To Eat and Beyond: A FrameNet-Inspired Annotation of Food and Its Uses Over Time

Teresa Paccosi Gauri Bhagwat Marieke van Erp
KNAW Humanities Cluster, DHLab, Amsterdam
{teresa.paccosi,gauri.bhagwat,marieke.van.erp}@dh.huc.knaw.nl

Abstract

We present an annotation scheme and a manually annotated dataset in English, grounded in Frame Semantics and its generative extension through qualia relations, developed specifically for the food domain. Our primary goal is to capture the diverse and often less frequent uses of food in historical English texts, with a particular focus on the various processes to which food is subjected and the contexts in which it is employed. We provide the annotation scheme, describe the annotation process and release the annotated dataset for food and its uses, along with some preliminary experiments assessing the capabilities of LLMs in applying this annotation scheme.

Keywords: Food Annotation, Frame Semantics, Historical Data

1. Introduction

Food plays a central role in everyday life, not only as a fundamental human necessity but also as a rich and multifaceted domain through which cultural norms, social practices, and personal identities are continuously shaped and expressed (Murcott, 2019). Food can for example be a diplomatic tool, think of lavish state banquets that are used to curry favours (Bendix and Fenske, 2014), a practice that in the Western world goes already back to the 17th century (Lamal, 2024). Foodstuffs are also medicinal tools: ginger, for example, was and is still used to soothe sore throats (Bode and Dong, 2011), while cloves were used to ward off moths (Marintan et al., 2018). While food as a source of nutrition has been extensively studied across various research fields (cf. Gerhardt et al., 2013; Jurafsky, 2014), its uses beyond mere consumption have received considerably less attention in the fields of Natural Language Processing (NLP) and Digital Humanities (DH). In this paper, we address this gap by representing the food domain beyond its nutritional use. We do so by taking into account its multiple identities as they emerge from different usage contexts in historical texts. Our interest lies both in examining how food-related practices and preparations have changed over time, as they reflect and evolve with the cultural identity of a community, and in capturing the uses of food beyond their dietary uses, including less frequent applications, with the aim to model a comprehensive approach to food that encompasses its full range of functions.

After discussing related work (Section 2) and the theoretical background underlying our annotation guidelines (Section 3), we present our annotation scheme, grounded in Frame Semantics and its extensions (Section 4). We then describe the annotation process carried out by human annotators and the final annotated dataset (Section 5). We furthermore present a comparison experiment with Large Language Models (LLMs) in annotating texts using this scheme (Section 6). We end with conclusion and future directions (Section 7). Our annotation guidelines, the dataset, the prompt and the code we used for the experiments can be found at: `https://github.com/trifecta-project/annotating-food-in-context`.

2. Related Work

Most NLP research on food identification has concentrated on representing food primarily through its nutritional properties, often in service of health-oriented, and more recently, also environmental applications (Popovski et al., 2019a; Cenikj et al., 2020; Stojanov et al., 2021; Van Erp et al., 2021; Agarwal et al., 2024). This perspective has contributed to the rise of *computational gastronomy* (Goel and Bagler, 2022; Cenikj et al., 2020; Bagler and Goel, 2024), a field that aims at modeling and extracting ingredients and their components (Stojanov et al., 2021; Agarwal et al., 2024). These efforts generally involve specific corpora that focus on nutritional values, based on annotated recipes sourced from modern culinary platforms and online databases (Popovski et al., 2019b; Wróblewska et al., 2022). In contrast, historical perspectives on food call for broader and more specific annotation criteria that go beyond nutritional information (Paccosi and Tonelli, 2024). Computational approaches to the semantic representation of specific conceptual domains, such as medicine (Dutra et al., 2023) and environmental studies (L'Homme et al., 2020), have adopted Frame Semantics (Fillmore et al., 1982) as their representational framework. This theoret-

Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026), pages 997–1008
11–16 May 2026.

ical framework has proved highly effective in capturing broad conceptual structures and the contextual relations among meanings. However, the generalising effort underlying its formulation, and even more so its large-scale implementation in the FrameNet repository (Ruppenhofer et al., 2016), has exposed certain limitations when dealing with domain-specific or temporally bound semantics. In particular, while Frame Semantics offers a powerful model for representing meaning in context, the general-purpose design of FrameNet prevents it from fully accounting for the fine-grained and context-dependent meanings that emerge within specialised domains. For instance, in the case of food-related semantics, FrameNet primarily models eating and nourishment, leaving aside other culturally and historically significant uses of food, highly relevant in approaching its diachronic evolution. Previous research on historical language within specialised linguistic domains, such as the sensory domain, has demonstrated the suitability of Frame Semantics for capturing meaning in a comprehensive and context-sensitive way. However, it has also revealed the need for ad hoc adaptations and extensions to account for domain-specific phenomena (Tonelli and Menini, 2021; Paccosi and Tonelli, 2024). Frame Semantics has also been employed to approach the study of food-related language; however, again, previous work has primarily focused on food as a nutritional source (Diederich, 2015), without considering its other uses, especially from a diachronic perspective.

3. Theoretical Background

Before describing our annotation scheme, we introduce the theory of Frame Semantics and its subsequent extensions in FrameNet (Ruppenhofer et al., 2016) and FrameNet Brasil (Belcavello et al., 2020), since they form the basis of our annotation scheme. Frame Semantics models the meaning of words by means of semantic frames, where each frame represents sets of words that evoke specific perspectives or participants within particular types of events. Within this theory, the meaning of a word is modeled as a mental representation, in which certain words, called Lexical Units (LUs), evoke specific situations called *frames*. Within each frame, a set of participants called Frame Elements (FEs) play distinct roles that contribute to the interpretation of the scenario. In a sentence such as “Giovanni eats a delicious apple”, *eats* is the verb that triggers the scenario of nutrition, namely one of the Lexical Units of the frame INGESTION, while *Giovanni* and *a delicious apple* are the *Ingestor* and *Ingestible* Frame Elements respectively. A frame representation is also able to include the different meanings a word can take on, as well as the different perspectives through which a concept can be described. Although its approach is designed to deal with polysemy, the effort was mainly directed towards generalisation. While FrameNet (Ruppenhofer et al., 2016) was in fact developed as a general-purpose resource, its application has become more specialised and it has become clear that word meanings must be interpreted in light of their specific usage contexts, including the specificity of particular domains (Robichaud and Rüggeberg, 2014). For this reason, FrameNet Brasil (FR-Br) introduced the concept of ternary qualia relations (Belcavello et al., 2020; Torrent et al., 2022, 2024), which allow for a more fine-grained modeling of word senses, linking different frames between them. In the qualia structure, the meaning of lexical items is structured on the basis of four generative factors called qualia roles. Each quale captures how human beings understand objects and relations in the world and provides a minimal explanation for the linguistic behavior of the lexical items. For instance, a given object can be observed and understood through its constituents or parts, such as material, weight or characteristics (*constitutive*), its aim or scope of use (*telic*), its coming into being (*agentive*), or in its basic categorical description (*formal*). Ternary qualia relations, a structure which includes all these relationships, introduced in FR-Br, are an extension of the original qualia structure from Generative Lexicon Theory (Pustejovsky, 1998), which is aimed at describing the meaning of words according to these four dimensions of representation. Qualia roles in FrameNet capture specific aspects of the meaning of a word by defining its relation to another evoked concept, as well as contextual meaning through FE-to-FE projections, i.e., frame-mediated ternary relations in which one FE or LU is linked to another. In this representation, a word can be both the Lexical Unit of a Frame and the Frame Element of another frame. Building on this idea, we propose annotation guidelines based on a FrameNet- and qualia-inspired formalisation, aimed at creating specialised resources for training systems to extract food mentions and their uses and manipulations in a diachronic perspective.

4. Capturing the Food Domain

As stated in Section 3, FrameNet was originally developed as a general-purpose resource focused on contemporary language. In the FrameNet repository the current FOOD frame considers food only in its consumptive dimension.[1] Other frames which include food in their FEs, such as INGESTION or COOKING_CREATION, still account exclusively for

[1] https://framenet.icsi.berkeley.edu/fnReports/data/frameIndex.xml?frame=Food Last visited: 22 October 2025

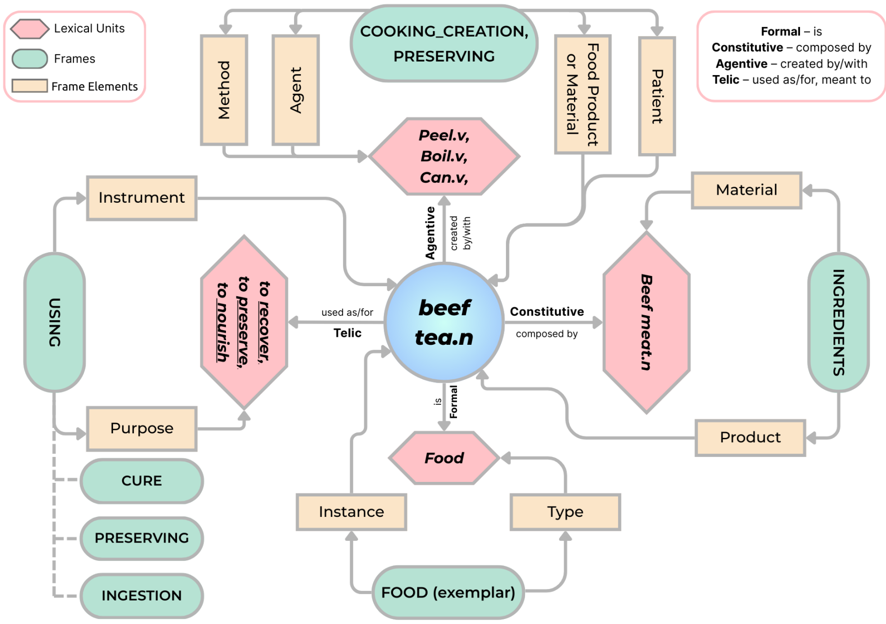

Figure 1: Food Domain through frames and semantic relation

the consumptive dimension of food, excluding other purposes for which food can be used. We argue that capturing the different contexts in which food takes on different roles is essential, especially in historical studies. To address this gap, we provide a formalisation to capture the different food usages, drawing on FrameNet frames while extending their scope. In this first phase, we mainly focus on an alternative use of food (medicine) and a particular type of process of manipulation of food (preservation). In the remainder of this section, we present the frames included in our annotation scheme to capture food-related descriptions in texts from a variety of perspectives. We define food exclusively in its literal dimension, leaving metaphorical or figurative uses of food expressions to future work. Fig.1 presents a schematic overview of our annotation scheme. The core of the scheme is represented by the LUs of the `FOOD` frame, of which *beef tea* is an example (the blue center of the structure), defined as the exemplar frame on which all the other frames depend. The list of LUs in the exemplar frame `FOOD` includes all the food items mentioned in a text, encompassing both contemporary and historical foods, along with their spelling variations. These LUs include all the words that refer or have referred to food items in the real world. The standard definition of food in the Oxford English Dictionary is "any nutritious substance that people or animals eat or drink in order to maintain life and growth" and serves here as a baseline for identifying food. However, our main expansion is that we do not restrict our notion of food to its act of consumption. For instance, substances that are normally ingested as nourishment but can also be applied externally, such as on the skin to treat a wound, are included in our conceptualisation of food. In other words, we consider food as any substance typically used for nutrition, even when it is employed for non-ingestive purposes. In the figure, we provide an example of LU in the center, namely *beef tea*, a beef broth used in the past to help convalescents.[2] Through the different relations, any LU can connect to various frames that represent different characterisations of food. As in FR-Br, our approach is inspired by an adaptation of the four relationships namely *formal*, *constitutive*, *agentive*, and *telic*, each of which will be described in detail below. With each of the qualia-like relationships, we present an example sentence from the annotation process. The example sentences are all annotated using the INCEpTION annotation environment version 38.3 (Klie et al., 2018).

4.1. Formal Frame

The **formal** characterisation introduces the exemplar frame and typically provides the basic categorical description of an object. In our scheme, this is represented by the `FOOD` frame, as defined in FrameNet and FR-Br. It includes a single core FE, *Food*, i.e. the LUs that refer to food items which we call *FOOD_LU*, and three non-core FEs: *Constituent_Parts*, *Descriptor*, and *Type*. Core FEs in

[2] https://en.wikisource.org/wiki/Mrs._Beeton%27s_Book_of_Household_Management/Chapter_XLV Last visited 10 March 2026.

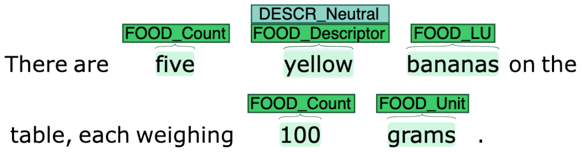

Figure 2: Example annotation of the FOOD exemplar frame

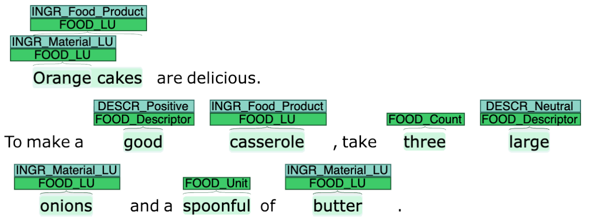

Figure 3: Example annotation of the INGREDIENTS constitutive frame

FrameNet are FEs which are essential to the meaning of a frame, while non-core FEs contribute to the description but are not fundamental. In this phase, we mainly focus on core FEs. For the purposes of our study, we do not distinguish between *Descriptor* and *Type*, as they represent closely related concepts; instead, we unify them under the FE *Descriptor*. To enrich future historical analyses, we add a secondary level of annotation that captures the context-dependent polarity of *Descriptor* (positive, negative, or neutral). Given our interest in the characterisation of food and its development over time, we also introduce two additional FEs to annotate quantity, *Count* and *Unit*. Prior research has shown that in trade logs, the quantity of a foodstuff can also be a possible marker for alternative uses of food (Bhagwat et al., 2025). These additional FEs are adopted from the MEASURES frame. An example of the FOOD annotation is given in Fig. 2.

4.2. Constitutive Frame

The INGREDIENTS (INGR) frame is activated by any LU referring to a food substance used to produce another food item, according to what in FR-Br they called the **constitutive** qualia relation, namely a relation expressing the relationship between a given object and its constituents, such as material, or characteristic parts (see Fig. 3 for an example of annotation). This frame is present in FR-Br, but not in FrameNet, and it is defined as comprising a material, i.e., any food that serves as an ingredient for another, used to create a product. The frame includes seven FEs. The core FEs we included in our scheme are *Material*, and *Product*. The food participant in terms of LU-to-FE projection in this frame is represented by *Product*, which in our scheme is explicitly called *Food_Product* to highlight its connection with the FOOD frame. In our scheme, in a sentence such as "beef tea is prepared by infusing beef meat in water", *beef meat* is the LU of INGREDIENTS (thus *Material*) and *beef tea* is both the LU of FOOD and the FE *Product* of INGREDIENTS. In our scheme, the INGREDIENTS frame is interpreted as expressing what a given FOOD is *composed of*.

4.3. Agentive/Manipulative Frame

Unlike the previous two (FOOD and INGREDIENTS), these frames model events rather than entities. In this case, we consider three frames in the **manipulative** (*created with*) and **agentive** (*created by*) characterisations, namely COOKING_CREATION, APPLY_HEAT and PRESERVING. These frames describe food and meal preparation and manipulation, including any type of manipulation performed, explicitly or implicitly, by an agent on the food, either to create or to modify it. For the COOKING_CREATION frame, exemplified in Fig. 4, we consider only the two core FEs, namely *Cook* and *Produced_Food* (the food participant in our scheme, which already carries the specification _*Food* in FrameNet). While APPLY_HEAT focuses primarily on the process of handling and transforming the ingredients, COOKING_CREATION emphasises the edible entity resulting from the process. We aggregate these two perspectives under COOKING_CREATION, which covers a broad range of food manipulation verbs. In our annotation scheme, rather than using the single Produced_Food category for both input and output of culinary processes, we distinguish between Food_Material and Food_Product: Food_Material represents the raw or initial food to which a culinary process is applied, while Food_Product corresponds to the final product resulting from a cooking effort.

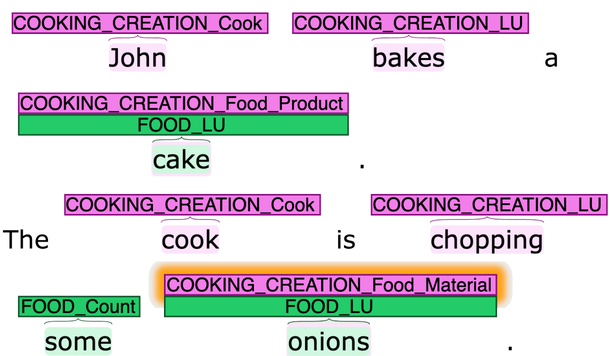

Figure 4: Example annotations of the COOKING_CREATION frame, illustrating the two possible Frame Elements.

For the PRESERVING frame, we focus on the role of the preserved *Patient*.[3] In our annotation, the *Patient* must be realised as food to trigger this frame, thus becoming *Food_Patient*, which is preserved to prevent deterioration, potentially using a *Medium* that can itself be food. While food can be annotated as either *Patient* or *Medium*, our primary focus in this realisation is on the Food-to-Patient relation, which captures the manipulation aspect. Nonetheless, as we are also interested in cases where food functions as a preserving tool, its telic realisation, presented in the next section, includes instances where it acts as a preserving medium.

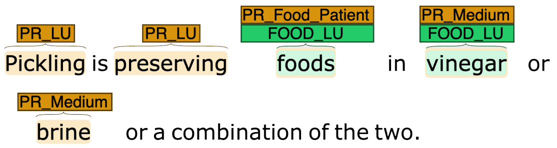

Figure 5: Example annotation of the PRESERVING frame

4.4. Telic Frame(s)

This last frame differs from the others in that it provides a prototypical structure for a set of frames that share the same purpose-oriented configuration, each capturing a different use or purpose of food. The **telic** relation refers to the specific function or intended purpose of an object, namely what it is *meant to* or what it is *used for*. In the case of food, this dimension of telicity is encoded in the core FEs of the frame USING, which is taken here as a reference model for examining how use-driven meanings of food are structured. These elements, recurring across the selected frames, support the conceptualisation of the various uses of foods. We used the core FEs of USING as a reference point to identify the shared FEs in the selected frames. Specifically, the relevant FEs are: *Agent*, the individual or entity using the instrument to achieve a goal; *Instrument*, the entity used by the Agent to fulfill their purpose which in our scheme is the *FOOD_LU*; and *Purpose*, which identifies the specific goal or reason for using the *Instrument* (here food), which in our scheme is represented by the specific LUs of the selected telic frames (e.g., *CURE_LU*, *PR_LU*, etc.). The frames activated by the telic relation are:

1. CURE: In this frame, the purpose of food is that to treat specific afflictions. It concerns a *Healer* treating or curing an *Affliction* (such as an injury, disease, condition, or pain) affecting a *Patient*, specifically when a food item is mentioned as a particular *Treatment* or *Medication* (labeled in Fig. 6 as *Food_Treatment*). As mentioned above, we focus solely on the core FEs, and we aggregated *Medication* and *Treatment* into the single FE *Food_Treatment*. This choice reflects the fact that food is conceptually closer to *Treatment*, while not excluding its potential overlap with *Medication*.

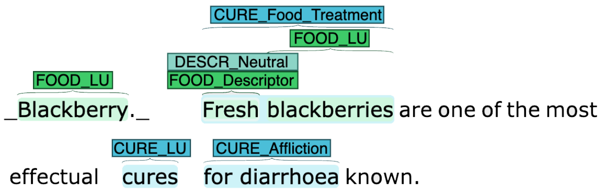

Figure 6: Example annotation for the CURE frame

2. PRESERVING (PR): Concerning this frame, the purpose of food is that of being the substance used to preserve another item from decay. An example is shown in Fig. 5 (see *PR_Medium*).

3. INGESTION (INGE): In this frame, the purpose of food is its most common one: consumption. The frame formalises this by representing an *Ingestor* who consumes a food or drink item, the *Ingestibles* (here, *Food_Ingestibles*), an action that involves placing the item in the mouth with the intent of delivering it to the digestive system. See Fig. 7.

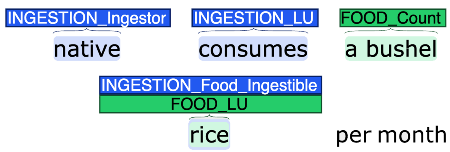

Figure 7: Example annotation for the INGESTION frame

The next section illustrates the annotation process for these frames and their semantic relations, following the guidelines we have just introduced.

5. Annotation Process

In this section, we describe the annotation process and a first evaluation of the annotation guidelines. Our efforts primarily focus on the frames that emerged from discussions with food historians of the PRESERVARE Project[4] at our institute.

[3]In this frame, we mean *Patient* in the grammatical sense, i.e. target or undergoer, not *Patient* in a medical sense as in the cure frame.

[4]https://preservare.eu/ Last visited: 10 March 2026.

Nonetheless, during the annotation phase, we also addressed the remaining frames.

5.1. Data Selection

The documents selected for annotation were identified through a semi-automatic process. We employed a food entity extractor[5] for English together with a list of words corresponding to LUs of the frames of interest to retrieve texts with a high density of references to food and its related uses, particularly those involving preservation and medicine (e.g., cure, dry, pickle, alleviate etc.). We harvested texts from Project Gutenberg,[6] the Internet Archive,[7] and a website containing a collection of historical cookbooks.[8] A food entity was annotated only when a LU of interest occurred within a ±5-word window around it. From this subset, we manually selected texts containing at least 100 food entities co-occurring with a relevant LU. This strategy ensured the selection of texts with the greatest number of references to the food-related uses of interest in this work. To construct a dataset that is balanced and suitable for diachronic analysis, we included texts from three centuries: the 18th, 19th, and 20th. For each century, we selected three to four documents, with a length of around 300 sentences each, representing different genres, including travel reports (*travel/ethnography*), medical and scientific texts (*medicine/science*), and cookbooks or household manuals (*recipe/household*), to capture a diverse range of food uses and descriptions distributed across time.

5.2. Annotation Workflow

Following the FrameNet annotation principles, the analysis of food-related situations is carried out in two main steps. Given a sentence, the first task is identifying and annotating frame-evoking words referred to as **Lexical Units** (LUs) to instantiate the FOOD frame. These LUs typically correspond to mentions of food items, encompassing both contemporary and historical foods. Once a *FOOD_LU* has been identified, the corresponding **Frame Elements** (FEs) participating in the event, such as *FOOD_Descriptor* or *FOOD_Count* should be annotated. Only after the core food-related situation has been established, the annotation proceeds by identifying any additional LU that may evoke alternative uses of the same food item. These are LUs

[5] https://huggingface.co/carolanderson/roberta-base-food-ner

[6] https://www.gutenberg.org/ Last visited 24 October 2025.

[7] https://archive.org/ Last visited: 24 October 2025

[8] https://www.foodsofengland.co.uk/references.html Last visited: 24 October 2025

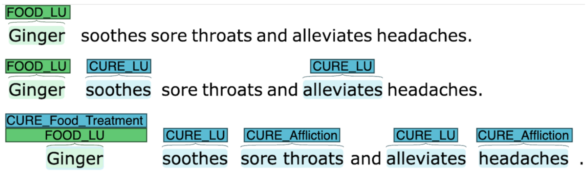

Figure 8: Annotation workflow

of the other proposed frames that contribute to the broader food domain, such as PRESERVING, or CURE. To sum up the process (see also Fig. 8), in a given sentence the annotator has to:

- Identify the words referring to food and annotate them as *FOOD_LU*.
- Annotate Frame Elements (FEs) related to each *FOOD_LU* in the FOOD frame (e.g., *FOOD_Descriptor*, *FOOD_Unit*, etc.);
- Check whether the food item participates in another event or usage that activates another frame (e.g., such as PRESERVING, or CURE), namely if any other LU or food realisation is present. If so: (1) Annotate the predicate (e.g., *alleviate* in the example sentence of Fig. 8) or any other argument of other frames (e.g., nouns for INGREDIENTS) as the LU of that frame, here *alleviate* as CURE_LU; (2) Re-annotate with a double span annotation the food item with its new role in the frame (e.g., *CURE_Food_Treatment* for *ginger* in the example sentence);
- Annotate the rest of the possible FEs of the newly found frame (e.g., *CURE_Affliction*).

A list of the initial LUs proposed for each frame, with the exception of FOOD and INGREDIENTS, whose LUs are too numerous to be included in a comprehensive list, is provided in the annotation guidelines on our GitHub repository. The list of LUs may be extended during the annotation process if new candidates are considered relevant for the specific frame evocation.

5.3. Adjudication Phase

We selected one document from *medicine/science* and one from *household/recipes*, and conducted an initial round of annotation on 30 sentences containing at least one *CURE_LU* and 30 sentences containing at least one *PR_LU*. Annotators were provided with the raw full sentences and asked to identify LUs and FEs, using INCEpTION version 38.3 (Klie et al., 2018), a web-based annotation platform customised to our guidelines. Following the annotation, a discussion was held to identify difficulties and refine the annotation guidelines accordingly. Annotators with diverse expertise were

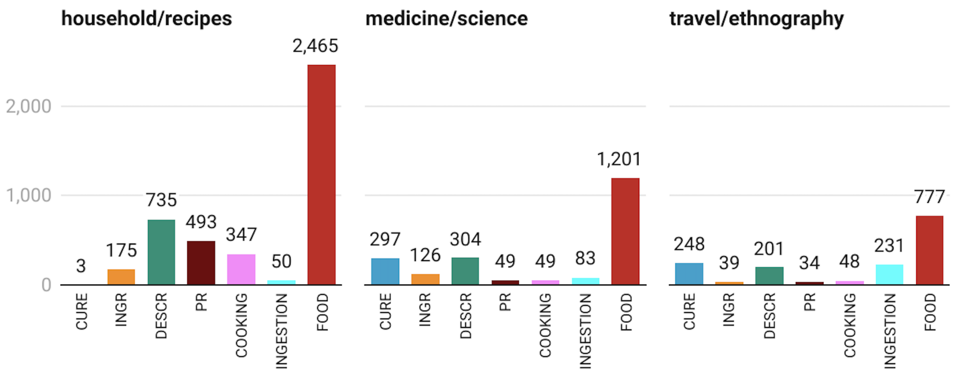

Figure 9: Distribution of frames across genres in our dataset

selected to ensure that the guidelines would support the creation of a valuable resource for historians and scholars investigating the evolution of food-related language. The adjudication phase involved a digital historian, two computational linguists, and a PhD student in Digital Humanities.

We computed the Inter-Annotator Agreement (IAA) using Krippendorff's α (Krippendorff, 2011), and found that CURE was the easiest frame to identify, while the others tend to have a general low agreement, likely due to their LUs' variety. In the first phase, the results indicate an overall low agreement in span selection, with scores ranging from 0.29 to 0.43. This first round of adjudication revealed that the lowest agreement occurred between the digital historian and the other annotators, underscoring the strong linguistic competence required to apply the scheme effectively. Upon reviewing the annotations, it became evident that the primary issue lay in the misunderstanding of the principles underlying Frame Semantics. In particular, a frame cannot be instantiated without the prior identification of a LU. Nevertheless, in many cases, annotators labeled only the FEs without marking any corresponding LU. Following a thorough discussion, we revised the annotation guidelines, guiding the steps of the annotation, as we presented in Section 5.2. Additionally, to account for potentially 'implicit' information, we introduced a dual annotation strategy, allowing a relevant FE to also be annotated as a LU when explicit LUs are absent (see Fig. 10).

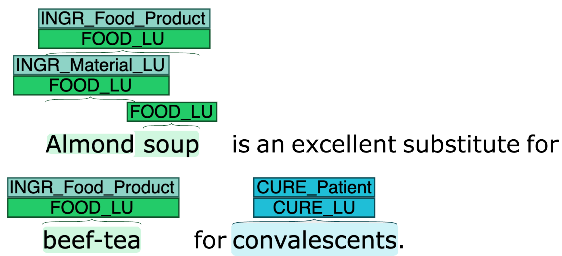

Figure 10: Example of double frame annotation for Frame Elements and Lexical Units

5.4. Agreement and Final Dataset

After this phase of consultation for the creation of the annotation guidelines, two expert linguists were trained on the annotation guidelines for the final annotation. The two annotators independently annotated 200 sentences randomly selected from *medicine/science* documents and 200 from *household/recipes* ones, drawn from the different centuries, for a total of 400 sentences. Initial agreement scores were 0.53 for the *household/recipes* documents (where we expect to find may mentions of the PRESERVING frame) and 0.72 for the second one (where we expect CURE frame to be the more frequent). Despite the notable improvement over earlier phases, a second round of adjudication was carried out. Disagreement primarily lies in the span mismatches and inconsistencies in the annotation of the *FOOD_Descriptor*'s polarity. We addressed these issues by manually correcting span mismatches and updating the annotation guidelines to include explicit span boundary definitions and a clearer definition of polarity. After this second adjudication phase, agreement improves considerably, reaching 0.94 for the first type of document and 0.97 for the second type.

Following this adjudication phase, the two annotators proceeded with the full annotation process, which resulted in a final dataset of approximately 6,000 annotated sentences ($\sim$300.000 tokens). We report the frames' annotation statistics in Table 1, and their distribution over genres in Fig. 9.

The distribution of frames confirms that our methodology and the selection of genres were effective for isolating specific uses of food. CURE tags mostly occur in medical and scientific texts, while for instance PRESERVING frame is predominantly found in household and recipe collections. By focusing on these genres, it becomes possible

Frame	Count
Cooking_Creation	444
Cure	545
Descriptor	1,240
Food	4,443
Ingestion	364
Ingredients	340
Preserving	576

Table 1: Total number of annotations in each frame in the final dataset

to capture more specialised or alternative uses of food and to construct a resource specifically designed for future analyses of such phenomena. Such a resource can also enable researchers to trace changes in food processing techniques (for instance, isolating *PR_LUs*) or medical practices (*CURE_LUs*) over time and to identify which foods were most frequently subjected to these processes, providing a valuable foundation for interdisciplinary research in the cultural and historical study of food from a quantitative perspective.

6. Assessing LLM Performance on Frame-Based Annotation of Food Uses

Building on this manually curated dataset, we tested several LLMs on our annotation scheme to evaluate whether it could be automatically applied for future annotations. The goal was twofold: first, to assess the clarity and robustness of the scheme in guiding automatic annotation, and second, to explore the potential for expanding the dataset beyond manual efforts. Recent studies have shown the effectiveness of LLMs in semantic annotation and other downstream annotation tasks (Reiss, 2023; Kuzman et al., 2023; Ding et al., 2023). While it remains challenging to generalize about their ability to fully replace human annotators, given that performance can vary considerably across tasks, LLMs have consistently shown strong performances in semi-automatic annotation scenarios (Giorgi et al., 2025; Aldeen et al., 2023). To this end, we randomly selected 150–200 examples per frame, namely a full sentence containing at least one of the frame (excluding `FOOD`, which is always present), balancing across genres and centuries, from the human-annotated texts. Models included a medium-size reasoning-focused (Deepseek-R1-Distil-Qwen-32B)[9] and two instruction-focused models, one medium (Mixtral-8x7B-Instruct-47B),[10] and one bigger (Llama-3.3-70B-Instruct),[11] as well as a larger instruction-based model, Deepseek_V3 (~671b),[12] to assess the impact of reasoning capabilities and model size. We opted for open-access models to ensure transparency and reproducibility of our results. We used the same prompt and generation settings for all models to ensure comparability. To make the models as deterministic as possible, we set the temperature to 0.0 and top-k to 1, so that the model always selects the most probable token. In this configuration, top-p (nucleus sampling) becomes irrelevant, as no sampling occurs. Therefore, top-p was kept at its default value, which does not affect the output, ensuring stable and reproducible token-level annotations. To promote fairness in comparison with human annotators, the prompt was designed to closely follow the annotation guidelines, combining few-shot examples with a chain-of-thought approach. Using the same prompt for all models, their performance was evaluated against the human-annotated gold standard by computing Precision, Recall, and F_1 for each frame over all entities, using an overall sequence-level evaluation. The results are reported in Table 2.

The results of the models' annotation show several patterns. Overall, larger and reasoning-focused models, consistently outperform the smaller instruction-focused models across most frames. This suggests that both model scale and the presence of reasoning capabilities are important factors for accurately annotating these texts. Frames such as `PRESERVING` (`PR`), `INGESTION` (`INGE`) and `CURE` are annotated quite well by the reasoning-focused and larger models, with higher F_1 scores and a more balanced P-R ratio. This indicates that these frames, which often require contextual inference, benefit from the reasoning capabilities and the increased parameter capacity of larger models. Conversely, smaller instruction-focused models show lower R or P depending on the frame, highlighting limitations in consistently extracting entities across diverse linguistic contexts. Some frames, notably `INGREDIENTS` (`INGR`) and `DESCRIPTOR` (`DESCR`), prove challenging for all models, with generally low F_1 scores for the smaller models and only a moderate performance for the larger model. This likely reflects the intrinsic difficulty of these frames, as noticed with human annotators, which may involve highly variable expressions, or a difficulty in clearly distinguishing the frame or its span boundaries. Notably, the `FOOD` frame presents a case where LLaMA, an instruction-focused model, achieves slightly higher F_1 than Deepseek-Distil-R1 and Deepseek_V3.

[9] https://huggingface.co/deepseek-ai/DeepSeek-R1-Distill-Qwen-32B

[10] https://huggingface.co/mistralai/Mixtral-8x7B-Instruct-v0.1

[11] https://huggingface.co/meta-llama/Llama-3.3-70B-Instruct

[12] https://huggingface.co/deepseek-ai/DeepSeek-V3

Frame	Deepseek-Distil-R1			Llama			Mistral			Deepseek_V3		
	P	*R*	*F1*	*P*	*R*	*F1*	*P*	*R*	*F1*	*P*	*R*	*F1*
PR	0.6182	0.6455	0.6205	0.2295	0.5791	0.3276	0.2164	0.3579	0.3204	0.7486	0.7459	**0.7472**
INGR	0.0220	0.0323	0.0262	0.0746	0.1667	0.1031	0.0495	0.1292	0.0641	0.2558	0.3667	**0.3011**
INGE	0.6254	0.6157	0.6214	0.4534	0.4310	0.4420	0.4505	0.3468	0.3114	0.7053	0.6667	**0.6855**
CURE	0.4635	0.4297	0.4589	0.3767	0.4861	0.4248	0.1786	0.4933	0.2518	0.5909	0.4432	**0.5062**
COOK	0.4221	0.3123	**0.3515**	0.1378	0.4598	0.2101	0.0711	0.3709	0.1192	0.1818	0.2292	0.2026
DESCR	0.3819	0.3433	**0.3046**	0.0933	0.2880	0.1426	0.027	0.3956	0.0425	0.2301	0.2886	0.2562
FOOD	0.5918	0.7543	0.6545	0.6148	0.7528	**0.6778**	0.1863	0.5487	0.2601	0.5864	0.7395	0.6540

Table 2: Frame-level metrics (Precision, Recall, F1) for each model, calculated over all entities of each semantic frame.

This suggests that for more straightforward extraction tasks, fine-tuning for instruction-following can provide an advantage even in medium-sized models. Nevertheless, Deepseek_V3 remains the top-performing model overall. Importantly, Deepseek_V3 emerges as the most reliable option also due to its more conservative prediction strategy. Unlike the other models, particularly LLaMA and Mistral, which tend to over-annotate by identifying substantially more spans than expected according to the gold standard, Deepseek_V3 maintains a stricter threshold for span detection. This reduces overgeneration and spurious predictions, resulting in more stable and robust performance across frames, even when it does not achieve the highest score on individual tasks. In summary, reasoning-focused and larger models perform quite well on frame-level annotation, although some frames remain challenging, suggesting the need for clearer instructions or more examples in the prompt. The Deepseek models appear promising for future silver-standard annotation, also because of their good balance between Precision and Recall, as refining their outputs could be less time-consuming than annotating from scratch. However, further investigation with different settings, along with additional systematic comparisons with human annotation, is needed to properly assess their suitability.

7. Conclusions and Future Directions

We presented an annotation scheme based on Frame Semantics, designed to capture the food domain from both historical and usage perspectives, with a particular focus on less frequent uses of food. Building on the qualia relations proposed in FR-Br, we modeled different characterisations of food, such as medicinal uses or preservative processes, through distinct semantic frames. To evaluate our guidelines, we annotated approximately 6,000 sentences from a selection of 18th, 19th, and 20th century texts. We assessed the scheme through human agreement and further tested several LLMs to explore the potential for automating the annotation process. The results are promising in both cases, confirming the quality of the guidelines and highlighting the potential of large, reasoning-capable LLMs for silver-standard annotation in this domain.

In future work, we plan to extend the annotation to other languages and text genres, and identify additional historical uses of food by continuing the ongoing collaboration with food historians. Ultimately, our goal is to compile a benchmark for the automatic extraction of food-related entities and their uses in historical texts, as well as extend the guidelines and annotations to other languages, providing a foundation for quantitative analyses and interpretations of food concepts in diachronic corpora.

Acknowledgments

This research was funded by the European Union under grant agreement 101088548 - TRIFECTA. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them. The authors would also like to thank Rik Hoekstra, Jiaqi Zhu, Louise Boyne, Veruska Zamborlini and the researchers of the project PRESERVARE, for the insightful discussions in the drafting of these guidelines, and their valuable feedback during the manual annotation phase.

Bibliographical References

Ayush Agarwal, Janak Kapuriya, Shubham Agrawal, Akhil Vamshi Konam, Mansi Goel, Rishabh Gupta, Shrey Rastogi, Niharika Niharika, and Ganesh Bagler. 2024. Deep learning based named entity recognition models for recipes. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 4542–4554.

Mohammed Aldeen, Joshua Luo, Ashley Lian, Venus Zheng, Allen Hong, Preethika Yetukuri, and Long Cheng. 2023. ChatGPT vs. human annotators: A comprehensive analysis of ChatGPT for text annotation. In *2023 International Conference on Machine Learning and Applications (ICMLA)*, pages 602–609. IEEE.

Ganesh Bagler and Mansi Goel. 2024. Computational gastronomy: capturing culinary creativity by making food computable. *NPJ Systems Biology and Applications*, 10(72):1–6.

Frederico Belcavello, Marcelo Viridiano, Alexandre Diniz da Costa, Ely Edison da Silva Matos, and Tiago Timponi Torrent. 2020. Frame-based annotation of multimodal corpora: Tracking (a)synchronies in meaning construction. In *Proceedings of the International FrameNet Workshop 2020: Towards a Global, Multilingual FrameNet*, pages 23–30, Marseille, France. European Language Resources Association.

Regina F. Bendix and Michaela Fenske. 2014. *Eating politically – Food and eating in politics*, pages 17 – 29. LIT Verlag Münster.

Gauri Bhagwat, Teresa Paccosi, and Marieke van Erp. 2025. Unpacking the weight of spices: a preliminary exploration of long-tail contexts in the VOC trade. In *DHBenelux2025*.

Ann M Bode and Zigang Dong. 2011. The amazing and mighty ginger. *Herbal medicine: Biomolecular and clinical aspects*, 2.

Gjorgjina Cenikj, Gorjan Popovski, Riste Stojanov, Barbara Koroušić Seljak, and Tome Eftimov. 2020. Butter: Bidirectional lstm for food named-entity recognition. In *2020 IEEE International Conference on Big Data (Big Data)*, pages 3550–3556. IEEE.

Catherine Diederich. 2015. *Sensory adjectives in the discourse of food*. John Benjamins Publishing Company.

Bosheng Ding, Chengwei Qin, Linlin Liu, Yew Ken Chia, Boyang Li, Shafiq Joty, and Lidong Bing. 2023. Is GPT-3 a good data annotator? In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 11173–11195, Toronto, Canada. Association for Computational Linguistics.

Lívia Dutra, Arthur Lorenzi, Lorena Larré, Frederico Belcavello, Ely Matos, Amanda Pestana, Kenneth Brown, Mariana Gonçalves, Victor Herbst, Sofia Reinach, et al. 2023. Building a frame-semantic model of the healthcare domain: Towards the identification of gender-based violence in public health data. In *Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana (STIL)*, pages 338–346. SBC.

Charles J Fillmore et al. 1982. Frame semantics. *Linguistics in the morning calm*, 3(5):111–137.

Cornelia Gerhardt, Maximiliane Frobenius, and Susanne Ley. 2013. *Culinary Linguistics: The chef's special*. John Benjamins Publishing Company.

Tommaso Giorgi, Lorenzo Cima, Tiziano Fagni, Marco Avvenuti, and Stefano Cresci. 2025. Human and llm biases in hate speech annotations: A socio-demographic analysis of annotators and targets. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 19, pages 653–670.

Mansi Goel and Ganesh Bagler. 2022. Computational gastronomy: A data science approach to food. *Journal of Biosciences*, 47(1):12.

Dan Jurafsky. 2014. *The language of food: A linguist reads the menu*. WW Norton & Company.

Jan-Christoph Klie, Michael Bugert, Beto Boullosa, Richard Eckart De Castilho, and Iryna Gurevych. 2018. The inception platform: Machine-assisted and knowledge-oriented interactive annotation. In *Proceedings of the 27th international conference on computational linguistics: system demonstrations*, pages 5–9.

Klaus Krippendorff. 2011. Computing krippendorff's alpha-reliability. Annenberg School for Communication Departmental Papers, Philadelphia, PA.

Taja Kuzman, Igor Mozetič, and Nikola Ljubešić. 2023. ChatGPT: Beginning of an end of manual linguistic data annotation? use case of automatic genre identification.

Nina Lamal. 2024. Eetbare nation branding? In *Wat schaft de pot? Eetcultuur in Nederland door de Jaren Heen*, pages 13–25. Sterck & De Vreese.

Marie-Claude L'Homme, Benoît Robichaud, and Carlos Subirats Rüggeberg. 2020. Building multilingual specialized resources based on FrameNet: Application to the field of the environment. In *Proceedings of the International FrameNet Workshop 2020: Towards a Global, Multilingual FrameNet*, pages 85–92.

Mega Alif Marintan, Tamara Adriani Salim, et al. 2018. The greatness of clove: Challenges in preserving historic newsprint collection in monumen pers nasional solo, indonesia. In *World Library and Information Congress 84th IFLA General Conference and Assembly*.

Anne Murcott. 2019. *Introducing the Sociology of Food and Eating*. Bloomsbury Publishing.

Teresa Paccosi and Sara Tonelli. 2024. 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*, pages 39–46.

Gorjan Popovski, Stefan Kochev, Barbara Korousic-Seljak, and Tome Eftimov. 2019a. Foodie: A rule-based named-entity recognition method for food information extraction. *ICPRAM*, 12:915.

Gorjan Popovski, Barbara Koroušić Seljak, and Tome Eftimov. 2019b. FoodBase corpus: a new resource of annotated food entities. *Database*, 2019:baz121.

James Pustejovsky. 1998. *The generative lexicon*. MIT press.

Michael V Reiss. 2023. Testing the reliability of chatgpt for text annotation and classification: A cautionary remark. *arXiv preprint arXiv:2304.11085*.

Benoît Robichaud and Carlos Subirats Rüggeberg. 2014. Discovering frames in specialized domains. In *LREC*.

Josef Ruppenhofer, Michael Ellsworth, Myriam Schwarzer-Petruck, Christopher R Johnson, and Jan Scheffczyk. 2016. Framenet ii: Extended theory and practice. Technical report, International Computer Science Institute.

Riste Stojanov, Gorjan Popovski, Gjorgjina Cenikj, Barbara Koroušić Seljak, and Tome Eftimov. 2021. A fine-tuned bidirectional encoder representations from transformers model for food named-entity recognition: Algorithm development and validation. *Journal of medical Internet research*, 23(8):e28229.

Sara Tonelli and Stefano Menini. 2021. Framenet-like annotation of olfactory information in texts. In *Proceedings of the 5th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pages 11–20.

Tiago Timponi Torrent, Ely Edison da Silva Matos, Frederico Belcavello, Marcelo Viridiano, Maucha Andrade Gamonal, Alexandre Diniz da Costa, and Mateus Coutinho Marim. 2022. Representing context in Framenet: A multidimensional, multimodal approach. *Frontiers in Psychology*, 13:838441.

Tiago Timponi Torrent, Ely Edison da Silva Matos, Alexandre Diniz Costa, Maucha Andrade Gamonal, Simone Peron-Corrêa, and Vanessa Maria Ramos Lopes Paiva. 2024. A flexible tool for a qualia-enriched framenet: the framenet brasil webtool. *Language Resources and Evaluation*, pages 1–29.

Marieke Van Erp, Christian Reynolds, Diana Maynard, Alain Starke, Rebeca Ibáñez Martín, Frederic Andres, Maria CA Leite, Damien Alvarez de Toledo, Ximena Schmidt Rivera, Christoph Trattner, et al. 2021. Using natural language processing and artificial intelligence to explore the nutrition and sustainability of recipes and food. *Frontiers in Artificial Intelligence*, 3:621577.

Anna Wróblewska, Agnieszka Kaliska, Maciej Pawłowski, Dawid Wiśniewski, Witold Sosnowski, and Agnieszka Ławrynowicz. 2022. TASTEset - recipe and food entities dataset. In *IEEE Dataport*.

A. Appendix

In Table 4, we provide a list of the initial LUs proposed for each frame, with the exception of FOOD and INGREDIENTS, whose LUs are too numerous to be included in a comprehensive list here. The list of LUs may be extended during the annotation process if new candidates are considered relevant for the specific frame evocation. In Table 4, we provide an overview of the frames and a short definition of their LUs and FEs.

Frames
COOKING_CREATION: bake.v, baking.n, barbecue.v, blanch.v, boil.v, boiling.n, braise.v, broil.v, broiling.n, brown.v, char.v, coddle.v, concoct.v, cook up.v, cook.n, cook.v, cooking.n, deep fry.v, fix.v, fry.v, frying.n, grill.v, grilling.n, make.v, melt.v, melting.n, microwave.v, parboil.v, plank.v, poach.v, preparation.n, prepare.v, put together.v, roast.v, roasting.n, roast.v, roasting.n, saute.v, scald.v, scorch.v, sear.v, simmer.v, simmering.n, singe.v, steam.v, steaming.n, steep.v, stew.v, stewing.n, toast.v, toasting.n, whip up.v
CURE: alleviate.v, alleviation.n, curable.a, curative.a, curative.n, cure.n, cure.v, ease.v, heal.v, healer.n, incurable.a, improve.v, nurse.v, palliate.v, palliation.n, palliative.a, palliative.n, rehabilitate.v, rehabilitation.n, rehabilitative.a, remedy.n, remedy.v, resuscitate.v, soothe.v, therapeutic.a, therapist.n, therapy.n, treat.v, treatment.n, subdue.v
PRESERVING: can.v, cure.v, dry.v, pickle.v, preservation.n, preserve.v, salt.v, tinned.a, smoke.v
INGESTION: breakfast.v, consume.v, devour.v, dine.v, down.v, drink.v, eat.v, feast.v, feed.v, gobble.v, gulp.n, gulp.v, guzzle.v, have.v, imbibe.v, ingest.v, ingestion.n, lap.v, lunch.v, munch.v, nibble.v, nosh.v, nurse.v, put away.v, put back.v, quaff.v, sip.n, sip.v, slurp.n, slurp.v, snack.v, sup.v, swig.n, swig.v, swill.v, tuck.v

Table 3: List of Lexical Units for Taste. They are followed by their PoS tag, either verb (v), or noun (n).

Frame	FE/LU	Label	Definition
`Food`	**Food** (LU)	FOOD_Food_LU	Words referring to food items (i.e., LUs in the `Food` frame).
	Constituent_parts	FOOD_Constituent_Parts	A component or a part of the FOOD_LU.
	Descriptor	FOOD_Descriptor	A characteristic or classification of the FOOD_LU, including both descriptive and typological information. It is further annotated for intrinsic polarity: (1) Positive [DESCR_Positive]; (2) Negative [DESCR_Negative]; (3) Neutral [DESCR_Neutral].
	Count	FOOD_Count	The number of units in which the FOOD_LU is quantified.
	Unit	FOOD_Unit	The standard unit of measurement used to quantify the FOOD_LU (e.g., grams, liters).
`Ingredients`	**Material** (LU)	INGR_Material_LU	The food entity which is used to make a Product.
	Product	INGR_Food_Product	The entity or substance that is created out of the Material, usually a complete dish.
`Cooking_Creation`	LU	COOKING_CREATION_LU	Every word which expresses a cooking or a culinary manipulation process.
	Cook	COOKING_CREATION_Cook	The person who prepares the Produced_Food.
	Food_Product	COOKING_CREATION_Food_Product	It is the result of a culinary effort .
	Food_Material	COOKING_CREATION_Food_Material	It is the raw food to which a culinary process is applied.
`Preserving`	LU	PR_LU	Every word expressing the act of preserving something from decaying.
	Agent	PR_Agent	The person performing the intentional act that leads to the preservation.
	Medium	PR_(Food)_Medium	The substance in which the Patient might be submerged or treated with to be preserved. Sometimes, it can be food itself.
	Patient	PR_Food_Patient	The organic or non-organic matter that undergoes preservation (usually, food).
`Cure`	LU	CURE_LU	Every word expressing the act of treating or curing someone
	Affliction	CURE_Affliction	It represent the injury, disease, condition, or pain felt by the patient
	Healer	CURE_Healer	It represents anyone who treats or cures the Patient.
	Treatment/ Medication	CURE_Food_Treatment	The ingested, applied, injected, etc. substance designed to cure the Patient. In this case, food.
	Patient	CURE_Patient	It is the sufferer of the injury, condition, disease or pain.
`Ingestion`	LU	INGESTION_LU	Every word expressing the act of eating or drinking.
	Ingestibles	INGESTION_Food_Ingestible	The Ingestibles are the entities that are being consumed by the Ingestor.
	Ingestor	INGESTION_Ingestor	The Ingestor is the person (or animal) eating or drinking.

Table 4: Frames for Food Domain, FEs and definitions. In bold, we underline the core FEs of each frame.