

## Semantic and Connotative Shifts in the Lexicon of Sustainable Transitions

Greta Zella<sup>1</sup>, Tommaso Caselli<sup>1</sup>, Saskia Peels-Matthey<sup>1</sup>, Jan Willem Bolderdijk<sup>2</sup>, and Gerry Wakker<sup>1</sup>

<sup>1</sup>University of Groningen

<sup>2</sup>University of Amsterdam

This contribution presents a study of semantic and connotation shifts in the English lexicon of sustainable transitions using Natural Language Processing (NLP) techniques, based on word embeddings.

Sustainable transitions challenge established conventions, requiring people to adopt innovations. Social psychological surveys show that the perception of conventions (e.g., consuming animal proteins) is changing: they are increasingly seen as immoral (Bolderdijk, 2023, Hassel, 2017). However, habits are not reflecting this shift: animal-based foods consumption is stable (Henchion et al., 2021). Does this mean that survey results are cheap talk and that sustainable innovations are not catching on? Not necessarily: there is a delay between change in people’s mind and in their behaviors, as choices are constrained by the system (e.g., innovations are expensive) and people tend to reject uncommon behaviors (Bolderdijk and Jans, 2021, Lindström et al., 2017).

Since language is one of the main tools to express ideas, it can be seen as a thermometer of social change: analyzing the lexicon of sustainable innovations can provide a clearer picture of what people think. Do changes in words used to express opinions align with survey results? For instance, is the word “milk” expanding its original meaning of “animal secretion” to systematically designate plant-based alternatives? Is the negative shift in the perception of dairy milk expressed in surveys also reflected in language?

We collect data from Reddit and we employ methods grounded in static word embeddings to capture variations in the meaning of lexical items. Our expectation is to detect changes (at semantic and connotative levels) that support evidence from psychological research, i.e., although sustainable innovations have not been widely adopted yet, their perception has shifted.

Two intrinsic dimensions of meaning are analysed: denotation and connotation. The former refers to the semantics of a word, i.e., its definition in the dictionary. The latter identifies a set of emotions that are not formally part of the semantics of an item but are implicitly associated with it (e.g., “vintage” elicits a positive perspective, while “decrepit” induces a negative one) (Basile et al., 2022).

For the identification of semantic shifts, we follow the method by (Gonen et al., 2020), which compares word embeddings of a target word from different corpora (representing different time frames or domains) by looking at the overlap between nearest neighbors. The assumption is that if the overlap between nearest neighbors is not consistent, the contexts in which the word occurs have changed and this indicates that its use - and thus, sense - have also changed. Instances of meaning expansions can be interpreted as cues of ongoing changes in behaviours by (part of) the public. If the word “milk” is systematically used to refer to plant-based drinks, this signals that the latter are considered as convincing substitutes to dairy milk.

To monitor connotation shifts we apply an unsupervised method that employs seed words representing a specific connotative dimension, as well as embedding representations of words from two corpora (Basile et al., 2022). These are used to train a regression model and obtain a hyperplane, i.e., a line that partitions data in two classes. The difference between the cosine distance of the target word from the hyperplane in the two corpora is interpreted as the measure of the connotative shift. In particular, we are interested in the polarity dimension, i.e., positive vs. negative sentiment associated with words. Shifts in polarity will represent key insights about how the perception of specific innovations is evolving.

We have identified **79** words from the domain of food by reviewing **12** EU and US policy documents published in 2020-2023. Comments containing the keywords are collected from 49 communities on Reddit (i.e., subreddits) across a 12-year time span (2010 - 2022). This data allows us to investigate whether there have been semantic and connotation shifts over time and/or across communities.

To showcase how our analysis framework will be extensively applied, we report on initial findings for the word “milk”. This word is particularly interesting given the growing consumption of plant-based alternatives, commonly addressed as types of “milk” in informal discourse, and the ban of this label for their advertisement in the EU. We therefore expect the use of “milk” to reflect a meaning expansion that would take this into account. For this word, we collected more than 1 million comments from 16 subreddits over the targeted period, which we divided in four blocks<sup>1</sup>. For each block, we have created static word embedding representations using Word2Vec (Mikolov et al., 2013).

The word embedding method shows no diachronic semantic shift, since the nearest neighbors in the four corpora considerably overlap. However, by exploring the actual contexts of occurrence, we have found an increase in frequency of modifiers specifying the source (i.e., animal vs. plant) of the product as reported in Figure 1. This suggests that there is an ongoing meaning expansion, with plant-based drinks being increasingly seen as alternatives. Concerning connotation shifts, we observe a slightly positive shift in time, but a negative one in specialised communities against generic ones.

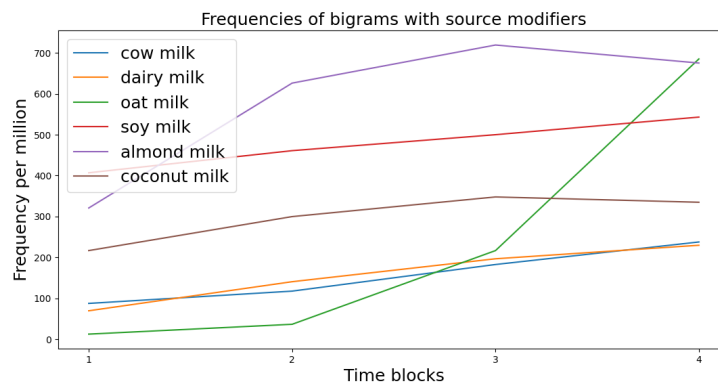


Figure 1: Most frequent bigrams across time blocks containing the noun “milk” and source modifiers.

Overall, the distributional contexts of “milk” are minimally affected and thus the static embeddings fail to capture the semantic expansion. At the same time, the

<sup>1</sup> Each block combines 3 years, except for the first one which consists of 4 years, as significantly less data is available for the period 2010-2013.

negative connotative shift across communities indicates that changes in the public are in place but still limited to niches. We are currently running experiments on the whole dataset. We will further compare the static embedding methods with contextualized ones from Large Language Models and develop a manually annotated, balanced and representative subcorpus to validate our findings.

The manually annotated dataset for validation will be randomly sampled from each subreddit and for each candidate word. Three annotators will indicate, for each comment, the connotation of the candidate word and will annotate its meaning. Correlation with results from computational methods will be computed.

## References

- Valerio Basile, Tommaso Caselli, Anna Koufakou, and Viviana Patti. *Automatically Computing Connotative Shifts of Lexical Items*, pages 425–436. 06 2022. ISBN 978-3-031-08472-0. doi: 10.1007/978-3-031-08473-7\_39. URL [https://link.springer.com/chapter/10.1007/978-3-031-08473-7\\_39](https://link.springer.com/chapter/10.1007/978-3-031-08473-7_39).
- Jan Willem Bolderdijk. Words speak louder than actions. Inaugural Speech, November 24 2023. URL <https://dare.uva.nl/search?identifier=abfb93ce-e867-4305-9cdd-b3f578f62447>.
- Jan Willem Bolderdijk and Lise Jans. Minority influence in climate change mitigation. *Current Opinion in Psychology*, 42:25–30, 2021. ISSN 2352-250X. doi: <https://doi.org/10.1016/j.copsyc.2021.02.005>. URL <https://www.sciencedirect.com/science/article/pii/S2352250X21000154>. *Psychology of Climate Change* (2021).
- Hila Gonen, Ganesh Jawahar, Djamel Seddah, and Yoav Goldberg. Simple, interpretable and stable method for detecting words with usage change across corpora. In *Association for Computational Linguistics*, pages 538–555, 2020. doi: 10.18653/v1/2020.acl-main.51. URL <https://aclanthology.org/2020.acl-main.51/>.
- Bodil Hassel. Perceptions of the milk industry: The role of social learning and the new media. Second cycle, a2e, Uppsala: SLU, Dept. of Urban and Rural Development, 2017. URL <http://urn.kb.se/resolve?urn=urn:nbn:se:slu:epsilon-s-6415>.
- Maeve Henchion, Aidan P. Moloney, John J. Hyland, Jesko Zimmermann, and Sinéad McCarthy. Review: Trends for meat, milk and egg consumption for the next decades and the role played by livestock systems in the global production of proteins. *Animal*, 15:100287, 2021. ISSN 1751-7311. doi: <https://doi.org/10.1016/j.animal.2021.100287>. URL <https://www.sciencedirect.com/science/article/pii/S1751731121001300>. Sustainable livestock systems for high-producing animals.
- Björn Lindström, Simon Jangard, Ida Selbing, and Andreas Olsson. The role of a "common is moral" heuristic in the stability and change of moral norms. *Journal of Experimental Psychology: General*, 147, 09 2017. doi: 10.1037/xge0000365. URL <https://psycnet.apa.org/doiLanding?doi=10.1037%2F%2F0000365>.
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in*

*neural information processing systems*, pages 3111–3119, 2013. doi: 10.48550/arXiv.1310.4546. URL <https://arxiv.org/abs/1310.4546>.