Effects of Task and Visual Context on Referring Expressions using Natural Scenes

Andreas Mädebach (a.maedebach@gmail.com)
Universitat Pompeu Fabra, Barcelona, Spain

Ekaterina Torubarova (ek.torubarova@gmail.com)
Universitat Pompeu Fabra, Barcelona, Spain

Eleonora Gualdoni (eleonora.gualdoni@upf.edu)
Universitat Pompeu Fabra, Barcelona, Spain

Gemma Boleda (gemma.boleda@upf.edu)
Universitat Pompeu Fabra / ICREA, Barcelona, Spain

Abstract

We explore contextual adaptation of referring expressions with respect to referential ambiguity and communicative intention. We focus not only on *whether* people adapt, but also on *how* by contrasting lexical specification (e.g., “batter”) and syntactic modification (e.g., “man in white pants”) when discriminating between objects in natural scenes (e.g., a batter wearing white pants and a referee). There are three main results. First, we replicate that speakers adapt their expressions to avoid ambiguity. Second, communicative intention has an effect: participants tended to use more specific names in a discrimination task than in a descriptive task, even without referential ambiguity in the context. Third, when given the choice, participants tended to prefer more specific words over adding modification – that is, using lexical rather than syntactic means to resolve ambiguity. This suggests that it may be less demanding to increase informativity of referring expressions with lexical specification than syntactic modification.

Keywords: object naming; referring expression; communication; context effects

Introduction

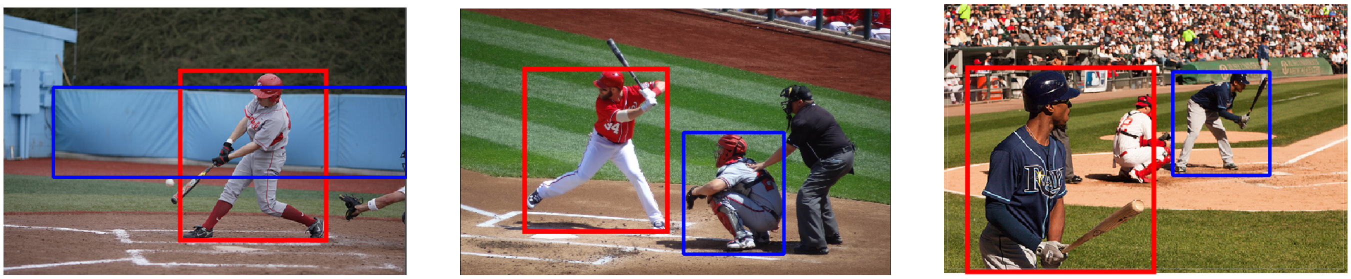
People use language to talk about the world, and to do that they need to produce adequate referring expressions for the objects and entities of interest. We are interested in how people use the different possibilities afforded by their languages when referring, depending on their communicative intention and the context. For instance, for the person highlighted with the red box in Figure 1a, one could use “the man”, “the batter”, and “the man in white pants”, among other options. The choice between these options depends, among other things, on the context an object occurs in. For instance, for Figure 1a any of the options is fine, but in Figure 1b “man” would not distinguish between the two men marked with the red and blue boxes, respectively. Someone wanting to uniquely refer to one of the two may thus prefer to use a different expression. Indeed, there is ample evidence that speakers adapt their referring expressions to the context, and, in particular, that they avoid expressions which are not informative enough to allow identification of a target object (e.g., Brennan & Clark, 1996; Graf, Degen, Hawkins, & Goodman, 2016; Jescheniak, Hantsch, & Schriefers, 2005). In the present study we explore contextual adaptation when people refer to objects in natural scenes. Unlike previous studies, we focus not only on *whether* people adapt, but also on *how*. In particular, we compare the choice between lexical specification (e.g., “batter”) and syntactic modification (e.g., “man in white pants”). We use English data.

We use a standard paradigm, involving visual stimuli and target objects presented together with other objects (e.g., Graf et al., 2016; Jescheniak et al., 2005; Van Der Wege, 2009). Participants are faced with a *discrimination* task, which asks them to produce a referring expression that distinguishes a target object from other objects in the context. Referential ambiguity is manipulated by differences in categorization between the target object and other objects in the scene.

We contrast three context conditions. In the *no-competitor* condition (Figure 1a), no other object is present which could be referred to with the same name as the target object. The two other contexts (panels (b) and (c)) both contain other objects that could be referred to with the same name as the target (e.g., “man” or “baseball player”) but differ in the way this ambiguity can be resolved. In panel (b), the *lexicon-sufficient* condition, this can be done by either lexical specification (e.g., “batter” or “hitter”) or syntactic modification (e.g., “man with the red shirt”). In panel (c), the *syntax-necessary* condition, only syntactic modification is possible (e.g., “the man looking left”).

Previous work using this paradigm has predominantly used artificial stimuli, involving side-by-side presentation of isolated objects (e.g., Graf et al., 2016; Jescheniak et al., 2005). In the present study we use natural scenes such as those in Figure 1. This provides more ecological validity, which is desirable in general; and, in particular, this allows us to include: (1) Objects of varying properties, as not all are equally prototypical, or prominent in an image; also, they exhibit different properties that correspond to variation found in reality, such as baseball players from different teams wearing different colors; (2) Scenes with objects that actually appear together in natural contexts (e.g., baseball players often appear in scenes with other baseball players, and with referees); (3) Bottom-up induced categories of objects (as a result of relying on a large-scale dataset; see *Methods* for details), which differs from the often limited set of basic-level categories used in psycholinguistic studies. Note that one could also use more varied categories with artificial contexts; how-

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society.*



(a) no-competitor (b) lexicon-sufficient (c) syntax-necessary

Figure 1: Examples of visual contexts with and without name competitors. Referential targets are indicated by the red boxes.

ever, "discovering" which kinds of objects afford alternative referring expressions in the first place is a more inductive process.

Furthermore, the chosen experimental design allows us to address three questions. The first is whether people use more informative expressions when the context induces referential ambiguity. For example, considering Figure 1, we expect participants to produce more informative expressions, using either lexical specification or syntactic modification, for the targets in panels (b) and (c) than for the target in panel (a). Previous work with artificial stimuli indicates that this should be the case and we expected this result to generalize to natural scenes. The main purpose of this part of our study is to ensure sensitivity of our materials and task to study variation in referring expressions.

The second question probes the linguistic means that people use when they need to adapt their expressions to the context. As mentioned above, people can felicitously refer to the target in Figure 1b by using either lexical specification ("batter" vs. "man") or syntactic modification ("man with the red shirt" vs. "man"); and the question is whether people prefer the lexical or the syntactic route if both are a viable option to avoid referential ambiguity.

The third question is whether the communicative intention modulates lexical specificity. We contrast the intention to uniquely refer to an object (our discrimination task) to merely providing a description (in an object naming task). Relatively specific names like "batter" are also used in descriptive object naming, although less specific names like "man" are typically preferred (e.g., Silberer, Zarrieß, & Boleda, 2020). We expect specific names to be used more in referential tasks when there is referential ambiguity; however, it is unclear whether this effect will also be found for unambiguous cases like Figure 1a, where specification is not required for discrimination. In addition, contrasting the rate of lexical specification across tasks allows us to evaluate whether (and to which degree) lexical choices in descriptive tasks are adapted when competitors are present in the context. Previous evidence suggests that this may indeed be the case, albeit to a smaller degree than in a referential setting requiring discrimination between objects (Van Der Wege, 2009).

Methods

Participants

We recruited 96 English native speakers from Amazon Mechanical Turk to participate in the study. They were paid $4.5. The results from 11 additional participants were excluded because they provided responses in less than 67% of the production trials.

Materials

We selected 72 images of natural scenes according to the following criteria. Twenty-four images were selected for each of the three context conditions: *no-competitor*, *lexicon-sufficient*, and *syntax-necessary* (see Figure 1). The potential for referential ambiguity was defined relative to a set of reference names and a set of specific names for each image. The same 24 pairs of reference and specific names were used in all three contexts. Reference names were defined as object names which are only informative in the no-competitor condition (e.g., "person", "man", "baseball player" for the images in Figure 1). Specific names were defined as object names which are informative in the no-competitor and lexicon-sufficient condition but not in the syntax-necessary condition (e.g., "batter" or "hitter" in Figure 1). Of note, for several name pairs (e.g., "bird" – "pelican"; "car" – "van") reference name and specific name correspond to a taxonomic classification into basic level and subordinate level names (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). However, not all name pairs were strictly taxonomical (e.g., "man" – "batter", "bottle" – "wine").

Images were selected from the ManyNames dataset (Silberer, Zarrieß, & Boleda, 2020). This dataset provides 36 crowd-sourced name annotations for target objects in 25K naturalistic images selected from VisualGenome (Krishna et al., 2017). We identified potential pairs of reference names and specific names based on the name variation attested in ManyNames. Image selection required a series of processing steps. We used a state-of-the-art Computer Vision model (Anderson et al., 2018) to detect objects in the images and label them. We first identified images in which multiple objects with the same name as the target object (henceforth called *competitors*) in ManyNames were detected (ca. 5K images).

We filtered out images for which the ManyNames data suggested that lexical specification may not be possible (i.e., a specific name was not used at least once) and considered only name pairs for which images with and without competitors were available. The remaining candidate images containing competitors were then manually checked to differentiate between the lexicon-sufficient and syntax-necessary conditions. The final goal of this selection procedure was to have triplets of images with highly similar objects that could be named with the same set of reference vs. specific names (see Figure 1).

In each image, the target object was marked with a red box, corresponding to the box used in ManyNames, and we additionally marked a second object with a blue box. In the no-competitor condition, the blue object shared neither reference nor specific name, meaning there was no potential for referential ambiguity (beyond generic terms like "object"; see Figure 1a). In the lexicon-sufficient condition the blue object shared the reference name but not the specific name with the target object, meaning that lexical specification could resolve ambiguity (see Figure 1b). In the syntax-necessary condition, the blue object shared reference and specific name with the target object, meaning that lexical specification could not resolve ambiguity (see Figure 1c). Three additional images, one for each context condition, showing target images from different categories than the critical items, were selected as warm-up trials.

An additional set of 36 images, from different objects categories, was selected to serve as stimuli in the identification trials (see below). As in the production set, two objects were marked in each image, one with a red box, and one with a blue box. We created an expression for each image which described either the object in the red box or the blue box (each for half of these images). Among these images, 12 had no competitor and 24 had a competitor object (corresponding to the distribution in the production set). For the images with competitors, referring expressions were chosen to be informative with a lexical specification (8 images), informative with a syntactic modification (8 images) or uninformative (i.e., not allowing identification of the target image; 8 images). Uninformative expressions were included to highlight the need to provide informative responses in the production task.

Design

There were two tasks, identification and production. In the identification task participants had to identify the target object in the image (i.e., red or blue box) based on the referring expression we provided. This task was included to make the communicative context more salient to the participants (see below) and was always conducted first. In the production task participants were asked to describe the target object so that another person would be able to identify which object they are referring to.

We used a fully crossed within-participants and within-items design, i.e., all participants saw all three conditions and all 72 images in the production task, and all 36 images in the identification task. Trials in the identification task were randomized but shown in the same order to all participants. For the critical production task, the sequence of conditions (per name pair) was counterbalanced across participants using a Latin square design. Six lists were created. In each of these lists, every 24 trials consisted of one image from each name pair, with all 3 context conditions appearing equally often. Lists were equally distributed across participants with different randomized trial sequences for each participant.

Procedure

The experiment was conducted online via Pavlovia (`https://pavlovia.org/`). There was no time limit for completing the study. At the beginning of the experiment, participants were informed that the goal of the experiment was to study how people talk about objects. First they completed the identification task. Images were presented centred on the screen with a description below. Participants were informed that this description had been provided by other people. They were asked to indicate whether the description refers to the object in the red or the blue box by pressing R or B on their keyboard and to choose the more likely target object if the description would not be clear enough.

After the identification trials, the production task started. Images were again shown centred on the screen, but instead of a description, now an input field for text was placed below them. This field contained already the definite determiner "the". Participants were instructed to type a description for the object marked by the red box, so that another person would be able to identify which object they are referring to. It was implied that the descriptions they provide could be shown to other participants in the identification task. They were instructed to describe the object itself and to avoid using the color of the box or the location of the target object in the image, because images may be presented differently for a person reading their description (e.g., the image could be mirrored). This was done to discourage the overuse of scarce expressions like "the left object". Participants could provide referring expression without a limitation in length. The production task started with 3 warm-up trials.

Data Processing

Preparation of the production data for analysis involved a series of processing steps to identify the object name and syntactic structure of the responses. This included spelling correction and homologisation of spelling variants based on the US-English dictionary of the enchant library (`https://abiword.github.io/enchant/`)) and syntactic parsing using the Stanford CoreNLP library (Manning et al., 2014). [1]

The spellchecked and parsed responses were categorized into different response types of interest using a list of expected reference and specific names for each name pair. Synonyms or superordinate names of a given reference name

[1] The data and the scripts used for preprocessing and analysis are provided here: `https://osf.io/p3jt5/`.

were treated as equivalent to the reference name (e.g., "man", "guy", "person"); synonyms and subordinate names of a given specific name were treated as a lexical specification (e.g., "batter", "slugger", "hitter"). The response types of interest were: (a) no specification (i.e., only a reference name was used; e.g., "the man"), (b) lexical specification (i.e., only a specific name was used; e.g., "the batter"), (c) syntactic modification (i.e., a reference name was used with some syntactic modification; e.g., "the man with the red shirt"), or (d) lexical specification and syntactic modification combined (i.e., a specific name was used with some syntactic modification; e.g., "the batter with the red shirt"). Any response with a more complex syntactic structure than a definite determiner followed by a noun was counted as including a syntactic modification. Noun-noun compounds were only treated as syntactic modifications if they were not included in a list of common compounds (Muraki, Abdalla, Brysbaert, & Pexman, 2022). Therefore, a response like "tennis player" was counted as a lexical specification whereas "front court player" was not.

Results and Discussion

Figure 2 shows the proportion of response types across context conditions.[2] In the three analyses reported below we analyze relative proportions of the different response types using generalized mixed effects models (binomial family) fitted in R (R Core Team, 2021) using the lme4-package (Bates, Mächler, Bolker, & Walker, 2015). A preregistration of our hypotheses and analysis plan can be found here: `https://aspredicted.org/gs9sb.pdf`. Responses in which the target object itself was not described were removed prior to the analyses. This included erroneous responses (i.e., missing responses, not referring to the object in the red box) as well as responses describing another concept than the intended target (e.g., "girl" instead of "the shirt worn by the girl"). Responses of the later kind partially reflect variation in how participants interpret the bounding box (for discussion of this problem see Silberer, Zarrieß, Westera, & Boleda, 2020). However, some of these responses may also be deliberate attempts by the participants to avoid referential ambiguity by focusing on distinctive features of the target objects rather than the object itself (e.g., "the red shirt" instead of "the player with the red shirt"). We leave it to future work to explore the use of this strategy in our task.

General Context Effect

In the first analysis, we tested whether the frequency of using any type of specification (i.e., lexical specification, syntactic modification or both combined) differed across context conditions. The fitted model included a fixed effect of context condition as well as corresponding random slopes (and intercepts) for participants and name pairs. As expected, the proportion of specific responses was larger in the two contexts with competitors than in the no-competitor context. There was no significant difference between the lexicon-sufficient and syntax-necessary contexts (see Table 1). This pattern replicates the finding, well attested in previous work using artificial stimuli, that referring expressions are adapted to avoid referential ambiguity. Most importantly, it demonstrates that our design, in particular using an imaginary interlocutor and natural scenes as stimuli, resulted in meaningful variation in referring expressions.

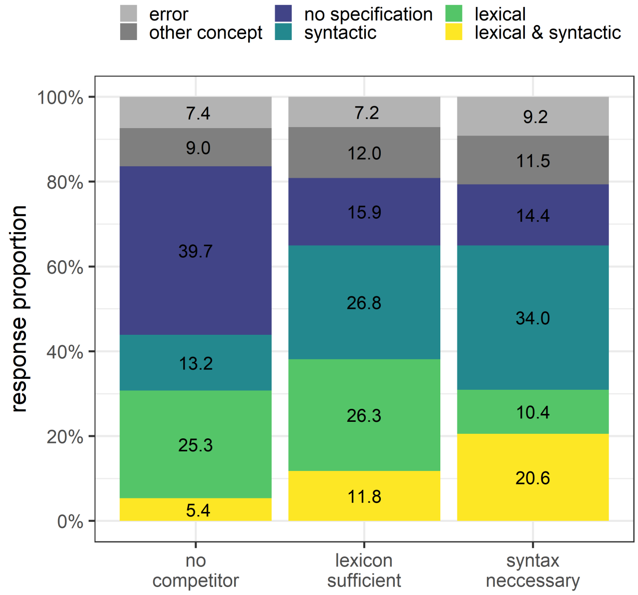


Figure 2: Response type proportions by context condition.

It is worth noting that participants produced a fair amount of responses without any specification even when the context demanded it (see Figure 2), which suggests insufficient attention or compliance with the instructions. However, we want to note that we did not formally test whether a given response would result in successful discrimination by an interlocutor. For instance, in the case of Figure 1c any lexical choice ("man", "baseball player" or "batter") seems insufficient to distinguish between the target and the competitor object. However, there may be a general bias towards the visually more salient object (here the batter in the front), which speakers take into account when choosing their expression. Similar considerations can be made for visual typicality with a bias towards the more typical candidate object for a given name (for an analysis of typicality effects on lexical choices see Gualdoni, Brochhagen, Mädebach, & Boleda, 2022).

Response Types in Competitor Contexts

The second analysis explored differences in specification type (lexical, syntactic, or both) between the two contexts with a competitor. The no-competitor condition was excluded from this analysis because variation in the relative proportions of lexical specification vs. syntactic modification is unlikely to reflect variation induced by the context but rather unspecific

[2] Analyses including trial block demonstrate highly consistent response type proportions across the experiment (see OSF-repository for details).

Table 1: Differences between context conditions in the overall rate of specification (regardless of type).

Contrast	Est.	SE	*z*	*p*
no-comp. vs. lex-suf.	-2.67	0.41	-6.47	<.001
no-comp. vs. syn-nec.	-2.98	0.36	-8.32	<.001
lex-suf. vs. syn-nec.	-0.32	0.38	-0.83	.406

Note. Estimates are on the log-odds scale; *p*-values are adjusted using the Holm-correction for multiple comparisons. Dependent variable is the proportion of any type of specification response (i.e., lexical, syntactic, or both) vs. no specification.

Table 2: Differences in response type proportions between the lexicon-sufficient and syntax-neccessary contexts

Response type	Est.	SE	*z*	*p*
syntactic	-1.83	0.31	-5.83	<.001
lexical	1.52	0.29	5.20	<.001
lexical&syntactic	-0.82	0.30	-2.75	.006

Note. Estimates are on the log-odds scale. Estimates reflect the probability of choosing the respective response type. For the estimates of purely syntactic modification or lexical specification data points with the combination of both were excluded from analysis.

variation across images. All fitted models included a fixed effect of context as well as corresponding random slopes (and intercepts) for participants and name pairs.

Participants combined lexical and syntactic modification more frequently in the syntax-neccessary context than in the lexicon-sufficient context. The combination of the lexical and syntactic route could – in principle – be considered over-informative, as one of the routes would be sufficient to avoid ambiguity. The phenomenon of over-informative responses has been well attested in the literature (Deutsch & Pechmann, 1982; Engelhardt, Bailey, & Ferreira, 2006; Koolen, Gatt, Goudbeek, & Krahmer, 2011; Degen, Hawkins, Graf, Kreiss, & Goodman, 2020). Over-informativity is a puzzle from a purely information-centered approach that assumes that speakers will be optimally informative (Frank & Goodman, 2012) but has been argued to serve efficiency in language by allowing for a faster identification of referents (Rubio-Fernández, 2016). In the context of the present study, it is worth noting that "over-informative" lexical specifications occur irrespective of context as part of natural naming variation (see for instance the no-competitor context in Figure 2). It seems likely that participants who would generally prefer a more specific name for a given object would opt to add syntactic modification to this name rather than opting for syntactic modification of a dispreferred (and less informative) object name. In other words, the higher rate of "over-informativity" in the syntax-necessary condition may largely reflect that participants needed to add a syntactic modification even if their name preference was relatively specific, whereas such addition was not necessary in the lexicon-sufficient condition.

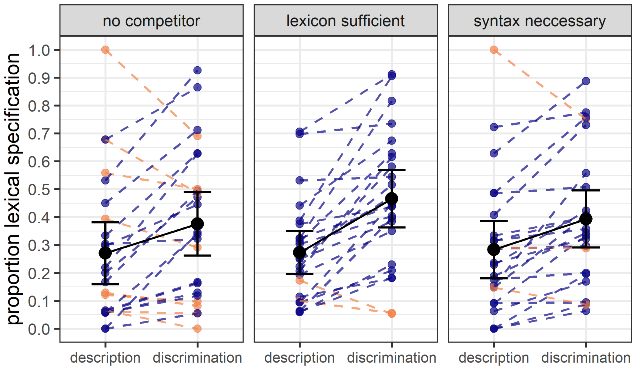


Figure 3: Proportion of using a lexical specification (vs. reference name) by task and condition. Small dots reflect individual images. Large black dots reflect mean and 95%CI across the images. Blue indicates an increase and yellow a decrease of lexical specificity in the discrimination task for this image.

Table 3: Fixed effect estimates for the effects of task and context on the lexical specification rate.

Fixed effect	Estimate	SE	*z*	*p*
(Intercept)	-1.03	0.26	-4.00	
task	-0.63	0.19	-3.26	.001
lexicon-sufficient	0.34	0.18	1.89	.058
syntax-necessary	0.16	0.18	0.89	.375
task:lexicon-sufficient	-0.43	0.18	-2.37	.018
task:syntax-necessary	0.01	0.18	0.04	.971

Note. Estimates are on the log-odds scale. The intercept reflects the average proportion of using a lexical specification (vs. a reference-level name) for the no-competitor context across tasks. The condition contrasts reflect the difference to the no-competitor condition across tasks. The interaction terms reflect the change of the respective condition contrast across tasks.

Of particular interest is the choice between two different forms of discriminating between two objects of the same category: using a more specific word (lexical specification), or a more complex noun phrase while keeping the head noun non-specific (syntactic modification). Here the main condition of interest is the lexicon-sufficient condition, in which subjects can actually choose between the two routes; in the syntax-necessary condition, the lexical route is not available. Correspondingly, we observed a higher rate of syntactic modification and a lower rate of lexical specification in the syntax-necessary condition. Importantly, the fact that the syntactic modification rate is lower and the lexical specification rate is higher in the lexicon-sufficient condition suggests that lexical specification may generally be preferred to some extent if both routes can be used to avoid referential ambiguity.

Task Effects

In the third analysis, we contrasted the use of a lexical specification (vs. using a reference-level name) in our data and in ManyNames. Our subjects carried out a discrimination task, whereas in ManyNames participants were asked to perform a descriptive task – to produce a name for a given object. For this analysis, we focus on the name, ignoring modification; that is, using the specific name and a modifier is counted as using the specific name, and using the reference name and a modifier is counted as using the reference name. This is for comparability, because in ManyNames subjects were constrained to use only names, not free referring expressions as in our experiment. Because of this restriction, we do not have information about people's preferences with respect to syntactic modification in a descriptive task.

This comparison with the descriptive task in ManyNames serves two purposes. First, it allows us to evaluate whether the task demands themselves induce a shift in lexical specificity when referring to objects. Second, it allows us to evaluate whether the frequency of lexical specification in the two competitor contexts (see above) is driven by uncontrolled differences in name preference for the images used in these conditions. If so, the same difference between contexts should be observed regardless of the task.

For this analysis, we aggregated the data to yield the lexical specification proportion for each image in each task. The statistical model included fixed effects for task and condition as well as their interaction. By-image and by-name pair intercepts, as well as by-image random slopes for the task effect were included as random effects in the model. Figure 3 illustrates the relative frequency of using a specific name (vs. a reference name) across tasks and context conditions. Table 3 shows the fixed effect estimates of the statistical model.

Lexical specification was generally more frequent in the discrimination task. Importantly, this effect was found for all three context conditions (ps $< .002$). This suggests that the communicative context of the discrimination task may shift name variation towards more specific object names even if the specification is not needed or not sufficient to avoid referential ambiguity. Notably, there was an interaction of task and condition for the contrast between the no-competitor and lexicon-sufficient condition. This reflects that these contexts only differed reliably in the discrimination task ($p = .015$), but not in the descriptive task. In fact, none of the conditions differed significantly in the descriptive task (ps $>$ .99). This shows that the higher rate of lexical specification in the lexicon-sufficient context (as compared to the syntax-necessary context) in the discrimination task is not driven by uncontrolled differences in name preference inherent to the specific images we chose for these contexts. Moreover, this result suggests that potential referential ambiguity impacts lexical choices in purely descriptive tasks much less than in discrimination, if at all (cf. Van Der Wege, 2009).

Summary and Conclusions

In the present study we have explored contextual adaptation when people refer to objects in natural scenes, with respect to referential ambiguity and communicative intention. There are three main findings. First, we replicate the previous finding that speakers adapt their expressions to avoid referential ambiguity in a given visual context. Second, we find that the communicative intention has an effect over and above the ambiguity of the context: subjects tended to use more specific object names in a discrimination task than in a descriptive task, even when they didn't need to because there was no referential ambiguity in the context. Third, the most novel finding is that, when given the choice, people tend to prefer to provide more information by using a more specific word rather than adding modification – that is, using the lexical rather than the syntactic means that language offers.

This last result suggests that the lexical route may be less costly for speakers. One possibility is that navigating the lexicon in search for a more specific name generally requires less effort than building a syntactic structure (which arguably involves additional processes in conceptualisation and grammatical encoding). Another possibility is that this result reflects merely a trade-off between retrieving a single vs. multiple lexical items when producing longer noun phrases. It is possible that our materials favored the lexical route, because we chose target objects for which specific names were already relatively frequently used in a purely descriptive task. More research is necessary to determine the boundary conditions for the bias towards lexical specification we observed. If this finding is indeed related to cognitive effort, then it should be modulated by factors affecting the cognitive demand imposed by choosing a lexical specification (e.g., lexical frequency of a specific name) and factors affecting the cognitive demand imposed by choosing a syntactic modification (e.g., syntactic complexity).

A related question is whether the choice between lexical specification and syntactic modification follows from a continuum of candidate expressions in which lexical specification tends to be less costly (at least for the present images) or whether there are other factors biasing referential expressions towards lexical specification beyond the cognitive demand for the speaker. One limitation of the present study is that expressions were always produced in the written modality. It seems likely that the effort required for lexical specification vs. syntactic modification will differ to some degree with different output modalities. It may be, for instance, that the (motoric) effort associated with the length of an expression (syntactic modifications are almost always longer than lexical specifications) has a larger impact when writing responses on a keyboard (as in the present study) compared to speaking or signing directly with an interlocutor.

We hope that future work will probe these issues and that the present study serves as a step towards a more comprehensive understanding of not only *whether* people choose to be more informative (the focus of most work in the area), but also *how*.

Acknowledgements

The authors thank the reviewers and the COLT research group for their useful feedback. 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) and the Spanish Research Agency (ref. PID2020-112602GB-I00). This paper reflects the authors' view only, and the funding agencies are not responsible for any use that may be made of the information it contains.

References

Anderson, P., He, X., Buehler, C., Teney, D., Johnson, M., Gould, S., & Zhang, L. (2018). Bottom-up and top-down attention for image captioning and visual question answering. *arXiv:1707.07998 [cs].*

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. doi: 10.18637/jss.v067.i01

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1482–1493. doi: 10.1037/0278-7393.22.6.1482

Degen, J., Hawkins, R. D., Graf, C., Kreiss, E., & Goodman, N. D. (2020). When redundancy is useful: A Bayesian approach to "overinformative" referring expressions. *Psychological Review*, *127*(4), 591–621. doi: 10.1037/rev0000186

Deutsch, W., & Pechmann, T. (1982). Social interaction and the development of definite descriptions. *Cognition*, *11*(2), 159–184. doi: 10.1016/0010-0277(82)90024-5

Engelhardt, P. E., Bailey, K. G., & Ferreira, F. (2006). Do speakers and listeners observe the Gricean Maxim of Quantity? *Journal of Memory and Language*, *54*(4), 554–573. doi: 10.1016/j.jml.2005.12.009

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084), 998–998. doi: 10.1126/science.1218633

Graf, C., Degen, J., Hawkins, R. X. D., & Goodman, N. D. (2016). Animal, dog, or dalmatian? Level of abstraction in nominal referring expressions. In *Proceedings of the 38th Annual Conference of the Cognitive Science Society.*

Gualdoni, E., Brochhagen, T., Mädebach, A., & Boleda, G. (2022). Woman or tennis player? Visual typicality and lexical frequency affect variation in object naming. In *Proceedings of the 44th Annual Conference of the Cognitive Science Society.*

Jescheniak, J. D., Hantsch, A., & Schriefers, H. (2005). Context effects on lexical choice and lexical activation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(5), 905–920. doi: 10.1037/0278-7393.31.5.905

Koolen, R., Gatt, A., Goudbeek, M., & Krahmer, E. (2011). Factors causing overspecification in definite descriptions. *Journal of Pragmatics*, *43*(13), 3231–3250. doi: 10.1016/j.pragma.2011.06.008

Krishna, R., Zhu, Y., Groth, O., Johnson, J., Hata, K., Kravitz, J., . . . Fei-Fei, L. (2017). Visual Genome: Connecting language and vision using crowdsourced dense image annotations. *International Journal of Computer Vision*, *123*(1), 32–73. doi: 10.1007/s11263-016-0981-7

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., & McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. In *Association for computational linguistics (acl) system demonstrations* (pp. 55–60). Retrieved from http://www.aclweb.org/anthology/P/P14/P14-5010

Muraki, E. J., Abdalla, S., Brysbaert, M., & Pexman, P. M. (2022, March). *Concreteness ratings for 62 thousand English multiword expressions* [Preprint]. doi: 10.31234/osf.io/m397u

R Core Team. (2021). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org/

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*(3), 382–439. doi: 10.1016/0010-0285(76)90013-X

Rubio-Fernández, P. (2016). How redundant are redundant color adjectives? An efficiency-based analysis of color overspecification. *Frontiers in Psychology*, *7*. doi: 10.3389/fpsyg.2016.00153

Silberer, C., Zarrieß, S., & Boleda, G. (2020). Object naming in language and vision: A survey and a new dataset. In *Proceedings of the 12th Language Resources and Evaluation Conference* (pp. 5792–5801). Retrieved from https://aclanthology.org/2020.lrec-1.710

Silberer, C., Zarrieß, S., Westera, M., & Boleda, G. (2020). Humans meet models on object naming: A new dataset and analysis. In *Proceedings of the 28th International Conference on Computational Linguistics* (pp. 1893–1905). doi: 10.18653/v1/2020.coling-main.172

Van Der Wege, M. M. (2009). Lexical entrainment and lexical differentiation in reference phrase choice. *Journal of Memory and Language*, *60*(4), 448–463. doi: 10.1016/j.jml.2008.12.003