UNDERSTANDING THE DYNAMICS OF CLIMATE-CRUCIAL FOOD CHOICE BEHAVIOURS USING DISTRIBUTIONAL SEMANTICS

Claudia Haworth *
Arts and Sciences BASc
University College London
London, United Kingdom
cghaworth1@sheffield.ac.uk

Gabriella Vigliocco
Department of Experimental Psychology
University College London
London, United Kingdom
g.vigliocco@ucl.ac.uk

ABSTRACT

Developed countries must make swift movements toward plant-based diets in order to mitigate climate change and maintain food security. However, researchers currently lack clear insight into the psychological dimensions that influence food choice, which is necessary to encourage the societal adaptation of new diets. In this project, we use Skip-gram word embeddings trained on the ukWaC corpus as a lens to study the implicit mental representations people have of foods. Our data-driven insights expand on findings from traditional, interview-based studies by uncovering implicit mental representations, allowing a better understanding the complex combination of conscious and sub-conscious processes surrounding food choice. In particular, our findings shed light on the pervasiveness of meat as the 'centre' of the meal in the UK.

1 INTRODUCTION

According to current projections, by 2050, the emissions budget available per capita under the IPCC's $1.5°$ C target will be swallowed entirely by diets high in ruminant animals (Ritchie et al., 2018). Dietary change will be forced by environmental and economic factors, and the food equity gap will widen (Garnett, 2013), meaning the developed world must adapt diets compatible with a "$1.5°$ world" (Schleusnner et al., 2016, p.832). However, by definition, 'sustainable diets' must not only have low environmental impacts, but be nutritionally complete, economically accessible, and culturally sensitive if they are to be widely adapted by society (Perignon et al., 2016; Macdiarmid & Whybrow, 2019). We must understand what drives food choice before we can strive to change it.

To understand decisions around food fully, we need an holistic approach which considers a range of factors. For example, consider the apparent cognitive dissonance between desires to eat sensorially indulgent foods (Graça et al., 2015; Olsen, 2008; Armstrong Soule & Sekhon, 2018) and intentions to eat healthily (Pieniak et al., 2010; Perignon et al., 2016), or in ways that satisfy social norms (Bogueva et al., 2017; Carlucci et al., 2015; Abbots & Coles, 2013; Pohjolainen et al., 2015). Most existing studies investigate a single influence on food choice using explicit methods such as consumer surveys or focus groups (for recent examples, see Morales & Higuchi (2018) or Markowski & Roxburgh (2019)), but these explicit methods rarely capture crucial implicit influences, such as cognitive and emotional associations between different foods (Köster, 2003; Dalenberg et al., 2014).

In this paper, we investigate what we believe to be the currently little-explored dimension of implicit determinants of food choice. To do so, we assume that language can be used as a window into how people think and feel as shaped by culture and habitual behaviours - thus providing insight into both the explicit and implicit knowledge people have about foods. We then use Machine Learning methods for analysis; specifically a Distributional Semantic Model. DSMs are not only valuable for Natural Language Processing, but also for modelling human cognitive relations and semantic memory (Jones et al., 2015). By examining the behaviour of food-word embeddings within this model, we are able to consider food choice as a mixture of explicit and implicit mental representations,

*Now at the Department of Computer Science, University of Sheffield rather than as the product of a single explicit factor. It is this data-driven approach that allows us to model how UK citizens implicitly think and feel about foods.

2 Model and study design

The basic design of our study was as follows: we defined a set of food-words comprehensively representative of diets across the UK. We trained the Skip-gram algorithm (Mikolov et al., 2013) on the ukWaC corpus (Baroni et al., 2008), and studied the behaviour of food-word embeddings using unsupervised learning, bootstrapping of psycholinguistic variables, and close textual analysis.

2.1 Choice of seed words

We obtained a total of 925 food terms (including all variants in spelling, pluralisations and synonyms) by cross referencing Appendix R to the National Diet and Nutrition Survey (PHE, 2018b)[1] with WordNet (Princeton University, 2010) and BBC Food. Words with fewer than twenty occurrences in the corpus were removed in line with the 'Sinclair cut-off'. (Baroni et al., 2008)). Words with polysemic meanings of very high frequency (i.e. 'date', 'Turkey') were removed. 14 Native English speakers were consulted over removal of words of more ambiguous polysemy (i.e. 'roll', 'chop'). Our final list contained 640 terms, including multi-word expressions like 'baked beans'.

2.2 Choice of Distributional Semantic Model and Corpus

Baroni et al. (2014) demonstrate that neural, context-predicting models (particularly Mikolov et al.'s Skip-gram (2013)) provide a very good fit to human performance in tasks such as analogy and context categorization. Skip-gram has also been used to accurately extrapolate psycholinguistic variables using a k-nearest neighbour approach (Mandera et al., 2015), suggesting the embeddings latently encode psychologically valid dimensions. We therefore assumed that the embeddings derived from Skip-gram could be considered a reasonable proxy for human semantic memory.

Our corpus needed to balance high-quality examples of UK English with the requirement of sufficient data to train Skip-gram for meaningful semantic representations. Our chosen corpus was the ukWaC, a web-crawled corpus containing 1.9 billion tokens extracted from 2.69 million documents (Baroni et al., 2008). The ukWaC comprises varied content extracted from .uk web domains (including academic literature, advertisements and public service documents), which was extensively linguistically post-processed to minimise the quantity of data 'noise'.

3 Results

3.1 Overall behaviour of food embeddings

To investigate how foods are represented and organised in the semantic memory of UK individuals, we looked for natural categories and groupings of the food-word embeddings using the unsupervised learning technique of k-means clustering (MacQueen, 1967).

Since the inherent randomness in the initialisation of k-means centroids can occasionally lead to a sub-optimal solution, we performed 100 tests of the optimum number of clusters using a combination of cluster validity indices (Silhouette, Davies-Bouldin and Caliński-Harabasz). We found the optimum number of clusters to be $k = 3$.

We performed Principal Co-ordinates Analysis on the 300-dimensional food word-embeddings *only*, to produce a 2-dimensional, visualisable space. Figure 1 shows how food-word embeddings split naturally into three categories: **Fish and Seafood**; **Edible Plants** i.e. fruit, vegetables, nuts, seeds; and **Miscellaneous**, which is a mixture of meat, savoury and sweet foods, and animal derivatives[2].

[1] Appendix R ('Main and subsidiary food groups and disaggregation categories') provides a detailed list of all foods recorded in four-day food diaries collected from a sample of UK individuals (PHE, 2018a)

[2] Due to space constraints, Figure 1 presents only a subset of all food words used in the model (for readability). See Figure 3 in the appendices for a larger-scale version of this visual representation with all food-word embeddings studied, and Figure 4 for a larger scale visual representation of the Miscellaneous category.

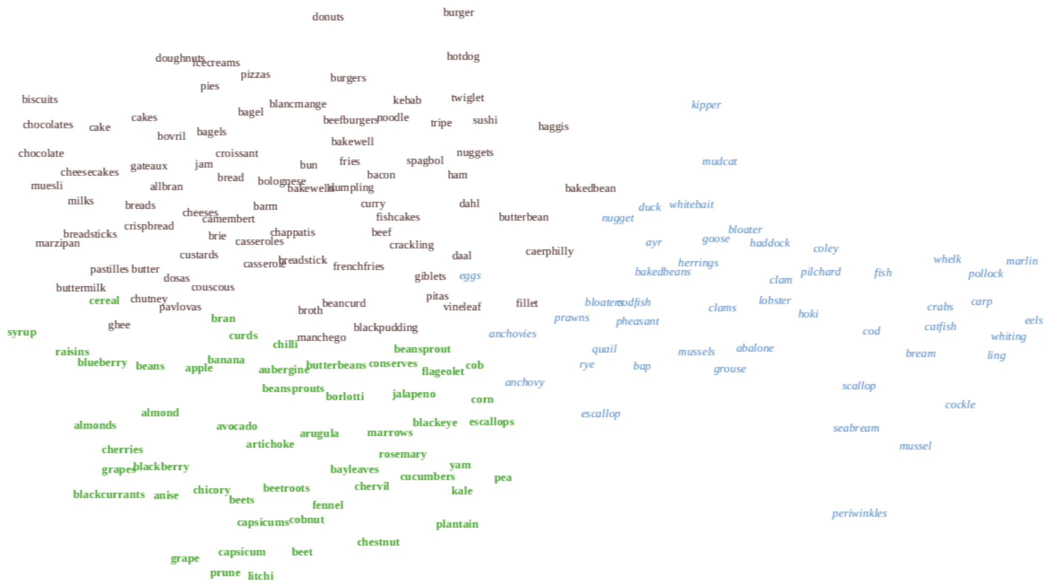

Figure 1: Arrangement of food-word embeddings according to k-means clustering, $k = 3$. Green, bold-face words represent the **Edible Plants** category; blue, italicised words are the *Fish and Seafood* category; and the brown, standard-face words are the 'Miscellaneous' category.

3.2 BOOTSTRAPPING VALENCES

To analyse the affect associated with the different foods, we followed Mandera et al.'s approach of bootstrapping valence scores (i.e., the extent to which a given word elicits positive, negative or no emotional associations) by averaging the valences of the k-nearest neighbours (2015), with $k = 10$ and neighbourhoods defined by cosine similarity. Figure 2 shows a box-plot of these estimated valence scores, grouped by k-means cluster; it is clear that across the board, foods have positive affective associations ($\bar{x} = 0.96, P(\mu \neq 0) < 0.001$). Miscellaneous foods have the strongest positive associations, mostly due to the presence of sweet foods in the category ($\mu_{sweet} = 1.33, \mu_{misc} = 1.12$). Fish and Seafood have the least positive associations, though averages are still above neutral ($\bar{x} = 0.50; P(\mu \neq 0) \leq 0.0001$). Both parametric (2-sample t-test) and non parametric (Mann-Whitney U-test) were applied as the distributions were unknown; in both tests, the differences in mean valence between the three categories was statistically significant ($P(\mu_1 \neq \mu_2) < 0.001$ for each pairwise comparison, after a Bonferronni correction for $b = 3$ tests).

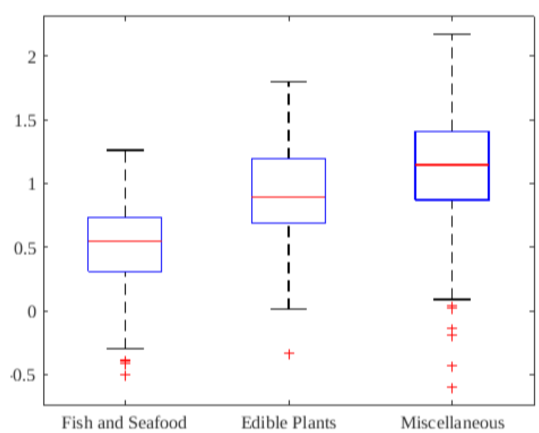

Figure 2: Box-plot of bootstrapped valences for the k-means clusters as defined in Section 3.2

3.3 Analysis of lexical neighbourhoods

The 10 nearest neighbour adjectives to each word in each k-means cluster were thematically coded using a similar scheme to Papies (2013), by two independent, native speakers of UK English with good knowledge of the project. Inter-coder agreement was 79%, with discrepancies in coding resolved through discussion between coders. Themes for coding were: 'sensory' (taste, texture); 'situational' (time/place of eating); 'hedonic' (judgement i.e. 'yummy', 'gross'); 'food preparation' (descriptions and verbs, including past-participles such as 'roasted,' 'fried'); 'nutrition'; 'other foods' (any food noun); and 'other - unrelated.' Results are presented in Table 1. Notable results (bold-faced) are that the Fish and Seafood category is associated with many non-food contexts but few sensory attributes, and Miscellaneous foods are the only group to be associated with hedonic language.

	Fish and Seafood	Edible plants	Miscellaneous
Sensory	**6.3**	26.8	22
Situational	0	0	0.9
Hedonic	0	0	**5.4**
Food preparation	32.6	26.8	40.9
Nutrition	0	0.1	0.4
Other foods	48.3	44.7	29.7
Other - unrelated	**12.8**	1.6	0.7

Table 1: Percentage of neighbour-adjectives in each description category for the three food clusters

4 Discussion and conclusions

Our headline result is the discovery that people in the UK mentally represent foods in three main categories: Fish and Seafood (FS), Edible Plants (P), and a Miscellaneous group (M) including meats, dairy products, and composite foods.

FS foods are described using a small proportion of sensory words, ($< 7\%$) and a low variety of food preparation terms (over 40% of these being 'breadcrumbed', 'grilled' and 'fried'), which indicates unfamiliarity with the food group. Given that unfamiliar foods are expected to be less sating (Brunstrom et al., 2008), and unfamiliarity with fish preparation associates fish with inconvenience (Olsen et al., 2007; Thorsdottir et al., 2012), FS foods forming their own category seems unsurprising.

With meat represented closely with composite foods like curries, pies and sandwiches (c.f. standard 'main meals'), the notion that "it's not a meal without meat in it" ((Macdiarmid et al., 2016)) appears implicitly in UK representations of foods. Matching with Yates & Warde's analysis of British eating habits (2015) we see evidence that meat is at the 'centre' of the meal; vegetables and fruits in their own, separate category relegates them to 'trimmings'. Indeed, with meat at the centre of the 'standard' foods category, we can see how the social environment is implicitly unsupportive of plant-based diets ((Markowski & Roxburgh, 2019; Macdiarmid et al., 2016) and why non-meat-eaters are perceived as "disrupting social conventions" ((Markowski & Roxburgh, 2019)).

Bootstrapping the valences of different foods revealed that in general, emotions toward food are positive ($\bar{x} = 0.96$) - the accuracy of these bootstrapped valences is validated by the known existence of positive hedonic asymmetry among consumer emotions (Schifferstein & Desmet, 2010). Climate crucial foods (meats) actually have a relatively low mean valence ($\mu_{meat} = 0.89, \mu_{misc} = 1.12$), which may suggest it is attachment to the implicit concept of the meal that keeps meat at the centre of the UK diet, rather than the desire for meat itself. Moreover, because only M foods are described using a wide range of hedonic and sensory attributes, a potentially useful strategy could be to increase the use of indulgent language for describing plant-based foods, given that style of description has been shown to not only encourage people to choose foods with more "indulgent" names (Turnwald et al., 2017), but to pre-bias them into actually perceiving the food as tastier, more satisfying and more caloric (Wansink et al., 2005).

ACKNOWLEDGMENTS

This research was supported by a European Research Council Advanced Grant (ECOLANG, 743035) and Royal Society Wolfson Research Merit Award (WRM\R3\170016) to GV.

REFERENCES

E-J Abbots and Benjamin Coles. Horsemeat-gate. *Food, Culture and Society*, 16(4):535–550, 2013. doi: 10.2752/175174413X13758634981976.

Catherine Anne Armstrong Soule and Tejvir Sekhon. Preaching to the middle of the road: Strategic differences in persuasive appeals for meat anti-consumption. *British Food Journal (online)*, 2018. doi: https://doi.org/10.1108/BFJ-03-2018-0209.

Marco Baroni, Silvia Bernadini, Adriano Ferraresi, and Eros Zanchetta. The WaCky Wide Web: A Collection of Very Large Linguistically Processed Web-Crawled Corpora. *Language Resources and Evaluation*, 43(3):209–226, 2008. doi: https://doi.org/10.1007/s10579-009-9081-4.

Marco Baroni, Georgiana Dinu, and Germán Kruszewski. Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, pp. 238–247, 2014. doi: 10.3115/v1/P14-1023.

Diana Bogueva, Dora Marinova, and Talia Raphaely. Reducing meat consumption: the case for social marketing. *Asia Pacific Journal of Marketing and Logistics*, 29(3):477–500, 2017. doi: https://doi.org/10.1108/APJML-08-2016-0139.

Jeffrey M Brunstrom, Nicholas G Shakeshaft, and Nicholas E Scott-Samuel. Measuring 'expected satiety' in a range of common foods using a method of constant stimuli. *Appetite*, 51:604–614, 2008. doi: 10.1016/j.appet.2008.04.017.

Domenico Carlucci, Guiseppe Nocella, Biagia De Devitiis, Rosaria Viscecchia, Francesco Bimbo, and Gianluca Nardone. Consumer purchasing behaviour towards fish and seafood products. Patterns and insights from a sample of international studies. *Appetite*, 84:212–227, 2015. doi: http://dx.doi.org/10.1016/j.appet.2014.10.008.

Jelle R Dalenberg, Swetlana Gutjar, Gert J ter Horst, Kees de Graaf, Remco J Renken, and Gerry Jager. Evoked Emotions Predict Food Choice. *PLoS ONE*, 9(12):1–16, 2014. doi: DOI:10.1371/journal.pone.0115388.

Tara Garnett. Food sustainability: problems, perspectives, and solutions. *Proceedings of the Nutrition Society*, 72:29–39, 2013. doi: doi:10.1017/S0029665112002947.

João Graça, Maria Manuela Calheiros, and Abílio Oliveira. Attached to meat? (Un)Willingness and intentions to adopt a more plant-based diet. *Appetite*, 95:113–125, 2015. doi: https://doi.org/10.1016/j.appet.2015.06.024.

Michael N. Jones, Jon Willits, and Simon Dennis. Models of Semantic Memory. In Jerome R. Busemeyer, James T. Townsend, Wang Zheng, and Ami Eidels (eds.), *Oxford Handbook of Mathematical and Computational Psychology*. Oxford University Press, Oxford: UK, 2015. URL `DOI:10.1093/oxfordhb/9780199957996.013.11`.

Egon P Köster. The psychology of food choice: some often encountered fallacies. *Food Quality and Preference*, 12(5-6):359–373, 2003. doi: https://doi.org/10.1016/S0950-3293(03)00017-X.

J. I. Macdiarmid and S. Whybrow. Nutrition from a climate change perspective. *Proceedings of the Nutrition Society*, pp. 1–8, 2019. doi: doi:10.1017/S0029665118002896.

J. I. Macdiarmid, Flora Douglas, and Jonina Campbell. Eating like there's no tomorrow: Public awareness of the environmental impact of food and reluctance to eat less meat as part of a sustainable diet. *Appetite*, 96:487–493, 2016. doi: http://dx.doi.org/10.1016/j.appet.2015.10.011.

J MacQueen. Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1: Statistics: 281–297, 1967. URL https://projecteuclid.org/euclid.bsmsp/1200512992. University of Califormnia Press, Berkeley, California.

Pawel Mandera, Emmanuel Keuleers, and Marc Brysbaert. How useful are corpus-based methods for extrapolating psycholinguistic variables? *The Quarterly Journal of Experimental Psychology*, 68(8):1623–1642, 2015. doi: 10.1080/17470218.2014.988735.

Kelly L Markowski and Susan Roxburgh. "If I became a vegan, my family and friends would hate me:" Anticipating vegan stigma as a barrier to plant-based diets. *Appetite*, 135:1–9, 2019. doi: https://doi.org/10.1016/j.appet.2018.12.040.

Tomas Mikolov, Kai Chen, G.S. Corrado, and Dean Jeffrey. Efficient Estimation of Word Representations in Vector Space. *Proceedings of Workshop at ICLR*, 2013. URL arXiv:1301.3781v3.

L Emilio Morales and Angie Higuchi. Is fish worth more than meat? - How consumers' beliefs about health and nutrition affect their willingness to pay more for fish than meat. *Food Quality and Preference*, 65:101–109, 2018. doi: https://doi.org/10.1016/j.foodqual.2017.11.004.

Svein Ottar Olsen. Antecedents of Seafood Consumption Behaviour. *Journal of Aquatic Food Product Technology*, 13(3):79–91, 2008. doi: https://doi.org/10.1300/J030v13n03_08.

Svein Ottar Olsen, Joachim Scholderer, Karen Brunsø, and Wim Verbeke. Exploring the relationship between convenience and fish consumption: a cross-cultural study. *Appetite*, 49:84–91, 2007. doi: doi:10.1016/j.appet.2006.12.002.

Esther K Papies. Tempting food words activate eating simulations. *Frontiers in Psychology*, 4:1–9, 2013. doi: 10.3389/fpsyg.2013.00838.

Marlène Perignon, Florent Vieux, Louis-Georges Soler, Gabriel Masset, and Nicole Darmon. Improving diet sustainability through evolution of food choices: review of epidemiological studies on the environmental impact of diets. *Nutrition Reviews*, 75(1):2–17, 2016. doi: doi: 10.1093/nutrit/nuw043.

Public Health England PHE. National Diet and Nutrition Survey: Appendix B Methodology for Years 7 and 8 of the NDNS RP, 2018a. URL https://www.gov.uk/government/statistics/ndns-results-from-years-7-and-8-combined.

Public Health England PHE. National Diet and Nutrition Survey: Appendix R Main and subsidiary food groups and disaggregation categories, 2018b. URL https://www.gov.uk/government/statistics/ndns-results-from-years-7-and-8-combined.

Z Pieniak, Wim Verbeke, and Joachim Scholderer. Health-related beliefs and consumer knowledge as determinants of fish consumption. *Journal of Human Nutrition and Dietetics*, 23:480–488, 2010. doi: doi:10.1111/j.1365-277X.2010.01045.x.

Pasi Pohjolainen, Markus Vinnari, and Pekka Jokinen. Consumers' perceived barriers to following a plant-based diet. *British Food Journal*, 117(3):1150–1167, 2015. doi: https://doi.org/10.1108/BFJ-09-2013-0252.

Princeton University. About wordnet. *Princeton University*, 2010.

Hannah Ritchie, David S Reay, and Peter Higgins. The impact of global dietary guidelines on climate change. *Global Environmental Change*, 49:46–55, 2018.

Hendrik N J Schifferstein and Pieter M A Desmet. Hedonic asymmetry in emotional responses to consumer products. *Food Quality and Preference*, 21:1100–1104, 2010. doi: 10.1016/j.foodqual.2010.07.004.

Carl-Friedrich Schleusnner, Joeri Rogelj, Michiel Schaeffer, Tabea Lissner, Rachel Licker, Erich M Fischer, Reto Knutti, Anders Levermann, Katja Frieler, and William Hare. Science and policy characteristics of the Paris Agreement temperature goal. *Nature Climate Change*, 6, 2016. doi: DOI:10.1038/NCLIMATE3096.

Fanney Thorsdottir, Kolbrun Sveinsdottir, Fridrik H Jonsson, Gunnthorum Einarsdottir, Inga Thorsdottir, and Emilia Martinsdottir. A model of fish consumption among young consumers. *Journal of Consumer Marketing*, 29(1):4–12, 2012. doi: DOI10.1108/07363761211193000].

B Turnwald, D Boles, and A Drum. Association between Indulgent Descriptions and Vegetable Consumption: Twisted Carrots and Dynamite Beets. *JAMA International Medicine*, 177(8):1216–1218, 2017. doi: doi:10.1001/jamainternmed.2017.1637.

Brian Wansink, Koert van Ittersum, and James E Painter. How descriptive food names bias sensory perceptions in restaurants. *Food Quality and Preference*, 16:393–400, 2005. doi: 10.1016/j.foodqual.2004.06.005.

Luke Yates and Alan Warde. The evolving content of meals in Great Britain. Results of a survey in 2012 in comparison with the 1950s. *Appetite*, 84:299–308, 2015. doi: 10.1016/j.appet.2014.10.017.

A APPENDIX

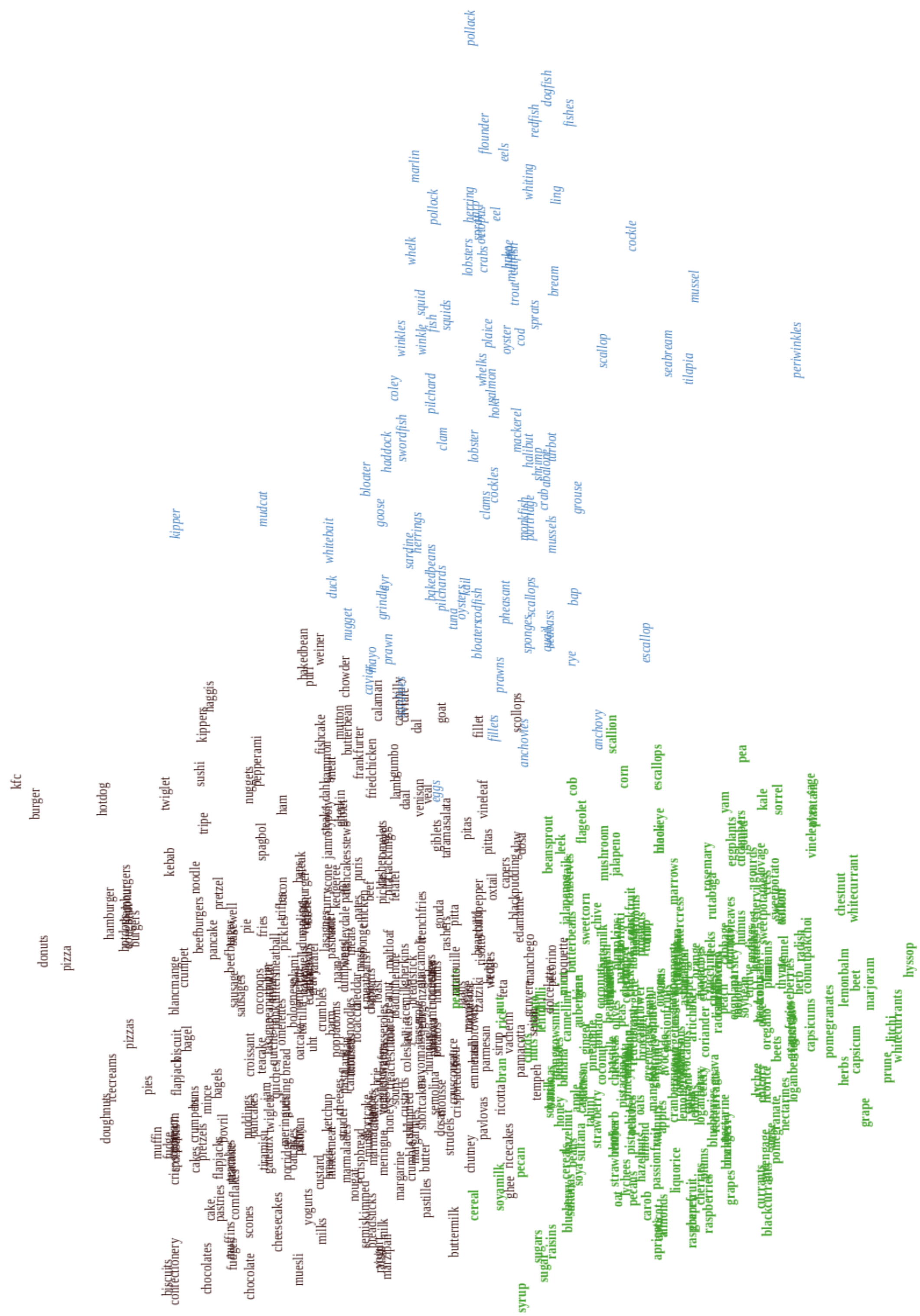

Figure 3: Arrangement of all food-word embeddings according to k-means clustering, $k = 3$. Green, bold-face words represent the **Edible Plants** category; blue, italicised words are the *Fish and Seafood* category; and the brown, standard-face words are the 'Miscellaneous' category.

Figure 4: Miscellaneous cluster, coloured by 'standard' food categories: red, bold-face words are Meat; black, standard-face words are Animal Derivatives (i.e. dairy, eggs); pink, italicised words are Sweet composite foods; blue, italicised words are Savoury composite foods; and green, italicised words are Other composite foods.