From Language to Action: Extraction and Disambiguation of Affordances in ModelAct

Irene Russo[1(✉)] and Livio Robaldo[2]

[1] ILC CNR Pisa, Pisa, Italy
irene.russo@ilc.cnr.it
[2] Dipartimento di Informatica, Universita' di Torino, Torino, Italy
robaldo@di.unito.it

Abstract. In this paper we focus on how information about concrete actions performed on food should be provided to IoT devices in terms of affordances extracted from corpora. Natural language processing has a role in defining which kind of knowledge devices interacting with machines and appliances should handle when humans send requests through natural language interfaces.

We propose a model for the extraction of affordances of food from corpora and their role in sequences of procedural (sub)actions. The food processor of the future can find helpful this knowledge to interact with users suggesting alternatives in food processing in recipes steps and basic reasoning about preconditions and consequences in making meals.

Keywords: Affordances · Natural language processing · Verbs' disambiguation

1 One Verb, More Action Types

The way human beings talk about basic and concrete actions can conceal the evidence that a verb like *to open* refers to different procedural sequences of subactions, dependent on the features of the objects: *open the window* is akin to *open the door*, but very different with respect to *open the nut*. In fact no one to one correspondence can be established between verbs and action concepts, causing huge problems for natural language understanding. Language processing required in tasks such as automatic translation and human machine interaction is difficult to achieve when reference to concrete actions is concerned, since the language labels used to indicate action concepts may pick up many different action types and each language shows a different variation potential.

This is one of the crucial reasons for which natural language instructions are not clear for a machine in an open domain communication scenario. The explicit understanding of the language conditions enabling the interpretation of simple sentences requires both a solid knowledge of the range of variation of verbs referring to action and a clear definition of the conceptual structures embodied by each language. Without these premises the overall language processing problem constituted by the linguistic categorization of action cannot be grounded. These issues

R. Giaffreda et al. (Eds.): IoT360 2014, Part I, LNICST 150, pp. 93–96, 2015.
DOI: 10.1007/978-3-319-19656-5_15

are relevant for the IoT paradigm, since this kind of knowledge should be provided to devices interacting with machines and appliances and with humans through natural language interfaces. How this knowledge should be represented? How can it be extracted from language? These are some of the questions that ModelAct (modelact.lablita.it), a FIRB project funded by MIUR, wants to address. ModelAct aims at providing a cognitive-based model for human categorization of actions. Since the same verb like *to open* can denote different sequences of (sub)actions - different sequences of body movements performed by an agent - it is essential for devices to disambiguate between these denotations and to reason on them, processing linguistic instructions with the awareness of movements implied. In particular, the definition of a natural language disambiguation module able to guess which sequences of (sub)actions is required to open a box instead of a nut is important. ModelAct exploits the ImagAct ontology (www.imagact.it), which identifies the action categories by means of prototypical scenes. Scenes are short 3D videos, they are language independent and their semantic value can be naturally understood, allowing the appropriate encoding in different languages (English, Italian, Spanish and Mandarin Chinese).

In this context, the notion of affordance can bridge the gap between ecological psychology studies and AI approaches to knowledge management. Affordances have been theorized by [3] as the possibilities for action that every environmental object offers. They are different and unique for every living being, in that they are not strictly related to objective properties; rather, they lie on possible ways in which living beings can interact with objects themselves. Affordance is defined as the quality of an object that enables an action: it concerns the relation between a perceptual property of the object and what an agent can do with it. In embodied robotics affordances are often conceptualized as a quality of an object, or of an environment, which allow an artificial agent to perform an action. Environment for artificial agents can be seen as a source of information that help the robot in performing actions, thus reducing the complexity of representation and reasoning [4], exactly in the same way an object showing such properties that afford sitting (e.g. a chair) will help us understand how we can use it (sitting).

Conceptual information concerning objects affordances can be acquired through language. Our focus is on verb-direct object pairs as the linguistic realizations of the relations between actions performed by an agent on objects. We distinguish between general affordances that can be potentially displayed by every objects - more generic verbs like *to take*, *to bring* etc. - and specific affordances as canonical/peculiar activities a specific object is involved in, like *open* for *bottle*. Affordance verbs as verbs that select a distinctive action for a specific object can be discovered through statistical measures in corpora [6].

2 A Possible Application: Food's Affordances for Food Processors

ModelAct analyses data from ImagAct, a previous project that, through manual annotation, clustered sets of sentences distinguishing senses of a verb in terms of

(sub)sequences of movements. 800 verbs, both for Italian and English, have been studied and more than 3000 objects' mentions are included in the dataset. In this paper we want to focus on the affordances of a specific class of objects, e.g. the one that are involved in making food. In the near future a food processor can be a complex appliance connected to our smartphones, able to make the best use of the available ingredients, assembling them following steps in a recipe that we choose [1]. Information from sensors monitoring the temperature, the weight and the moisture level of the food can interact with information in recipe's steps. It will be possible to monitor the processor from a smartphone and food's affordances can suggest alternatives at a certain stage (e.g. "Now they are boiled. Do you prefer your zucchini as soup or in puree?"). For this reason, a knowledge base covering a wide range of possibilities and a basic set of reasoning rules on them should be provided to devices; acquiring it through language means helping in the settings of naturale language interfaces interacting with devices.

From ImagAct we can extract basic affordances for food nouns (see Fig. 1):

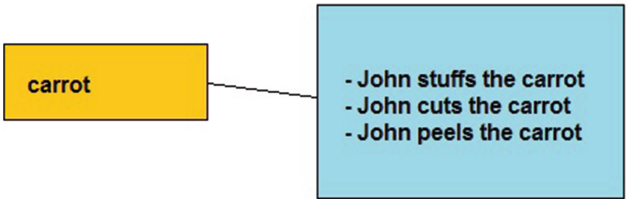

Fig. 1. ImagAct affordances for carrot

However, this knowledge base should be improved including more possibilities. From a corpus of more than 5000 recipes crawled from instructables.com and parsed with Stanford CoreNLP tools [5] we can add valuable knowledge to our model, discovering affordances and extracting sequences of actions, expressing the kind of action(s) that precede or follow the affordance we are focusing on.

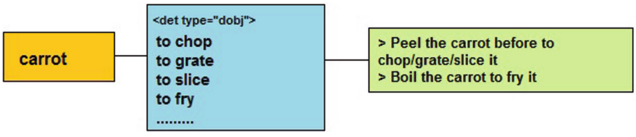

Fig. 2. Corpus extracted affordances for carrot

With respect to the knowledge we have in ImagAct, we can add that a carrot can be the patient of other actions, both generic and specific affordances (see Fig. 2). Intersecting the list of nouns in ImagAct that have food as hypernyms in WordNet 3.0 [2] and selecting the ones with frequency more than 100 in our

recipes corpus, we obtained a list of 78 nouns. For each of them basic affordances, local affordances sequences and global affordances sequences can be extracted. With basic affordances we mean the full list of action verbs the noun can be direct object of (e.g. for carrot *to cut, to chop, to grate, to slice*; for milk *to add, to pour, to stir, to whisk*; for sugar *to add, to burn, to dissolve, to mix, to sprinkle*). With local affordances sequences we mean a pair of action verbs that in the same sentence are coordinated and have the noun as direct object (e.g. *Take the carrot and dip it* or *Add a pinch of sugar and stir it*). With global affordances sequences we mean a temporal ordered list of action verbs that have the noun as direct object in the context of the same recipe (e.g. *Take the carrot, clean the carrot, shave the carrot on the salad*). From recipes it's possible to locate each action in a sequence, enabling reasoning on actions sequences in the future in terms of preconditions and consequences (e.g. a carrot should boiled before to make a pure). This knowledge is not static but deeply influence by saliency and probability; with a different corpus (e.g. a corpus composed by just Asian recipes) it could change radically, implying that the information can be adapted for special needs.

References

1. Beetz, M., Klank, U., Kresse, I., Maldonado, A., Msenlechner, L., Pangercic, D., Rhr, T., Tenorth, M.: Robotic roommates making pancakes. In: 2011 11th IEEE-RAS International Conference on Humanoid Robots (2011)
2. Fellbaum, C. (ed.): WordNet: An Electronic Lexical Database. MIT Press, Cambridge (1998)
3. Gibson, J.J.: The Ecological Approach to Visual Perception. Houghton Mifflin, Boston (1979)
4. Horton, T.E., Chakraborty, A., St. Amant R.: Affordances for robots: a brief survey. Avant **3**(2), 70–84 (2012)
5. Manning, C.D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S.J., McClosky, D.: The Stanford CoreNLP natural language processing toolkit. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pp. 55–60 (2014)
6. Russo, I., De Felice, I., Frontini, F., Khan, F., Monachini, M.: (Fore)seeing actions in objects. Acquiring distinctive affordances from language. In: Proceedings of the 10th International Workshop on Natural Language Processing and Cognitive Science - NLPCS 2013, pp. 151–161 (2013)