Benchmarking the Semantics of Taste: Towards the Automatic Extraction of Gustatory Language

Teresa Paccosi[1,2,3], Sara Tonelli[1]

[1]*Fondazione Bruno Kessler, Via Sommarive, 18, Trento*

[2]*Università degli studi di Trento, Via Calepina, 14, Rovereto*

[3]*DHLab / KNAW Humanities Cluster, Oudezijds Achterburgwal 185 1012 DK Amsterdam, The Netherlands*

Abstract

In this paper, we present a benchmark containing texts manually annotated with gustatory semantic information. We employ a FrameNet-like approach previously tested to address olfactory language, which we adapt to capture gustatory events. We then propose an exploration of the data in the benchmark to show the possible insights brought by this type of approach, addressing the investigation of emotional valence in text genres. Eventually, we present a supervised system trained with the taste benchmark for the extraction of gustatory information from historical and contemporary texts.

Keywords

Sensory semantics, gustatory language, information extraction, digital humanities

1. Introduction

Despite the central role of nutrition in our lives, taste has been often classified as an inferior sense in the Western philosophical tradition. This downplayed role is reflected in the vocabulary used to describe the gustatory experience, which, together with smell, is characterized by a scarcity of domain-specific terms [1]. The difficulty in capturing the semantics of taste could help explain why there are few works in the fields of Natural Language Processing (NLP) and Digital Humanities (DH) that deal with this sense and, in particular, the language used to describe its experience. While there has been renewed interest in the automatic extraction of nutrients and ingredients from texts for health and medicinal purpose [2], less attention has been devoted to the development of tools and models focused on capturing the semantics of sensory experiences, especially in a diachronic fashion.

In this paper, we present an English benchmark for the study of gustatory language and a supervised system for the automatic extraction of taste-related events in English, which we trained using this benchmark. The benchmark was built to be a counterpart to the olfactory one presented in [3], with the idea of making the study of the language of these two senses comparable. The system is designed as a means to study the language used to describe the experience of tasting from both synchronic and diachronic perspectives. The selected formal representation for the semantics of taste is based on Frame Semantics [4], and the system is trained to identify the lexical units and the possible semantic roles contributing to the construction of a gustatory event. We present the results of the experiments and an exploration of the benchmark data, aiming to demonstrate the potential of frame-based analysis for sensory studies.

2. Related Work

In recent years, there has been a growing interest within the NLP community in developing resources designed to capture the sensory content of language [5]. In particular, in the framework of the three-year European Project "Odeuropa"[1] aimed at preserving intangible cultural heritage, several works have focused on analyzing smell descriptions [6] and extracting olfactory information from texts. For instance, [3] created a manually annotated benchmark with smell events, which has been subsequently used to train a system for olfactory information extraction [7, 8]. The benchmark focuses on the language used to describe olfactory experiences and covers a period of four centuries (1600-1900), making it useful for historical research. An extension in this direction is SENSE-LM, a system for extracting sensory information from texts, which shows that combining language models with lexical resource-based approaches yields better results in extracting sensory references from texts compared to systems that do not integrate these two components [9]. The authors were the first to combine sensorimotor representations with the textual features of language models for the task of sensory information extraction in text documents. Even if they propose the system for all the 5 senses, they only tested it on olfactory

[1]https://odeuropa.eu/

Frame Element	Definition
Taste_Source	The food items that are ingested
Quality	Any property used to describe the taste (usually adjectives)
Taste_Carrier	Anything that can contain the taste source
Taster	The person/animal who ingests the food
Evoked_Taste	The taste that is evoked but it is not present (e.g., it tastes like onions)
Location	The place in which the food is tasted
Taste_Modifier	An ingredient that can modify the perception of the taste of a taste source
Circumstances	The condition or circumstance in which the taste event occurs
Effect	Any effect provoked by the tasting experience

Table 1
List of Gustatory Frame Elements

and auditory language, using respectively the benchmark of [3] and an artificial dataset they generated with GPT-4 [10]. Most existing work on food representation in the field of NLP focuses on health-related applications. A notable work with a linguistic focus is [2], where the authors concentrate on identifying noun-compound headnouns for developing conversational agents in the e-commerce domain. They propose a supervised approach based on a neural sequence-to-sequence model to identify the most informative token in Italian food compound-nouns, obtaining promising results despite the complexity of the task. Taste has been also addressed from a diachronic point of view in [11], in which the author reconstructs the evolution of food language focusing on the history of some dishes and ingredients across continents using computational linguistic tools. Several studies have developed named-entity recognition (NER) models to automatically extract food entities for medicinal purposes and food science applications [12, 13], creating domain-specific corpora by sourcing data from culinary websites and online recipe books [14, 15].

3. Benchmark for Taste

The training data we use for the models in this paper is a benchmark created according to the annotation guidelines presented in [16]. The formalization adopted to annotate the benchmark is inspired by Frame Semantics [4] and their implementation through the FrameNet annotation project [17]. In FrameNet, events and situations are constructed as *frames*, structures that represent the knowledge necessary to understand the meaning of words. Frames include two main components, namely **lexical units**, domain-specific words or expression that trigger the frame, and **frame elements**, domain-specific semantic roles usually attached as dependents to the lexical unit. In our case, taste events are captured through a so-called *Gustatory frame*, which is triggered in a document by Taste_Words (i.e., domain-specific lexical units). Each lexical unit is annotated in the bench-

mark together with the frame elements associated with it, which the taste extraction system should then identify automatically. For instance, in the sentence "[Slimy milk]$_{Taste\_Source}$ has an [unpleasant]$_{Quality}$ taste", the system has to identify the Taste_Word ('taste'), and then the possible frame elements (in this case, Taste_Source and Quality). A list of the possible frame elements and their definition is provided in Table 1. The documents annotated in the benchmark cover 5 different domains or genres, almost evenly distributed with 3/4 documents for century in every domain for a total of 72 documents. The genres are: *Literature*, *Science & Philosophy*, *Household & Recipes*, *Travel & Ethnography*, and *Medicine & Botany*. To select the documents we automatically search for texts presenting a greater density of lexical units (taste words) [2] spanning through several English corpora and taste-related websites. The corpora form which we extract the documents we annotated are: (1) *Early English Books Online (EEBO)*[3], a collection of documents published between 1475 and 1700 covering different domains such as literature, philosophy, politics, religion, geography, history, politics, and mathematics; (2) *Project Gutenberg*[4], a digitized archive of cultural works, containing different repositories, mainly in the literary domain; (3) *medievalcookery.com*[5] a list of texts freely available online relating to medieval food and ancient cooking recipes; (4) *foodsofengland.co.uk*[6] an online library which holds the complete texts of several cook books from 1390 to 1974; (5) *Wikisource*[7], an online digital library of free-content textual sources managed by the Wikimedia Foundation; (6) *British Library*[8], a collection of 65,227 digitised volumes from the 16th to the 19th Century; (7) *London Pulse*

---

[2] The list of lexical units is provided in Appendix A
[3] https://textcreationpartnership.org/tcp-texts/eebo-tcp-early-english-books-online/
[4] https://www.gutenberg.org/
[5] https://www.medievalcookery.com/etexts.html?England
[6] http://www.foodsofengland.co.uk/references.htm
[7] https://en.wikisource.org/wiki/Main_Page
[8] https://data.bl.uk/digbks/

Frame Elements (FEs)	1500	1600	1700	1800	1900	Overall
Taste_Words	440	2417	500	1498	803	5,648
Taste_Source	372	1627	375	1081	599	4,393
Quality	197	1495	255	881	489	1,732
Taste_Modifier	135	142	66	154	78	1,357
Taster	65	173	85	185	100	638
Evoked_Taste	20	127	31	53	16	247
Location	11	44	12	24	16	116
Taste_Carrier	9	38	9	26	12	98
Circumstances	19	206	38	228	82	656
Effect	24	56	32	34	31	174

Table 2

Statistics of the Taste Benchmark

Medical Reports[9], a collection of 5800 Medical Officer of Health reports from the Greater London area from 1848 to 1972.

In Table 2 we report the statistics of the annotated benchmark (note that in [16] we presented only a preliminary version of the benchmark containing around 1,400 Taste_Words). The most frequent frame element is the Taste_Source, followed by Quality and Taste_Modifier, which represent the core frame elements, while the rest of the frame elements are much sparser. Even if the distribution of the frame elements is not balanced, the system is trained to extract the taste words and all the 9 frame elements. Two expert linguists, trained on [16]'s guidelines, annotated three documents from 1670, 1720, and 1920 to assess Inter Annotator Agreement (IAA). The Krippendorff's alpha score [18] at span level was 0.70, indicating a moderate agreement.

4. Exploration of olfactory and gustatory benchmarks

It has been observed that words used to describe olfactory and gustatory experiences tend to appear more frequently in emotionally charged contexts and carry a stronger evaluative content compared to words related to other senses [19]. By 'evaluative content', we refer in this paper to the concept of 'emotional valence', which is defined as "the pleasantness of a word in terms of positive and negative meaning" ([1], p. 201). We therefore conducted an exploration of the gustatory benchmark to investigate the positive and negative connotations of gustatory events *across different text genres*. We perform the same analysis for olfactory events, using the olfactory benchmark of [3] in order to compare the outcome for the two senses. To perform this analysis, we first divide Taste_Words and Smell_Words into *positive* and *negative*.

To this purpose, we use the categories proposed in the Historical Thesaurus of English of `Savouriness` and `Unsavouriness` for Taste and `Fragrant/Fragrance` and `Stench` for Smell[10]. This thesaurus contains almost every recorded word in English from medieval times to the present day, ordered into detailed hierarchies of meaning. In the Thesaurus, every category of the hierarchy is divided per part of speech (PoS). For our analysis, we manually selected all the nouns, adjectives and adverbs used in the period we cover with our documents, namely from 16th century to 20th century. We then assigned the words labeled as Taste_Words and Smell_Words in the documents to one of the two categories (positive or negative) and calculated the normalized frequency of each category across different text genres. As reported in Section 3, the genres represented in the gustatory benchmark are: Literature, Science & Philosophy, Household & Recipes, Travel & Ethnography, Medicine & Botany. In the olfactory benchmark presented in [3], there are instead 10 different genres: Household & Recipes, Law & Regulations, Literature, Medicine & Botany, Perfumes & Fashion, Public health, Religion, Science & Philosophy, Theatre, Travel & Ethnography.

We display the output of this analyses in Fig. 1 (for taste words) and Fig. 2 (for smell words), aimed at showing which emotional valence prevails in each genre for the two senses. We observe that two genres exhibit opposite tendencies: `medicine/botany` shows a more negative orientation in the smell benchmark and a more positive one in the taste benchmark, whereas `travel/ethnography` is more positive concerning smell and more negative for taste (see Fig. 1 and Fig. 2, where the light blue refers to negative valencies and the dark blue to positive ones). We then analyzed the most frequent smell / taste sources in the two selected genres to motivate why they exhibit

---

[10]In the categories at https://ht.ac.uk/category/: The world>physical sensation>Taste/Flavour>Savouriness&Unsavouriness; The world>physical sensation>Smell/Odour>Fagrant/Fragrance&Stench
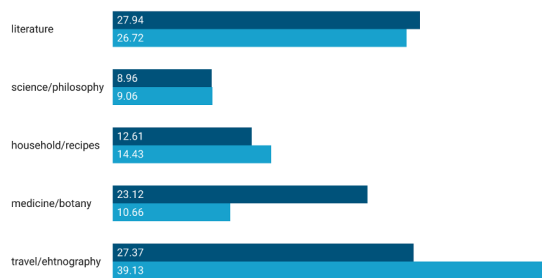
Figure 1: `Savoury` (dark blue) and `Unsavoury` (light blue) frequencies of taste words in genres
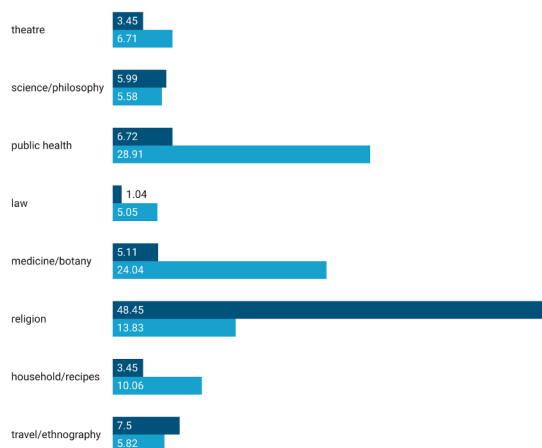

Figure 2: `Fragrant/Fragrance` (dark blue) and `Stench` (light blue) frequencies of smell words in genres

such difference in emotional valence. We notice that smell sources in `medicine/botany` tend to be common to hospital and disease-related domains having words such as 'urine' and 'fetid bronchitis', while taste sources more easily belong to the realm of common food, with words such as 'almonds' and 'apples'. For what concerns `travel/ethnography` instead, among the most frequently described taste sources there are exotic and rare foods such as 'coconut' and 'plantain', likely resulting unpleasant to the palates of foreign travelers. Smell sources tend to refer instead to plants, like 'flowers' or 'roots', hence usually pleasant or neutral to the noses of the writers. This analysis of categories and sources' distribution in the genres underlines the importance of a frame-base analysis for understanding and comparing sensory descriptions, in particular their emotional valence.

5. System for Gustatory Information Extraction

The benchmark introduced in the previous sections is used to train a classifier whose goal is to detect gustatory information in English texts. The system is based on multi-task learning (Section 5.1), and is then compared with a "single task" classifier, which we consider our baseline (Section 5.2).

5.1. Multitask configuration

To build our system for gustatory information extraction, we adopted a multitask learning approach [20, 21], a configuration successfully tested for olfactory information extraction in [7, 8]. This approach treats the classification of lexical units and each frame element as different tasks. Additionally, we explored a "single task" classification approach, where both lexical units and frame elements are classified within a multiclass token classification task. The results of these experiments served as a baseline for evaluating the effectiveness of the multitask approach. In both configurations, we employed a transformer-based model fine-tuned for a token classification task [22]. This methodology has proved effective across various NLP tasks, including olfactory information extraction [8] and the extraction of food-related ingredients [13]. We experiment the two configurations with monolingual (English) and multilingual versions of BERT and RoBERTa and with an English historical model, MacBERTh. The models we use are listed below:
- **English BERT**: bert-base-cased [11] [23]
- **Multilingual BERT (mBERT)**: bert-base-multilingual-cased [12][23]
- **English historical model**: MacBERTh [13] [24]
- **English RoBERTa**: roberta-base [14][25]
- **Multilingual RoBERTa (RoBERTa xlm)**: xlm-roberta-large[15] [26]

We fine-tuned each model using the same data, maintaining identical training, validation, and test splits, and evaluated them using 5-fold cross-validation. Each fold contained 80% of the lexical units and their related frame elements for training, 10% for validation (dev), and 10% for testing. These splits were consistent across all configurations and not entirely random. This configuration ensured a balanced distribution of frame elements and comparability in every run. For labeling the data, we adopted the IOB (Inside-Outside-Beginning) labeling format, as used in [7, 8]. This method facilitates a comprehensive analysis of sentences and lexical expressions by

Model	T_Word	T_Source	Quality	Circum.	Effect	Evoked_T	Loc.	T_Carr.	T_Modif.	Taster
BERT	0.917	0.537	0.780	0.413	0.196	0.457	0.379	0.111	0.781	0.518
BERT	*0.903*	*0.530*	*0.712*	*0.308*	*0.019*	*0.254*	*0.206*	*0.0*	*0.681*	*0.434*
mBERT	0.919	0.554	0.784	0.402	0.180	0.466	0.357	0.087	0.763	0.511
mBERT	*0.910*	*0.557*	*0.740*	*0.284*	*0.0*	*0.304*	*0.162*	*0.0*	*0.694*	*0.434*
MacBERTh	**0.943**	0.580	0.799	0.444	**0.285**	**0.501**	0.338	0.093	0.783	0.512
MacBERTh	*0.909*	*0.548*	*0.720*	*0.366*	*0.021*	*0.226*	*0.242*	*0.0*	*0.688*	*0.455*
RoBERTa	0.913	0.558	0.786	0.414	0.219	0.473	0.406	0.094	0.772	0.508
RoBERTa	*0.891*	*0.553*	*0.726*	*0.343*	*0.0*	*0.33*	*0.228*	*0.0*	*0.726*	*0.5*
RoB.-xlm	0.932	**0.587**	**0.817**	**0.452**	0.279	0.497	**0.416**	**0.105**	**0.784**	**0.563**
RoB.- xlm	*0.903*	*0.601*	*0.777*	*0.4*	*0.021*	*0.409*	*0.25*	*0.0*	*0.743*	*0.539*

Table 3
Results (F1) of the classifiers on the lexical unit (T_Word) and 9 frame elements with single (italics) and multitask configurations. The results are the average of the f1 results of each label across the 5 folds.

labeling each token with either Inside, Outside, or Beginning labels as appropriate. To fine-tune the models, we used MaChAmp [27], a specialized toolkit designed for multi-task fine-tuning scenarios. In this approach, each label classification is treated as a distinct task. This setup ensures that simpler tasks, such as recognizing lexical units, contribute as auxiliary tasks to more complex label classifications like "Circumstances" or "Effect" which include entire sentences rather than individual words. MaChAmp enables the choice of different parameters, such as loss weight, epochs and batch size, and we tested different configurations [16]. The results in Table 3 for the multitask approach share the configuration which yielded the best results. The configuration is the same for all the models and it is reported in Appendix A.

5.2. "Single Task" configuration as Baseline

Similar to the system for smell information extraction presented in [8], we designed our baseline approach as a single-task multiclass classification, where the model assigns one of 21 possible labels to each token. These labels include 20 representing either "begin" or "inside" of each lexical unit and frame element, and 1 label representing "outside". As we did for the multitask approach, each model is fine-tuned with a token classification head on top [17]. During the training of each model, a hyperparameter search was conducted on the first fold of our data. The search space included learning rates $[1e-5, 2e-5, 3e-5, 4e-5, 5e-5]$, batch sizes $[8, 16, 32]$, and training epochs up to 20, with warmup applied for 10% of the training steps. After determining the optimal hyperparameters for each model, it is fine-tuned

five times, each time with a different data fold, and the average scores were computed. We present the results of for the single task approach of each model in italics in Table 3. We observe high performance variations across different frame elements, with the best results obtained for "Quality" and "Taste_Modifier". This is probably due to the fact that their syntactic realization tends to be consistent in the different documents, with "Quality" mainly expressed by adjectives and "Taste_Modifier" by prepositional phrases introduced by *with*. On the contrary, classification results for "Taste_Source" are quite low despite it being the most frequent FE in the training set, probably because they can be expressed by many different role fillers and syntactic constructions. Upon reviewing the test and prediction results, we find that most mistakes concerning Taste_Source are due to a wrong span extent, for instance the system predicts "the taste of [lollilop]" while the gold standard is "the taste [of lollipop]". This issue is also likely reflected in the inter-annotator agreement (IAA) of the benchmark. In the future, we will consider alternative ways to evaluate text spans beside exact match, for instance by computing the cosine similarity between gold instances and system predictions.

Overall, MacBERTh is the best model for Taste_Word detection, but the different FEs are mostly detected with higher accuracy using RoBERTa xlm. For this reason, we plan to adopt this model for our future research on gustatory language.

6. Conclusions and Future Direction

In this paper, we presented a benchmark for gustatory events containing manually annotated taste-related information, built as a counterpart to the one proposed in [3]. The benchmark is constructed with the same approach adopting a frame-based methodological framework to

---

[16]Loss weight with different combinations over the labels $[1, 0.75]$, epochs $[10, 20, 30]$, and batch size $[16, 32]$
[17]https://huggingface.co/docs/transformers/tasks/token_classification

analyze sensory language. We emphasized the importance of frame-based analysis to capture sensory events by exploring the characterization of positive and negative valence in the benchmarks through the analysis of taste and smell words and sources. The analysis based on frames seems to bring relevant insights into capturing sensory valence from different perspectives, likely supporting the suitability of this approach to deal with humanistic inquiries. We then presented a supervised system to automatically extract taste-related frames, trained on this benchmark. This preliminary exploration and the results obtained with our experiments seem promising for future exploration with automatically extracted data. Indeed, the limited data of the benchmark are not enough to draw relevant conclusions, and for this reason we plan to use our system to extract more data and conduct large-scale analyses of the evolution of sensory information over time. The limited number of documents is likely a contributing factor to the significant discrepancies in accuracy among the different frame elements, necessitating more instances to enable a good generalization. Future steps should involve increasing the number of documents and providing less sparse annotations, aiming for better temporal balance. The focus should be on annotating frame elements with lower scores and fewer instances in the benchmark, such as Taste_Carrier and Location. Additionally, alternative metrics and techniques should be employed to capture and explain performance variations across different models. As a further comparison, we plan also to assess the performance of general-purpose frame semantic parsers like LOME [28] on our benchmark.

7. Aknowledgments

References

[1] B. Winter, Sensory linguistics: Language, perception and metaphor, volume 20, John Benjamins Publishing Company, 2019.

[2] B. Magnini, V. Balaraman, S. Magnolini, M. Guerini, F. B. Kessler, T. Povo, What's in a food name: Knowledge induction from gazetteers of food main ingredient, in: Proceedings of CLiC-it 2018, 2018, p. 241.

[3] S. Menini, T. Paccosi, S. Tonelli, M. Van Erp, I. Leemans, P. Lisena, R. Troncy, W. Tullett, A. Hürriyetoğlu, G. Dijkstra, et al., A multilingual benchmark to capture olfactory situations over time, in: Proceedings of the 3rd Workshop on Computational Approaches to Historical Language Change, 2022, pp. 1–10.

[4] C. J. Fillmore, Frame semantics and the nature of language, Annals of the New York Academy of Sciences 280 (1976) 20–32.

[5] S. S. Tekiroğlu, G. Özbal, C. Strapparava, A computational approach to generate a sensorial lexicon, in: Proceedings of the 4th Workshop on Cognitive Aspects of the Lexicon (CogALex), Association for Computational Linguistics and Dublin City University, Dublin, Ireland, 2014, pp. 114–125. URL: https://aclanthology.org/W14-4716. doi:10.3115/v1/W14-4716.

[6] R. Brate, P. Groth, M. van Erp, Towards olfactory information extraction from text: A case study on detecting smell experiences in novels, in: Proceedings of the The 4th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature, International Committee on Computational Linguistics, Online, 2020, pp. 147–155. URL: https://aclanthology.org/2020.latechclfl-1.18.

[7] S. Menini, T. Paccosi, S. S. Tekiroğlu, S. Tonelli, Scent mining: Extracting olfactory events, smell sources and qualities, in: Proceedings of the 7th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature, 2023, pp. 135–140.

[8] S. Menini, Semantic frame extraction in multilingual olfactory events, in: Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), 2024, pp. 14622–14627.

[9] C. Boscher, C. Largeron, V. Eglin, E. Egyed-Zsigmond, Sense-lm: A synergy between a language model and sensorimotor representations for auditory and olfactory information extraction, in: Findings of the Association for Computational Linguistics: EACL 2024, 2024, pp. 1695–1711.

[10] O. AI, Gpt-4 technical report, arXiv preprint arXiv:2303.08774 (2023).

[11] D. Jurafsky, The language of food : a linguist reads the menu / Dan Jurafsky., first edition. ed., W.W. Norton Company, New York, 2014 - 2014.

[12] G. Cenikj, G. Popovski, R. Stojanov, B. K. Seljak, T. Eftimov, Butter: Bidirectional lstm for food named-entity recognition, 2020.

[13] R. Stojanov, G. Popovski, G. Cenikj, B. Koroušić Seljak, T. Eftimov, A fine-tuned bidirectional encoder representations from transformers model for food named-entity recognition: Algorithm development and validation, Journal of Medical Internet Re-

search 23 (2021) e28229.

[14] G. Popovski, B. K. Seljak, T. Eftimov, Foodbase corpus: a new resource of annotated food entities, Database 2019 (2019) baz121.

[15] A. Wróblewska, A. Kaliska, M. Pawłowski, D. Wiśniewski, W. Sosnowski, A. Ławrynowicz, Tasteset–recipe dataset and food entities recognition benchmark, arXiv preprint arXiv:2204.07775 (2022).

[16] T. Paccosi, S. Tonelli, A new annotation scheme for the semantics of taste, in: Proceedings of the 20th Joint ACL-ISO Workshop on Interoperable Semantic Annotation@ LREC-COLING 2024, 2024, pp. 39–46.

[17] J. Ruppenhofer, M. Ellsworth, M. Schwarzer-Petruck, C. R. Johnson, J. Scheffczyk, FrameNet II: Extended theory and practice, Technical Report, International Computer Science Institute, 2016.

[18] K. Krippendorff, Computing krippendorff's alpha-reliability, 2011.

[19] B. Winter, Taste and smell words form an affectively loaded and emotionally flexible part of the english lexicon, Language, Cognition and Neuroscience 31 (2016) 975–988.

[20] R. Caruana, Multitask learning: A knowledge-based source of inductive bias1, in: Proceedings of the Tenth International Conference on Machine Learning, Citeseer, 1993, pp. 41–48.

[21] R. Caruana, Multitask learning, Machine learning 28 (1997) 41–75.

[22] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Advances in neural information processing systems 30 (2017).

[23] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 2019, pp. 4171–4186.

[24] E. Manjavacas Arévalo, L. Fonteyn, MacBERTh: Development and evaluation of a historically pretrained language model for English (1450-1950), in: Proceedings of the Workshop on Natural Language Processing for Digital Humanities (NLP4DH), Association for Computational Linguistics, 2021, pp. 23–36. URL: https://aclanthology.org/2021.nlp4dh-1.4.pdf.

[25] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, Roberta: A robustly optimized BERT pretraining approach, CoRR abs/1907.11692 (2019). URL: http://arxiv.org/abs/1907.11692. arXiv:1907.11692.

[26] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, V. Stoyanov, Unsupervised cross-lingual representation learning at scale, CoRR abs/1911.02116 (2019). URL: http://arxiv.org/abs/1911.02116. arXiv:1911.02116.

[27] R. Van Der Goot, A. Üstün, A. Ramponi, I. Sharaf, B. Plank, Massive choice, ample tasks (machamp): A toolkit for multi-task learning in nlp, arXiv preprint arXiv:2005.14672 (2020).

[28] P. Xia, G. Qin, S. Vashishtha, Y. Chen, T. Chen, C. May, C. Harman, K. Rawlins, A. S. White, B. Van Durme, LOME: Large ontology multilingual extraction, in: D. Gkatzia, D. Seddah (Eds.), Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations, Association for Computational Linguistics, Online, 2021, pp. 149–159. URL: https://aclanthology.org/2021.eacl-demos.19. doi:10.18653/v1/2021.eacl-demos.19.

Part of Speech	Lexical Units
Nouns	Acidity, aftertaste, aroma, bitterness, dainty, delicacy, disgust, distaste, flavor, flavour, flavorful, flavourful, flavoring, flavouring, flavorsome, flavoursome, flavorous, flavourous, gustation, insipidity, mistaste, over-eating, palatableness, piquancy, pungency, rancidity, relish, rellish (obsolete), saltness, sapidity, sapor, savor, savoriness, savour, sharpness, smack, smatch, sourness, sowreness (archaic form of sourness), sweetness, tang, tarage, tartness, tast (obsolete), taste, tastelessness, tasting, unsavoriness, unsavouriness
Adjectives	Acid, acidic, appetizing, appetizing, bitter, bitter-sweet, bland, dainty, delectable, delicious, delightsom(e), disgusting, flavorless, flavorful, flavourful, flavourless, flavoursome, gamy, indigestible, insipid, juicy, mellow, palatable, piquant, pungent, racy, rancid, rank, salt/salty, sapid, savory, savoury, savourly, seasoned, sharp, sour, soured, sower (archaic form of sour), spicy, stale, sweet, tangy, tart, tasteless, tasty, toothsome, unpalatable, unsavor, unsavour, unsavoury, unsavory, unseasoned, unsweet, unsweetened, wearish, wersh, yummy
Verbs	Drink (up), drinking (up), drank (up), drunk (up), eat (up), ate (up), eateth (archaic), eaten (up), eating (up), distaste, distasting, distasted, mistaste, mistasted, mistasting, partake, partaking, partook, partaken, relish, relisheth (archaic), relishing, relished, season, seasoning, seasoned, smack, smacking, smacked, smatch (obsolete), sweeten, sweetening, sweetened, taste, tasting, tasted
Adverbs	Sweetly, sourly, tastefully, bitterly, tastingly, unsavourily, unsavourly, insipidly, savourously, savourily, flavourfully

Table 4
Lexical units for Taste

Hyperparameter	Value
$\beta1$, $\beta2$	0.9, 0.99
Dropout	0.2
Epochs	20
Batch Size	32
Learning Rate (LR)	0.0001
Decay Factor	0.38
Cut Fraction	0.3
All tasks loss weight	1

Table 5
Hyperparameter value used for the experiments which yield the best results

Appendices

A. Lexical Units and Frame Elements

In Table 4, we display the list of lexical units or taste words presented in [16].

B. Hyperparameter Values

The hyperparameter setting for all our models is presented in Table 5. The setting is the default MaChAmp's hyperparameter values, with the addition of loss weights at 1, and 20 epochs of training.