Svetla KOEVA (*Sofia*)

MULTILINGUAL CULINARY LEXICON

Abstract

The article presents the **Multilingual Culinary Lexicon** – an online repertory for looking up information related to food, cooking, and culinary matters in twenty five languages. A Culinary Lexicon has been developed for Bulgarian, and based on the relations of equivalence with English, it has been connected with 23 languages for which (lexical-)semantic networks are available. This resulted in an online Multilingual Culinary Lexicon which contains synonyms, translation equivalents, (for some languages) definitions and examples, and for some synonym sets (synsets) – the respective ingredients. The synsets in Bulgarian were supplied with explanatory definitions, grammatical and stylistic labels, and examples in a similar way as in the regular explanatory dictionaries. The Bulgarian Culinary Lexicon comprises 3,222 synsets which contain 5,338 synonyms. The article offers an overview of the way the Multilingual Culinary Lexicon is being developed with a special focus on the additional lexicographic information it contains in comparison with the (Princeton) WordNet.

Keywords: Bulgarian Culinary lexicon, Multilingual Culinary lexicon, WordNet

1. Introduction

The **Multilingual Culinary Lexicon** (CulNet)[1] is an online repertory for looking up information related to food, cooking, and culinary matters in twenty five languages. The Multilingual Culinary Lexicon is primarily based on the WordNet structure (namely the Princeton WordNet and the Bulgarian WordNet) – a (lexical)semantic network whose nodes represent synonyms (synonym sets, synsets) denoting concepts and whose arcs, connecting the nodes, encode different types of relations (semantic: genus-kind, part-whole, etc.; morphosemantic: predicate-agent, predicate-instrument, etc.; derivational; extralinguistic; inter-language: translation equivalents) (Miller 1986; Fellbaum 1998).

A Culinary Lexicon has been developed for Bulgarian, and based on the relations of equivalence with English, it has been connected with 23 languages for which (lexical)semantic networks are available. For example, the noun {брашно} (flour) is described by the definition in the Princeton WordNet as a 'fine powdery foodstuff obtained by grinding and sifting the meal of a cereal grain'.[2] This word is connected to a more general term – the hypernym {хранителен продукт} (food product, foodstuff) 'a substance that can be used or prepared for use as food'; to a number of more specific words – hyponyms: {пшенично брашно} (wheat flour) 'flour obtained from ground and sifted grains', {соево брашно} (soybean flour) 'flour made from ground soybeans', etc.; with words denoting the whole of which it is a portion – holonyms: {хляб} (bread), {сладкиш} (pastry), etc.; and with equivalent words or synsets from other languages (if they are already included in the respective wordnets): {irin} in Basque, (farina) in Catalan, {fariña} in Galician, {brašno} in Croatian, {mel} in Danish, {bloem; meel; meelbloem} in Dutch, {jauho} in Finish, {farine} in French, {αλεύρι} in Greek, {קֶמַח} in Hebrew, {hveiti} in Icelandic, {farina} in Italian, {miltai} in Lithuanian, {mąka} in Polish, {farinha} in Portuguese, {făină} in Romanian – with a definition, {múka} in Slovak, {moka} in Slovene, {harina} in Spanish, {mjöl} in Swedish, etc.

We built on the existing resources for Bulgarian (such as the Bulgarian WordNet[3] (Koeva 2021), explanatory and synonymous dictionaries,[4] encyclopaedias, the Bulgarian National Corpus[5] (Koeva et al. 2012), etc.) and for other languages (wordnets) and:

i) selected the appropriate synsets from the Princeton WordNet;[6]

ii) represented the chosen synsets in the Bulgarian WordNet and enriched them with lexicographic data: explanatory definitions, examples, grammatical and stylistic labels, in this way constructing the **Bulgarian Culinary lexicon**;

iii) classified the synsets in the Bulgarian Culinary Lexicon into a system of categories related to food, cooking and culinary matters;

iv) Extracted information from the Bulgarian definitions and examples for the ingredients where possible;

v) linked the Bulgarian synsets with some of the available wordnets in a common resource – the **Multilingual**

Culinary Lexicon.[7]

This resulted in an online Multilingual Culinary Lexicon which contains synonyms, translation equivalents, (for some languages) definitions and examples, and for some synsets – the respective ingredients. The synonyms from the Bulgarian Culinary Lexicon are free for download;[1] the online interface[1] allows browsing the synsets, their immediate hypernyms and hyponyms (if part of the Multilingual Culinary lexicon), their definitions, ontology-based classification, and the extracted ingredients (for Bulgarian and English and the available data for the other languages). All the information is accessible in a non-structured way through the web interface of the Bulgarian WordNet.[3,8]

The paper offers an overview of the way the Multilingual Culinary Lexicon is being developed, with a special focus on the additional lexicographic information it contains in comparison with the (Princeton) WordNet. The structure of the paper is as follows: in Section 2 the general structure of the Multilingual Culinary Lexicon is presented; Section 3 is focused on how the information is organised for Bulgarian: selection of synonyms, construction of definitions, assignment of grammatical and stylistic labels, choice of examples; while Section 4 discusses the lexical, conceptual, and extralinguistic relations linking the selected synonyms and the concepts they represent.

2. General Overview of the Multilingual Culinary Lexicon

The **Multilingual Culinary Lexicon** contains nouns, verbs, and adjectives, and the organisation of the selected groups of nouns, verbs, and adjectives follows the structure of the Princeton WordNet. In WordNet "nouns are organised as topical hierarchies, verbs are organised by a variety of entailment relations, and adjectives and adverbs are organised as N-dimensional hyperspaces" (Miller et al. 1990/1993: 3). Nouns and verbs build hierarchical structures based on the relations of **hypernymy** and **hyponymy** among the synsets, called **hypernymy** and **troponymy** in the case of verbs (Fellbaum 1990/1993: 40). Descriptive adjectives are organised into structures which have a "central" synset, connected via the relation of **antonymy**, and satellite synsets, linked through the relations **similar to** and **also see** (Fellbaum et al. 1990/1993: 28). Relational and participial adjectives do not follow the typical "star-like" structure; rather, they are linked via derivational relations with a noun or verb respectively.

Bulgarian is the "core" language in the Multilingual Culinary Lexicon. The vocabulary for Bulgarian was initially selected based on the synsets included in the Princeton WordNet. The synsets under some hypernyms related to food, cooking, and culinary matters were selected: {food; solid food} 'any solid substance (as opposed to liquid) that is used as a source of nourishment'; {food; nutrient} 'any substance that can be metabolised by an animal to give energy and build tissue'; {vitamin} 'any of a group of organic substances essential in small quantities to normal metabolism'; {home appliance; household appliance} 'an appliance that does a particular job in the home; {cook} 'transform and make suitable for consumption by heating', etc. The chosen groups of thematically related words were further extended with derivationally related adjectives and synsets linked to them with morphosemantic and appropriate extralinguistic relations. The name of the resource – Culinary Lexicon – relates with the thematic scope of its content, although not all words that are included in it can strictly be classified as culinary, i.e. vitamins.

The initial list of potential candidates was manually evaluated[9] and some inappropriate synsets were excluded: for example, cultivated plant spices and their fruits are left, but not the plant genus and families. Some of the targeted synsets already have their counterparts in the Bulgarian WordNet. The synsets which were missing in the Bulgarian WordNet were automatically translated and manually evaluated as follows: a) words that do not express the target meaning are removed; b) words that signify the meaning but are in an inappropriate form (i.e. with spelling errors or with omitted or redundant parts) are edited; c) new synonyms that correspond to the meaning are included. The relations between synsets were automatically transferred from the Princeton WordNet and then manually evaluated, whereby the approach to the changes of relations (removal, addition, transformation) was conservative.

The node for an English synset for which the concept has not been lexicalised in Bulgarian is preserved; the meaning is presented descriptively and the definition is marked with the expression 'no lexicalisation'. For example, the English concept {hot cereal} is not lexicalised in Bulgarian and it is translated as {топла зърнена закуска} (hot breakfast cereal) with the definition 'no lexicalisation – a cereal that is served hot for breakfast'. At the moment the **Bulgarian CulNet** includes 521 synsets which do not have lexicaliztation in Bulgarian and 116 Bulgarian language specific synsets.

The lack of a lexicalisation in English is indicated within the structure of the **Bulgarian CulNet** as follows: i) the language-specific Bulgarian synsets are related to unique identification numbers not present at the Princeton WordNet; the language specific synonyms are transliterated into the Latin alphabet within notes to the

synonyms, and the definition is translated into English within a note to the synset; ii) the relations with other synsets show the position of the language-specific synset within the semantic network, and through these relations the position of the corresponding empty node in the Princeton WordNet can be identified. For example, the Bulgarian multiword expression {зелева сарма} represents a language specific concept that does not exist in the Princeton WordNet and it is attached a transliteration note: *zeleva sarma,* as well as a note with the English translation of the definition: 'rolled in a whole cabbage leaf that has been pickled in brine'. The hypernym {сарма} is also language specific and it is also attached a transliteration note: *sarma*, and a translation note: 'a kind of dish made of minced meat and rice, rolled in cabbage or vine leaves', etc. The hypernym of {сарма} (sarma) {ястие; блюдо; гозба; манджа}(dish) 'food prepared in a certain way by thermal or other type of cooking' is linked with its translation equivalent in English, and all of these relations outline the position of a corresponding empty node in the Princeton WordNet.

At the moment the **Bulgarian Culinary Lexicon** comprises 3,222 synsets, from which 2,936 synsets represent nouns, 224 synsets – verbs, and 62 synsets – adjectives. There are 5,338 synonyms altogether (some synsets contain only one word).

The synsets for Bulgarian were supplied with explanatory definitions, grammatical and stylistic labels, and examples in a similar way as in the regular explanatory dictionaries (described in more detail in the Section 3).

The synsets are additionally organised in a shallow hierarchy containing 48 categories and subcategories. The main categories are: Plant Food, Food Ingredients, Dairy, Fish and Seafood, Meat, Pasta, Condiments, Drinks, Dishes, Kitchen Appliances and Utensils, Eating and Diet, Restaurants, Cooking, Professions Related to Culinary, Adjectives Related to Food and Cooking, and Verbs Related to Food and Cooking. The following example illustrates the shallow hierarchy:

Category: Plant food

Subcategory: Vegetables

Subcategory: Fruits

Subcategory: Fruit type

Subcategory: Pulses and grains

Subcategory: Nuts

Category: Food ingredients

Category: Dairy

Subcategory: Cheese

…

For example, within the subcategory Vegetables are classified the following words: {репичка} (radish); {ряпа} (turnip); {савойско зеле} (savoy cabbage); {сладък картоф; батат} (sweet potato); {сладък пипер} (sweet pepper); {спанак} (spinach), and many others. The members of the categories are automatically predicted on the basis of the definitions and further evaluated manually.

The ingredients, if they exist and could be automatically identified by the definition and examples, are explicitly listed. For example, {бешамел} (white sauce, béchamel) with the definition '**milk** thickened with a **butter** and **flour** roux' is associated with the ingredients: *milk*, *butter*, *flour*.

Each synset in the Bulgarian WordNet is labeled with an identification number corresponding to the Inter-Lingual Index (ILI), which provides the option for linking wordnets via the relation of interlingual equivalence.

After Princeton WordNet, many (lexical-)semantic networks were developed for many languages, and this became a prerequisite for the creation of the Open Multilingual WordNet[10] (Bond et al. 2012; Bond et al. 2013), which is built around a shared set of "Base" Concepts (Bond et al. 2016). The notion of Base Concepts targets maximum overlap and compatibility across wordnets of different languages, allowing at the same time the distributive development of semantic networks with language-specific structure and lexicalisation patterns (Vossen et al. 1998, 53). A Collaborative Inter-Lingual Index (CILI) is developed that integrates the data from different wordnets so as to include salient and frequent lexicalised concepts in all languages and relations that mean the same things for all languages (Bond et al. 2016). Although some cross-language links might not express total equivalence between the meaning of the synsets in different languages (similar to the synonyms in a given synset, which in most cases are not interchangeable in all contexts), the Inter-Lingual Index is used for linking the Bulgarian WordNet with some of the other freely available wordnets.

The Bulgarian Culinary Lexicon has been correlated through the Inter-Lingual Index with the (lexico)semantic networks of 23 languages (and English): Albanian, Basque, Catalan, Croatian, Danish, Dutch, French, Finnish, Galician, Greek, Hebrew, Italian, Icelandic, Lithuanian, Polish, Portuguese, Romanian, Russian, Serbian, Slovak, Slovene, Spanish, and Swedish. Not all concepts are present for all languages; for example {бешамел} (white sauce, béchamel) is present for: Italian – {bechamel; besciamella}, Dutch – {sous bechamel); Lithuanian – {baltasis padažas} Spanish – {bechamel; besamel; besamela; salsa bechamel; salsa besamel; salsa blanca) and the definition 'leche espesada con un roux de mantequilla y harina';[11] Basque – {bexamel}; Finish – {valkokastike; bechamel-kastike; bechamel}; Catalan – {beixamel; salsa beixamel; salsa blanca}; Slovak –{biela omáčka; bešamelová omáčka; bešamel}; Romanian – {sos alb} and the definition 'Sos preparat cu lapte şi unt'; Polish – {beszamel; sos beszamelowy}. The following information is accessible online: synonyms, definition, immediate hypernym, category, ingredients for Bulgarian and English; synonyms for the other 23 languages and where available for these languages – the definitions. The Culinary WordNet can be used as a dictionary of synonyms or as a bilingual or multilingual dictionary.

3. Structure in the Bulgarian Culinary Lexicon

The **Bulgarian Culinary Lexicon** includes synonyms, grammatical and stylistic labels, explanatory definitions, and examples. Furthermore, the synsets for Bulgarian are linked with their translation equivalents in English (through the Inter-Lingual Index) and through English, with 23 more languages. Synsets are related to each other by means of conceptual and extralinguistic relations; however, in the online version of CulNet only **hypernymy** – **hyponymy** relations are presented.

A typical entry consists of a synset, and in some cases only one word represents the entry. The synonyms (called literals) can be simple words, compound words, or multiword expressions. Each entry is related to exactly one sense and also includes information about the word class, any other relevant grammatical information, and information for the register, style, and origin of the synonyms. Entries also give one or more examples of how the word is used in context.

The Bulgarian Culinary Lexicon combines the best features of WordNet and traditional dictionaries: the hierarchical organisation of synsets and the lexicographic definitions and examples, respectively.

The descriptions of selected words related to food, cooking, and culinary matters that already existed or are newly included in the Bulgarian WordNet were enhanced in numerous ways: the synonyms were evaluated and, if necessary, edited; the definitions were elaborated or re-constructed to become explanatory; where appropriate, a grammar or a stylistic label was assigned to synonyms; examples illustrating the sense and the usage of synonyms were added.

3.1. Synonyms

A scale of synonymity can be established, reflecting the degree of synonymy: absolute synonymy, cognitive synonymy, and near-synonymy (Cruse 2000: 157).

The absolute synonyms are considered to be interchangeable in any context, but in practice, they are limited to a timespan or a group of people, and in some cases the absolute synonyms are just spelling alternations. For example, in Bulgarian such synonyms are the words {кервиз, керевиз, целина} (celery), where different names of the plant are used in different parts of the country and the words {кервиз, керевиз} are spelling alternations.

Cognitive synonyms are words (varying in their evaluative or connotative meaning) that can be interchangeable in grammatical declarative sentences (Cruse 1986: 86). Cognitive synonymy is also described as incomplete synonymy (Lyons, 1981: 148), or non-absolute or partial synonymy (Lyons, 1995: 60). For example, the pairs {карфиол; конопида} (cauliflower); {наяждам се; нахранвам се} (fill up); {питателен; хранителен} (alimentary; nutritious) are cognitive synonyms. The words {провизии; продоволствия; хранителни запаси} (provisions; commissariat; food stocks) are not interchangeable in some contexts: in the collocation with the word *военни* (military), the word most preferred in Bulgarian is *провизии* (provisions); and in the expression *изграждане на хранителни запаси* (building food stocks), the expression *хранителни запаси* (food stocks) could not be replaced with either of the other two words.

Near-synonyms can be specified as words whose meaning is relatively close or similar; their senses overlap to a great degree, but not completely (Murphy 2003: 155). Near-synonyms are presented separately (in different synsets) and they are linked with one of the similarity relations: **similar to, also see**, and **verb group**. For example, the adjective synset {вкусен} (tasty) 'pleasing to the sense of taste' is linked with the relation **similar to** the synset {отбран} (choice) 'appealing to refined taste' and with the relation **also see** to the synset {апетитен} (appetizing; appetising) 'appealing to or stimulating the appetite especially in appearance or aro-

ma'.

The names of foods and beverages largely share the characteristics of terms (although they are used in everyday life), and for this reason the registered synonymy is usually absolute. Cognitive synonymy can be observed more often in nouns, adjectives, and verbs related to nutrition, food preparation, and characteristics of food, and near synonymy is encountered only with some verbs and adjectives from the selected domain.

3.2. Multiword Expressions

Multiword expressions (words denoting a single concept, but constituted by two or more orthographic words) can be identified on the basis of a number of tests: whether they have as a synonym a simple word {пека на грил; **гриловам**} (grill); {сладък лимон; **цитрон**} 'citron'; whether they have as a translation equivalent a simple word in another language {пазя диета} (**diet**) 'follow a regimen or a diet, as for health reasons'; whether they have as a synonym a term {маслен боб, **Phaseolus lunatus**} (butter bean; Phaseolus lunatus).

There are numerous classifications of multiword expressions (Sag et al. 2002; Baldwin and Kim 2010; etc.). We present here a short classification with the view of illustrating the types of multiword expressions that are present in the Bulgarian Culinary Lexicon: i) expressions that are semantically and syntactically compositional but are used to denote a single concept, and ii) expressions that are fully lexicalised with different degrees of semantic and syntactic compositionality (non-transparency). The first type of multiword expressions is difficult to determine: their constituents express the meaning with which they participate both in the particular multiword expression and separately, but as a multiword expression they denote compositionally the meaning of a single concept and have been conventionalised for its naming: {зелен пипер}(green pepper), {консервна кутия} (can). The lexicalised expressions on their part are: fixed (fully lexicalised expressions which do not change either in terms of word order or grammar (благодарение на 'thanks to'); semi-fixed (the order and constituency of the parts of the compositional word are fixed but the constituents can undergo certain paradigmatic changes within certain grammatical categories: {гола праскова} (naked peach) 'nectarine' and in plural *голи праскови* 'nectarines'; non-fixed (they change morphologically, can undergo word order changes, as well as tolerate mutable elements in their composition: {споделям хляба си} (break bread) and with an insertion *споделям с радост хляба си* 'gladly share my bread'.

The base form of the multiword expressions is recorded, and the head word and dependent parts are additionally marked. The multiword expressions are marked with two types of labels: **phraseological expressions** for cases like {формичка за сладки} {cookie cutter) and **sentence** if the multiword expression formally represents a sentence for cases like {ям и пия} (wine and dine) 'eat sumptuously'.

3.3. Definitions

Unlike the definitions in (Princeton) WordNet and in some other wordnets, where the convention has been adopted to present a short definition which indicates but does not describe the meaning, the construction of the definitions for Bulgarian synonyms is subject to the following principles: i) representation of the meaning in the manner adopted for the definitions in monolingual general dictionaries; ii) originality of the definition, i.e. lack of literal repetition of the definition provided in existing dictionaries (this is achieved mainly through the expansion of the definitions by the addition of further information); iii) use of a specific structure not only for different parts of speech, but also for the specific lexical sets (Atkins and Rundell 2008: 124) within a single part of speech class.

The usual way to construct definitions is to use the traditional genus and differentia definition, which is recognised as one of the most useful methods (Atkins and Rundell 2008: 415). Typically this kind of definition consists of a superordinate term (genus) (the hypernym) and a specifying part (differentia), which renders the meaning unique and shows how it differs from other members of the same category (co-hyponyms); for example, the definition of {макарони} (macaroni) is '**dry pasta** in the form of thing long sticks of wheat dough, which are boiled in water until soft and seasoned with various sauces'. In taxonomic classifications, it is possible for a holonym to be used in the generalising part – for example, {мечо грозде} (bearberry) and the definition 'representative of evergreen shrubs of the **genus Arctostaphylos**, which are found in the northern regions; has small red fruits'. The advantage of combining the WordNet structure with lexicographic information is that the meanings of the superordinates (hypernyms) are always explicitly defined and can be traced within the WordNet hypernymy tree.

Atkins and Rundell (2008: 124) define a lexical set as any "group of words that share a common element of meaning". For the specific lexical sets such as Plant Food (Fruits, Vegetables, Nuts, Pulses and Grains), Meat, Fish and Seafood, Drinks, and Dishes, the procedure was to compile a corpus-based definition template following Atkins and Rundell (2008: 127–128). Furthermore, only the template elements for which reliable information is collected are filled in. For example, the definition template for **Plant Food** includes the following

elements: (a) hypernym, (b) holonym from an existing taxonomy, (c) physical characteristics, (d) geographic distribution, (e) economic significance; the definition of the word {райска ябълка} (oriental persimmon) contains three elements of the template: 'a representative of the **tree plants** of the **family Ebenaceae** with **edible, sweet, yellow to orange fruits**'.

The thematic domain of the terms is indicated within the definition only if all members of a synset are terms; for example, the word {кайсия (за дърво)} (apricot, tree) is described with the definition 'in **botanics** – a kind of low-growing tropical woody plant of the genus Chrysobalanus, found in America; have edible fruits'. If there are other restrictions in use, these are explicated before the definition; for example, {алкохолен} (alcoholic) is defined as '**for beverages, certain groceries**, etc. – which contains a certain amount of ethyl alcohol'.

The definition of verbs is subject to the same structure and, when possible, contains information on the arguments. The definitions of verbs that represent one and the same **semantic frame** – encyclopaedic knowledge structures or conceptualisations underlying the meaning of sets of lexical items (Fillmore 2003: 288) – are related by explicating one and the same (mental) model (Atkins 1996: 540–541). Although the number of verbs in CulNet is relatively small, such an approach provides semantic coherence in the description of verbs evoking a given semantic frame (which is not the case with traditional dictionaries). For example, the verb [закусвам] (have breakfast) has the definition 'take the first meal of the day (usually early in the morning)' and the verb {вечерям} (have dinner, dine) – the definitions 'take the last meal of the day, usually in the evening or early night'.

The definition with adjectives starts with "which is; which is characterised by"; for example, {апетитен} (appetising) has the definition 'which stimulates the appetite through appearance, aroma, colour, etc.'. The definition of relational adjectives contains the noun from which the adjective has been derived and the standard definition includes "which refers to; which is related to; which is characteristic of" and other similar definitional descriptions. For example, {маслен} (butter) has the definition "which has the characteristics of an oil, is obtained from or contains a substance of the oil type".

3.4. Grammatical and Stylistic Labels

When there is a specific use – a substandard word, figurative meaning, phraseological meaning, an obsolete word, informal use, and so on – that indicates whether the synonyms are cognitive or absolute, a label is attached to the respective synonym. The system of labels developed for Bulgarian reflects the following differences in the usage of synonyms (Koeva 2021: 55 – 56): **belonging to non-standard lexis** – a **dialectal word**: {конопида} (cauliflower) instead of *карфиол* (cauliflower); **folk word*:*** {бабина душица} (grandma's soul) 'thyme' instead of *мащерка* 'thyme'; a **word with an undesired use*:*** {фиш енд чипс} (fish and chips) instead of *риба с пържени картофи* (fish and chips); **the historical period of use** – an **obsolete word**: {дървено масло} (olive oil) instead of *зехтин* (olive oil); a **new word**: {брънч} (brunch) instead of *късна закуска и ранен обяд* (brunch); **the expressive properties of the literals** – a **collective word**: {кухня} (gastronomy) 'a particular style of cookery (as of a region)'; a **diminutive word**:[12] {курабийка} (cookie); **the frequency of use of the literals** – a **rare word**: {всеядец} (omnivore) 'a person who eats all kinds of foods'.

The **grammatical labels** indicate a constant grammatical feature of a particular literal. Perfective and imperfective verbs in Bulgarian express different meanings, although the verb aspect pairs are closely related. In the Bulgarian WordNet verb aspect pairs are included in one and the same synonymous set, although the perfective and imperfective members of a pair are not cognitive synonymous, and as a consequence only some of the literals are translation equivalents to the respective synonyms in English. Such a representation is fully synchronous with the tradition in dictionary compilation of imperfective and perfective verbs being included in a single dictionary entry. However, the main motivation for this representation is preserving the structure of the Princeton WordNet. For the differentiation of verbs of different aspect, a label is attached to each verb indicating its aspect: **perfective verb**: {препържа (frizzle), **imperfective verb**: {препържвам} (frizzle), a verb that can function as **perfective or imperfective** depending on the context: {филетирам} (fillet), an **exclusively imperfective verb with no perfective equivalent**: {ям} (eat), an **exclusively perfective verb with no imperfective equivalent**: no examples within the culinary domain.

The grammatical labels also describe some paradigmatic restrictions: a **word used exclusively in the singular**: {целина} (celery); a **word typically used in the singular**: {захар} (sugar); a **word exclusively used in the plural**: {броколи} (broccoli); a **word typically used in the plural**: {зелева сърма} (cabbage roll).

3.5. Examples

At least one usage example is added, which showcases the meaning of the synonymous set. The examples are in the form (in the usual case) of full sentences and created in a few specific ways: via the search and choice

within corpora or through dedicated construction. The multitude of examples in the different languages (examples are not provided for all languages) can be thought of as a multilingual corpus with corresponding sentences with semantically annotated meanings of the respective words.

4. Lexical and Conceptual Relations

Two types of semantic relations are included (following WordNet): among literals (called lexical) and among synsets (called conceptual). The conceptual relations apply to all synonyms in both connected synonymous sets, while the lexical relations link words within one or two synonymous sets. The conceptual relations among synonymous sets are: **hypernymy:** {зеле} 'cabbage' **is a hypernym of** {китайско зеле} (Chinese cabbage); {нарязвам} (carve, cut up) **is a hypernym of** {филетирам} (filet); **hyponymy:** {филетирам} (filet) **is a hyponym / troponym of**[13] {нарязвам}(carve, cut up); **holonymy** (holo part, holo portion): {джинджифил, the plant} (common ginger; Zingiber officinale) 'tropical Asian plant widely cultivated for its pungent root; source of ginger root and powdered ginger' **is a holonym (holo part) of** {джинджифил} (ginger) 'pungent rhizome of the common ginger plant; used fresh as a seasoning especially in Asian cookery'; {хляб} (bread) **is a holonym (holo portion) of** {брашно} (flour); **meronymy** (mero part, mero portion): {джинджифил} (ginger) **is a meronym (mero part)** of {джинджифил, the plant} (common ginger; Zingiber officinale); {брашно}(flour) **is a meronym (mero portion) of** {хляб} (bread); **subevent:** {ям} (eat) 'eat a meal; take a meal' **has a subevent** {нагъвам} (tuck in; put away) 'eat up; usually refers to a considerable quantity of food;' {нагъвам} (tuck in; put away) **is a subvert of** {ям} (eat); **cause:** {храня; нахранвам} (feed; give) 'give food to' **causes** {ям; изяждам} (eat) 'take in solid food', {ям; изяждам}(eat) **is caused by** {храня; нахранвам} (feed; give); **verb group** {ям} (eat) 'eat a meal; take a meal' is **in a verb group with** {изяждам} (eat) 'take in food'), **similar to** {ароматен; пикантен}(flavourful; flavorous) **is similar to** {вкусен} (tasty), **also see** {храня се; ям} (eat) 'eat a meal; take a meal' is an **also see relation with** {ям навън, храня се навън} (eat out; dine out) 'eat at a restaurant or at somebody else's home').

The **lexical relations** are semantic: **synonymy** and **antonymy** {храня се вкъщи} (eat in; dine in) and {храня се навън} (eat out; dine out) and **derivational**. The derivational relations that have been transferred from the Princeton WordNet are at the level of the syn sets, and the most frequent general relation is labeled as **derivative in English**. Part of the derivational relations have been checked, and if the relation has proven valid for Bulgarian, the relation **derivative in Bulgarian** has been introduced. For example, between the noun {тостер} (toaster) with the meaning 'a kitchen appliance (usually electric) for toasting bread' and the verbs {препичам; препека} (toast) with the meaning 'for bread – make brown and crisp by heating', there is no derivational relation, unlike in English, but between the noun {препичане} (toasting) and the verb {препичам} (toast) from the synset {препичам; препека} (toast) in Bulgarian, there is a derivational relation.

Since derivational relations are important for determining morphosemantic relations, the derivational relations between nouns and verbs are introduced. These derivational relations do not reflect the semantics of the derivation; rather, they reflect the word-formation mechanism: **suffixation**: {ям} (eat) – {яде**не}** (eating); **conversion**: {**вечеря**м} (dine) – {**вечеря**} (dinner); **prefixation**: {вкус} (taste) – {**о**вкуся} (flavour); **substitution** verbal suffix – nominal suffix: {ферме**нтирам}** (ferment) – {ферме**нтация}** (fermentation); **derivation** (a non-specified derivational relation, which is not immediately obvious from a contemporary point of view or is not part of the Bulgarian system of word formation): {мириша} (smell) – {миризма} (smell).

If there exists a derivational relation between two synsets connecting some of the synonyms that comprise them, then there exists a semantic relation of a given type between these two synsets (called a morphosemantic relation); for example: **agent** (*a cook cooks*), **instrument** (*a chisel chisels*) and so on. These relations are grounded in the idea that meanings of word-formation affixes can be exhaustively classified into a small number of semantic categories of the type 'semantic role' (Fellbaum et al. 2007). For example, morphosematic relations are: **has agent** / **is agent of** between {ловя риба} (fish) and рибар (fisherman); **has event / is event of** between {ловя риба} (fish) and {риболов} (fishing); **has undergoer / is undergoer of** between {ловя риба} (fish) and {риба} (fish). Other morphosemantic relations are: **is means to / by means of** between {храна} (nutriment; nutrition) and {храня} (nutrify; nourish); **uses / is used for** between {прехранвам се} (nourish) and between храна (nutriment; nutrition); **has result / is result of** between {недохранвам} (undernourish) and недохранване (undernourishment); **has location / is location of** between {вечерям} (dine) and {вагон-ресторант} (dining car); **has instrument / is instrument for** between {варя; сварявам} (boil) {чайник} (kettle; boiler).

Extralinguistic relations have also been introduced between synsets: **domain category** {филетирам} (fillet)

and {готвене; готварство} (cooking), **domain region** {плодова пита} (tart) and {Съединени американски щати} (USA)), **usage domain** {nosh-up} (without lexicalisation in Bulgarian) 'a large satisfying meal' and {slang}.

The advantage of the proposed resource is that links that usually are omitted in traditional printed dictionaries because of space issues are provided, and each sense of the polysemous word is related to its correct hypernym. As has already been proved, the WordNet can be used as an ontology by treating the hypernymy relation between synsets as subsumption between concepts (Gangemi et al. 2003).

5. Conclusion

There are not many bilingual and multilingual dictionaries (Castro 2009; Marando 2020; Sinclair 2009; among others) or encyclopaedias (for example, Katz 2004) devoted to culinary matters, and there are only a few attempts to link lexicographic sense descriptions in and across languages via existing ontologies like WordNet (for example, BabelNet[14], Cornetto[15]).

Most of the information for food, cooking, and culinary matters is available on specialised websites and portals. Usually they provide recipes, culinary articles, or information about culinary history and cooking techniques. Along with these, some of the websites contain monolingual or bilingual culinary dictionaries offering culinary definitions and information about ingredients and cooking techniques. Most of them are for English (French, Spanish) and are written by journalists or professional chefs.

Recently, an ontology that models the domain of dishes as they are presented in 112 menus, ΑΜΑΛΘΕΙΑ, was released. It provides trilingual (Greek, English and Russian) translations of menus, dietary and cultural information about the dishes and their ingredients, as well as information about the geographical dispersion of the dishes (Markantonatou et al. 2021). The ontology is based on Langual – a reference thesaurus for food, published in basic English and in English, Czech, Danish, French, German, Italian, Portuguese, and Spanish (Møller and Ireland 2017). Another ontology that is based on Langual is FoodOn – a farm and nutrition ontology in which food products are categorised by cultural origins, food transformation process, food container and wrapping, part of plant or animal, food product type, etc. (Dooley et al. 2018).

The **Multilingual Culinary Lexicon** which is presented in this paper is oriented towards the Bulgarian language as its core. It is very difficult to describe the different foods, methods of food preparation, eating habits, etc. that are typical for a given group of people, and at the same time to find the correct translation equivalents (or their absence) in other languages. For this reason, we approached this problem by providing a reliable lexicographic description for Bulgarian for over 2,200 lexical units related to food, cooking, and culinary matters and linked the data for Bulgarian to the lexical-semantic networks of 23 more languages (through English).

The linked data combining synonyms, hierarchical information (categorisation in domains related to food, cooking, and culinary matters), lexical and semantic relations from WordNet, elaborated dictionary definitions and examples for Bulgarian, and translation equivalents provide options for cross-lingual checking and comparing the information for language pairs for which bilingual dictionaries do not exist. Information for the ingredients that is not explicitly presented in WordNet is extracted from (Bulgarian and English) definitions and examples and attached to the synsets.

An attempt at proposing food ontology is made in the form of a shallow hierarchy with 48 subsuming categories and subcategories. The additional ontological information and the ontological structure of WordNet may provide further options for the generation of cross-lingual sense clusters organised by different criteria. To maximise the lexicalised intersection, the core lexicon of the Multilingual Culinary Lexicon could be further elaborated on the basis of more languages.

References

Atkins 1996: Atkins, S. B. T. Bilingual Dictionaries: Past, Present and Future. – Gellerstam, M., J. Järborg, S.-G. Malmgren, K. Norén, L. Rogström and C. R. Papmehl. (eds.) *Euralex'96 Proceedings,* (1996), Gothenburg: Gothenburg University, Department of Swedish. 515–590.

Atkins and Rundell 2008: Atkins, S. B. T., Rundell, M. *The Oxford Guide to Practical Lexicography,* (2008). Oxford: Oxford University Press.

Baldwin and Kim 2010: Baldwin, T., S. N. Kim. Multiword expressions. – *Handbook of Natural Language Processing*, Second Edition, (2010), Chapman and Hall/CRC press, 267–292.

Bond et al. 2012: Bond, F. and Paik, K. (2012). A survey of wordnets and their licenses. – *Proceedings of the*

6th Global WordNet Conference (GWC 2012), Matsue, (2012), 64–71.

Bond et al. 2013: Bond, F. and R. Foster. Linking and extending an open multilingual wordnet. – *51st Annual Meeting of the Association for Computational Linguistics: ACL-2013*. Sofia, (2013), 1352–1362 <https://aclanthology.org/P13-1133/> retrieved on 30.04.2022.

Bond et al. 2016: Bond, F., P. Vossen, J. McCrae, C. Fellbaum. CILI: The collaborative interlingual index. – *Proceedings of the 8th Global Wordnet Conference* (*GWC 2016*), (2016), 50–57. <https://aclanthology.org/2016.gwc-1.9.pdf> retrieved on 30.04.2022.

Castro 2009: Castro, L. *Eat, Drink, Think in Spanish: A Food Lover's English-Spanish/Spanish-English Dictionary,* (2009), Ten Speed Press.

Cruse 1986: Cruse, D. A. *Lexical Semantics* (1986). Cambridge: Cambridge University Press.

Dooley, D. M.; Griffiths, E. J.; Gosal, G. S.; Buttigieg, P. L.; Hoehndorf, R.; Lange, M.C.; Schriml, L. M.; Brinkman, F. S. L.; Hsiao, W. W. L. FoodOn: A harmonized food ontology to increase global food traceability, quality control and data integration. – *NPJ Science of Food (*2018), *2*, 23. <https://www.nature.com/articles/s41538-018-0032-6> retrieved on 30.04.2022.

Gangemi et al. 2003: Gangemi, A., Navigli, R., Velardi, P. The OntoWordNet Project: Extension and Axiomatization of Conceptual Relations in WordNet. – Meersman, R., Tari, Z., Schmidt, D. C. (eds) *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE*. OTM 2003. *Lecture Notes in Computer Science*, (2003), vol 2888. Springer, Berlin, Heidelberg, 820–838.

Fellbaum 1990/1993: Fellbaum, C. English Verbs as a Semantic Net. – In *International Journal of Lexicography (*1990), Volume 3, Issue 4, 278–301; reprinted in 1993, 40–51.

Fellbaum 1998: Fellbaum, C. (Ed.). WordNet: An Electronic Lexical Database, (1998), Cambridge, MA: MIT Press.

Fellbaum et al. 1990/1993: Fellbaum, C., D. Gross, and K. Miller. Adjectives in WordNet. – *International Journal of Lexicography,* (1990) , Volume 3, Issue 4, 268–277; reprinted in 1993, 26–39.

Fellbaum et al. 2007: Fellbaum, C., A. Osherson, P. E. Clark. Putting Semantics into WordNet's "Morphosemantic" Links. – *Responding to Information Society Challenges: New Advances in Human Language Technologies. Springer Lecture Notes in Informatics,* (2007), vol. 5603, 350–358.

Fillmore 2003: Fillmore, C. Double-decker Definitions: The Role of Frames in Meaning Explanations. – *Sign Language Studies,* (2003), 3(3): 263–295.

Katz 2004: Katz, S. (Ed.). *Encyclopedia of Food and Culture*. Volume 1: *Acceptance to Food Politics*, (2004), New York: Charles Scribner's Sons.

Koeva 2021: Koeva, S. The Bulgarian WordNet: Structure and specific features. – *Papers of Bulgarian Academy of Sciences* (2021), 8, 1, 47–70. <https://www.papersofbas.eu/images/Papers_2021-1/Koeva_optimized.pdf> retrieved on 30.04.2022.

Koeva et al. 2012: Koeva, S., I. Stoyanova, S. Leseva, T. Dimitrova, R. Dekova, and E. Tarpomanova. The Bulgarian National Corpus: Theory and Practice in Corpus Design. – *Journal of Language Modelling* (2012), Vol. 0, No. 1, pp. 65–110. retrieved on 30.04.2022.

Lyons 1981: Lyons, J. *Language and Linguistics*, (1981). Cambridge: Cambridge University Press.

Lyons 1995: Lyons, J. *Linguistic Semantics*, (1995), Cambridge: Cambridge University Press.

Marando 2020: Marando, C. S. *The Indispensable Multilingual Dictionary of Food: 10 ways translation tool for Kindle*: *with over 26.000 Culinary Terms in French, English, German, Spanish, Italian, Portuguese, Russian, Chinese*, (2020), Kindle.

Markantonatou et al. 2021: Markantonatou, S.; Toraki, K.; Minos, P. Vacalopoulou, A.; Stamou, V.; Pavlidis, G. ΑΜΑΛΘΕΙΑ: A dish-driven ontology in the food domain. – *Data*, (2021), vol. 6, no. 41. <https://www.mdpi.com/2306-5729/6/4/41> retrieved on 30.04.2022.

Miller 1986: Miller. G. Dictionaries in the Mind. – *Language and Cognitive Processes* (1986), 1, 171–185.

Miller 1990/1993: Miller, G., R. Beckwith, C. Fellbaum, D. Gross, and K. J. Miller. Introduction to WordNet: An On-Line Lexical Database. – *International Journal of Lexicography* (1990), 3(4), 235–244; reprinted in 1993, 1–9.

Møller and Ireland 2017: Møller, A.; Ireland, J. LanguaLTM 2017 – *Multilingual Thesaurus* (*English–Czech–Danish–French–German–Italian–Portuguese–Spanish*). Technical Report. Danish Food Informatics, (2018).

Murphy 2003: Murphy, M. L. *Semantic Relations and the Lexicon*, (2003), Cambridge: Cambridge University Press.

Piasecki and Koeva 2017: Piasecki, M., S. Koeva. WordNet Relations in the Bulgarian-Polish Bilingual Perspective. – Blagoeva, D., T. Aleksandrova (red.). *Dokladi ot Mezhdunarodnata yubileyna konferenciya na Instituta za balgarski eizk "Prof. Lyubomir Andreychin" (Sofia, 2017)*, chast I, Izdatelstvo na BAN "Prof. Marin Drinov", 285–297.

Sag et al. 2002: Sag, I., T. Baldwin, F. Bond, A. Copestake, D. Flickinger. Multiword expressions: A pain in the neck for NLP. – *Proceedings of the 3rd International Conference on Intelligent Text Processing and Computational Linguistics* (CICLing-2002), (2002), 1–15.

Sinclair 2005: Sinclair, C. G. *Dictionary of Food: International Food and Cooking Terms,* (2005), Bloomsbury Publishing.

Vossen et al. 1998: Vossen, P., L. Bloksma, H. Rodriguez, S. Climent, N. Calzolari, A. Roventini, F. Bertagna, A. Alonge, W. Peters. *The EuroWordNet Base Concepts and Top-Ontology*. Deliverable D017D034D036 EuroWordNet LE2-4003, (1998).

PhD Svetla Koeva
Institute for Bulgarian Language
Bulgarian Academy of Sciences
Sofia, Bulgaria
e-mail: svetla@dcl.bas.bg

[1] http://dcl.bas.bg/culnet-2020/

[2] The synonyms, the definitions, and the remaining related information come from either the Bulgarian Culinary Lexicon based on the Bulgarian WordNet or the Princeton WordNet and where appropriate these have been adapted by the author.

[3] https://dcl.bas.bg/bulnet/

[4] The main lexicographic resource was the Dictionary of the Bulgarian Language: http://ibl.bas.bg/rbe/ .

[5] http://search.dcl.bas.bg

[6] https://wordnet.princeton.edu

[7] The presented resource is bilingual through the links to Princeton WordNet, and multilingual through the subsequent links to 23 other languages wherever the data was available.

[8] The Bulgarian data is released under the Creative Common Attribution ShareAlike 4.0 International license (https://creativecommons.org/licenses/by-sa/4.0/) and the data from the other wordnets is subjects to separate licenses listed at: http://compling.hss.ntu.edu.sg/omw/ or are used as the result of signed agreements with the copyright holders (for Romanian and Serbian).

[9] The Multilingual Culinary Lexicon was developed at the Department of Computational Linguistics (Institute for Bulgarian Language) by Svetla Koeva (project coordinator), Tsvetana Dimitrova, Hristina Kukova, Svetlozara Leseva, Valentina Stefanova, Ivelina Stoyanova, and Maria Todorova.

[10] http://compling.hss.ntu.edu.sg/omw/

[11] The definitions and examples are linked if they are presented in the wordnets. The definitions should describe the equivalent concept in a similar way to how it is described by the definitions in Bulgarian and English.

[12] The formation of diminutive words in Bulgarian is regular. In the Bulgarian Culinary Lexicon diminutive words are included if they have a new meaning built on the diminutive one.

[13] Some Princeton WordNet relations have been renamed following the conventions of the BalkanNet project, for example **substance homonymy** / **meronymy** is labeled **portion homonymy** / **meronymy**, **troponymy** in verbs is labeled **hyponymy**, while **entailment** (proper inclusion) is defined as **subevent,** and **entailment** (backward presupposition) has not been transferred (Piasecki and Koeva 2017); **antonymy** and derivational relations in English are included in the Bulgarian WordNet at the level of the synonymous set and **antonymy** can be characterised as **near antonymy (antonymy** is a literal relation and when defined at the level of synsets, it also relates synonyms of the antonymy pair)**.**

[14] https://babelnet.org

[15] http://cornetto.clarin.inl.nl/index.html