Horse or pony?
Visual typicality and lexical frequency affect variability in object naming

Eleonora Gualdoni[*] **Andreas Mädebach**[*] **Thomas Brochhagen**[*] **Gemma Boleda**[*] [†]

[*]Universitat Pompeu Fabra
[†]ICREA
{firstname.surname}@upf.edu

1 Introduction

We successfully refer to objects in most interactions, and in particular choose a word in our lexicon to name them (e.g., "horse" or "pony" in Figure 1A). This requires complex cognitive processing that allows us to link the properties of the object with our lexicon. Moreover, the mapping between our representation of the object and the lexicon is not one-to-one, and often different names can be used for the same object. In the present study, we explore factors that affect naming variation for visually presented objects. We focus on two variables: visual typicality of the image and lexical frequency of the name. The latter serves as a proxy for ease of lexical access. By analysing objects in realistic scenes, we explore the role of typicality not only of the object (as was done previously), but also of the visual context.

Previous psycholinguistic studies focused on relatively small datasets and simple images of isolated objects (e.g., Snodgrass and Vanderwart, 1980). We expand on this by analysing a large object naming dataset collected in the context of Language&Vision research (Silberer et al., 2020): ManyNames[1]. ManyNames provides up to 36 naming annotations for 25K objects in realistic scenes. We will call the most frequently annotated name *top name* ("horse" in Fig. 1A), and the second most frequently annotated name *alternative name* ("pony" in Fig. 1A). Previous work only took top names into consideration, and used subjective ratings of visual typicality, operationalising them as the similarity between a given visual object and the prototypical mental representation associated with this object's top name. We include alternative names in the analysis, and define a computational procedure to assess visual typicality of objects and contexts (see Methods section below).

Our measure of naming variation is agreement on the top name. We do so because there is a direct relationship between naming variation and agreement on the top name: higher agreement indicates lower variation, and vice versa.

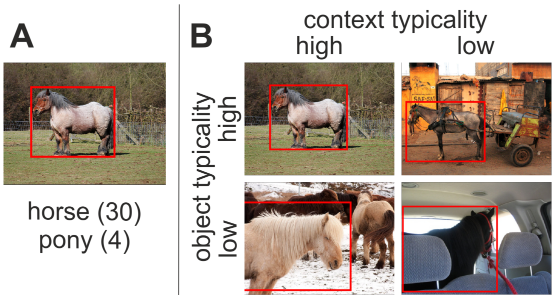

Figure 1: A: Example image with annotated names and response count. B: Illustration of target and context typicality variation for the top name "horse".

Based on previous studies, we expect higher name agreement with increasing typicality of the object for the top name (e.g., Snodgrass and Vanderwart, 1980). The analysis of context typicality is more exploratory. Previous work has shown that placing other objects than the target in the context affects naming (Graf et al., 2016); however, more general aspects of context (including whether the object is in, say, a beach or a home) have not been studied. We can generally extend our prediction for object typicality to the visual context, expecting higher agreement for objects in more typical visual contexts. However, effects may be less pronounced: Contexts are likely less informative for a given name than the object itself. When it comes to frequency, we also expect higher agreement for more frequent top names.

For alternative names, we hypothesize opposite effects compared to top names: Higher object or context typicality for an alternative name, as well

[1] Available at https://github.com/amore-upf/manynames.

as higher frequency of an alternative name, should result in lower name agreement, due to increased competition between the alternative name and the top name when choosing a name. Again, the effect of context typicality may be less pronounced, because context prototypes may often be more similar across candidate names than object prototypes.

2 Methods

Data We analyse naming data for 16K images from ManyNames – those that had at least two names. To estimate visual prototypes for a given name, we select 30-500 objects with that name from VisualGenome (Krishna et al., 2016), ensuring that these objects were not included in ManyNames (VisualGenome is the dataset from which the ManyNames images were selected, and also contains object names). We average the vectorial representations of these objects, obtained with the bottom-up-attention model by Anderson et al. (2018). This average representation is the visual prototype for a name. We compute object typicality for a given ManyNames object as the cosine similarity between the object's features –which we obtain in the same way as for VisualGenome objects– and the prototype of its names; this results in two typicality estimates, one for the top name, one for the alternative name.

We obtain context prototypes by averaging the features of all context objects (as detected by Anderson et al., 2018). Note that "context objects" includes what people would commonly call an object (like a cat or a table), but also background elements like patches of grass or sky. Anderson et al. 2018 use this procedure as a representation of the global context of an object, which is then used by an image captioning model. Analogously, we here use it to represent the context in which an object appears. As with object typicality, we compute context typicality by using the cosine similarity between the features of the object's context and the context prototypes of its names. Frequency estimates for the names are from a subtitle corpus of American English (Brysbaert and New, 2009).

Statistical Model We fit a binomial mixed-effects model with name agreement on the top name (in %) as the outcome variable and fixed effects for standardised object typicality, context typicality, and log-frequency, each relating to the top name and the alternative name. Top names and alternative names are treated as random factors with corresponding random slopes for all predictors.

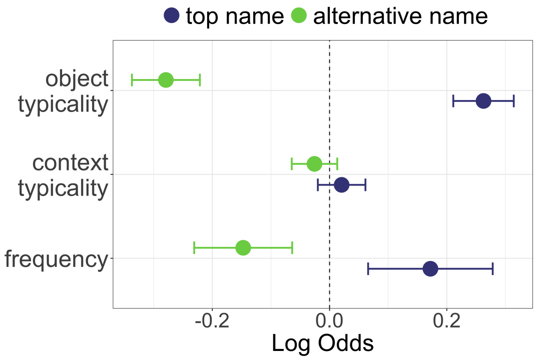

Figure 2: Fixed effect estimates. Error bars reflect the 95% CI. Positive vs. negative estimates show, respectively, the increase and decrease in name agreement for 1 *SD* increase in the predictor variable.

3 Results and Discussion

Fixed effect estimates are shown in Figure 2. Object typicality for top name and second name affect agreement on the top name as we expected: Name agreement is higher the more typical an object is for its top name, and lower the more typical it is for the alternative name. A similar pattern is found for frequency: higher frequency of the top name relates to higher name agreement, whereas higher frequency of the alternative name relates to lower name agreement. In other words, people tend to choose the same name for an object when the object is very typical for that name, or that name is very frequent. In contrast, naming variation increases the more typical the object is for an alternative name, or when the alternative name is relatively easy to access.

However, we find no clear fixed effect for context typicality. That being said, including context typicality as a random effect significantly improves the model fit. This suggests meaningful variation of this effect across names. One reason for this meaningful variation may be that different causes of naming variation, e.g. perceptual ambiguity ("jaguar/leopard") *vs* categorical ambiguity ("mug/cup") *vs* the availability of cross-classifying alternatives ("man/teacher"), interact differently with context typicality effects. Moreover, this issue may also be related to differences in the *informativity* of context prototypes: relatively unspecific names, like "man/woman", likely do not have particularly informative context prototypes because they appear in a diverse array of scenes. This contrasts to names like "teacher/skier", for which the scene setting may be more diagnostic (e.g., a classroom or a snowy outdoor environment). Further research is needed to look into these factors, as well as to assess the sensitivity of our computational quantification of context typicality.

In sum, our large scale computational analysis strengthens previous findings about object naming and expands the general picture, suggesting that different candidate names jointly affect name agreement: Visual and lexical characteristics relating to name candidates beyond the top name are informative for predicting variability in object naming. On a methodological level, our results demonstrate the potential of using large scale datasets with realistic images in conjunction with computational methods to inform models of human object naming.

Acknowledgements

We thank Carina Silberer for advice regarding Computer Vision models, and the SCiL anonymous reviewers. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 715154). This paper reflects the authors' view only, and the EU is not responsible for any use that may be made of the information it contains.

References

Peter Anderson, Xiaodong He, Chris Buehler, Damien Teney, Mark Johnson, Stephen Gould, and Lei Zhang. 2018. Bottom-up and top-down attention for image captioning and visual question answering.

Marc Brysbaert and Boris New. 2009. Moving beyond kucera and francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for american english. *Behavior Research Methods*, 41:977–90.

Caroline Graf, Judith Degen, Robert X D 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*, pages 2261–2266, Austin, TX. Cognitive Science Society.

Ranjay Krishna, Yuke Zhu, Oliver Groth, Justin Johnson, Kenji Hata, Joshua Kravitz, Stephanie Chen, Yannis Kalantidis, Li-Jia Li, David A. Shamma, Michael S. Bernstein, and Fei-Fei Li. 2016. Visual genome: Connecting language and vision using crowdsourced dense image annotations.

Carina Silberer, Sina Zarrieß, and Gemma Boleda. 2020. Object naming in language and vision: A survey and a new dataset. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 5792–5801, Marseille, France. European Language Resources Association.

Joan G. Snodgrass and Mary Vanderwart. 1980. A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, 6(2):174–215.