

Healthy Food Depiction on Social Media: The Case of Kale on Twitter

Maija Kāle¹[0000-0002-6951-9009] and Ebenezer Agbozo²[0000-0002-2413-3815]

¹ University of Latvia, Riga, Latvia

² Ural Federal University, Yekaterinburg, Russian Federation
maijakale@gmail.com

Abstract. This food computing study analyzes Twitter microblog entries related to food. A particular attention here is paid to one specific healthy food – kale. Such an approach is chosen due to kale’s popularity as a healthy food. By applying sentiment measurement, authors contribute to the understanding of text-based stories behind healthy food, and conclude that kale offers clear benefits for consumers via its specific health-related aspects (anti-inflammatory, immune system boosting, etc.) while taste is being left outside the discourse. The lack of any references to taste in kale-related social media stories leads to the conclusion that healthy food descriptions in general are short of hedonistic manifestations. Subsequently, this can be seen as one of the drawbacks of healthy food presentation textually vis-a-vis comfort foods. With this study authors add to the knowledge of the most efficient methods in shaping healthier consumer behaviors and inspire for further research of text-based information related to food.

Keywords: Food Computing, Kale, Sentiment-Token Bigram, NLP, Health, Taste, Word Association, Food Blog, Social Media.

1 Introduction

The idea of a human as a rational wealth maximizer [1] can be best dissected when looking at the mechanisms lying behind the food choice, which is a complex phenomenon impacted by factors such as context, culture, social class, emotions, mood et al. but where the usually heavy-weight factor of health remains somewhat outside the decision making. The rationality of a human can be questioned, considering the global data on obesity and other long-term consequences of lifestyle-related illnesses [2]. Lifestyle-induced obesity, Type 2 diabetes and cardiovascular diseases are on the rise, leaving public health policy makers with little success to report.

Meanwhile, the world has increasingly become a macroscopic sensor of our eating habits and food cultures, evidenced via multiple photos, hashtags and documentations of food and its background stories. “As food is central in our life, a significant fraction of online content is about food” [3]. The evolution of homo sapiens from hunter-gatherers to super-consumers [4] has led to a chicken-and-egg conundrum regarding the main culprit – whether it is the consumer who gives preference to sweet and fatty

fast food or whether we should put the blame on multinational corporations, i.e. food industry actors that produce a plethora of increasingly sugary, salty and fat-intense foods, widely advertising them on social media.

Food consumption is a complex phenomenon involving many aspects, identifiable to a varying degree. When it comes to the consumer's choice of food, the control over decision-making is limited, leaving the food choice dependant on unpredictable outside triggers that appear in many formats, including those of food images and texts. By analyzing texts on social media related to food, we can see that food choice can serve as a marker of belonging to a certain social class, because "popular stereotypes about who shops at farmer markets, makes their own kombucha, and drinks shade-grown pour-over coffee reflect the reality that how we eat communicates important information about our identities, including class background and status." [5].

Through the social media text corpus, we can also look into the ways the food is described and derive sentiments associated with it. At the same time, we bear in mind that the field of food computing itself is a challenging area requiring a cross-sectoral research approach where cognitive science and social science interlink with computer science. This adds to the complexity of food computing choice analysis because the notions prevalent in these science fields are often conflicting or at best not self-explanatory. We will illustrate these challenges in our work where we analyze the consumption of a particular food - kale, which is increasingly popular worldwide and widely considered to be healthy.

2 Can Healthy Food Be Tasty?

In terms of healthy vs unhealthy food choices, the dichotomy between health and taste is pronounced, as "many consumers still tend to over-consume energy-dense foods because of two factors that work in combination: (1) foods that are "unhealthy" are widely associated with being tasty" [6], and (2) taste is the main driver of food decisions [7]. One of the key hindrances towards consuming healthy food is the unhealthy=tasty intuition: "Previous attempts on the part of policy makers to induce healthy food choices by educating consumers have not yet led to major improvements or fundamental changes in actual behavior [...] the objectives of health and taste often conflict - and [...] taste usually prevails in food decision making." [8]

This intrinsically implies that the level of hedonism is higher in relation to unhealthy foods, which are considered tastier than their healthy counterparts. Choosing chocolate as one of the foods best depicting the level of hedonism, we can see that "a 'reduced fat' label, for example, negatively affects the acceptance of chocolate bars" [8]. Meanwhile, descriptions of desserts and wine abound in sensuality. Taking into account the central role food plays in human lives, "...it is no surprise the hashtag #foodporn is one of the most popular hashtags on social media, especially on visual media such as Instagram, often accompanied with a close-up of a tantalizing dish." [9]. In the world of #foodporn the arena of sensuality is dominated by cakes and chocolate, and research reveals that "in the 1980s wine reviewers began to increase their use of the body as a metaphor, starting to use words like fleshy, muscular, sin-

ewy, big-boned, or broad-shouldered.” [10] Not many of these connotations are found in the healthy food discourse.

Imbuing healthy food with the desired portion of hedonism and presenting it as a solution to unhealthy diets has its grounding in the multisensory field. “Research on overeating assumes that pleasure must be sacrificed for the sake of good health. Contrary to this view, the authors show that focusing on sensory pleasure can make people happier and willing to spend more for less food, a triple win for public health, consumers, and companies alike” [11]. Thus, pleasure and hedonism matter, and there is potential for embedding and highlighting those qualities in healthy food. Using text as a proxy for such embeddings could be one of the solutions for enhancing more healthy eating habits globally. Currently, however, we argue that the overall narrative surrounding healthy food is not sufficiently engaging on an emotional level and lacking in hedonistic connotations. As illustrated by Kāle and Agbozo [12], healthy food is more associated with ease of preparation and seasonality while lacking any hedonistic attributes.

3 Cross-Sectoral Challenges of Discussing Food Choice Determinants

There is a growing disagreement among the researchers across varying interlinked research fields on what actually matters when consuming food. Much of the research focus has lied on the juxtaposition between taste and healthiness since “the objectives of health and taste often conflict - and (...) taste usually prevails in food decision making.” [13]. Taste as the right combination of flavors has dominated almost all recipe-related data analysis - “to test the “food pairing hypothesis,” a recent theory proposes that tasty recipes are more likely to combine ingredients that share flavor compounds.” [10]. Assumptions behind such food pairing research are that 1) taste is determined by the chemical profile of the food, 2) taste is the most important determinant in food choice. As the research group concludes: “Although many factors such as colors, texture, temperature, and sound play an important role in food sensation, palatability is largely determined by flavor, representing a group of sensations including odors (due to molecules that can be bind olfactory receptors), tastes (due to molecules that stimulate taste buds), and freshness or pungency (trigeminal senses). Therefore, the flavor compound (chemical) profile of the culinary ingredients is a natural starting point for a systematic search for principles that might underlie our choice of acceptable ingredient combination.” [14].

Another stream of research, which is more qualitative and based on cognitive science research methods rather than statistical analysis of big data, prioritizes all other factors over taste in categorizing the food as tasty or not. “It is time to accept that growing body of gastrophysics evidence demonstrating that the environment, not to mention the plateware, dish-naming, cutlery and so on, all exert an influence over the tasting experience.” [15]. It turns out that taste is a rudimentary attribute of the whole food consumption process and is primarily related to the so-called food safety test, which is the most important instinctive test humans perform when choosing to con-

sume food. “We mostly do not pay attention to what we taste. Rather, our brains just do a quality check first, to ensure that there is nothing wrong with the food or drink and that it tastes pretty much as we expected (or predicted) that it should” [15]. Here the following assumptions are put into operation: 1) many other factors are more important than taste, 2) the perceived content of food prevails over the chemical content (“perceived content” here means contextual factors - visual representation of the food and its surrounding narrative).

The controversy between these theoretical standpoints and the assumptions they operate on stem first and foremost from the challenge of cross-sectionality *per se*. While the research on recipes or any available large food-related datasets focuses on statistical analysis and views the existing data in terms of their analytical utility, the experiment-based research focuses instead on such cognitive aspects that are not yet well-documented in large-scale data. This discrepancy raises an important question - to what extent can we pair theories and ideas from one field (in this case cognitive science) with another field’s methodology (big data analytics, statistics)? As illustrated via data analysis of food micro blogging entries on Twitter, the notion of complexity as a proxy for food’s hedonic attributes was scantily represented in healthy food discourse [12]. It shows that food computing can serve, albeit to a limited degree, as an intersection point between cognitive science and data analytics.

4 Healthy Food Choice: the Case of Kale

25 BC Apicius stated that “we eat first with our eyes” (Apicius, 25 BC) [4], which could be reformulated into “we eat first with our mind”. Here we mean that the modern human is consuming texts and stories *about* food as eagerly as images *of* it. Recent studies have indicated the extent to which textual information or storytelling about food influences the multi-sensory experience of food intake. “In particular, according to one analysis of online restaurant menu descriptions, the average price of a dish in the US was found to go up by 6 cents for each additional letter in the dish description” [16]. The increasing impact of words is also well documented in the growing number of food blogging and food-related texts on social media. Taking into account the growing popularity of food blogging, we base our analysis on the textual information available about food, which allows us to make conclusions about the way food is being depicted in those texts.

We aim to dig into stories/texts available on social media and appraise to what extent the quest for hedonism is reflected in the healthy food domain judging from the sentiments described in the texts. The object of our focus here is kale, which has experienced a dramatic hike in its consumption in a relatively short period of time – e.g. in Denmark, where the consumption of kale has increased by 900% over the last ten years [17]: “When telling the story of food policy in action, it makes sense to start with the humble kale. Kale is a real superfood: sustainable, relatively cheap, packed with vitamins and possible to grow in conditions as low as -20°C . However, for decades kale consumption was in decline, abandoned by consumers. Until recently. (..) This is just one of many areas where the Nordic countries seem to have turned a cor-

ner and begun changing food habits – to the benefit of public health, the environment and the local economies” [18].

Tracing the triggers for kale consumption in the Nordic countries and globally is beyond the ambition of this article. We rather aim to understand the sentiments behind kale consumption by looking at Twitter food microblog entries related to kale, trying to pinpoint the factors behind its popularity and tracing the expressions of hedonism that could appear in texts related to the consumption of this particular food.

5 Methodology

We identified text mining as the best technique for discovering patterns in unfamiliar data (Visa, 2001) – as illustrated in figure 2. We have also employed Latent Semantic Analysis (LSA), as it is known for extracting and representing the meaning of words’ contextual use by using statistical computations applied to large corpora of text, as well as due to its ability to illustrate the number of word (term) occurrences in input documents [19] [20]. Data representing food blogs was collected using R programming language and Twitter API with the following keywords: “kale” and “food”. A total of $N = 10,000$ observations was gathered spanning tweets from January 2020 to June 2020 (figure 1). Only the English language tweets were used for analysis. As seen in figure 1, there is a decline in the use of “kale” and “food”, which we attribute to the SARS-CoV-2/COVID-19 pandemic that hit the globe during the first half of 2020.

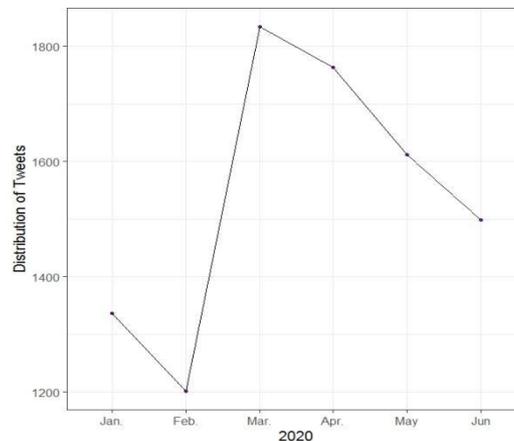


Fig. 1. Distribution of Dataset over Time.

The dataset on kale was then preprocessed by lemmatization (where words are transformed into their original form – e.g. Eating is transformed to its lemma, Eat) and normalization (where stopwords, numbers, whitespaces, unnecessary symbols such as

urls are removed). With duplicates deleted, we obtained N=9504 observations in total. A cleaned dataset provided us with N= 9246 observations. The observations were further tokenized into individual words/tokens followed by the part of speech (POS) tagging in order to extract the adjectives required for analysis using the UDPipe Model [21]. SQL was used to query frequent tokens greater than two (2) for the purpose of dimensionality reduction. Frequent bigrams with adjectives as the leading token followed by a second token were then obtained from the dataset. From the tokenized clean dataset a sentiment analysis of the adjective bigram was performed.

To evaluate the general emotions expressed per observation/Twitter post, the study employed the Loughran-McDonald lexicon (Loughran and McDonald Sentiment Word Lists), which was initially built for sentiment analysis in the financial context [22]. The Loughran and McDonald Sentiment Word Lists were previously adopted in the following studies:

- understanding sentiment disagreements with respect to bitcoin fluctuations [23],
- analyzing sentiments in financial microblogs [24],
- identifying moral hazards in restaurant hygiene inspections in New York [25],

The dictionary consists of tokens classified into “negative”, “positive”, “uncertain”, “litigious”, “constraining”, “superfluous”, “interesting”, “modal words strong” and “modal words weak”. The second semantic lexicon utilized was the AFFIN semantic lexicon, which assigns a polarity score to each word [26] and thus consists of words classified into positive or negative.

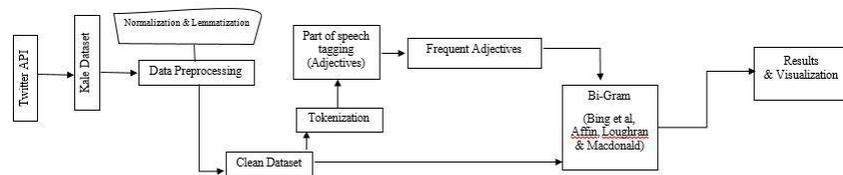


Fig. 2. Methodology and Research Workflow.

The following analytical techniques were employed:

- a. discerning the most frequent bigrams of adjectives and tokens (i.e. adjective + token)
- b. discerning the most frequent bigrams of sentiments and tokens (i.e. sentiment + token).

Choosing to focus on bigrams was justified by the advantage of bigrams over unigrams since bigrams are better suited for handling negated phrases [27]. Bigrams contain more explicit information than words. They also have a relatively big semantic granularity and can deal with unknown words [28]. Furthermore, bigrams provide some understanding of the context since they have a better description capacity than e.g. unigrams. [29]

Hereby we introduce a novel text analysis technique for representing co-occurrence, which we term as Sentiment-Token Bigrams. Here, our algorithm matches the occurrence of a term from the AFFIN and Loughran-McDonald lexicons respectively with the tokens in the ‘kale dataset’ (as the first component of the bigram). This is followed by finding in the dataset (corpora) the subsequent token after the sentiment-token. See the algorithm for Sentiment-Token Bigrams in Table 1.

Table 1. Algorithm for Sentiment-Token Bigram.

Algorithm: Sentiment-Token Bigram
Data: Twitter Posts Dataset (N observations)
Sentiment Lexicon: Dictionary of terms associated with sentiments
Result: Bigram with a Sentiment Lexicon Term as the first word
<pre> load SentimentLexicon for each observation do extract LexiconTerm and SubsequentToken repeat until no observations left end count bigrams delete bigrams with low frequency (user-defined threshold) visualize top Sentiment-Token Bigram </pre>

6 Data Analysis

The Twitter microblog analysis of sentiments regarding kale shows the following results:

- Bing sentiment analysis reveals that ‘anti-inflammatory’ (30), ‘benefit kale’ (15), ‘amaze food’ (15) and ‘amaze kale’ (12) are the most frequent bigrams of the kale-related dataset
- Loughran-McDonald sentiment analysis yields ‘boost immune’ (30), ‘boost food’ (25), ‘benefit kale’ (15) and ‘boost immunity’ (8).

Both sentiment graphs offer a wide range of bigrams containing the words ‘beautiful’, ‘benefit’ and ‘awesome’ among other most frequent bigrams, and both of them abound in verbs – e.g. ‘to benefit’ and ‘to boost’ – promising impressive results that kale can deliver to one’s health.

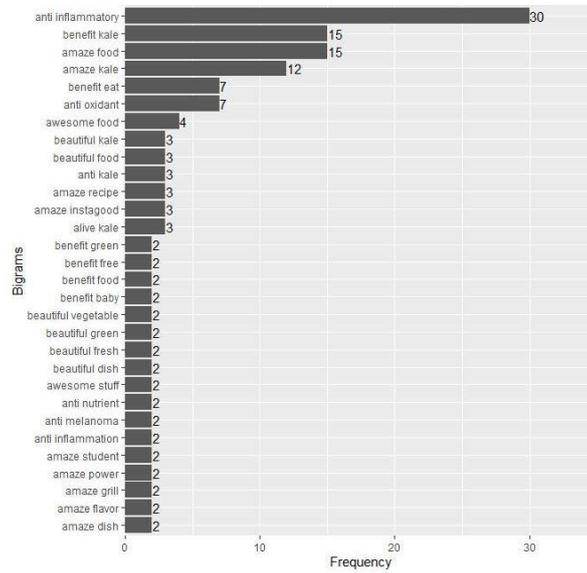


Fig. 3. Most Frequent Bigrams (Bing Sentiment + Token).

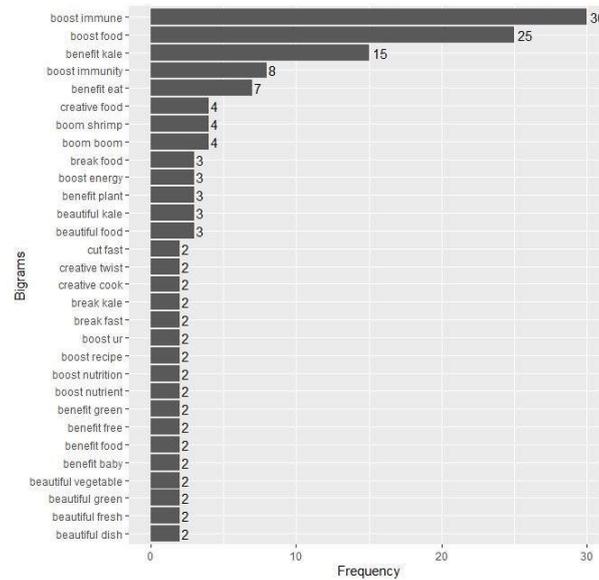


Fig. 4. Most Frequent Bigrams (Loughran-McDonald Sentiment + Token).

Analyzing the most frequent bigrams with adjectives, we see that ‘sweet potato’ (249) and ‘healthy food’ (245) dominate the discourse, implying that kale might only serve as an ‘anti-inflammatory’ and ‘immunity boosting’ healthy addition to other foods.

If we look at the bigrams determining composition or the context of other foods kale is consumed with, we can notice certain food-related value-based statements, such as ‘fast food’ (136), ‘comfo food’ (116) and ‘vegan food’ (114). None of these statements reflect any hedonistic attitudes associated with consuming this particular food.

We also see surprisingly little reference to tasty food. ‘Favorite food’ (101), ‘fresh food’ (53) and ‘delicious food’ (49) could be related to taste aspects, but those are not prevalent.

All in all, bigrams provide a meta-story of kale. By looking at the sentiment analysis and the most frequent bigrams of adjectives and nouns, we are able to follow the most frequently told ‘stories’ about kale on a more detailed, micro-level as well.

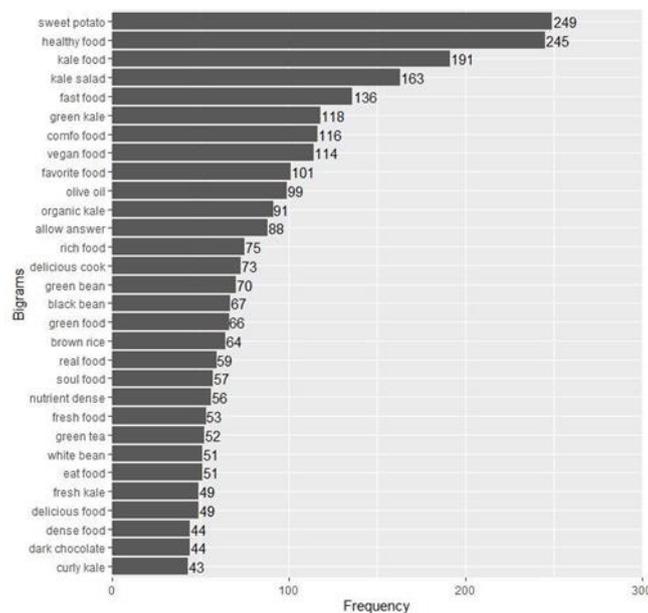


Fig. 5. Most Frequent Bigrams (Adjective + Token).

In conclusion, this analysis reveals that kale offers clear-cut and rather ambitious promises for the immune system, thereby turning the focus on health. Hedonism and tastiness are left out of the discourse, putting kale into the healthy food domain where food consumption is governed by functional rather than sensual benefits.

7 Conclusions

With this research, we aimed to define some of the methods useful for facilitating the consumption of healthy foods. Having made an analysis of texts describing food, we have ascertained the lack of hedonism-denoting terms in healthy food descriptions. We have also tried to distinguish the triggers behind preferences for healthy foods as

depicted in the texts of microblog entries available on Twitter. For the analysis unit we chose kale - a food whose consumption over the last decade has grown significantly.

Kāle and Agbozo [12] concluded that contemporary food blog entries on Twitter related to healthy food do not focus on taste aspects and contain few references to hedonic expressions, accentuating rather “simple and easy” qualities and to a lesser, or barely discernible, extent - “complex and enjoyable”. The same holds true for the case study analysis of kale, although the bigrams related to kale do feature more concrete promises for health improvement. Still, the notions of hedonism and tastiness are left outside the sentiments associated with consuming the healthy food kale.

Though it may not have adversely influenced the result of the analysis, a minor limitation of the study was the lack of a wider scope of dataset from previous years (i.e. before 2020).

The study’s theoretical contribution is evident in the introduction of a novel Sentiment-Token Bigram algorithm which could be applied in the computational linguistics.

Lastly, this research also contributes to the discussion of the extent to which concepts that are used in cognitive science domain are operational in the quantitative analysis of texts. Due to the challenge of cross-sectionality, this research has aimed to refine the questions that could be posed by the research community with regard to food and human relationships. Potentially, instead of asking ‘how tasty the food is?’ we should rather ask: ‘how interesting or entertaining the food is?’ or ‘to what extent food consumption is related to social status?’ Instead of narrowing down the focus on taste, as has been customary in computer science to date, the focus should be broadened and encompass e.g. food’s associations with entertainment, sensuality and hedonism, as well as its ability to elevate one’s social status or manifest other important identity markers. The growing research of food computing which raises cross-sectionally relevant questions brings us ever closer to understanding the phenomenon of the food choice.

References

1. Amadae, S.M.: Rational choice theory. Encyclopaedia Britannica: <https://www.britannica.com/topic/rational-choice-theory>, last accessed 2020/01/02.
2. Min, W., Jiang, S., Jain, R.C.: Food Recommendation: Framework, Existing Solutions and Challenges. *IEEE Transactions on Multimedia* (2019).
3. Min, W., Jiang, S., Liu, L., Rui, Y., Jain, R.: A survey on food computing. *ACM Computing Surveys (CSUR)* 52 (5), 1–36 (2019).
4. Spence, C., Okajima, K., Cheok, A. D., Petit, O., Michel, C.: Eating with our eyes: From visual hunger to digital satiation. *Brain Cogn.* 110, 53-63 (2016). doi:10.1016/j.bandc.2015.08.006
5. Finn, S.: Can “Taste” Be Separated from Social Class? In: Ludington, C. & Booker, M. (Eds.): *Food Fights: How History Matters to Contemporary Food Debates*, pp. 81-99. Chapel Hill: University of North Carolina Press (2019).

6. Raghunathan, R., Walker Naylor, R., Hoyer, W.D.: The unhealthy= tasty intuition and its effects on taste inferences, enjoyment, and choice of food products. *Journal of Marketing* 70 (4), 170–184 (2006).
7. Tepper, B. J., Trail, A.C.: Taste or health: A study on consumer acceptance of corn chips. *Food Quality and Preference* 9 (4), 267–272 (1998).
8. Mai, R., Hoffman, S., Helmert, R. J., Velichkovsky, M. B., Zahn, S., Jaros, D., Schwarz E.H.P., Rohm, H.: Implicit food associations as obstacles to healthy nutrition: the need for further research. *Br. J. Diabetes Vasc. Dis.*, vol. 11 (4), Jul/Aug 2011, 182-186 (2011). doi:10.1177/1474651411410725
9. Mejova, Y., Abbar, S., Haddadi, H.: Fetishizing food in digital age: #foodporn around the world. In: Tenth International AAAI Conference on Web and Social Media (2016).
10. Jurafsky, D.: *The language of food: A linguist reads the menu*. WW Norton & Company (2014).
11. Cornil, Y., Chandon, P.: Pleasure as a substitute for size: how multisensory imagery can make people happier with smaller food portions. *J. Mark. Res.* Vol. LIII, 847-864 (2016). doi:10.1509/jmr.14.0299
12. Kāle, M., Agbozo, E.: Tracing complexity in food blogging entries. In: CEUR Workshop Proceedings 2612, pp. 51–62 (2020).
13. Mai, R., Hoffmann, S.: How to combat the unhealthy= tasty intuition: The influencing role of health consciousness. *Journal of Public Policy & Marketing* 34 (1), 63–83 (2015).
14. Ahn, Y., Ahnert, S.E., Bagrow, J.P., Barabási, A-L.: Flavor network and the principles of food pairing. *Scientific reports* 1, 196, (2011).
15. Spence, C.: *Gastrophysics: The new science of eating*. Penguin UK (2017).
16. Spence, C.: Complexity on the menu and in the meal. *Foods* 7, 158 (2018). doi:10.3390/foods7100158
17. Nordic Food Policy Lab. Collecting, curating and co-producing policy tools for a consumer powered food transition, https://www.wur.nl/upload_mm/1/9/d/6f259c39-dbbb-453d-9f46-290a7b721999_OECD%20-%20WUR%20-%20050919%20-%20PerssonM.PDF, last accessed 2020/07/15.
18. Halloran, A., Fischer-Møller, M.F., Persson, M., Skylare, E.: *Solutions Menu - A Nordic guide to sustainable food policy*. Nordic Council of Ministers (2018).
19. Wei, C. P., Yang, C. C., & Lin, C. M.: A latent semantic indexing-based approach to multilingual document clustering. *Decision Support Systems*, 45(3), 606-620, (2008).
20. Miles, I., Saritas, O., & Sokolov, A.: *Intelligence: Environmental and Horizon Scanning*. Foresight for Science, Technology and Innovation, pp. 63-93. Springer, Cham. (2016).
21. Straka, M. 2018. UDPipe 2.0 prototype at CoNLL 2018 UD shared task. In: Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies, 197–207, (2018).
22. Loughran, T., McDonald, B.: When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance* 66 (1), 35–65 (2011).
23. Ahn, Y., Kim, D.: Sentiment disagreement and bitcoin price fluctuations: a psycholinguistic approach. *Applied Economics Letters*, 27(5), 412–416 (2020).
24. Cortis, K., Freitas, A., Daudert, T., Huerlimann, M., Zarrouk, M., Handschuh, S., Davis, B.: Semeval-2017 task 5: Fine-grained sentiment analysis on financial microblogs and news. Association for Computational Linguistics (ACL). (2017)
25. Mejia, J., Mankad, S., Gopal, A.: A for Effort? Using the Crowd to Identify Moral Hazard in NYC Restaurant Hygiene Inspections (Jan 1, 2018). Kelley School of Business Research Paper, 18-58 (2018).

26. AArup Nielsen, F.: A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. arXiv preprint arXiv:1103.2903 (2011).
27. Chen, J.: Search engine For Twitter sentiment analysis (Doctoral dissertation). (2015).
28. Li, J., Sun, M.: Non-Independent term selection for Chinese text categorization. *Tsinghua Science and Technology*, 14(1), 113-120 (2009).
29. Kaptein, R., Hiemstra, D., Kamps, J.: How different are language models and word clouds? In: *European conference on information retrieval*, pp. 556-568. Springer, Berlin, Heidelberg (2010).