# A Bayesian Model of Joint Category and Feature Learning

Lea Frermann
ILCC, University of Edinburgh
l.frermann@sms.ed.ac.uk

From day one, infants are exposed to a complex world, and they need to acquire an extraordinary amount of knowledge in order to be able to understand their environment and react meaningfully to it. How do they acquire and represent this knowledge? Structured mental representations, in terms of categories (e.g., `animal`, `furniture`) of concepts (e.g., DOG, CHAIR) have been shown to underlie fundamental cognitive abilities such as learning and using language, and influence the way humans perceive and react to their environment.

We develop the first computational model which investigates process with which children acquire categories, and their associated features. Computational models of cognitive phenomena allow to systematically investigate the influence of the input and processing constraints, and to draw conclusions about human cognitive processing in general [1], and category learning in particular [2, 3].

Our model captures three important characteristics of child category acquisition. First, categories and their features are acquired *jointly*, and the two aspects mutually influence each other [4]. Second, features of categories are *structured* into feature types, which are shared across categories (e.g., `animals` have characteristic behaviors; `tools` have characteristic functions; and both categories have characteristic appearance) [5]. Finally, learning proceeds *incrementally*: children immediately integrate, and utilize, novel information of the input they receive from their environment [6].

We formalize the above characteristics in a Bayesian model which acquires (a) categories, (b) feature types, and (c) category-feature type associations from linguistic input. We approximate the learning environment of the child with child-directed language [7]. While this ignores other modalities (e.g., visual or pragmatic), learning from text corpora allows us to train and test our model on a large scale. We learn categories (e.g., `animal`) of concepts (e.g., DOG, CAT) from linguistic mentions of concepts in their local context, which serves as an approximation of the concepts' features. Concepts with similar features are assigned the same category. Our model represents both categories and feature types as clusters of words (cf. Figure 1).

Given a corpus of concept mentions in context, for each input we (1) draw the category of the observed concept from a global distribution over concepts $k \sim p(\theta)$; (2) draw a feature type from the category-specific distribution over feature types $g \sim p(\phi|k)$; and (3) draw a set of features from the feature type-specific distribution over features $f_i \sim p(\psi|g)$. Importantly, we estimate the parameters of our model incrementally with particle filters [8], an incremental Monte Carlo method. Our model sequentially observes input data (ordered wrt. the age of the addressed child), and updates its parameter estimates on-the-fly with the novel information. Concretely, it maintains a set of parameter samples ('particles') which are individually updated (propagated through time).

Evaluation shows that our model captures important characteristics of the learning process. Intuitively meaningful categories, feature types and their associations emerge (Figure 1). The quality of the learnt clusters improves steadily with incoming data (Figure 2). We also show that our model learns successfully under processing constraints which approach the cognitive constraints that humans are subjected to (individual curves in Figure 2).

## References

[1] T. L. Griffiths, C. Kemp, and J. B. Tenenbaum. Bayesian models of cognition. *Cambridge Handbook of Computational Cognitive Modeling*, 2008.

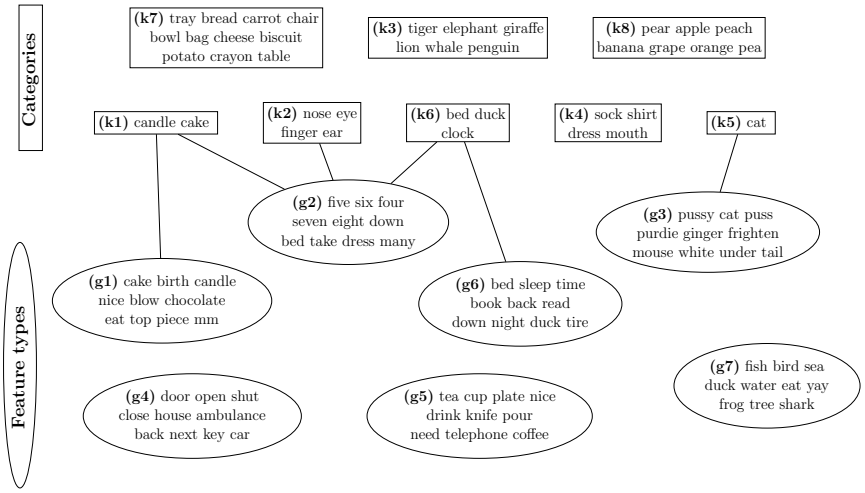


Figure 1: Examples of model induced categories (k1 – k8; top) and feature types (g1 – g7; bottom). Connecting lines indicate a strong association between the category and the respective feature type.

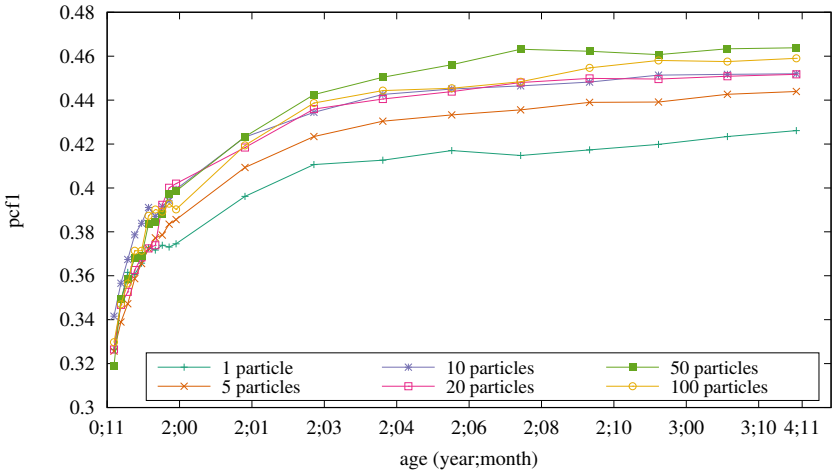


Figure 2: Learning curves of category quality (pcf1; interpolated precision and recall) for models with varying numbers of particles. Fewer particles correspond to more severe processing constraints.

[2] Lea Frermann and Mirella Lapata. A Bayesian Model for Joint Learning of Categories and their Features. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1576–1586, Denver, Colorado, May–June 2015. Association for Computational Linguistics.

[3] Lea Frermann and Mirella Lapata. Incremental Bayesian Category Learning from Natural Language. *Cognitive Science*, 40(6):1333–1381, 2016.

[4] Robert L. Goldstone, Yvonne Lippa, and Richard M. Shiffrin. Altering Object Representations through Category Learning. *Cognition*, 78:27–43, 2001.

[5] Ken McRae, George S. Cree, Mark S. Seidenberg, and Chris McNorgan. Semantic feature production norms for a large set of living and nonliving things. *Behavioral Research Methods*, 37(4):547–59, 2005.

[6] Marc H. Bornstein and Clay Mash. Experience-based and on-line categorization of objects in early infancy. *Child Development*, 81(3):884–897, may 2010.

[7] Brian MacWhinney. *The CHILDES project: Tools for analyzing talk.* Lawrence Erlbaum Associates, Hillsdale, NJ, USA, third edition edition, 2000.

[8] Arnaud Doucet, Nando de Freitas, and Neil Gordon. *Sequential Monte Carlo Methods in Practice.* Springer, New York, 2001.