

# Steady Formulas, Shifting Spells: Estimating Unseen Folktale Diversity

Folger Karsdorp  
KNAW Meertens Institute

Mike Kestemont  
University of Antwerp

Melvin Wevers  
University of Amsterdam

Cultural heritage offers rich accounts of diversity in cultural production. Despite this varied diversity, traditional approaches to quantifying diversity often adopt a simplistic view, treating each unique item as equally different from others. Such methods fall short in capturing the subtleties of inter-item differences, leading to an underestimation of the true diversity. Our study proposes an innovative approach to diversity measurement that accounts for the functional attributes of items. Leveraging the concept of ‘Functional Attribute Diversity’ (FAD) (Walker, Kinzig, and Langridge 1999), we aim to uncover layers of unseen variation in cultural artifacts. In this paper, we present an application of FAD to a collection of Dutch folktales, illustrating the method’s potential to provide a more accurate understanding of cultural diversity. However, FAD only quantifies the observed functional diversity within the database. Given the inherent incompleteness of folktale collections, this observed diversity likely under-estimates the true diversity of folktales. Therefore, we explored the concept of ‘Unseen Functional Diversity’, which aims to estimate the scale of the unobserved diversity in the data and thus enhance our understanding of the full range of variation.

Building on the premise that each folktale carries a unique set of attributes that contribute to its distinctive identity, we operationalized the concept of FAD. The FAD takes into account the distribution of attributes across folktales, considering not just the presence of an attribute but its relative occurrence within the entire collection. We applied this measure to the Dutch Folktale Database (Meder et al. 2016), a collection of stories categorized according to the Aarne-Thompson-Uther index (Uther 2004) – an international system for classifying folktales based on recurring motifs. The motifs in each tale served as functional attributes, allowing us to compute the FAD for the collection and for folktale categories. Applying the FAD measure revealed differences in functional diversity across different folktale categories. ‘Anecdotes and Jokes’ emerged as the most functionally diverse category, while others like ‘Tales of Magic I’ and ‘Animal Tales’ exhibited lower functional diversity.

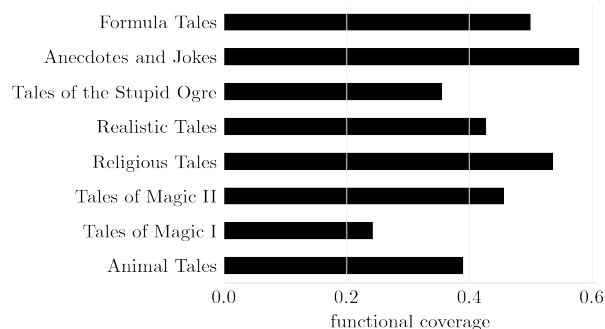


Figure 1: Coverage scores for each category in the Dutch Folktale Database, showing the ratio of observed Functional Attribute Diversity (FAD) to estimated FAD. The graph illustrates the proportion of potential functional diversity that is captured in the observed data for each category, highlighting the significant unseen diversity in most categories.

To estimate the unseen functional diversity, we adopted an estimator that provides a lower bound for the true FAD (Chao et al. 2017). This estimator considers pairs of items in which one or both items are missing from the sample, thus accounting for the unseen diversity. Applying this estimator to the Dutch Folktale Database, we found that the observed diversity only represented 46% of the estimated total diversity, suggesting a significant amount of unobserved functional diversity in the database. Notably, the coverage of functional diversity varied substantially across folktale categories (see Figure 1): ‘Anecdotes and Jokes’ showed the highest coverage (58%), whereas ‘Tales of Magic I’ exhibited the lowest coverage (24%). These findings illustrate the potential of the unseen FAD estimator in providing a less biased understanding of diversity in cultural artifact collections and highlight the need for further data collection and preservation efforts to capture the unobserved diversity. Looking ahead, we believe that our methodology could be extended beyond textual data to other types of cultural artifacts, including image and video collections. For instance, applying this approach to collections of artistic images could help uncover unseen diversity and biases therein, thereby contributing to a more comprehensive understanding of our cultural heritage. Such investigations have profound implications for fields such as computational folklore, digital humanities, and computer vision, where understanding the complexity and diversity of cultural artifacts is of paramount importance.

## Bibliography

- Chao, A., Chiu, C., Colwell, R. K., Magnago, L. F. S., Chazdon, R. L., & Gotelli, N. J. (2017, November). Deciphering the enigma of undetected species, phylogenetic, and functional diversity based on Good-Turing theory. *Ecology*, 98(11), 2914–2929. <https://doi.org/10.1002/ecy.2000>
- Meder, T., Karsdorp, F., Nguyen, D., Theune, M., Trieschnigg, D., & Muiser, I. (2016). Automatic Enrichment and Classification of Folktales in the Dutch Folktale Database. *Journal of american folklore*, 129(511), 78–96. [https://pure.knaw.nl/portal/files/2337691/JAF\\_VVB.pdf](https://pure.knaw.nl/portal/files/2337691/JAF_VVB.pdf)
- Uther, H.-J. (2004). *The types of international folktales: a classification and bibliography, based on the system of Antti Aarne and Stith Thompson*. Suomalainen Tiedeakatemia, Academia Scientiarum Fennica.
- Walker, B., Kinzig, A., & Langridge, J. (1999). Plant Attribute Diversity, Resilience, and Ecosystem Function: The Nature and Significance of Dominant and Minor Species. *Ecosystems*, 2, 95–113.