Greek culinary tourism is lost in translation

Stella Markantonatou[1] and George Pavlidis[1]

[1] ILSP/Athena RIC, Artemidos 6, 15125 Paradisos Amarousiou, Greece
stiliani.markantonatou@gmail.com
[2] ILSP/Athena RIC, University Campus at Kimmeria, 67100 Xanthi, Greece
gpavlid@gmail.com

Abstract. In the framework of the ongoing development of an application that both offers a multilingual, comprehensive (ingredients, nutritional properties, cultural context, location) presentation of the culinary wealth of Central and East Macedonia and Thrace and draws on OCR and Machine Translation techniques, we develop an ontology of national and local foods and wines supported by a standards compatible multilingual thesaurus. We talk about the problems encountered in the particular terminological domain and outline our methodology of populating the thesaurus.

Keywords: Greek Culinary Tourism, Thesauri, Food Terminology

1 Introduction

Today, the culinary aspect of tourism is prominent and it seems that it is going to take more ground in the next decade [1]. At the same time, mass tourism seems to reach a tipping point as the quest of individual experiences is taking over [1]. This implies that the tourist of the next decade will seek the pleasure of discovering tastes together with the cultural traits of foods in their local settings rather than being directed to pre-selected restaurants. However allowing oneself to savouring foods in (medium or low) cost local restaurants and, at the same time, enjoying their cultural content and context, requires multi-dimensional knowledge. Naturally, knowledge of details of a foreign language and scripture is absolutely necessary [2]. However, more is required such as knowledge of the materials and their dietary properties, knowledge of the cooking techniques and, of course, knowledge of the cultural settings of the foods.

Greece has a language with own scripture and boasts a wealth of local foods and drinks that form part of a rich cultural context and could play a major role in culinary tourism. At the moment, however, only guided tours are available with gastronomy as the sole topic while gastronomy enhanced with the cultural traits of food seems to be a rare theme for guided tours. No relevant electronic applications are available at the moment, to the best of our knowledge. This paper describes the methodology that the project GRE-Taste will adopt in order to face this challenge for the areas of Central and East Macedonia and Thrace.

Certain IT solutions that have recently been proposed for solving the linguistic problem of culinary touristic experience provide the restaurants with menu translations [3]. The advantage of this approach is that the presentation of the food can be controlled by the restaurants but the disadvantage is that restaurants have to pay the

relevant costs; this fact implies that small/medium restaurants will be reluctant to invest this money. Other solutions provide Machine Translation of menus through mobiles and typing in/OCR-ing, such as the Word Lens [4] and the Purdue Menu Translator [5]. This approach seems better suited for the experience seeking tourist because, apart from ensuring freedom of movement, it allows for enhancements such as personalization and infinite enrichment of the information.

For our discussion here, it is important to stress that despite their technical differences, the so far mentioned applications require a special database for each language and none of them includes Greek (as yet). Greek has to make do with solutions such as Google Images, Google Translate, BabelNet and the Wikipedia. Google Images works best because it harvests the Greek food sites for images using powerful transliteration algorithms. However, images are not enough for the purposes of culinary tourism, for instance a Muslim visitor cannot know whether a meat dish is made of veal or pork. BabelNet and Wikipedia require some minimal knowledge of Greek as the names of the foods have to be searched either in Greek or be transliterated. However, a machine that would simply provide transliteration and then direct to these resources would not be a good solution because, unfortunately, it is a fact that neither satisfactory translations of Greek food terms nor sufficient multilingual descriptions exist out there in the web and the aforementioned resources cannot generate them. In addition to the above, it should be stressed that understanding a menu is not a terminological problem only, it has to do with the way menus are written and on the contextual knowledge required to fully appreciate a type of food.

We will expand on the linguistic problems in Section 2. In Section 3 we will talk about international efforts to offer a structured presentation of the combined knowledge that is necessary for supporting culinary tourism. Lastly, in Section 4 we will outline our methodology for developing multilingual thesauri and ontologies in order to support Machine Translation of menus and searches on culinary issues.

2 The language problem

Below, we present the features of Greek menus that increase the difficulty of understanding and make them a challenging target for Machine Translation. Additional evidence for our presentation we draw from 15 menus of restaurants, fast food places and taverns in Macedonia (8 menus from the Municipality of Serres) and Thrace (7 menus from the Municipality of Xanthi).

It should be mentioned that most of the menus do not provide an English translation (13 out of 15). One of the two menus providing an English translation has an addition with dishes of the date that is in Greek only. The provided English translations are generally reasonable given the challenges involved in the translation of culinary texts.

As said before, understanding a menu requires some non-trivial knowledge of Greek. The problems we have identified are listed below.

Terminology. For several characteristic dishes of Greece in general and of Macedonia and Thrace in particular there is no description or even no hint in the Wikipedia. On the other hand, Google translations, when they exist, are not always reliable.

An indicative list of very well known dishes offered in the menus that have no Wikipedia lemma follows: κλέφτικο, εξοχικό, στιφάδο ('stew', unsatisfactory Google translation also used in the menus), τζιεροσαρμάς, χουνκιάρ, κολοκυθοκεφτέδες ('pumpkin balls', wrong Google translation), μπουγιουρντί.

Furthermore, types of food that feature in the menus are not found in Wikipedia such as λαδόκολλα (=cooked or served in parchment paper), της ώρας (=grilled food), παντρεμένη (=something is combined with something else), τουρλού (=mixture of ingredients, often a creative approach of the chef). Types of meat such as κόντρα φιλέτο ('sirloin', Google translation), καβουρμάς, κοψίδι, πανσέτα, ψαρονέφρι[1] ('tenderloin', Google translation) are also missing from Wikipedia.

Given that most of the terms referred so far are about meat dishes and ways of cooking meat, all the contents of the menus from ψησταριές (=taverns and restaurants serving only roasted meat) could not be found in Wikipedia.

Compounding is a major word creation process in Greek and it is compounding that Greek uses to denote a range of foods that are not standardized, for instance creative versions of well-known types of food featuring in the menus such as κεφτέδες, κροκέτες, τηγανιά made with less-standard prevailing materials, e.g. κασερο-/φαβο-/κολοκυθο-κεφτέδες ('yellow cheese'/'split pea'/'zucchini'- 'balls'). These terms are unlikely to be found in databases such as Wikipedia or BabelNet while Google Translate returns good translations if both the parts of the compound are fairly common, for instance κρεμμυδοκεφτέδες 'onion balls'.

Furthermore, Greek uses multiword expressions to denote foods such as αγκινάρες αλά πολίτα, χοιρινό με σέλινα ('pork with celery', reasonable Google translation) μπακαλιάρος σκορδαλιά ('salt cod and garlic dip', reasonable Google translation, not in Wikipedia).

Lastly, dialectic names should also be taken into account. For instance, βλήτα is the name of a very popular summer dish based on the plant Amaranthus blitum. However, on some of the Ionian islands the term βλήτρα is used in the menus instead.

Independent searches in Wikipedia and BabelNet show that very common Greek dishes such as λεμονάτο, λαδερά (the most prominent family of Greek dishes), σπετζοφάι, αγκινάρες αλά πολίτα are either not mentioned at all or, more rarely, the corresponding Wikipedia lemma is in Greek only. Google translations do not fare well on this front either, for instance, λαδερά is translated as 'greasy'. In general, foods with multiword expression names receive a literal translation in Google, for instance, μπακαλιάρος σκορδαλιά is reasonably translated as 'salt cod and garlic dip' but μελιτζάνες παπουτσάκια is translated as 'eggplant shoes'.

Morphosyntactic issues. Several food names are multiword expressions, however, for some of them the order of the constituent words may vary freely e.g. πατάτες τηγανητές/τηγανητές πατάτες (=French fries). Furthermore, spelling of certain terms may vary widely, e.g. τηγανιτός, τηγανητός (=fried) while spelling mistakes are not rare. Lastly, terms that are either well known e.g. τηγανητές (=fried) or are used fre-

[1] The terms mostly describe the part of the animal; some of them are animal specific, e.g. pork or buffalo.

quently in the same menu, e.g. χειροποίητες (=handmade) may be abbreviated in irregular ways, for instance τηγ. and χειρ., χειροπ. respectively.

In the overall, the English speaking tourist, let alone tourists that speak other languages, has to switch between Google Translate and the Wikipedia in the hope to find a translation of food terms---this hope is often forlorn. In most cases it is unlikely that s/he will have some illuminating description of the food.

As expected, the main linguistic problem of translating menus is terminological, some of the terms being fixed and others being creative. Misspellings, idiosyncratic abbreviations and normal syntactic phenomena of Greek contribute to the complexity of the problem.

3 Structuring culinary, dietary and cultural knowledge

The international literature offers several examples of structuring culinary, dietary and cultural knowledge, most often than not independently of one another. Here we briefly describe two ontologies that adopt a comprehensive approach.

The YAMO+ ontology [6]. YAMO+ is a core ontology for food aimed to support application ontologies. Its creators have combined interviews of experts and extraction of knowledge from a wealth of resources including courses on food science, recipes, culinary dictionaries, nutrition glossaries, other food oriented ontologies, encyclopedias, relevant technical reports, the AROVOC controlled vocabulary [7] and large electronic lexica (WordNet [8]). The facets of the developed ontology reveal some of the main types of information on food that is required by application ontologies: Food, Diet, Ingredient, Recipe, Meal. All facets are structured with the ISA relation. In the Food facet, two main subclasses inherit from the root class Food, namely the classes Edible Food and Drinkable food. In the Diet facet, the top node Diet is subsumed by Regular diet, Special diet and Vegan diet, therefore the ontology covers the possible dietary preferences. In the Ingredient facet, the top node Ingredient is subsumed by Plant origin ingredient and Animal origin ingredient. Properties of the foods that are encoded as such are Colour, Nutritional Information, Taste, Temperature, Flavour, Texture, Appearance, Freshness, Recipe yield, Course Type, Diet type, Recipe category and Ingredient used. In short, the ontology describes all the foods from the point of ingredients, cooking techniques and nutritional properties and can support answers to questions like "shall I have this food for breakfast or for lunch", "is this food compatible with my diet", "does this food contain ingredients that I should avoid".

Europeana Food and Drink (EFD) ontology [9]. European Commission has funded the project EFD that has developed a multilingual ontology in order to "facilitate search (our comment: for cultural objects, in particularly in Europeana) and semantically enrich Cultural Heritage items pertaining to the 'food and drink' theme" [10: 1]. The project used Wikipedia as a primary resource for structured knowledge on food and its relations to material and immaterial culture such as objects of cooking and history of foods and drinks respectively. In addition, EFD used a host of resources such as food ontologies, AGRIVOC and the Art and Architecture Thesaurus [11] and

the TNG [12] for defining the ontology it has delivered. The project has produced "drink and food gazetteers" as foods and drinks are related with information about their geographical distribution, their role in social events (such as festivities, weddings etc) and with other cultural traits (such as instruments for cultivating their ingredients that can be accessed through Europeana).

Structured multidimensional knowledge on food will be the backbone of the GRE-Taste work as it is outlined in the next section.

4 Outlining our approach

GRE-Taste aims to provide the visitor of Central & East Macedonia and Thrace with a multilingual application that will offer culinary, nutritional and cultural information about the Greek and regional foods and drinks. The application, through keying in or an OCR plus Machine Translation or an optical recognition facility or combinations of them, will offer multidimensional knowledge about Greek food in Greek, English and Russian.

Such multidimensional knowledge has to be structured in order to be useful both for Machine Translation and for searching purposes. We develop rich multilingual thesauri that are at the heart of the enterprise. The thesauri capture the names of foods and ingredients, in addition to nutritional terminology and names of items, places, events and dates (instances of the last three classes are denoted by the so-called named entities) and their dialectic or other variants and map them on English and Russian names. Several food names, e.g. λαδερά, or cooking processes (e.g. αυγοκόβω) are unlikely to have translational equivalents in the other languages; these terms are transliterated in the corresponding languages in order to guide the visitors in pronouncing them. The thesauri offer a description of the transliterated word, for instance, for στιφάδο the description explains that it is a type of stew ('stew' is the translation provided by Google and used in various menus and recipes) and proceed in clarifying the exact nature of the dish. Vocabularies are being built from monolingual and parallel corpora consisting of menus and recipes collected with dedicated crawlers [13] as well as from local enterprises (the last ones are OCRed and edited), as well as existing food ontologies (including the ontologies presented in Section 3), various lexica (both printed and online), the AAT, the AGRIVOC and the Wikipedia. Experts in nutrition and local producers of meat products, wines and patisserie products support the enterprise. The relevant standards will be adhered to (Table 2 offers an overview of relevant standards).

To develop the multilingual thesauri, we use the in-house thesaurus editing environment and the thesauri for Greek folk art museums described in [14]. The thesaurus editing environment has been designed to allow for both building a thesaurus with a structure defined by the editor and mapping the terms on established thesauri such as the AAT.

To support searches, the overall information will be structured as an ontology that will share the features of EFD and YAMO+ since GRE-Taste will offer multidimen-

sional information about food extending over the culinary, the nutritional and the cultural domain. We will use Protégé [15] as an ontological editor.

Table 2. Basic ISO standards for terminology.

Code	Name and field
ISO 1087-1:2000	Terminology work - Vocabulary - Part 1: Theory and application
ISO 704:2009	Terminology work - Principles and methods
ISO 860:2007	Terminology work - Harmonization of concepts and terms
ISO 29383:2010	Terminology policies - Development and implementation
ISO 23185:2009	Assessment and benchmarking of terminological resources - General concepts, principles and requirements
ISO 15188:2001	Project management guidelines for terminology standardization
ISO 1951:2007	Presentation/representation of entries in dictionaries - Requirements, recommendations and information
ISO 10241-1:2011	Terminological entries in standards - Part 1: General requirements and examples of presentation
ISO 10241-2:2012	Terminological entries in standards - Part 2: Adoption of standardized terminological entries
ISO 12620:2009	Terminology and other language and content resources - Data category specifications

The parallel corpora are in their majority Greek-English ones and will support the development of the Machine Translation system.

Acknowledgements

The present work was supported by the project "GRE-Taste: Savouring Greece'', MIS code 5002437, co-financed by Greece and European Union in the framework of the Operational Programme "Competitiveness, Entrepreneurship and Innovation" 2014-2020.

References

1. The Ontario Culinary Tourism Alliance+Skift (2015) The Rise of Food Tourism (Special Report)

2. Jurafsky D (2014) The language of food: A linguist reads the menu. W.W. Norton & Company, New York, London

3. Translate my menu https://translatemymenu.com/en_US/

4. Word Lens https://en.wikipedia.org/wiki/Word_Lens

5. Purdue Menu Translator http://www.purdue.edu/newsroom/research/2011/110908BoutinMenutranslate.html

6. Madalli D, Chatterjee U, Dutta B (2017) An analytical approach to building a core ontology of food. Journal of Documentation, 73:123-144, Emerald Publishing Limited

7. AGROVOC http://aims.fao.org/vest-registry/vocabularies/agrovoc

8. WordNet https://wordnet.princeton.edu/

9. Europeana Food and Drink https://foodanddrinkeurope.eu/

10. Tugarev A, Tolosi L, Alexiev V (2015) Domain-specific modeling: Towards a Food and Drink Gazetteer. First International Keystone Conference, Coimbra, Portugal, Sep 2015. http://vladimiralexiev.github.io/pubs/Tagarev2015DomainSpecific Gazetteer. pdf

11. Art and Architecture Thesaurus (AAT) http://www.getty.edu/research/tools/vocabularies/aat/

12. Thesaurus of Geographical Names (TNG) http://www.getty.edu/research/tools/vocabularies/tgn/index.html

13. Mastropavlos N, Papavassiliou V (2011) Automatic Acquisition of Bilingual Language Resources. Proceedings of the 10th International Conference of Greek Linguistics, Komotini, Greece

14. Vacalopoulou A, Markantonatou S, Minos P, Toraki K (2018) Openly Available Resource for the Management and Promotion of Greek Museums with Folk Exhibits. Paper presented at ICSIMAT, Athens 2018

15. PROTÉGÉ ontology editor https://protege.stanford.edu/