An object deserves more than a single name

Anonymous ACL submission

Abstract

1 Introduction

Categorizing and naming real-world objects is a fundamental ability of human cognition and communication. In NLP, object naming is a core phenomenon which is part of virtually every Language & Vision task. Since objects are often simultaneously a member of multiple categories (e.g., a young *beagle* is at once a DOG, a BEAGLE, an ANIMAL, a PUPPY etc.), the act of naming an object amounts to that of selecting a lexical concept for it among the various potential alternatives.

Research on categorization and language production has used object naming as a basic paradigm for investigating the processes that underly formation, organization and retrieval of concepts in the human mind (**?**) *//sz: cite more here//*. Research in computer vision has focused on automatic object recognition where, recently, powerful models have been developed that classify visual objects in real-world images into thousands of different categories Szegedy et al. (2015). In NLP (L&V), however, research on object naming is relatively scarce. While there has been an explosion of interest in various, sometimes complex, L&V tasks, ranging from image captioning (**???**), referring expression resolution and generation (Kazemzadeh et al., 2014; **?**; **?**), to multi-modal summarization or visual dialogue (**??**), there has hardly been any work looking at linguistic questions involved in object naming. As the starting point of this work, we argue that existing resources in L&V constitute an excellent basis for empirical, large-scale investigations into object naming. At the same time, we show that we need to systematically extend them and collect naming data in a more ??? way, so as to arrive at a solid empirical approach.

Shortcomings of existing resources:

- most corpora record a single label or a small set of descriptions for each object
- no systematic inventory of object labels
- existing taxonomies (WordNet) are not really useful

2 Related Work

Cognition: Concepts and categorization Seminal work on concepts by Rosch suggests that object names typically exhibit a preferred level of specificity called the **entry-level**. This typically corresponds to an intermediate level of specificity, i.e., **basic level** (e.g, *bird*, *car*) (**?**), as opposed to more generic (i.e., **super-level**; e.g., *animal*, *vehicle*) or specific categories (i.e., **sub-level**; e.g., *sparrow*, *convertible*). However, less prototypical members of basic-level categories tend to be instead identified with sub-level categories (e.g., a PENGUIN is typically called a *penguin* and not a *bird*) (**?**). While the traditional notion of entry-level categories suggests that objects tend to be named by a *single* preferred concept, research on pragmatics has found that speakers are flexible in their choice of the level of specificity. Scenarios where multiple objects (of the same category) are present induce a pressure for generating names which uniquely identify the target (**?**), such that sub-level names can be systematically elicited in these cases (**?**)(Graf et al., 2016).

Vision: Object Recognition State-of-the-art computer vision systems are able to classify images into thousands of different categories (e.g. Szegedy et al. (2015)). These object recognition systems are now widely used in vision & language research. Nevertheless, the way the treat object recognition is conceptually very simple (if not to say, naive): standard object classification schemes are inherently "flat", and treat object labels as mutually exclusive (Deng et al., 2014), ignoring all kinds of linguistic relations between these labels and ignoring the fact that an object can easily be an instance of several categories.

Vision & language: Naming and Referring Ordonez et al. (2016) have studied the problem of deriving appropriate object names, or so-called entry-level categories, from the output of an object recognizer. Their approach focusses on linking abstract object categories in ImageNet to actual words via translation procedures that e.g. involve corpus frequencies. Zarrieß and Schlangen (2017) learn a model of object naming on a corpus of referring expressions paired with objects in real-world images, but focus on combining visual and distributional information and on zero-shot learning. Thus, object naming is an important task for referring expression generation, though most research in this area has focussed on content and attribute selection (Kazemzadeh et al., 2014; Gkatzia et al., 2015; Zarrieß and Schlangen, 2016; Mao et al., 2015).

3 Data Collection

describe the YouNameIt task here

3.1 Materials

describe sampling of images, category selection

3.2 Procedure

describe the crowdsourcing set-up and the task

3.3 Data

give an overview of the collected data

4 Analysis

4.1 Agreement, basic-level and entry-level names

analyse data mainly from Phase 0

- to what extent do people agree when their task is to give the most straightforward name they can think of to a visual object?
- is the level of agreement the same for all categories?
- how specific are the most familiar names? link names to WordNet, show that WordNet might not be ideal to assess specificity
- how does agreement evolve in the later rounds of the game? (when people have to avoid taboo names), does agreement increase as the set of names becomes more narrow, or does agreement decrease as people do creative, clever, unexpected things?

4.2 Cases of disagreement

when and why do people give different names to the same object? this will probably happen in phase 0, and even more so in the later round *//sz: this is what I expect//*

- analyse naming disagreement using WordNet, how do names for the same object relate to each other according to WordNet?
- can we identify instances of cross-classification? so objects that are systematically part of several classes (e.g. cake/dessert)
- we might need to do some manual annotation here and try to carefully describe the phenomena

4.3 Taxonomic relations

can we elicit natural sub-ordinate, super-ordinate concepts?

5 Model

5.1 Model 1

Train a simple naming model (classifiers) on original VisualGenome names. Test it on our data. How well does the model predict the most familiar name and the set of available names?

5.2 Model 2

Train a simple naming model (classifiers) on our data, maybe combined with VisualGenome data. How well does it work?

5.3 Model 3

can we model taxonomic/conceptual knowledge more directly? induce a taxonomy? or a multimodal space for objects + names?

6 Conclusion

We have presented a systematic, large-scale study on object naming with real-world images and crowdsourced data.

References

Jia Deng, Nan Ding, Yangqing Jia, Andrea Frome, Kevin Murphy, Samy Bengio, Yuan Li, Hartmut Neven, and Hartwig Adam. 2014. Large-scale object classification using label relation graphs. In *European Conference on Computer Vision*. Springer, pages 48–64.

Dimitra Gkatzia, Verena Rieser, Phil Bartie, and William Mackaness. 2015. From the virtual to the realworld: Referring to objects in real-world spatial scenes. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Lisbon, Portugal, pages 1936–1942. http://aclweb.org/anthology/D15-1224.

Caroline Graf, Judith Degen, Robert XD Hawkins, and Noah D Goodman. 2016. Animal, dog, or dalmatian? level of abstraction in nominal referring expressions. In *Proceedings of the 38th annual conference of the Cognitive Science Society*. Cognitive Science Society.

Sahar Kazemzadeh, Vicente Ordonez, Mark Matten, and Tamara L Berg. 2014. ReferItGame: Referring to Objects in Photographs of Natural Scenes. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)*. Doha, Qatar, pages 787–798.

Junhua Mao, Jonathan Huang, Alexander Toshev, Oana Camburu, Alan L. Yuille, and Kevin Murphy. 2015. Generation and comprehension of unambiguous object descriptions. *ArXiv / CoRR* abs/1511.02283. http://arxiv.org/abs/1511.02283.

Vicente Ordonez, Wei Liu, Jia Deng, Yejin Choi, Alexander C. Berg, and Tamara L. Berg. 2016. Learning to name objects. *Commun. ACM* 59(3):108–115.

Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. 2015. Going deeper with convolutions. In *CVPR 2015*. Boston, MA, USA.

Sina Zarrieß and David Schlangen. 2016. Easy things first: Installments improve referring expression generation for objects in photographs. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Berlin, Germany, pages 610–620. http://www.aclweb.org/anthology/P16-1058.

Sina Zarrieß and David Schlangen. 2017. Obtaining referential word meanings from visual and distributional information: Experiments on object naming. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Vancouver, Canada, pages 243–254. http://aclweb.org/anthology/P17-1023.