

## Let's talk about smells: mapping the semantic organization of the English odor vocabulary

Thomas Hörberg  
Gösta Ekman Laboratory, Department of Psychology,  
Department of Linguistics,  
Stockholm University

Most people find it difficult to name familiar odors (e.g. Herz & Engen 1996; Jönsson & Stevenson 2014). With some notable exceptions (e.g. Burenhult & Majid 2011; Wnuk & Majid 2014), most languages lack a vocabulary that primarily is used for odor qualities. Olfaction has been found to be one of the senses that is hardest to verbalize (i.e., in terms of poor linguistic codability, see Majid et al. 2018). Instead of using devoted, abstract terms for describing odors (similar to, for instance, the abstract terms used for describing color qualities), speakers often resort to source-based descriptions (e.g. 'it smells like a lemon') (e.g. Majid et al. 2018).

Relatively little is still known about the vocabulary that is used to describe odors. Yet, in many industries, there is a need for a standardized vocabulary for classifying the perceptual space of odors and facilitating objective communication about smells. Although various olfactory taxonomies have been proposed within specific domains (Burlingame et al. 2004; Fisher et al. 2018; Suffet & Rosenfeld 2007), no descriptive system for describing a wider variety of odor qualities has been agreed upon (Kaeppler & Mueller, 2013).

In this talk, I first give a selective overview of studies investigating the perceptual-semantic organization of odors, many of which have been based on perceptual data in the Dravnieks odor-term rating data set (Dravnieks 1985). I will then present results from two studies investigating the semantic organization of the odor vocabulary in English. In the first study, participants were exposed to 12 odors, differentiated in pleasantness and edibility, and provided ratings of each odor as well as verbal descriptions. In order to explore the semantic variability of these descriptors, we performed multiple factor analysis (MFA), using data on odorant type, odor ratings, and different aspects of the semantics of the descriptors semantics (i.e., semantic category, gustatory / olfactory association and specificity, and imageability / concreteness). Our analyses indicate that the descriptors used by the participants primarily are differentiated in terms of pleasantness and edibility, semantic category and abstractness (e.g. source-based vs. abstract), as well as in terms of vividness and strength.

In the second study, we investigated the semantic organization of the English odor vocabulary using natural language data collected from the Internet. Based on the distribution of words in olfactory and gustatory contexts in a three-billion-word corpus of written English, we identified olfactory descriptors on the basis of their olfactory-semantic association (Iatropoulos et al. 2018). The semantic organization of the identified descriptors was then derived from a distributional-semantic word embedding model (Mikolov et al. 2013), which represents semantic distances as vector distances in a multidimensional space. We analyzed the vector distances with dimensionality reduction and clustering techniques (i.e., PCA and hierarchical clustering) to derive the principal dimensions and clusters of the semantic spaces within and across descriptor types. Our analyses show that odor descriptors primarily are differentiated in terms of pleasantness and edibility, in line with previous studies. They also identify descriptor categories that primarily are related to olfactory and gustatory qualities, similar to those found in earlier studies. In order to further validate our method, we finally derived a perceptual/semantic descriptor space from the Dravnieks perceptual data set (Dravnieks 1985). The perceptual space matched qualitatively with a corpus-based space, thereby showing the viability of our corpus-based approach.

Our method provides a fully data-driven way to identify a large and potentially arbitrary set of olfactory descriptors and their semantic organization. It is applicable to any language for which there is a large enough data set available.

## References

- Burenhult, N., & Majid, A. (2011). Olfaction in Aslian Ideology and Language. *The Senses and Society*, 6(1), 19–29.
- Burlingame, G. A., Suffet, I. H., Khiari, D., & Bruchet, A. L. (2004). Development of an odor wheel classification scheme for wastewater. *Water Science and Technology*, 49(9), 201–209.
- Dravnieks, A. (1985). Atlas of odor character profiles.
- Fisher, R. M., Barczak, R. J., Suffet, I. H., Hayes, J. E., & Stuetz, R. M. (2018). Framework for the use of odour wheels to manage odours throughout wastewater biosolids processing. *Science of The Total Environment*, 634, 214–223.
- Herz, R. S., & Engen, T. (1996). Odor memory: Review and analysis. *Psychonomic Bulletin & Review*, 3(3), 300–313.
- Iatropoulos, G., Herman, P., Lansner, A., Karlgren, J., Larsson, M., & Olofsson, J. K. (2018). The language of smell: Connecting linguistic and psychophysical properties of odor descriptors. *Cognition*, 178, 37–49.
- Jönsson, F. U., & Stevenson, R. J. (2014). Odor Knowledge, Odor Naming, and the “Tip-of-the-Nose” Experience. In B. L. Schwartz & A. S. Brown (Eds.), *Tip-of-the-Tongue States and Related Phenomena* (p. 305–326).
- Kaeppler, K., & Mueller, F. (2013). Odor Classification: A Review of Factors Influencing Perception-Based Odor Arrangements. *Chemical Senses*, 38(3), 189–209.
- Majid, A., Roberts, S. G., Cilissen, L., Emmorey, K., Nicodemus, B., O’Grady, L., ... Levinson, S. C. (2018). Differential coding of perception in the world’s languages. *Proceedings of the National Academy of Sciences*, 115(45), 11369–11376.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. arXiv:1301.3781 [cs].
- Suffet, I. H., & Rosenfeld, P. (2007). The anatomy of odour wheels for odours of drinking water, wastewater, compost and the urban environment. *Water Science and Technology*, 55(5), 335–344.
- Wnuk, E., & Majid, A. (2014). Revisiting the limits of language: The odor lexicon of Maniq. *Cognition*, 131(1), 125–138.