Instructions for EMNLP-IJCNLP 2019 Proceedings

Anonymous EMNLP-IJCNLP submission

Abstract

1 Introduction

The real-world objects that we interact with in our every-day life can be categorized into many thousands and maybe millions of categories. And even a single object can be member of many categories, i.e. at different taxonomical levels or in different parts of a taxonomy. For instance, both objects in Figure 1 are at once instances of CAKE, CHEESECAKE, DESSERT, SWEET, PASTRY, FOOD etc. Hence, when speakers name objects, e.g. when referring, they have to select a lexical item from a complex network of concepts and competing lexical alternatives.

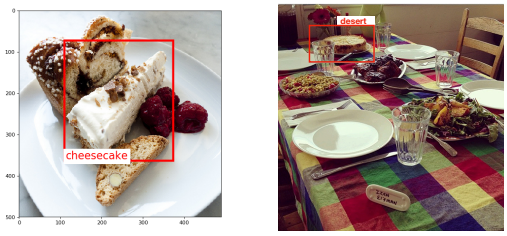

Figure 1: Two objects of the same type of cake, with different names in VisualGenome

To date, research in NLP has surprisingly little to say about object naming, despite the fact that there has been a recent explosion of interest in various, and even complex, language & vision tasks ranging from image captioning (**???**) to e.g. visual dialogue (**??**). In contrast, closely related areas, such as computer vision and cognitive science, have investigated very related tasks in quite some depth: object recognition systems developed in the area of computer vision are now able to classify images into thousands of different categories (e.g. **?**). Furthermore, work on concepts, following the seminal work by Rosch, suggests that objects are typically conceptualized at a preferred level of specificity called the **entry-level**. Psycho-linguistic studies have been able to support this theory based on collections of so-called object naming norms.

This paper aims at addressing the genuinely linguistic questions revolving around the phenomenon of object naming by (i) presenting a collection of high-quality, large-scale naming data, and (ii) analysis methods for this data and (iii) a first baseline model that accounts for the semantic flexibility of names for objects in real-world images. From computer vision, we borrow the idea of modeling realistic visual objects in realistic scenes (real-world images), but go beyond the simplistic assumption that object names correspond to unambiguous labels in a flat classification scheme (with no conceptual relations between the labels). From psycholinguistics, we borrow the idea of eliciting natural, representative naming data from many subjects, but go beyond using artificial, highly stylized objects.

2 Related Work

3 Analysis

3.1 Agreement

We compute the following agreement measures:

- **top %**: for each object, we calculate the relative frequency of the most common name, and then average over all objects
- **SD %**: for each object, we calculate the Snodgrass agreement measure, and then average over all objects
- **=VG**: the proportion of objects where the most frequent name coincides with the name annotated in VisualGenome

	all synsets			max synset				min synset			
domain	% top	SD	=VG	id	% top	SD	=VG	id	% top	SD	=VG
person	0.52	2.14	0.50	professional.n.01	0.61	2.02	0.20	athlete.n.01	0.34	2.67	0.36
tableware	0.52	1.92	0.40	crockery.n.01	0.52	1.92	0.40	crockery.n.01	0.52	1.92	0.40
clothing	0.64	1.59	0.70	neckwear.n.01	0.79	0.91	0.77	footwear.n.01	0.47	2.55	0.40
instruments	0.66	1.52	0.79	furnishing.n.02	0.67	1.50	0.80	kitchen_utensil.n.01	0.60	1.85	0.56
solid food	0.67	1.43	0.56	baked_goods.n.01	0.67	1.43	0.56	baked_goods.n.01	0.67	1.43	0.56
structure	0.67	1.55	0.73	bridge.n.01	0.75	1.21	0.87	place_of_worship.n.01	0.46	2.26	0.08
vehicle	0.72	1.13	0.71	train.n.01	0.93	0.43	0.99	aircraft.n.01	0.52	1.50	0.41
food, nutrient	0.72	1.27	0.68	edible_fruit.n.01	0.80	0.89	0.79	vegetable.n.01	0.52	1.99	0.15
plants	0.79	0.86	0.73	flower.n.01	0.79	0.86	0.73	flower.n.01	0.79	0.86	0.73
ware	0.82	0.96	0.94	cutlery.n.02	0.82	0.96	0.94	cutlery.n.02	0.82	0.96	0.94
tool	0.86	0.73	0.94	tool.n.01	0.86	0.73	0.94	tool.n.01	0.86	0.73	0.94
animal	0.91	0.43	0.94	feline.n.01	0.95	0.29	0.99	fish.n.01	0.39	2.53	0.55
all	0.70	1.35	0.73								

Table 1: Agreement in object names for objects of different domains, if applicable, synsets with maximal and minimal agreement (top %) are shown

Table 1 shows that, overall, our annotators achieve a fair amount of agreement in the object naming choices. The domain where annotators agree most is the animal domain, which, interestingly, happens to be the domain that has been mostly discussed in the object naming literature. *//sz: ... much more to say//*

Why is naming more flexible in certain domains than in others?

3.2 Disagreement

Why and when do speakers disagree in their object naming choices. We first propose a qualitative analysis here.

- **synonyms**: e.g. aircraft vs. airplane
- **taxonomic levels**: e.g. man vs. person
- **cross-classification**: e.g. chicken vs. dinner
- **conceptual disagreement**: e.g. swan vs. goose

Can we tease these types of disagreements apart automatically, using WordNet?