Learning Adjectives and Nouns from Affordances on the iCub Humanoid Robot

Onur Yürüten[1], Kadir Fırat Uyanık[1,3], Yiğit Çalışkan[1,2], Asil Kaan Bozcuoğlu[1], Erol Şahin[1], and Sinan Kalkan[1]

[1] Kovan Res. Lab, Dept. of Computer Eng., Middle East Technical University, Turkey
{oyuruten,kadir,asil,erol,skalkan}@ceng.metu.edu.tr,
[2] Dept. of Computer Eng., Bilkent University, Turkey caliskan@cs.bilkent.edu.tr
[3] Dept. of Electrical and Electronics Eng., Middle East Technical University, Turkey

Abstract. This article studies how a robot can learn *nouns* and *adjectives* in language. Towards this end, we extended a framework that enabled robots to learn affordances from its sensorimotor interactions, to learn nouns and adjectives using labeling from humans. Specifically, an iCub humanoid robot interacted with a set of objects (each labeled with a set of adjectives and a noun) and learned to predict the effects (as labeled with a set of verbs) it can generate on them with its behaviors. Different from appearance-based studies that directly link the appearances of objects to nouns and adjectives, we first predict the affordances of an object through a set of Support Vector Machine classifiers which provided a functional view of the object. Then, we learned the mapping between these predicted affordance values and nouns and adjectives. We evaluated and compared a number of different approaches towards the learning of nouns and adjectives on a small set of novel objects.
The results show that the proposed method provides better generalization than the appearance-based approaches towards learning adjectives whereas, for nouns, the reverse is the case. We conclude that affordances of objects can be more informative for (a subset of) adjectives describing objects in language.

Keywords: affordances, nouns, adjectives

1 Introduction

Humanoid robots are expected to be part of our daily life and to communicate with humans using natural language. In order to accomplish this long-term goal, such agents should have the capability to perceive, to generalize and also to communicate about what they perceive and cognize. To have the human-like perceptual and cognitive abilities, an agent should be able (i) to relate its symbols or symbolic representations to its internal and external sensorimotor data/experiences, which is mostly called *the symbol grounding problem* [1] and (ii) to conceptualize over raw sensorimotor experiences towards abstract, compact and general representations. Problems (i) and (ii) are two challenges an embodied agent faces and in this article, we focus on problem (i).

The term concept is defined by psychologists [2] as the information associated with its referent and what the referrer knows about it. For example, the concept of an apple is all the information that we know about apples. This concept includes not only how an apple looks like but also how it tastes, how it feels etc. The appearance related aspects of objects correspond to a subset of noun concepts whereas the ones related to their affordances (e.g., edible, small, round) correspond to a subset of adjective concepts.

Affordances, a concept introduced by J. J. Gibson [3], offers a promising solution towards symbol grounding since it ties perception, action and language naturally. J. J. Gibson defined affordances as the action possibilities offered by objects to an agent: Firstly, he argued that organisms infer possible actions that can be applied on a certain object directly and without any mental calculation. In addition, he stated that, while organisms process such possible actions, they only take into account relevant perceptual data, which is called as perceptual economy. Finally, Gibson indicated that affordances are relative, and it is neither defined by the habitat nor by the organism alone but through their interactions with the habitat.

In our previous studies [4,5], we proposed methods for linking affordances to object concepts and verb concepts. In this article, we extend these to learn nouns and adjectives from the affordances of objects. Using a set of Support Vector Machines, our humanoid robot, iCub, learns the affordances of objects in the environment by interacting with them. After these interactions, iCub learns nouns and adjectives either (i) by directly linking appearance to noun and adjective labels, or (ii) by linking the affordances of objects to noun and adjective labels. In other words, we have two different approaches (appearance-based and affordance-based models) for learning nouns and adjectives, which we compare and evaluate. Later, when shown a novel object, iCub can recognize the noun and adjectives describing the object.

2 Related Studies

The symbol grounding problem in the scope of noun learning has been studied by many. For example, Yu and Ballard [6] proposed a system that collects sequences of images alongside speech. After speech processing and object detection, objects and nouns inside the given speech are related using a generative correspondence model. Carbonetto et al. [7] 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. On another side, Saunders et al. [8] suggested an interactive approach to learn lexical semantics by demonstrating how an agent can use heuristics to learn simple shapes which are presented by a tutor with unrestricted speech. Their method matches perceptual changes in robot's sensors with the spoken words and trains k-nearest neighbor algorithm in order to learn the names of shapes. In similar studies, Cangelosi et al. [9,10] use neural networks to link words with behaviours of robots and the extracted visual features.

Based on Gibson's ideas and observations, Şahin et al. [11] formalized affordances as a triplet (see, e.g., [12,13,14] for similar formalizations):

$$(o, b, f), \tag{1}$$

where f is the effect of applying behaviour b on object o. As an example, a behaviour b_{lift} that produces an effect f_{lifted} on an object o_{cup} forms an affordance relation $(o_{\text{cup}}, b_{\text{lift}}, f_{\text{lifted}})$. Note that an agent would require more of such relations on different objects and behaviours to learn more general affordance relations and to conceptualize over its sensorimotor experiences.

During the last decade, similar formalizations of affordances proved to be very practical with successful applications to domains such as navigation [15], manipulation [16,17,18,19,20], conceptualization and language [5,4], planning [18], imitation and emulation [12,18,4], tool use [21,22,13] and vision [4]. A notable one with a notion of affordances similar to ours is presented by Montesano et al. [23,24]. Using the data obtained from the interactions with the environment, they construct a Bayesian network where the correlations between actions, entities and effects are probabilistically mapped. Such an architecture allows action, entity and effect information to be separately queried (given the other two information) and used in various tasks, such as goal emulation.

In this article, our focus is linking affordances with nouns and adjectives. In addition to directly linking the appearance of objects with nouns and adjectives, we learn them from the affordances of objects and compare the two approaches.

3 Methodology

3.1 Setup and Perception

We use the humanoid robot iCub to demonstrate and assess the performance of the models we develop. iCub perceives the environment with a Kinect sensor and a motion capture system (VisualEyez VZ2). In order to simplify perceptual processing, we assumed that iCub's interaction workspace is dominated by an interaction table. We use PCL[25] to process raw sensory data. The table is assumed to be planar and is segmented out as background. After segmentation, the point cloud is clustered into objects and the following features extracted from the point cloud represent an object o (Eq. 1):

- *Surface features:* surface normals (azimuth and zenith angles), principal curvatures (min and max), and shape index. They are represented as a 20-bin histogram in addition to the minimum, maximum, mean, standard deviation and variance information.
- *Spatial features:* bounding box pose (x, y, z, theta), bounding box dimensions (x, y, z), and object presence.

3.2 Data Collection

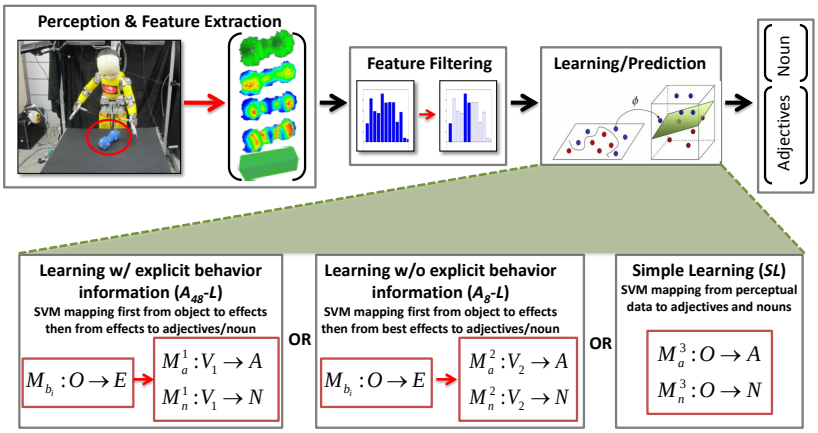
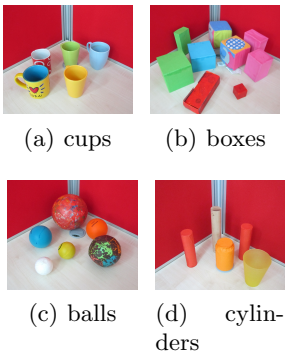


Fig. 1. Overview of the system. iCub perceives the environment and learnes the affordances. From either the perceptual data or the affordances, it learns different models for learning nouns and affordances.

The robot interacted with a set of 35 objects of variable shapes and sizes, which are assigned the nouns "cylinder", "ball", "cup", "box" (Fig. 2).

The robot's behaviour repertoire $\mathcal{B}$ contains six behaviors ($b_1, ..., b_6$ - Eq. 1): *push-left*, *push-right*, *push-forward*, *pull*, *top-grasp*, *side-grasp*. iCub applies each behaviour b_j on each object o_i and observes an effect $f_{o_i}^{b_j} = o_i' - o_i$, where o_i' is the set of features extracted from the object after behaviour b_j is applied. After each interaction epoch, we give an appropriate effect label $E_k \in \mathcal{E}$ to the observed effect $f_{o_i}^{b_j}$, where E can take values *moved-left*, *moved-right*, *moved-forward*, *moved-backward*, *grasped*, *knocked*, *disappeared*, *no-change*[4]. Thus, we have a collection of $\{o_i, b_j, E_{o_i}^{b_j}\}$, including an effect label $E_{o_i}^{b_j}$ for the effect of applying each behaviour b_j to each object o_i.



(a) cups (b) boxes (c) balls (d) cylinders

Fig. 2. The objects in our dataset.

3.3 Learning Affordances

Using the effect labels $E \in \mathcal{E}$, we train a Support Vector Machine (SVM) classifier for each behavior b_i to learn a mapping $\mathcal{M}_{b_i} : \mathcal{O} \rightarrow \mathcal{E}$ from the initial representation of the objects (i.e., $\mathcal{O}$) to the effect labels ($\mathcal{E}$). The trained SVMs

[4] The *no-change* label means that the applied behavior could not generate any notable change on the object. For example, iCub cannot properly grasp objects larger than its hand, hence, the *grasp* behaviour on large objects do not generate any change.

can be then used to predict the effect (label) $E_{o_l}^{b_k}$ of a behavior b_k on a novel object o_l using the trained mapping $\mathcal{M}_{b_k}$. Before training SVMs, we use ReliefF feature selection algorithm [26] and only use the features with important contribution (weight > 0) to training.

3.4 Adjectives

We train SVMs for learning the adjectives of objects from their affordances (see Fig. 1). We have six adjectives, i.e., $\mathcal{A} = \{$'*edgy*'-'*round*', '*short*'-'*tall*', '*thin*'-'*thick*'$\}$, for which we require three SVMs (one for each pair). We have the following three adjective learning models:

- *Adjective learning with explicit behavior information (**$\boldsymbol{A}_{48}$-AL**)*:
  In the first adjective learning model, for learning adjectives $a \in \mathcal{A}$, we use the trained SVMs for affordances (i.e., $\mathcal{M}_b$ in Sect. 3.3) to acquire a **48-dimensional** space, $\mathcal{V}_1 = (\hat{E}_1^{b_1}, ..., \hat{E}_8^{b_1}, ..., \hat{E}_1^{b_6}, ..., \hat{E}_8^{b_6})$, where $\hat{E}_i^{b_j}$ is the confidence of behaviour b_j producing effect E_i on the object o. We train an SVM for learning the mapping $\mathcal{M}_a^1 : \mathcal{V}_1 \rightarrow \mathcal{A}$.
- *Adjective learning without explicit behavior information (**$\boldsymbol{A}_8$-AL**)*:
  In the second adjective learning model, for learning adjectives $a \in \mathcal{A}$, we use the trained SVMs for affordances to acquire an **8-dimensional** affordance vector, $\mathcal{V}_2 = (p(E_1), ..., p(E_8))$, where $p(E_i)$ is the maximum SVM confidence of a behaviour b_j leading to the effect E_i on object o. From $\mathcal{V}_2$, we train an SVM for learning the mapping $\mathcal{M}_a^2 : \mathcal{V}_2 \rightarrow \mathcal{A}$.
- *Simple adjective learning (**SAL**)*:
  In the third adjective learning model, we learn $\mathcal{M}_a^3 : \mathcal{O} \rightarrow \mathcal{A}$ directly from the appearance of the objects.

After learning, iCub can predict the noun and adjective labels for a novel object (Fig. 3).

3.5 Nouns

We train **one SVM** for nouns $\mathcal{N} = \{$'ball', 'cylinder', 'box', 'cup'$\}$, for which we have 413 instances. Similar to adjectives, we have three models:

- *Noun learning with explicit behavior information (**$\boldsymbol{A}_{48}$-NL**)*:
  Similar to A_{48}-AL, we train an SVM for learning the mapping $\mathcal{M}_n^1 : \mathcal{V}_1 \rightarrow \mathcal{N}$.
- *Noun learning without explicit behavior information (**$\boldsymbol{A}_8$-NL**)*:
  Similar to A_8-AL, we train an SVM for learning the mapping $\mathcal{M}_n^2 : \mathcal{V}_2 \rightarrow \mathcal{N}$.
- *Simple noun learning (**SNL**)*:
  Similar to SAL, we train an SVM for learning the mapping $\mathcal{M}_n^3 : \mathcal{O} \rightarrow \mathcal{N}$ directly from the appearance of the objects.

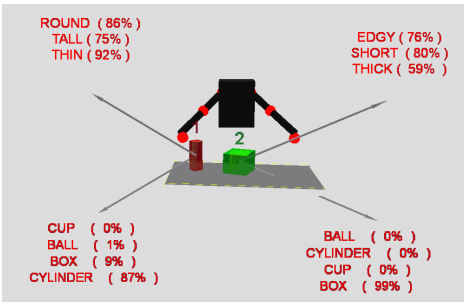


Fig. 3. After learning nouns and adjectives, iCub can refer to an object with its higher level representations or understand what is meant if such representations are used by a human.

4 Results

The prediction accuracy of the trained SVMs that map each behaviour b_i on an object to an effect label (i.e., $\mathcal{M}_{b_i} : \mathcal{O} \rightarrow \mathcal{E}$) is as follows: 90% for *top-grasp*, 100% for *side-grasp*, 96% for *pull*, 100% for *push-forward*, 92% for *push-left* and 96% for *push-right*.

4.1 Results on Adjectives

Table 1. The dependence between adjectives and affordances for the model A_{48}-AL ($\mathcal{M}_a^1$). TG: Top Grasp, SG: Side Grasp, PR: Push Right, PL: Push Left, PF: Push Forward, PB: Pull. For each behavior, there are eight effect categories: a: Moved Right, b: Moved Left, c: Moved Forward, d: Pulled, e: Knocked, f: No Change g: Grasped, h: Disappeared.

Adjective	TG	SG	PR	PL	PF	PB
	abcdefgh	*abcdefgh*	*abcdefgh*	*abcdefgh*	*abcdefgh*	*abcdefgh*
Edgy	-----+--	-----**-	*---**-+	-*--**-+	---***-+	---*++-+
Round	-----**-	-----+--	*---+*-+	-*--+*-+	---***-*	---***-*
Short	-----**-	-----+--	+---**-+	-+--**-+	---+**-+	---+**-+
Tall	-----**-	-----**-	*---+*-+	-*--+*-+	---*++-*	---*++-*
Thin	-----**-	-----**-	*---+*-+	-*--+*-+	---*+*-+	----++-+
Thick	-----+--	-----**-	*---**-*	-*--**-*	---***-*	---+*+-*

Using Robust Growing Neural Gas [27], we clustered the types of dependence between each adjective and the effects of the behaviours into *Consistently Small* (-), *Consistently Large* (+) and *Highly Variant* (*). These dependencies allow iCub to relate adjectives with what it can and cannot do with them. Table 1 shows these dependencies for the model A_{48}-AL ($\mathcal{M}_a^1$) introduced in Sect. 3.4. We see from the table what behaviours can consistently generate which effects on which types of objects (specified with their adjectives). For example, with

a consistently large probability, the robot would generate *no change* effect on *edgy* or *thick* objects when *top grasp* behavior was applied. Furthermore, the *short* and *tall* objects show a clear distinction in response to pushing behaviors (*tall* objects have a high probability to be *knocked* while *short* objects simply get pushed).

The dependencies for the no-explicit-behavior model A_8-AL ($\mathcal{M}_a^2$) is in Table 2. We see from the table that *round* objects have a consistently high probability to generate *disappeared* effect, whereas *edgy* objects do not have such consistency. Furthermore, *tall* objects have consistently low probabilities in obtaining *moved-left, -right, -forward* or *pulled* effects. Almost all effects can be generated on *thin* objects with consistently high probability.

Table 2. The dependence between adjectives and affordances for the model A_8-AL ($\mathcal{M}_a^2$). MR: Moved Right, ML: Moved Left, MF: Moved Forward, P: Pulled, K: Knocked, NC: No Change, G: Grasped, D: Disappeared.

Adjective	MR	ML	MF	P	K	NC	G	D
Edgy	$*$	$*$	$*$	$*$	$+$	$+$	$*$	$*$
Round	$*$	$*$	$*$	$*$	$*$	$+$	$+$	$+$
Short	$*$	$*$	$*$	$*$	$*$	$+$	$+$	$+$
Tall	$-$	$-$	$-$	$-$	$+$	$+$	$+$	$+$
Thin	$*$	$+$	$+$	$+$	$+$	$+$	$+$	$+$
Thick	$*$	$*$	$*$	$*$	$*$	$*$	$+$	$+$

The comparison between the different adjective learning methods is displayed in Table 3, which displays the average 5-fold cross-validation accuracies. We see that the explicit-behavior model (A_{48}-AL) performs better than A_8-AL and SAL models. The reason that A_8-AL is worse than the other methods is eminent in Table 2, where we see that different adjective categories end up with similar descriptor vectors, losing distinctiveness. On the other hand, the A_{48}-AL model that has learned adjectives from the affordances of objects performs better than directly learning SAL model.

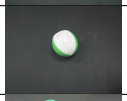
Table 3. Avg. prediction results for the three adjective models in Sect. 3.4.

	A_{48}-AL $\mathcal{M}_a^1$	A_8-AL $\mathcal{M}_a^2$	SAL $\mathcal{M}_a^3$
Edgy-Round	87%	72%	89%
Short-Tall	93%	95%	89%
Thin-Thick	95%	72%	91%

An important point is whether adjectives should include explicit behaviour information (i.e., A_{48}-AL vs. A_8-AL). Theoretically, the performance of these models should converge while one-to-one, unique behavior-to-effect relations dominate the set of known affordances. In such cases, the behavior information would be redundant. On the other hand, with a behavior repertoire that may pose many-to-one-effect mappings, behavior information must be taken into account to obtain more distinguishable adjectives.

Results on Adjectives of Novel Objects Table 4 shows the predicted adjectives from the different models on novel objects. We see that, for adjectives, $\mathcal{M}_a^1$ is better in naming adjectives than $\mathcal{M}_a^2$. For example, $\mathcal{M}_a^2$ mis-classifies object-5 as *edgy*, object-7 as *thin* and object-1 as *thick* whereas $\mathcal{M}_a^1$ correctly names them. On some objects (e.g., object-3), where there are disagreements between

Table 4. Predicted adjectives for novel objects using 3 different models (bold labels denote correct classifications).

ID	Object	A_{48}-AL $\mathcal{M}_a^1$	A_8-AL $\mathcal{M}_a^2$	SAL $\mathcal{M}_a^3$
1		**edgy (54 %)** **short (97 %)** **thin (59 %)**	**edgy (89 %)** **short (91 %)** **thick (52 %)**	**edgy (89 %)** **short (55 %)** **thin (52 %)**
2		**round (77 %)** **short (77 %)** **thin (89 %)**	**round (90 %)** **short (91 %)** **thin (67 %)**	edgy (79 %) **short (42 %)** **thin 67 %**
3		**edgy (63 %)** short (94 %) **thin (96 %)**	round (72 %) short (92 %) **thin (72 %)**	**edgy (64 %)** **tall (67 %)** **thin 84 %**
4		**round (84 %)** **short (98 %)** **thick (91 %)**	edgy (%94) **short (% 87)** thin (% 68)	**round (77 %)** **short (68%)** thin (62 %)
5		**round (84 %)** **short (97 %)** **thick (95 %)**	edgy (% 81) **short (% 93)** **thick (% 59)**	**round (89 %)** **short (67 %)** **thick (58 %)**
6		**edgy (84 %)** **short (98 %)** **thin (92 %)**	**edgy (79 %)** **short (80 %)** **thin (79 %)**	**edgy (79 %)** tall (45 %) thick (62 %)
7		**edgy (62 %)** **short (98 %)** **thick (78 %)**	**edgy (52 %)** **short (93 %)** thin (53 %)	round (84 %) **short (54 %)** **thick (68 %)**
8		**round (72 %)** **short (98 %)** **thick (79 %)**	**round (69 %)** **short (95 %)** **thick (64 %)**	edgy (89 %) **short (67 %)** **thick (52 %)**

the models, correctness cannot be evaluated due to the complexity of the object. If we look at the direct mapping from objects' appearance to adjectives ($\mathcal{M}_a^3$), we see that it misclassifies object-7 as *round*, object-6 as *tall* and objects 2 and 8 as *edgy*.

4.2 Results on Nouns

For the three models trained on nouns (Sect. 3.5), we get the following 5-fold cross-validation accuracies: A_{48}-NL: 87.5%, A_8-NL: 78.1% and SNL: 94%. We see that, unlike the case in adjectives, directly learning the mapping from appearance to nouns performs better than using the affordances of objects. This suggests that the affordances of the objects (used in our experiments) are less descriptive for the noun labels we have used. The dependency results for nouns (similar to the ones in adjectives shown in Tables 1 and 2) are not provided for the sake of space.

Results on Nouns of Novel Objects Table 5 shows the results obtained on novel objects. Unlike the case in adjectives, the simple learner (SNL) significantly outperforms the A_{48}-NL and A_8-NL models. Hence, we conclude that the set of nouns (cup, cylinder, box, ball) we have are more of appearance-based.

5 Conclusion

We proposed linking affordances with nouns and adjectives. Using its interactions with the objects, iCub learned the affordances of the objects and from these, built different types of SVM models for predicting the nouns and the adjectives for the objects. We compared the results of learning nouns and adjectives with classifiers that directly try to link nouns and adjectives with the appearances of objects.

Table 5. Noun prediction for novel objects using 3 different models (see Table 4 for pictures of the objects).

ID	A_{48}-NL	A_8-NL	SNL
1	box (74 %)	cylinder (42 %)	box (97 %)
2	ball (83 %)	ball (44 %)	ball (97 %)
3	cylinder (87 %)	cylinder (39 %)	cylinder (95 %)
4	box (94 %)	cylinder (38 %)	cylinder (86 %)
5	box (89 %)	cylinder (35 %)	box (94 %)
6	cup (89 %)	cylinder (44 %)	box (46 %)
7	box (89 %)	box (32 %)	box (93 %)
8	cup (89 %)	cylinder (44 %)	cup (98 %)

We showed that, by using learned affordances, iCub can predict adjectives with more accuracy than the direct mode. However, for the nouns, direct methods are better. This suggests that a subset of adjectives describing objects in a language can be learned from the affordances of objects. We also demonstrated that explicit behavior information in learning adjectives can provide better representations. It is important to note that these findings are subject to the sensorimotor limitations of the robot, which are maintained by the number and the quality of the behaviors and the properties of the perceptual system. A sample video footage can be viewed at http://youtu.be/DxLFZseasYA

6 Acknowledgements

This work is partially funded by the EU project ROSSI (FP7-ICT-216125) and by TÜBİTAK through projects 109E033 and 111E287. The authors Onur Yuruten, Kadir Uyanik and Asil Bozcuoglu acknowledge the support of TÜBİTAK 2210 scholarship program.

References

1. S. Harnad. The symbol grounding problem. *Physica D: Nonlinear Phenomena*, 42(1-3):335 – 346, 1990.
2. A.M. Borghi. Object concepts and embodiment: Why sensorimotor and cognitive processes cannot be separated. *La nuova critica*, 49(50):90–107, 2007.
3. J.J. Gibson. *The ecological approach to visual perception*. Lawrence Erlbaum, 1986.

4. N. Dag, I. Atil, S. Kalkan, and E. Sahin. Learning affordances for categorizing objects and their properties. In *IEEE ICPR*. IEEE, 2010.
5. I. Atil, N. Dag, Sinan Kalkan, and Erol Şahin. Affordances and emergence of concepts. In *Epigenetic Robotics*, 2010.
6. C. Yu and D. H. Ballard. On the integration of grounding language and learning objects. In *Proc. 19th Int. Conf. on Artifical Intelligence*, AAAI'04, pages 488–493. AAAI Press, 2004.
7. P. Carbonetto and N. de Freitas. Why can't jose read? the problem of learning semantic associations in a robot environment. In *Proc. the HLT-NAACL 2003 workshop on Learning word meaning from non-linguistic data*, pages 54–61, 2003.
8. C. L. Nehaniv J. Saunders and C. Lyon. Robot learning of lexical semantics from sensorimotor interaction and the unrestricted speech of human tutors. In *Proc. 2nd Int. Symp. New Frontiers HumanRobot Interact. ASIB Convent.*, 2010.
9. A. Cangelosi. Evolution of communication and language using signals, symbols, and words. *Evolutionary Computation, IEEE Tran. on*, 5(2):93–101, 2001.
10. A. Cangelosi. Grounding language in action and perception: From cognitive agents to humanoid robots. *Physics of life reviews*, 7(2):139–151, 2010.
11. E. Şahin, M. Çakmak, M.R. Doğar, E. Uğur, and G. Üçoluk. To afford or not to afford: A new formalization of affordances toward affordance-based robot control. *Adaptive Behavior*, 15(4):447–472, 2007.
12. L. Montesano, M. Lopes, A. Bernardino, and J. Santos-Victor. Learning object affordances: From sensory–motor coordination to imitation. *Robotics, IEEE Tran. on*, 24(1):15–26, 2008.
13. A. Stoytchev. Learning the affordances of tools using a behavior-grounded approach. *Towards Affordance-Based Robot Control*, pages 140–158, 2008.
14. D. Kraft, N. Pugeault, E. Baseski, M. Popovic, D. Kragic, S. Kalkan, F. Wörgötter, and N. Krüger. Birth of the object: Detection of objectness and extraction of object shape through object action complexes. *International Journal of Humanoid Robotics*, 5(2):247–265, 2008.
15. E. Ugur and E. Şahin. Traversability: A case study for learning and perceiving affordances in robots. *Adaptive Behavior*, 18(3-4):258–284, 2010.
16. P. Fitzpatrick, G. Metta, L. Natale, S. Rao, and G. Sandini. Learning about objects through action - initial steps towards artificial cognition. In *IEEE ICRA*, 2003.
17. R. Detry, D. Kraft, A.G. Buch, N. Kruger, and J. Piater. Refining grasp affordance models by experience. In *IEEE ICRA*, pages 2287 –2293, 2010.
18. E. Ugur, E. Oztop, and E. Şahin. Goal emulation and planning in perceptual space using learned affordances. *Robotics and Autonomous Systems*, 59(7-8), 2011.
19. E. Ugur, E. Şahin, and E. Oztop. Affordance learning from range data for multi-step planning. *Int. Conf. on Epigenetic Robotics*, 2009.
20. L. Montesano, M. Lopes, A. Bernardino, and J. Santos-Victor. A computational model of object affordances. In *Advances In Cognitive Systems*. IET, 2009.
21. J. Sinapov and A. Stoytchev. Learning and generalization of behavior-grounded tool affordances. In *Development and Learning, 2007. ICDL 2007. IEEE 6th International Conference on*, pages 19 –24, july 2007.
22. J. Sinapov and A. Stoytchev. Detecting the functional similarities between tools using a hierarchical representation of outcomes. In *Development and Learning, 2008. ICDL 2008. 7th IEEE International Conference on*, pages 91 –96, aug. 2008.
23. L. Montesano, M. Lopes, A. Bernardino, and J. Santos-Victor. Learning object affordances: From sensory–motor coordination to imitation. *Robotics, IEEE Tran. on*, 24(1):15 –26, feb. 2008.

24. Luis Montesano, Manuel Lopes, Francisco Melo, Alexandre Bernardino, and Jos Santos-Victor. A computational model of object affordances. *Advances in Cognitive Systems*, 54:258, 2009.
25. R.B. Rusu and S. Cousins. 3d is here: Point cloud library (pcl). In *IEEE ICRA*, pages 1–4. IEEE, 2011.
26. K. Kira and L. A. Rendell. A practical approach to feature selection. In *Proc. 9th Int. Workshop on Machine Learning*, pages 249–256, 1992.
27. A. K. Qin and P. N. Suganthan. Robust growing neural gas algorithm with application in cluster analysis. *Neural Networks*, 17(8-9):1135–1148, 2004.