Towards Understanding Situated Text: *Concept Labeling and Weak Supervision*

Antoine Bordes[*], Nicolas Usunier[*], Ronan Collobert[†], Jason Weston[†]
∗ LIP6, Université Paris 6, Paris, France.
†NEC Laboratories America, Princeton, NJ.

Much of the focus of the natural language processing community lies in solving syntactic or semantic tasks with the aid of sophisticated machine learning algorithms and the encoding of linguistic prior knowledge. One of the most important features of natural languages is that their real-world use (as a tool for humans) is to communicate something about our physical reality or metaphysical considerations of that reality. This is strong prior knowledge that is simply ignored in most current systems. For example, in current parsing systems there is no allowance for the ability to disambiguate a sentence given knowledge of the physical reality of the world. If one happened to know that Bill owned a telescope while John did not, then this should affect parsing decisions given the sentence "John saw Bill in the park with his telescope." Similarly, one can improve disambiguation of the word bank in "John went to the bank" if one happens to know whether John is out for a walk in the countryside or in the city. In summary, many human disambiguation decisions are in fact based on whether the current sentence agrees well with one's current *world model*.

Concept Labeling In this work, we propose a general framework for using *world knowledge* called the *concept labeling* task. The knowledge we consider can be viewed as a database of physical entities existing in the world (e.g. people, locations or objects) as well as abstract concepts, and relations between them. Our task thus consists of labeling each word of a sentence with its corresponding *concept* from the database as is illustrated in Figure 1.

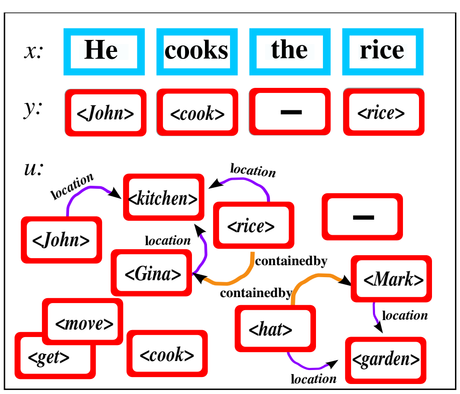

Figure 1: **An example of a training triple for *concept labeling.*** The universe u contains all the known concepts that exist, and their relations. The label y consists of the concepts that each word in the sentence x refers to, including the empty concept "-".

The solution to this task does not provide a full semantic interpretation of a sentence, but we believe is a first step towards that goal. Indeed, in many cases, the meaning of a sentence can only be uncovered after knowing exactly which concepts, e.g. which unique objects in the world, are involved. If one wants to interpret "He passed the exam", one has to infer not only that "He" refers to a "John", and "exam" to a school test, but also exactly which "John" and which test it was. In that sense, concept labeling is more general than traditional tasks like word-sense disambiguation, co-reference resolution, and named-entity recognition, and can be seen as a unification of them.

Learning We then go on to propose a tractable algorithm for this task inspired by the LaSO algorithm [2]. It can *learn* to use world knowledge and linguistic content of a sentence seamlessly *without the use of any hand-crafted rules or features.*

The experimental evaluation of our algorithm uses a novel simulation procedure to generate natural language and concept label pairs: the simulation generates an evolving world of interactions between actors, objects and locations, together with sentences describing the successive evolutions. This provides large labeled data sets with ambiguous sentences without any human intervention. The results demonstrate that our algorithm can *learn* to use world knowledge for word disambiguation and reference resolution when standard methods cannot.

Related Works Our work concerns learning the connection between natural language and another symbolic system. It is referred to as *grounded* (or *situated*) *language processing* [6] in the literature.

Some of the earliest works that used world knowledge to improve linguistic processing involved hand-coded parsing and no learning at all, perhaps the most famous being situated in blocks world [9]. More recent works on grounded language acquisition have focused on *learning* to match language with some other representation. Grounding text with a visual representation also in a blocks-type world was tackled [10, 3]. Other works also use visual grounding [8, 12, 4, 1], or a representation of the intended meaning in some formal language [13, 5, 11]. Example applications of such grounding include using the multimodal input to improve clustering [7], word-sense disambiguation [1, 4], or to make the machine predict one representation given the other [13]. Although these learning systems can deal with some ambiguities in natural language, the representations that they consider, to the best of our knowledge, do not take into account the changing environment.

Extensions We conclude by describing two extensions to our framework: (i) the supervised signal is much weaker and only consists of observations of the world and speech within that world; and (ii) we explore the ability of a learning algorithm to understand sentences and update its own internal model of the world based on its understanding.

References

[1] K. Barnard and M. Johnson. Word Sense Disambiguation with Pictures. *Artificial Intelligence*, 167(1-2), 2005.

[2] H. Daumé III and D. Marcu. Learning as Search Optimization: Approximate Large Margin Methods for Structured Prediction. In *ICML 2005*, 2005.

[3] J. Feldman, G. Lakoff, D. Bailey, S. Narayanan, T. Regier, and A. Stolcke. L 0-The first five years of an automated language acquisition project. *Artificial Intelligence Review*, 10(1):103–129, 1996.

[4] M. Fleischman and D. Roy. Intentional Context in Situated Language Learning. In *9th Conference on Computational Natural Language Learning*, 2005.

[5] R.J. Kate and R.J. Mooney. Learning Language Semantics from Ambiguous Supervision. In *Proc. of the 22th National Conference on Artificial Intelligence*, 2007.

[6] D. Roy and E. Reiter. Connecting Language to the World. *Artificial Intelligence*, 167(1-2):1–12, 2005.

[7] J.M. Siskind. Grounding Language in Perception. *Artificial Intelligence Review*, 8(5):371–391, 1994.

[8] R. Thibadeau. Artificial perception of actions. *Cognitive Science*, 10(2):117–149, 1986.

[9] T. Winograd, M.G. Barbour, and C.R. Stocking. *Understanding natural language.* Academic Press New York, 1972.

[10] P.H. Winston. The psychology of computer vision. *Pattern Recognition*, 8(3):193–193, 1976.

[11] Y.W. Wong and R. Mooney. Learning Synchronous Grammars for Semantic Parsing with Lambda Calculus. In *Annual Meeting - ACL*, volume 45, 2007.

[12] C. Yu and D.H. Ballard. On the Integration of Grounding Language and Learning Objects. In *Proc. of the 19th National Conference on AI*, 2004.

[13] L.S. Zettlemoyer and M. Collins. Learning to Map sentences to Logical Form: Structured Classification with Probabilistic Categorial Grammars. In *UAI 2005*, volume 5, 2005.

Topic: learning algorithms
Preference: oral