testing set), we evaluated the prediction accuracies of nouns, adjectives and whether or not co-occurrences helped at all. Moreover, we demonstrated the usefulness of the co-conceptualization via a game called "What object is it?", which involves predicting nouns from a set of adjectives.



(a) Adjective prediction accuracy

(b) Noun prediction accuracy

Fig. 5: Noun and adjective prediction accuracies for the testing set with respect to weighted contribution of co-occurrence.

TABLE VI: Predicted adjectives for some objects from the test set (bold denotes correct classification). The co-occurrence weight w_{na} is taken as 0.2 where prediction performance is maximized.

	Adjectives		
Objects	**Perceptual Similarity** ($s^{\text{a}}_{\text{perc}}$)	**Perc. Similarity and Cooccurence** $s^{\text{a}}_{\text{comb}}$	**SVM**
O_1	**hard (61%)** **noisy (67%)** tall (54%) **thick (55%)** **round (54%)**	**hard (65%)** **noisy (70%)** **short (51%)** **thick (60%)** **round (59%)**	**hard (95%)** **noisy (93%)** **short (92%)** **thick (74%)** **round (76%)**
O_2	**hard (55%)** **silent (67%)** **tall (64%)** thin (54%) **edgy (55%)**	**hard (59%)** **silent (66%)** **tall (57%)** **thick (51%)** **edgy (60%)**	**hard (75%)** **silent (89%)** short (69%) thick (54%) round (59%)
O_3	**hard (54%)** **silent (61%)** short (56%) **thick (53%)** **edgy (57%)**	**hard (59%)** **silent (60%)** short (60%) **thick (58%)** **edgy (61%)**	**hard (82%)** **silent (89%)** short (92%) **thick (96%)** **edgy (89%)**
O_4	**soft (60%)** **silent (58%)** **short (52%)** **thick (53%)** **round (54%)**	**soft (59%)** **silent (63%)** **short (57%)** **thick (57%)** **round (59%)**	**soft (99%)** **silent (93%)** **short (88%)** **thick (54%)** **round (98%)**
O_5	**hard (56%)** **silent (73%)** short (53%) **thin (51%)** **round (52%)**	**hard (61%)** **silent (70%)** short (53%) **thin (51%)** **round (56%)**	**hard (73%)** **silent (83%)** short (78%) thick (62%) **round (75%)**

A. *Effect of Co-occurrence on Prediction*

We have previously shown that learning adjectives is more difficult than learning nouns [5], which is also reflected by findings and hypotheses in Psychology [30] and Language [31]. This is mainly due to the fact that adjectives are mostly related to changes in only a few dimensions (such as height or width) whereas nouns depend on many more dimensions [31]; for this reason, it is more difficult to capture relevant changes for adjectives in a high-dimensional feature space, making learning of adjectives more difficult [5].

This leads to more mistakes in predicting adjectives when the same learning method is used. However, wrong predictions can be rectified by using the co-occurrences between nouns and adjectives, as we show in Fig. 5a. The figure displays the effect of the co-occurrence weight w_{na} (Eq. 7) on the prediction accuracy. We see that the predicted noun category can contribute and correct wrong adjective predictions coming from the perceptual features. See also tables VI and VII, which show rectification of some wrong adjective predictions.

Noun prediction using ($s^{\text{n}}_{\text{perc}}$) performs 100% on both the training and the test sets. Since this leaves no room for improvement by co-occurrence, we have added noise to noun prediction based on perceptual features (s_{perc}) with 40% probability by subtracting $s^{\text{n}}_{\text{perc}}$ from one. As shown in Fig. 5b, co-occurrence can improve wrong noun predictions similar to the case for adjectives.

Comparing Fig. 5a and Fig. 5b, we see that the contribution of co-occurrence (e.g., when $w_{\text{na}} = w_{\text{an}} = 1$) from nouns to adjectives is bigger than the reverse. The reason is that nouns share more adjectives, than different adjectives share nouns, as visible in Table II.

TABLE VII: Predicted nouns for some objects from the test set (bold denotes correct classification). The co-occurrence weight w_{an} is taken as 0.2.

	Nouns		
Objects	**Perceptual Similarity** ($s^{\text{n}}_{\text{perc}}$)	**Perc. Similarity and Cooccurence** $s^{\text{n}}_{\text{comb}}$	**SVM**
O_1	Box (22%) Cylinder (24%) **Cup (37%)** Ball (17%)	Box (23%) Cylinder (24%) **Cup (35%)** Ball (18%)	Box (25%) Cylinder (23%) **Cup (45%)** Ball (7%)
O_2	**Box (32%)** Cylinder (30%) Cup (19%) Ball (19%)	**Box (36%)** Cylinder (34%) Cup (15%) Ball (15%)	Box (38%) Cylinder (44%) Cup (3%) Ball (15%)
O_3	**Box (34%)** Cylinder (25%) Cup (21%) Ball (20%)	**Box (32%)** Cylinder (25%) Cup (22%) Ball (21%)	**Box (67%)** Cylinder (16%) Cup (4%) Ball (13%)
O_4	Box (22%) Cylinder (23%) Cup (20%) **Ball (35%)**	Box (22%) Cylinder (23%) Cup (22%) **Ball (33%)**	Box (3%) Cylinder (3%) Cup (1%) **Ball (93%)**
O_5	Box (24%) **Cylinder (47%)** Cup (16%) Ball (13%)	Box (24%) **Cylinder (43%)** Cup (18%) Ball (15%)	Box (34%) **Cylinder (44%)** Cup (6%) Ball (16%)

B. *The "What object is it?" Game*

It is a game where the robot is exposed to a set of adjectives and expected to predict the noun category which is best described by the adjectives. This game demonstrates the impact of the interaction between adjective and noun concepts.

Table VIII shows some example turns of the game where each row consists of a number of adjectives and noun concepts having two highest confidences. Predictions that the robot makes depend on the robot's past interactions with objects, i.e., on its subjective perception of categories. For example, our training set is composed of boxes which are *hard*, *short*, and *thick* in addition to being inherently *edgy*. Therefore, if the list *hard, short, thick, edgy* is presented as adjectives, they anticipate the characteristics of the *box* category which is learned in the training. As another example, *cups* and *cylinders* are *hard* and *round* objects according to the training set. Therefore, adjectives *hard* and *round* together lead to *cup* and *cylinder* being predicted. On the other hand, if a combination of some adjectives which does not describe any noun category is given, the prediction confidences are low (e.g., when *round, thin, soft, tall* in Table VIII).

IV. Conclusion

We proposed a method for co-learning nouns and adjectives. Our method uses both (i) the perceptual similarity based predictions of adjectives and nouns from the perceptual features, which is based on a prototype-based conceptualization method that we have previously proposed [5] and (ii) the co-occurrences of adjectives and nouns. We demonstrated that prediction of adjectives becomes more accurate with the contribution of co-occurrences between adjectives and nouns whereas this effect is not visible for nouns since they are predicted with 100% accuracy.

TABLE VIII: "What object is it?" game: Determine noun based on given adjectives.

Given Adjectives				**Predicted Nouns**
a_1	a_2	a_3	a_4	
hard	short	thick	edgy	Box (73%) Cup (53%)
hard	round	-	-	Cup (72%) Cylinder (70%)
silent	short	thick	round	Ball (70%) Cup (52%)
short	thick	-	-	Cup (69%) Ball (69%)
round	thin	soft	tall	Ball (18%) Cylinder (17%)
soft	silent	thick	-	Ball (45%) Cup (23%)
hard	noisy	round	-	Cup (73%) Cylinder (46%)
short	thick	round		Ball (70%) Cup (69%)
edgy	-	-	-	Box (100%) Others (0%)

ACKNOWLEDGMENT

This work is partially funded by the Turkish National Science and Technology Organization (TÜBITAK) as a project (no: 111E287).

REFERENCES

[1] A. Cangelosi, "Grounding language in action and perception: From cognitive agents to humanoid robots," *Physics of Life Reviews*, vol. 7, no. 2, pp. 139–151, 2010.

[2] J. Sinapov, C. Schenck, K. Staley, V. Sukhoy, and A. Stoytchev, "Grounding semantic categories in behavioral interactions: Experiments with 100 objects," *Journal of Robotics and Autonomous Systems (to appear)*, 2013.

[3] F. Xu and J. B. Tenenbaum, "Word learning as bayesian inference." *Psychological review*, vol. 114, no. 2, p. 245, 2007.

[4] M. C. Frank, N. D. Goodman, and J. B. Tenenbaum, "A bayesian framework for cross-situational word learning," *Advances in neural information processing systems*, vol. 20, pp. 20–29, 2007.

[5] O. Yuruten, K. F. Uyanik, Y. Caliskan, E. Bozcuoglu A. K., Sahin, and S. Kalkan, "Development of adjective and noun concepts from affordances on the icub humanoid robot," *12th International Conference on Adaptive Behaviour (SAB)*, 2012.

[6] C. Yu and D. H. Ballard, "On the integration of grounding language and learning objects," *19th Int. Conf. on Artifical Intelligence*, pp. 488–493, 2004.

[7] S. Griffith, V. Sukhoy, T. Wegter, and A. Stoytchev, "Object categorization in the sink: Learning behavior–grounded object categories with water," in *Proceedings of the ICRA Workshop on Semantic Perception, Mapping and Exploration*, 2012.

[8] P. Carbonetto and N. de Freitas, "Why can't jose read? the problem of learning semantic associations in a robot environment," in *The HLT-NAACL Workshop on Learning word meaning from non-linguistic data*, 2003, pp. 54–61.

[9] A. F. Morse, P. Baxter, T. Belpaeme, L. B. Smith, and A. Cangelosi, "The power of words," *Int. Conference on Epigenetic Robotics*, 2011.

[10] A. Cangelosi and D. Parisi, "The processing of verbs and nouns in neural networks: Insights from synthetic brain imaging," *Brain and Language*, vol. 89, no. 2, pp. 401–408, 2004.

[11] A. Cangelosi, "Evolution of communication and language using signals, symbols, and words," *IEEE Transactions on Evolutionary Computation*, vol. 5, no. 2, pp. 93–101, 2001.

[12] A. Cangelosi, E. Hourdakis, and V. Tikhanoff, "Language acquisition and symbol grounding transfer with neural networks and cognitive robots," in *International Joint Conference on Neural Networks*, 2006, pp. 1576–1582.

[13] C. Yu and L. B. Smith, "Rapid word learning under uncertainty via cross-situational statistics," *Psychological Science*, vol. 18, no. 5, pp. 414–420, 2007.

[14] J. M. Siskind, "A computational study of cross-situational techniques for learning word-to-meaning mappings," *Cognition*, vol. 61, no. 1, pp. 39–91, 1996.

[15] I. McMahon, V. Chu, L. Riano, C. G. McDonald, Q. He, J. M. Perez-Tejada, M. Arrigo, N. Fitter, J. C. Nappo, and T. Darrell, "Robotic learning of haptic adjectives through physical interaction," *IROS workshop on Advances in Tactile Sensing and Touch based Human-Robot Interaction*, 2012.

[16] A. Petrosino and K. Gold, "Toward fast mapping for robot adjective learning," in *Dialog with Robots: AAAI Fall Symposium Series*, 2010.

[17] H. Dindo and D. Zambuto, "A probabilistic approach to learning a visually grounded language model through human-robot interaction," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2010, pp. 790–796.

[18] A. Chella, H. Dindo, and D. Zambuto, "Grounded human-robot interaction," in *Biologically Inspired Cognitive Architectures: AAAI Fall Symposium Series*, 2009.

[19] A. Chauhan and L. S. Lopes, "Using spoken words to guide open-ended category formation," *Cognitive Processing*, vol. 12, no. 4, pp. 341–354, 2011.

[20] P. Haazebroek, S. van Dantzig, and B. Hommel, "A computational model of perception and action for cognitive robotics," *Cognitive Processing*, vol. 12, no. 4, pp. 355–365, 2011.

[21] Y. Sugita, J. Tani, and M. V. Butz, "Simultaneously emerging braitenberg codes and compositionality," *Adaptive Behavior*, vol. 19, no. 5, pp. 295–316, 2011.

[22] A. M. Glenberg and V. Gallese, "Action-based language: A theory of language acquisition, comprehension, and production," *Cortex*, vol. 48, no. 7, pp. 905–922, 2011.

[23] K. Gold, M. Doniec, C. Crick, and B. Scassellati, "Robotic vocabulary building using extension inference and implicit contrast," *Artificial Intelligence*, vol. 173, no. 1, pp. 145–166, 2009.

[24] R. B. Rusu and S. Cousins, "3D is here: Point Cloud Library (PCL)," in *IEEE International Conference on Robotics and Automation (ICRA)*, Shanghai, China, May 9-13 2011.

[25] A. Qin and P. Suganthan, "Robust growing neural gas algorithm with application in cluster analysis," *Neural Networks*, vol. 17, no. 8-9, pp. 1135 – 1148, 2004.

[26] J. Koenderink and A. van Doorn, "Surface shape and curvature scales," *Image and vision computing*, vol. 10, no. 8, pp. 557–564, 1992.

[27] E. Rosch, "Reclaiming concepts," *Journal of Consciousness Studies, 6*, vol. 11, no. 12, pp. 61–77, 1999.

[28] A. K. Qin and P. N. Suganthan, "Robust growing neural gas algorithm with application in cluster analysis," *Neural Networks*, vol. 17, no. 8-9, pp. 1135–1148, 2004.

[29] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[30] C. Sandhofer and L. B. Smith, "Learning adjectives in the real world: How learning nouns impedes learning adjectives," *Language Learning and Development*, vol. 3, no. 3, pp. 233–267, 2007.

[31] G. Sassoon, "Adjectival vs. nominal categorization processes: The rule vs. similarity hypothesis," *Belgian Journal of Linguistics*, vol. 25, no. 1, pp. 104–147, 2011.