Quantifying informativeness of names in visual space

Eleonora Gualdoni (eleonora.gualdoni@upf.edu)
Department of Translation and Language Sciences, Universitat Pompeu Fabra, Barcelona, Spain

Charles Kemp (c.kemp@unimelb.edu.au)
School of Psychological Sciences, University of Melbourne, Melbourne, Australia

Yang Xu (yangxu@cs.toronto.edu)
Department of Computer Science, Cognitive Science Program, University of Toronto, Toronto, Canada

Gemma Boleda (gemma.boleda@upf.edu)
Department of Translation and Language Sciences, Universitat Pompeu Fabra / ICREA, Barcelona, Spain

Abstract

The human lexicon expresses a wide array of concepts with a limited set of words. Previous work has suggested that semantic categories are structured compactly to enable informative communication. Informativeness is typically quantified with respect to an entire semantic domain and not at the level of individual names. We develop a measure of name informativeness using an information-theoretic framework grounded in visual object representations derived from natural images. Our approach uses computer vision models to characterize informativeness of individual names with respect to large-scale data in a naturalistic setting. We show that our informativeness measure predicts degrees of specificity in lexical categories more precisely than alternative measures based on entropy and frequency. We also show that name informativeness jointly captures within-category similarity and distinctiveness across categories. Our analyses suggest how the variability of names from a broad part of the lexicon may be understood through the lens of information theory.

Keywords: lexicon; naming; visual object representation; informativeness; information theory

Introduction

The lexicon uses a limited set of names to describe a potentially infinite range of objects. How names encode concepts can vary from word to word, and the link between the lexicon and our conceptual system is not one-to-one (Malt, Sloman, Gennari, Shi, & Wang, 1999; Murphy, 2004). For example, classic work on concepts and categories suggests that words encode categories with different levels of specificity (e.g., "sparrow" is more specific than "bird", which in turn is more specific than "animal") and that specificity interacts with the structure of the categories themselves (Brown, 1958; Rosch & Mervis, 1975; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976; Jolicoeur, Gluck, & Kosslyn, 1984). More specific words encode more specific categories, and thus provide more information about the intended referents. However, the more specific a word is, the less usable it is in communication, since fewer objects fall under its denotation. How do we determine the information capacity of words in the lexicon? An attempt to tackle this problem is found in Kireyev (2009), where word informativeness is estimated through text by means of Latent Semantic Analysis. Here we offer an information-theoretic approach to quantify the informativeness of individual words based on data about naming visually presented objects.

man (10), batter (10), baseball player (5),
player (5), person (3)

Figure 1: Example image from the ManyNames dataset (Silberer, Zarrieß, & Boleda, 2020), with corresponding naming responses and counts in parentheses.

Our theoretical starting point is a recent line of work that explores semantic category structures using an information-theoretic approach that captures the notion of efficient communication (Corter & Gluck, 1992; Kemp, Xu, & Regier, 2018; Zaslavsky, Kemp, Regier, & Tishby, 2018; Zaslavsky, Regier, Tishby, & Kemp, 2019; Xu, Liu, & Regier, 2020; Zaslavsky, Maldonado, & Culbertson, 2021). This line of work is built on a simple communicative scenario involving a speaker and a listener, and defines informativeness as the precision at which the listener can successfully reconstruct the intended meaning from the speaker's utterance. Minimizing the speaker-listener meaning difference means maximizing the average informativeness of the transmitted message, that is, the informativeness of the lexical system. Existing research under this framework has found evidence for the hypothesis that semantic categories attested in the world's languages are more informative than alternative possible category systems. However, these studies have typically focused

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society.*

on analyzing small, isolated, semantic domains and on measuring informativeness of the entire naming system associated with a domain (Regier, Kay, & Khetarpal, 2007; Zaslavsky et al., 2018, 2019, 2019; Xu et al., 2020; Zaslavsky et al., 2021).

We develop a novel measure of *name informativeness*, which can be used to quantify the information capacity of individual names and to compare these names on the basis of their informativeness. We ground our measure of name informativeness in visual representations of natural objects by combining computer vision models with a large-scale free naming dataset. Figure 1 shows an example object in the resource we use, which was given a variety of names by participants. To quantify name informativeness in this naturalistic setting, we use object representations generated by computer vision models that have been shown to capture human classification of real-world objects (e.g., Krizhevsky, Sutskever, & Hinton, 2012).

We provide an initial exploration of our measure of name informativeness in a series of analyses pertaining to category specificity, as well as category homogeneity and distinctiveness, which are properties that are extensively discussed in the categorization literature (Mervis & Rosch, 1981; Medin, 1983). Our analyses suggest that the information-theoretic measure accurately captures the informativeness of names, and can therefore contribute to a better understanding of the relationship between words and concepts across a large part of the lexicon.

Materials and methods

Theoretical framework

Our measure of name informativeness is formulated using a framework that extends previous information-theoretic work on efficient communication (Regier et al., 2007; Zaslavsky et al., 2018). The framework is based on a hypothetical communicative scenario involving a *speaker* and a *listener*. We represent the environment, or universe, as a set of objects $\mathcal{U}$.[1] We assume that the speaker wants to communicate a certain target object in the universe $t \in \mathcal{U}$. This target is sampled from a prior distribution $P(t)$ reflecting communicative need, or the frequencies with which objects need to be communicated. The speaker is uncertain about the target and her mental representation of the target is therefore a distribution $m_t(u)$ over $\mathcal{U}$, the set of objects that speaker and listener can potentially talk about. We refer to this speaker representation as a *meaning*, and define it as a similarity-based distribution over $\mathcal{U}$ (Eisape, Levy, Tenenbaum, & Zaslavsky, 2020; Regier, Kemp, & Kay, 2015):[2]

$$m_t(u) \propto exp(Sim(t,u)), \tag{1}$$

The speaker communicates m_t by producing a name n, and selects this name according to the distribution $P(n|m_t)$. The listener hears n and tries to reconstruct the speaker's meaning by forming a mental representation $\hat{m}_n(u)$ that reflects an inference about the speaker's intended meaning. In particular, for a given name n we define the listener's representation as:

$$\hat{m}_n(u) = \sum_{m_t \in \mathcal{M}} P(m_t|n) m_t(u), \tag{2}$$

Here $\mathcal{M}$ is the meaning space, or the set of meanings induced by all objects in $\mathcal{U}$. In other words, the listener's mental representation $\hat{m}_n(u)$ for a specific name n, such as "cat", is a weighted average of the meanings induced by all objects in $\mathcal{U}$. The weights of all these meanings are computed via Bayesian inference:

$$P(m_t|n) \propto P(n|m_t)P(m_t), \tag{3}$$

Here the prior $P(m_t)$ is induced by the need distribution $P(t)$, and $P(n|m_t)$ is high for meanings m_t that are commonly expressed by generating the name "cat."

The name n produced for the target t is informative for the listener to the extent that the listener representation for that name $\hat{m}_n(u)$ is similar to the speaker meaning m_t. We formulate the difference between the speaker's intended meaning and the listener's reconstructed meaning as the Kullback-Leibler divergence between m_t and $\hat{m}_n$. We thus define the distortion created by name n for a target t in meaning space as follows:

$$D[m_t||\hat{m}_n] = \sum_u m_t(u) \log\left(\frac{m_t(u)}{\hat{m}_n(u)}\right) \tag{4}$$

How suitable a name n is for a target object t depends on the visual similarity between t and other objects that are commonly called n. For instance, if t is visually similar to objects that are commonly called n, then n is a highly informative name for t, and using it to label t will create little distortion.

The average distortion of a name over the meaning space is then defined as follows:

$$\sum_{m_t \in \mathcal{M}} P(m_t|n) D[m_t||\hat{m}_n] \tag{5}$$

Finally, we define the informativeness of a name as the inverse of its average distortion: a more informative name should create less distortion in the communication between the speaker and the listener.

Dataset

To ground and evaluate our framework comprehensively and in a naturalistic setting, we work with the ManyNames dataset (Silberer, Zarrieß, & Boleda, 2020). This dataset contains 25K real-world images sampled from the VisualGenome dataset (Krishna et al., 2017) and annotated in a free-naming

[1] Our notation is inspired by previous information-theoretic work on semantic typology (Regier et al., 2007; Zaslavsky et al., 2018, 2021).

[2] In the original formulation, a sensitivity parameter γ is included in the argument of the exponential function (Eisape et al., 2020). However, for our initial analyses, we decide to work with a simplified version of Equation 1, leaving to further work the option of optimizing the sensitivity parameter.

task in English, where participants were asked to "name the object in the red frame with the first name that comes to mind" (see Figure 1 for an example). No instructions were given to the annotators about whether the names produced had to unambiguously distinguish the target from the context objects in the images, making the task descriptive, as opposed to discriminative, in its nature. On average 31 names were collected for each object, making this dataset suitable for studying how multiple names suit the same object. Objects are organized in 7 domains: *animals / plants*, *buildings*, *clothing*, *food*, *home*, *vehicles*, and *people*. In the current study, we work with the 238 names in ManyNames that have been produced for at least 20 images. We set this threshold in order to obtain reliable representations of objects in visual space.

Model specification in visual space

The objects in the ManyNames dataset constitute the universe of objects $\mathcal{U}$. In order to apply the information-theoretic framework described above, the first step is to decide how to represent the space of meanings $\mathcal{M}$. To leverage visual information about natural objects in real-world settings, we use state-of-the-art deep learning models of computer vision. We follow Eisape et al. (2020) and use visual features derived from a deep-learning object classification model. In particular, following the existing work of Silberer, Zarrieß, Westera, and Boleda (2020) and Gualdoni, Brochhagen, Mädebach, and Boleda (2022b), we extract features (2048-d vectors) for the image areas inside the red frame with the Bottom-up attention model of Anderson et al. (2018) trained on VisualGenome (Krishna et al., 2017). Then, given a target t, we compute the speaker meaning induced by the target using Equation 1, where *Sim* is defined as the cosine similarity between the objects' visual features. In the rest of the paper, we call our proposed measure *NV-Informativeness*, where the prefix indicates that the measure exploits both Naming data and Visual information, in contrast to the baseline measure presented in the following Section. From here on, *visual space* refers to the vector space populated by the visual features of the ManyNames objects.

Baseline measure

To evaluate the effectiveness of our measure of name informativeness, we also consider a baseline measure of informativeness that does not use any visual information. To do so, we follow Xu et al. (2020) and define $m_t(u)$ based on the assumption that the speaker has no uncertainty about the target:

$$m_t(u) = \begin{cases} 1 & \text{if t =u} \\ 0 & \text{if t} \neq \text{u.} \end{cases} \quad (6)$$

After making this adjustment, we define a baseline informativeness measure using Equations 4 and 5 in exactly the same way as we defined our NV-Informativeness measure. This approach amounts to defining informativeness as the inverse of the entropy of a name: lower entropy corresponds to higher informativeness. We call this baseline measure *N-Informativeness* throughout our analyses to contrast it with NV-Informativeness. N-Informativeness does not assume any uncertainty in the speakers' beliefs about the target and does not make use of visual information, deriving name informativeness only from the speakers' naming choices. We expect this aspect to be a key limitation of N-Informativeness with respect to capturing name informativeness of objects in the world.

When evaluating both NV-Informativeness and N-Informativeness, we make the simplifying assumption that all objects in $\mathcal{U}$ have the same probability of being sampled, which means that the prior $P(m_t)$ is constant.

Results

We carry out two analyses that evaluate the effectiveness of our informativeness measure and examine some of its properties. The first analysis asks how well the measure accounts for category specificity, a psycholinguistic property of names that is characterized without direct reference to visual space. The second analysis explores the relationship between our measure and homogeneity and distinctiveness, two properties of categories that are characterized here in terms of distributions in visual space.

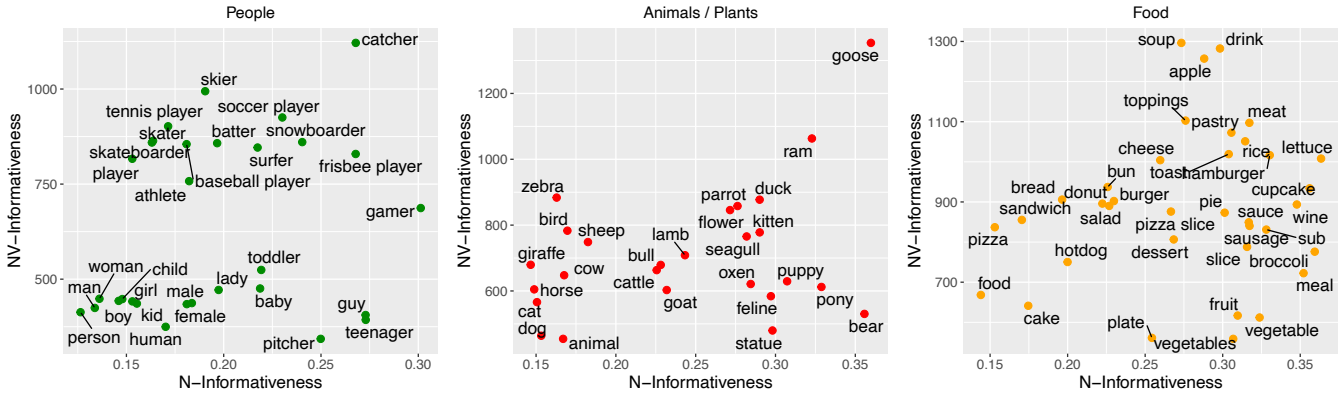
Analysis I: Category specificity

Figure 2 shows the relationship between NV-Informativeness and N-Informativeness for three of the semantic domains in ManyNames: *people*, *food*, and *animals / plants*.

NV-Informativeness captures name specificity to a large extent. Consider, for instance, the domain *people*: with only a few exceptions, our measure gives high scores to specific names that highlight the sport played by the person depicted in the image, such as "catcher" or "soccer player", clearly separating these cases from more general taxonomic names, such as "person", "girl", or "man". In the domain *animals / plants*, "animal" correctly receives the lowest informativeness score, while names like "kitten", "duck", or "goose" are labeled as more informative. Similarly, in the domain *food*, generic names such as "vegetables", "fruit" and "food" receive low informativeness scores, while "apple", "cheese", or "rice" receive higher scores.

Still, our measure is not perfect. For instance, if we consider some hierarchically structured word pairs, such as "wine"-"drink" or "sausage"-"meat", we see that NV-Informativeness fails to identify the first, more specific member of each pair as more informative. Below we describe a quantitative analysis that estimates the performance of our measure on hierarchically structured word pairs of this kind.

The baseline measure N-Informativeness is not as good as NV-Informativeness at measuring specificity. For instance, in the domain *people*, N-Informativeness fails to separate general taxonomic names, such as "lady" or "child", from more specific names, such as "player" or "skater". This shows the advantage of our NV-Informativeness measure: a simple measure of name entropy like N-Informativeness does not have access to visual information, while NV-Informativeness



(a) (b) (c)

Figure 2: NV-Informativeness plotted against N-Informativeness for names belonging to three domains: (a) *people*, (b) *animals / plants*, and (c) *food*.

is sensitive to important distinctions by factoring in how suitable names are for visually represented objects. Incorporating visual information in this way seems useful for separating specific names, such as "baseball player" and "skier", from taxonomic names, that can be valid alternatives when referring to the same object, such as "man" and "woman".

The case of "pitcher" sheds light on what happens with a polysemous word (see Figure 2(a), lower right quadrant, for the position on "pitcher"). In the ManyNames dataset, "pitcher" is used for both baseball players and water jars, that is, visually very dissimilar objects belonging to different categories. Accordingly, NV-Informativeness places it among the least informative words. Instead, N-Informativeness –blind to visual information– assigns a high score to it. On the other hand, there are cases where visual information may be misleading. Consider, for instance, "bear" and "statue":[3] In the ManyNames dataset, these names have been used to label visually diverse objects, such as wild bears and teddy bears for "bear", and artifacts representing different things for "statue". NV-Informativeness, leveraging the objects' visual appearance, concludes counter-intuitively that these are very general names that can label very diverse objects. In contrast, N-Informativeness assigns high informativeness scores to both names, showing more robustness to these kind of cases.

Qualitative inspection thus suggests that NV-Informativeness is promising as a measure of category specificity. We next perform a quantitative test of how well the informativeness measures perform on the task of hypernymy detection, which requires identifying which of the two words in pairs like "jeep"-"car" or "woman"-"person" is the most general one (i.e. the hypernym; the converse relation is named hyponymy). Because hypernyms can be applied to more diverse objects, they are less informative than hyponyms (Lyons, 1977). We extract 118 hyponym-hypernym pairs from our lexicon based on WordNet relationships (Miller, 1994). We compute the accuracy of the measures in predicting the hypernym according to a rule-based model that classifies the word with the lowest score as the hypernym.

Table 1 summarizes the results. Both N-Informativeness and NV-Informativeness classify hyponyms as more informative than hypernyms, with an accuracy score much higher than chance: the baseline measure N-Informativeness reaches an accuracy of 0.67, while NV-Informativeness reaches an accuracy of 0.80. Thus, both measures effectively capture name specificity, with a clear advantage for NV-Informativeness. We further test an additional baseline, the inverse of word frequency. This baseline obtains an accuracy of 0.64 on the task, very close to the performance of N-Informativeness and again well below the performance of NV-Informativeness

Chance	N-Informativeness	NV-Informativeness
0.50	0.67	**0.80**

Table 1: Model accuracy in hypernymy identification.

Analysis II: Category homogeneity and distinctiveness

As described earlier, the literature on concepts and categorization has proposed that the structure of categories jointly maximizes homogeneity, or within-category similarity, and distinctiveness, or between-category difference (Medin, 1983; Mervis & Rosch, 1981). Previous work provides separate formalizations of these two aspects of category structure, and relies on an explicit combination function for combining the two into a single measure of category goodness (Regier et al., 2007). It seems possible, however, that homogeneity and distinctiveness are two facets of informativeness, and that a single informativeness measure can cap-

[3] In the ManyNames dataset, "statue" belongs to the "animals / plants" domain. Probably this is due to the fact that these images were labeled with animal names in VisualGenome (e.g. "horse" for a horse that is part of a statue).

ture them both. To explore this idea we computed homogeneity and distinctiveness based on the denotations of names in our visual space, and examined how both relate to our informativeness measures.

We expect category homogeneity to be higher for more specific names: instances of more specific categories, such as DALMATIAN or BATTER, are visually more similar to each other than instances of more general categories, such as DOG or BASEBALL PLAYER. The expectation is less clear for distinctiveness. For names with similar levels of specificity, like "penguin" and "robin", the expectation is that distinctiveness will correlate positively with informativeness. A name like "penguin" is informative because after hearing this label the listener's probability distribution is peaked over a region in the image space that includes only penguins. "Robin", instead, is less informative because after hearing this label the listener's distribution covers a region of image space that also includes sparrows and members of other categories, as robins are similar to other kinds of birds whereas penguins are not.

In other words, more informative names are expected to be more isolated with respect to other names at the same hierarchical level. However, predictions are less clear for names that capture hierarchical relations between concepts ("robin" - "songbird" - "bird") or are related but not in a taxonomic way ("boy"-"baseball player").

	N-Inform	**NV-Inform**
corr. w. homogeneity	-0.02	**0.67***
corr. w. distinctiveness	-0.03	**0.48***

Table 2: Pearson's correlation coefficient *r* between name informativeness and category homogeneity, and between name informativeness and category contrast. Values marked with * are significantly different from 0 (α level: 0.05).

We compute the homogeneity of a name as the average pairwise cosine similarity between objects in our visual space that have been called by that name. To compute distinctiveness, we use the visual prototypes for names we made available in previous work (Gualdoni et al., 2022b).[4] The prototype of a name is the centroid (average) of the visual representations of the objects labeled with that name in VisualGenome, the resource from which ManyNames images were sampled. It is a surrogate for a prototype, in Rosch and Mervis (1975)'s sense, of the category denoted by a name (one that only takes into account visual properties). Distinctiveness is then defined as the average cosine distance between the prototype of a name and the prototypes of all the other names in our visual space. This measure is related to an index of crowdedness defined in Gualdoni, Brochhagen, Mädebach, and Boleda (2022a), with the difference that it applies to names as opposed to individual objects.

Table 2 shows the correlation between informativeness and homogeneity, for both N-Informativeness and NV-Informativeness. NV-Informativeness performs as expected, and achieves a solid positive correlation of 0.67 with homogeneity and a more moderate correlation of 0.48 with distinctiveness. The lower correlation with distinctiveness may suggest that, indeed, hierarchically related and/or non-taxonomically related concepts have a more nuanced relationship with informativeness, a possibility that further work should explore. The baseline N-Informativeness shows no correlation with either homogeneity and distinctiveness. Because both of these properties were defined in visual terms, it may not be surprising that an informativeness measure blind to visual information struggles to capture either one.

[4]This analysis excludes 15 names in our space that do not have prototypes in Gualdoni et al. (2022b)'s data.

Discussion

We developed a measure of name informativeness, that we call NV-Informativeness, based on visual representations of objects and human naming behavior, and evaluated its ability to account for properties of categories across a substantial part of the lexicon. Our evaluation included two analyses that examined three prominent phenomena related to conceptual aspects of the lexicon. The first focused on category specificity as an example of a psycholinguistic property of names not directly related to visual information. The second examined category homogeneity and distinctiveness, which were formulated in terms of distributions in visual space. The NV-Informativeness measure is superior to a baseline (N-Informativeness) formulated in the same information-theoretic framework that relies on naming data but not on visual object representations. This finding suggests that accounting for object properties (and visual properties in particular) is crucial for measures of name informativeness; and that state of the art models in computer vision provide useful representations for objects, consistent with recent work on other aspects of naming and categorization (Günther, Marelli, Tureski, & Petilli, 2021; Gualdoni et al., 2022b).

Using computer vision models allowed us to examine stimuli that are more naturalistic than those typically used in previous work on categorization and efficient communication – for instance, color chips (Regier et al., 2007; Zaslavsky et al., 2018) or stylized pictures of containers (Zaslavsky et al., 2019). Our image-based approach also allows our method to be applied on a much wider scale than previous analyses of individual domains such as colors and containers, and opens up the possibility of exploring the role of visual context (e.g. the context surrounding the red frame in Figure 1). Our previous work shows how computer vision representations of visual context can be successfully used for research in cognitive science (Gualdoni et al., 2022b), and developing an informativeness measure that exploits these representations is a promising avenue for future work.

An important advantage of the proposed measure is that it is grounded in a principled framework that allows flexible extensions. A natural direction for future work is to explore informativeness not only at the level of individual words, as

done, for instance, in Kireyev (2009), but at the level of individual objects. For example, the name "batter" may be more informative when applied to some pictures of batters –for instance, to pictures that show the stereotypical features that people associate to batters– than to others. Our approach also opens up future possibilities for the formalization and comprehensive evaluation of notions such as "basic level", widely discussed in the cognitive and psychological literature.

Our measure is related to metrics for name informativeness developed in computational linguistics, in the context of language modeling (or word prediction). In these cases, the notion of name informativeness is related to how predictable words are given a linguistic context (Aina, Liao, Boleda, & Westera, 2021; Pimentel, Maudslay, Blasi, & Cotterell, 2020; Orita, Vornov, Feldman, & Daumé III, 2015). Our measure instead relates name informativeness to how predictable words are given a visually represented object, that is, to the visual properties of the object. Another promising avenue for further research is to compare (and eventually combine) informativeness derived from textual and visual data. This will enable the exploration of further issues related to meaning such as polysemy, which is pervasive in natural language but occurs less in visual stimuli (cf. however the discussion on "pitcher" above), and non-visual properties such as the origin and function of an object in the case of "statue".

Conclusion

Our work suggests how properties of object names across a substantial part of the lexicon can be characterized from a visually grounded and information-theoretic perspective. Our approach draws on both visual object representations and naming data, and allows us to quantify the variation in informativeness across individual names in the lexicon. The initial results reported here seem promising, but future work is needed to explore and assess the potential of our approach to account for the organization and use of the human lexicon.

Acknowledgments

The authors thank the reviewers and the COLT research group for their useful feedback. This research is an output of grant PID2020-112602GB-I00/MICIN/AEI/10.13039/501100011033, funded by the Ministerio de Ciencia e Innovación and the Agencia Estatal de Investigación (Spain), of grant agreement No. 715154 funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, and of NSERC Alliance International-Catalyst Grant ALLRP 576149-22. The second author was supported by grant FT190100200 from the Australian Research Council. 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

Aina, L., Liao, X., Boleda, G., & Westera, M. (2021, November). Does referent predictability affect the choice of referential form? A computational approach using masked coreference resolution. In *Proceedings of the 25th conference on computational natural language learning* (pp. 454–469). Online: Association for Computational Linguistics. Retrieved from https://aclanthology.org/2021.conll-1.36 doi: 10.18653/v1/2021.conll-1.36

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. In *Proceedings of CVPR*.

Brown, R. (1958). How shall a thing be called? *Psychological review*, *65 1*, 14-21.

Corter, J. E., & Gluck, M. A. (1992). Explaining basic categories: feature predictability and information. *Psychological Bulletin*, *111*(2), 291–303.

Eisape, T., Levy, R., Tenenbaum, J. B., & Zaslavsky, N. (2020). Toward human-like object naming in artificial neural systems. In *Bridging AI and Cognitive Science workshop at the international conference on representation learning (iclr)*.

Gualdoni, E., Brochhagen, T., Mädebach, A., & Boleda, G. (2022a). *What's in a name? A large-scale computational study on how competition between names affects naming variation.* Retrieved from https://psyarxiv.com/t84eu/

Gualdoni, E., Brochhagen, T., Mädebach, A., & Boleda, G. (2022b). Woman or tennis player? Visual typicality and lexical frequency affect variation in object naming. In J. Culbertson, A. Perfors, & V. Rabagliati H. & Ramenzoni (Eds.), *Proceedings of the 44th annual conference of the cognitive science society.* Cognitive Science Society. doi: 10.31234/osf.io/34ckf

Günther, F., Marelli, M., Tureski, S., & Petilli, M. A. (2021, 10). Vispa (vision spaces): A computer-vision-based representation system for individual images and concept prototypes, with large-scale evaluation. doi: https://doi.org/10.31234/osf.io/n4dmq

Jolicoeur, P., Gluck, M. A., & Kosslyn, S. M. (1984). Pictures and names: Making the connection. *Cognitive Psychology*, *16*(2), 243-275.

Kemp, C., Xu, Y., & Regier, T. (2018). Semantic typology and efficient communication. *Annual Review of Linguistics*(4), 109-128.

Kireyev, K. (2009, 01). Semantic-based estimation of term informativeness. In (p. 530-538). doi: 10.3115/1620754.1620831

Krishna, R., Zhu, Y., Groth, O., Johnson, J., Hata, K., Kravitz, J., ... Li, F.-F. (2017, 05). Visual Genome: Connecting Language and Vision Using Crowdsourced Dense Image Annotations. *International Journal of Computer Vision*, *123*.

Krizhevsky, A., Sutskever, I., & Hinton, G. (2012). Imagenet classification with deep convolutional neural networks. *Neural Information Processing Systems*, *25*.

Lyons, J. (1977). *Semantics* (Vol. 1). Cambridge University Press. doi: 10.1017/CBO9781139165693

Malt, B. C., Sloman, S. A., Gennari, S., Shi, M., & Wang, Y. (1999). Knowing versus naming: Similarity and the linguistic categorization of artifacts. *Journal of Memory and Language*, *40*(2), 230-262. doi: https://doi.org/10.1006/jmla.1998.2593

Medin, D. (1983). Structural principles in categorization. In T. Tighe & B. Shepp (Eds.), *Perception, cognition, and development: Interactional analyses.* Erlbaum.

Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology*, *32*, 89-115.

Miller, G. A. (1994). WordNet: A lexical database for English. In *Human Language Technology: Proceedings of a workshop held at Plainsboro, New Jersey, March 8-11, 1994.* Retrieved from https://aclanthology.org/H94-1111

Murphy, G. L. (2004). *The Big Book of Concepts* (2nd ed.). Cambridge, MA (etc.): The MIT Press.

Orita, N., Vornov, E., Feldman, N., & Daumé III, H. (2015, July). Why discourse affects speakers' choice of referring expressions. In *Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing (volume 1: Long papers)* (pp. 1639–1649). Beijing, China: Association for Computational Linguistics. Retrieved from https://aclanthology.org/P15-1158 doi: 10.3115/v1/P15-1158

Pimentel, T., Maudslay, R. H., Blasi, D., & Cotterell, R. (2020, November). Speakers fill lexical semantic gaps with context. In *Proceedings of the 2020 conference on empirical methods in natural language processing (emnlp)* (pp. 4004–4015). Online: Association for Computational Linguistics. Retrieved from https://aclanthology.org/2020.emnlp-main.328 doi: 10.18653/v1/2020.emnlp-main.328

Regier, T., Kay, P., & Khetarpal, N. (2007). Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences*, *104*, 1436 - 1441.

Regier, T., Kemp, C., & Kay, P. (2015, 01). Word meanings across languages support efficient communication. In (p. 237-263). doi: 10.1002/9781118346136.ch11

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*(4), 573-605.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382-439.

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). Marseille, France: European Language Resources Association.

Silberer, C., Zarrieß, S., Westera, M., & Boleda, G. (2020, December). Humans meet models on object naming: A new dataset and analysis. In *Proceedings of the 28th international conference on computational linguistics* (pp. 1893–1905). Barcelona, Spain (Online): International Committee on Computational Linguistics.

Xu, Y., Liu, E., & Regier, T. (2020, 08). Numeral Systems Across Languages Support Efficient Communication: From Approximate Numerosity to Recursion. *Open Mind*, *4*, 57-70. Retrieved from https://doi.org/10.1162/opmi_a_00034

Zaslavsky, N., Kemp, C., Regier, T., & Tishby, N. (2018). Efficient compression in color naming and its evolution. *Proceedings of the National Academy of Sciences*, *115*(31), 7937–7942. Retrieved from http://www.pnas.org/content/115/31/7937 doi: 10.1073/pnas.1800521115

Zaslavsky, N., Maldonado, M., & Culbertson, J. (2021). Let's talk (efficiently) about us: Person systems achieve near-optimal compression. In *Proceedings of the 43rd annual meeting of the cognitive science society.*

Zaslavsky, N., Regier, T., Tishby, N., & Kemp, C. (2019). Semantic categories of artifacts and animals reflect efficient coding. In *41st annual meeting of the Cognitive Science Society.*