

A Bilingual Lexicosemantic Network of Bread Based on a Parallel Corpus

Ivan Derzhanski

Olena Siruk

Institute of Mathematics and Informatics

Bulgarian Academy of Sciences

iad58g@gmail.com

olebosi@gmail.com

Abstract

We present an experiment in using a corpus of Bulgarian and Ukrainian parallel texts for the automatised construction of a bilingual lexicosemantic network representing the semantic field of BREAD. We discuss the extraction of the relevant material from the corpus, the production of networks with varying parameters, some issues of the interpretation of these networks, and possible ways of making them more accurate and informative.

Keywords: bread, lexical semantics, semantic net, wordnet, parallel corpus, Bulgarian language, Ukrainian language

1. Introduction

Systems for thesaurus representation of vocabulary such as WordNet (Fellbaum, 1998) are widely used for storing lexical and semantic (including ontological) data. They typically are hierarchical networks which reflect synonymy by grouping lexemes into synsets, and other lexical and semantic relations by labelled directed edges. The knowledge encoded in them can come from various sources. Resources of this type have been developed for many languages.

In this study we present an experiment in the automatised construction of a bilingual lexicosemantic network on the basis of a corpus of parallel texts. Our working languages are Bulgarian and Ukrainian. They are related, though not nearest of kin, spoken in regions neither adjacent nor really distant, moderately close typologically, and with similar history of substantial lexical borrowing from Western European and other languages.

We chose to focus our attention on the field of BREAD. The concept of bread is extremely important in Western civilisation: as bread and other bread-like products have been baked and eaten for millennia, bread vocabulary is highly developed everywhere; on the other hand, for many centuries it has been developing separately, making for very complex and interesting relationships between words of different languages. Besides, bread and bread-like goods are produced by people and (mostly) for people; as such this semantic and lexical field is part of the anthropocentric image of the world and the anthropocentric vocabulary, which forms an ideal basis for setting and solving any general linguistic problems. The apparatus of theoretical notions and technical approaches developed on such material has the best chances of being extrapolated to other nominal lexicosemantic fields.

2. The Corpus

The bilingual Bulgarian–Ukrainian corpus (CUB) (Derzhanski and Siruk, 2019) consists of parallel texts available in electronic libraries or obtained by us from paper editions through scanning, optical character recognition and error correction by *ad hoc* software tools and by hand. For this reason the corpus is composed of fictional works, mostly of novels, which dominate in such sources.

Because original and translated parallel texts for Ukrainian and Bulgarian are hard to come by, especially in online-accessible computer-readable form, we also use Bulgarian and Ukrainian literary translations from other languages as corpus material. The current version of CUB includes eleven sectors, each of which covers parallel Bulgarian and Ukrainian texts with the same original language:

- original Bulgarian and Ukrainian texts, as well as translations from English-1 (by authors from the British Isles), English-2 (by authors from the United States), French, German, Italian, Russian-1 (stories about the past and present), Russian-2 (stories about the future), and French—approx. 2 million words in each of the ten sectors (in the two corpus languages counted together; for various reasons the ratio tends to be about 53:47);
- the Bible, in canonical translations from Hebrew, Aramaic and Greek into Bulgarian and Ukrainian —1⅓ million words.

The total size of CUB is 10 million words in Ukrainian (and 11½ million in Bulgarian). The Bible is aligned by verse, and the other texts (mostly) by sentence.

3. Data Collection and Preliminary Analysis

The collection of the material took place in the following way. We started with the two languages' principal 'bread' words, Bg *хляб* and Uk *хліб*, and their near-synonym derivatives Bg *хлебец* '(a little) bread (hypocoristic)', *хлебче* 'small bread, roll', Uk *хлібець* 'little bread (dimin. or hypocor.)', *хлібина*¹ 'loaf of bread'. All occurrences of these words in CUB were located. The numbers are summarised in Table 1, separating the singular and the plural (and, in Bulgarian, count) forms of the words *хляб/хліб*, as well as the cases when they cooccur with a specification of quantity such as 'a loaf', 'a piece' or '200g'. The label 'others' means other words for bread-like substances, on which more anon:

Bg \ Uk	<i>хліб</i> (sg.)	(Q) <i>хліба</i>	<i>хліб</i> (pl.)	<i>хлібина</i>	<i>хлібець</i>	others	Σ
<i>хляб</i> (sg.)	1111	53	6	13	5	61	1249
(Q) <i>хляб</i>	22	153	0	0	0	5	180
<i>хляб</i> (count)	8	3	18	5	0	8	42
<i>хляб</i> (pl.)	22	1	18	3	0	18	62
<i>хлебец</i>	2	1	0	1	4	0	8
<i>хлебче</i>	0	0	0	0	8	41	49
others	22	3	0	16	3		44
Σ	1187	214	42	38	20	133	1634

Table 1: Occurrences of correspondences of *хляб/хліб* and cognate words

We see that Bg *хляб* is somewhat more readily used as a count noun and Uk *хліб* as a mass one, that it is more common for Bg *хляб* than for Uk *хліб* to correspond to a word with a different root, and that for Bg *хлебче* and Uk *хлібина* this is even more common.

At the next stage of the research the 'other' corresponding words were similarly sought out, then the correspondences of their correspondences, and so on until no new words were found. Only words denoting kinds of cooked dough (baked, boiled or fried) or their parts, products or subproducts – but not just dough, flour, grain or gruel – were considered. Words meaning 'piece (of anything)' were included only when it was clear that bread or another relevant substance was meant (and usually mentioned in the same sentence). Where in serious doubt, we preferred to err on the side of inclusion. For example, while both Bg *настет* and Uk *наштет* normally mean 'meat or fish paste', they do occur in the meaning of 'pie, pâté (with a crust)', especially in translations from languages where the related word (De *Pastete*, Fr *pâté*, Ru *наштет*) has this as one of its regular meanings; accordingly we took all cases where Bg *настет* or Uk *наштет* corresponds to a word for a bakery product in the other language, as well as the occurrences of the pair *настет* : *наштет* from *The Black Obelisk* by E. M. Remarque, where the *Leberpastete* 'liver pâté' is cut in slices and eaten as a dish in its own right, which suggests a crust. Uk *товченик* 'fish or meat dumpling' (Hrinchenko, 1958: v. 4, 270), where flour is an ingredient but not a major one, is on the fringe; its entry in *The Geese and Swans Are Flying* by M. Stelmakh was taken because the Bulgarian translation is *тузаница* 'pancake'. But we stopped short of including Uk *зедфип*

¹ A Bulgarian counterpart of this word (*хльбина, хльбинка*) was in use as late as the 19th c. and is registered in N. Gerov's dictionary (Panchev, 1904: 501; Panchev, 1908: 320), but has been lost in the contemporary language.

'marshmallow', which is what is left of *marshmallow cookies* in I. Shaw's *Bread Upon the Waters* (contrariwise, in the Bulgarian translation of the book only *бисквити* 'cookies' is preserved).

Several cases of idiomatic use of bread words whose literal meaning was not very close were discarded: the sentence pair *Със сух комат да се задавиш дано* || *Бодай ти немащеним млинцем удавися* 'May you choke on a dry chunk of bread || an unbuttered pancake' (V. Shishkov, *Gloomy River*) was not considered an occasion to postulate a word pair *комат* : *млинецъ*. Exceptionally a word was counted twice if it had two counterparts on the other side: the pair of sentences *Има ли — попитал той — такъв човек под слънцето, който не би предпочел пшеничната питка пред ечемичната?...* || *Чи ж е, — спитав він, — така людина під сонцем, яка б, маючи ячний корж, не хотіла б пшеничної паляниці?* 'Is there,' he asked, 'any man under the sun who would not prefer a wheat roll₂ over of a barley cake₁?' (B. Prus, *Pharaoh*; tr. into English by Christopher Kasparek) yielded the two word pairs *питка* : *корж* and *питка* : *паляниця*. Adjectives derived from the words *хляб* and *хліб* were not taken,² but from other words they were: the pair of phrases *тоя козуначен Казанова* || *цей тістечковий женолюб* 'old Pastry-Casanova' (E. M. Remarque, *Three Comrades*; tr. into English by A. W. Wheen) produced the word pair *козунак* : *тістечко*.

The total number of word pairs thus found was 3240.

As is typical in Slavic languages, many Bulgarian and Ukrainian bread words have diminutives (and rediminutives), which sometimes acquire different meanings—an extreme case is Ukrainian *батончик* 'stick of confectionery, candy bar'³ from *батон* 'long bread loaf'. To avoid making a judgement in each case, we considered lexemes that only differ in diminution as separate if each appeared three or more times (e. g., Bg *кора* 2, *корица* 1 and *коричка* 16 'crust' were all counted jointly).

In Bulgarian the words *кора*₁ 'crust (of bread)' and *кора*₂ 'sheet of phyllo pastry', as well as *пита*₁ 'round bread' and *пита*₂ 'fruit pie', which have diverged semantically to a considerable degree, were considered different lexemes. In Ukrainian *сандвич* and *сендвич* 'sandwich' were counted together, as were *баница* and *баниця* < Bg *баница*⁴ '(Bulgarian) layered pastry', being simply different adaptations of the same foreign (and infrequent) words.

In all 91 Bulgarian and 110 Ukrainian lexemes were found.⁵ They are listed in Appendices A and B, along with the number of their occurrences in our data, all in citation form, which is the plural in two cases (Bg *равиоли* 'ravioli' and Uk *потанці* 'dunked bread'). The vast majority denote baker's goods specified for shape (elongated, round, crescent), size, grain (oats, rye), presence of leaven, taste (savoury, sweet), presence of a topping or filling, etc.

A few words are exceedingly rare or appear in rare forms. Bg *книш* can hardly be called a regular word of the language, but appears as a rendering of the identical Ukrainian word in P. Zahrebelnyi's novel *Let's Come to Love*, where it couldn't have been translated because it happens to be a character's surname. Bg *такос* 'taco' appears in R. Bradbury's *Death is a Lonely Business* in the un-Bulgarian plural form *такос*, which indicates that it has not (yet) been adopted. Uk *гостія* 'host, Catholic Eucharist wafer' is not registered by SUM, though found in some dictionaries of loanwords (e. g., Vojkiv et al., 1955: 114). Uk *оцинок* 'flat unleavened bread' (Hrinchenko, 1958: v. 3, 84) is an uncommon variant of *оцинок* (SUM, 1974: v. 5, 840). Uk *пляцинда* 'sweet pie' is found in Hrinchenko (1958: v. 3, 199), but SUM (1975: v. 6, 572) knows only the variant *плагинда*. Uk *попряничок*, variant of *пряник* 'gingerbread', is a *hapax legomenon* in M. Lukash's translation of *The Decameron* by G. Boccaccio, but still registered by SUM-20. Uk *тетерка* appears in our data as a diminutive of *тетеря* 'bread soup, sop', but SUM (1979: v. 10, 102) only registers the homonym *тетерка* 'greyhen'.

Let us formulate several semantic categories with the aid of the two languages' main interpreting dictionaries, RBE and SUM, and illustrations chosen among the words found in CUB:

- I. bread proper: *хляб* and *хліб* themselves, as well as the hypocoristic Bg *хлебци*;
- II. a whole unit (i. e., loaf), able to appear in combination with *хляб/хліб* (Bg *пита* and *самун*; Uk only *буханецъ*);

² On two occasions an exception is made for the Ukrainian adjective, which obviously corresponds to the Bulgarian noun in *коричка хляб* : *хлібна скоринка* 'bread crust' and *трошичка хляб* : *хлібна крихта* 'bread crumb'.

³ Used for 'Nestlé's Crunch bar' in the translation of R. Bradbury's *Death Is a Lonely Business*. We discard this word when its match is Bg *шоколадче* 'chocolate bar', but when it is *вафла* 'wafer', we take the pair for want of certainty that we should not: a Nestlé Crunch (despite containing crisped rice) is not really a bread product, but a *батончик* in general can be.

⁴ This word actually has a Ukrainian version in the person of *баник*, interpreted as *род ватрукики* 'a kind of cheesecake' (Hrinchenko, 1958: v. 1, 26), but absent from SUM.

⁵ The idea that the set of words had to be fully connected had the corollary that some semantically appropriate words were not taken: such were Bg *аньолоти* 'agnolotti' and *пельмени* '(Russian) dumplings' and Uk *пельмені*, which only correspond to one another in the corpus.

- III. a piece of bread or other baker's goods (Bg *комат, кора₁, крайче, крайциник, къшией, порязаница, среда, троха, филия*⁶; Uk *м'якуш⁷, м'якушка, окраець, скорина, шкурунка*);
- IV. a piece of food, or of anything but usually bread (Bg *залък, резен, хапка*; Uk *крихта, кришка, скиба*);
- V. a piece of anything (Bg *къс, парче*; Uk *кавалок, кусень, кус(ок), луста, шмат(ок)*);
- VI. a whole item, except for those accounted for under II.

The correspondences look as shown in Table 1, in which, when two values symmetrical with respect to the main diagonal differ significantly (e. g., Bg words for pieces of bread correspond to Uk words for pieces of food 74 times, whilst the opposite happens only 3 times), the greater one is in boldface and the smaller one in italics:

Bg \ Uk	I	II	III	IV	V	VI	Σ
I	1418	13	3	0	0	107	1541
II	3	15	0	1	0	49	68
III	6	1	36	74	44	44	205
IV	2	0	3	11	31	4	<i>51</i>
V	0	3	6	<i>13</i>	70	1	93
VI	<i>13</i>	6	<i>1</i>	0	3	1259	<i>1282</i>
Σ	<i>1442</i>	38	<i>49</i>	99	148	1464	3240

Table 2: Correspondences between the semantic categories

From this table one sees that Bulgarian has a greater fondness of the word *хляб* and the terms for units and pieces of bread big and small, and Ukrainian of special terms for baker's goods and general words for pieces of food or of any stuff.⁸

Finally, let us look at the individual lexemes. Table 3 shows the upper left corner of the correspondence table sorted by the overall frequency of the words in our data, that is, featuring the most frequent words. The values which are greatest in both their rows and their columns are in boldface:

	<i>хліб</i>	<i>бутерброд</i>	<i>торт</i>	<i>пиріг</i>	<i>корж</i>	<i>шматок</i>	<i>тістечко</i>	<i>булочка</i>	<i>сухар</i>	<i>печиво</i>	...	Σ
<i>хляб</i>	1415	6	—	—	8	—	—	—	1	1	...	1533
<i>сандвич</i>	1	117	—	—	—	—	1	1	—	—	...	171
<i>питка</i>	2	—	—	1	62	1	—	—	—	1	...	139
<i>торта</i>	—	—	104	6	—	—	3	—	—	2	...	116
<i>сладкиш</i>	—	—	10	19	—	—	31	2	—	16	...	96
<i>парче</i>	—	—	—	—	—	44	—	—	—	—	...	71
<i>сухар</i>	—	—	—	—	2	—	—	—	54	—	...	63
<i>баница</i>	—	—	—	39	—	—	—	—	—	7	...	60
<i>филия</i>	4	4	—	—	—	3	—	—	—	—	...	53
<i>кифла</i>	—	1	—	1	—	—	—	38	—	—	...	51
...
Σ	1442	137	118	116	83	81	73	64	60	56	...	3240

Table 3: Correspondences between the most frequent lexemes

⁶ This word is not in fact restricted to bread (there is *филияка салам* 'slice of sausage', for example), but bread is always understood unless something else is specified.

⁷ The dictionary interprets this word as "Підшкірна частина плодів, ягід тощо" 'The part under the skin of fruit, berries etc.' (SUM 1973: 838), but in the corpus it behaves in the same way as *м'якушка*.

⁸ There is another circumstance which is outwith our scope here but still relevant: Uk *хліб* is more readily used in the meaning of 'grain'.

Uk *хліб* is found 27 times outside of the most frequent pair *хляб* : *хліб* (and no other Bulgarian word corresponds to it more than 4 times), whereas Bg *хляб* occurs 118 times (among its Ukrainian correspondences are *хлібина* ‘loaf of bread’ 23, *опрісок* ‘unleavened bread’ 20, *буханець* ‘loaf of bread’ 13, *грінка* ‘toast’ 10). Also, Bg *филия* ‘slice of bread’ is the 9th most frequent word in the list, whilst its most frequent Uk counterpart *скибка* is only the 18th.

4. A Bilingual Lexicosemantic Network

We can build a lexicosemantic network on the basis of the correspondence table, ignoring the numbers and drawing an edge between a Bulgarian and a Ukrainian node if there is at least one match. Figure 1 presents such a network.⁹ The triangles and the stars are the Bulgarian and Ukrainian words, respectively. In the centre is Bg *питка* ‘bread roll’; from this word any other can be reached in five moves at most (the farthest ones being Uk *батончик*, *луства*, *м’якушка*, *тортик*, *трубочка* and *шкурунка*). Other choices might have been Bg *хляб* and Uk *калач*. The number of edges is 395.

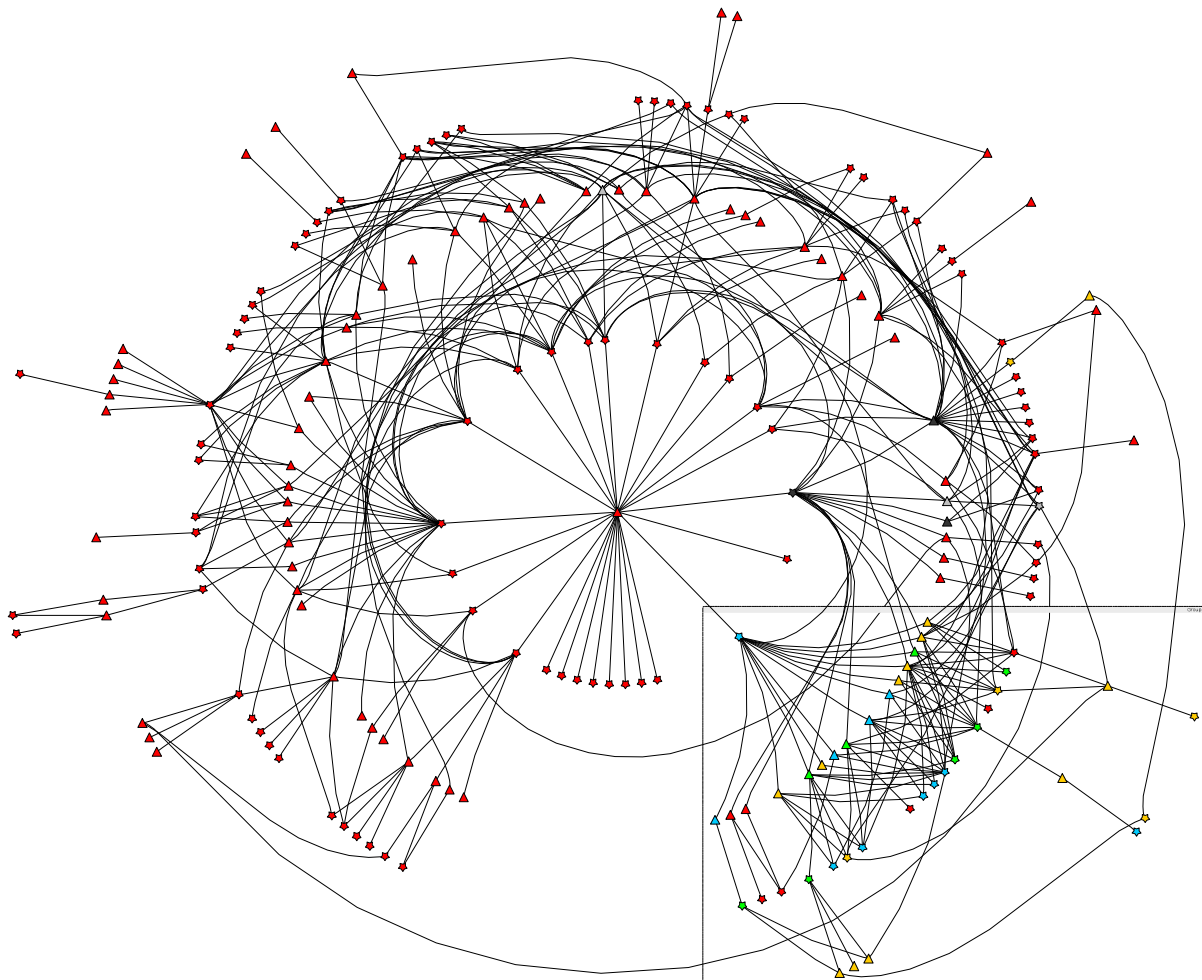


Figure 1: The unabridged lexicosemantic network of bread

The greatest distance between two nodes is 9 edges; such a one separates Bg *кекс* ‘sweet cake, cupcake’ and Uk *луства* ‘piece’, or Bg *маршал* ‘Marshall cake’ and Uk *батончик* ‘stick of confectionery, candy bar’ and *трубочка* ‘puff’, or Bg *троха*, *трохица/трошица* and *трохичка/трошичка* ‘crumb’ and Uk *батончик* and *трубочка* again. Uk *батон* and *батончик* are separated by 6 edges; Uk *калач* and *калачик*, *сухарик* and *сухарець*, *торт* and *тортик* by 4; the remaining Ukrainian pairs of cognate

⁹ The figures are drawn with the graph editor yEd (<https://www.yworks.com/products/yed>).

words, as well as all Bulgarian ones, are 2 edges apart, which here means that there is a word in the other corpus language to which each corresponds at least once.

We would draw attention to the lower right corner of the image, connected to the rest of the graph through Uk *шматок*, where almost all piece words are located, with the exception of Uk *м'якуш* 'soft inner part of bread' (adjacent to Bg *хляб* in the graph) and Bg *среда* 'middle, inner part (incl. of bread)', but with the addition – for reasons of geometry – of Bg *нафора* and *просфора* and Uk *гостія* and *облатка* 'communion wafer', as well as Uk *глетяк* 'underbaked bread' (an infrequent word which only occurs once in CUB, in Oles Honchar's *Guide-on Bearers*, in the plural and with the meaning 'chunks of ~', and this has prompted its being translated into Bulgarian as *клицави хапки* 'sodden mouthfuls'), *грінка* (in Bulgarian mostly *препечена филийка* 'toasted slice of bread') and *тост* 'toast'.

Conceivably some edges are an artefact of translation from third languages and actually connect words with substantially different meanings. The network in Figure 2 is limited to the 764 word pairs found in texts where Bulgarian or Ukrainian is the original language (somewhat less than ¼ of the 3240 in the whole corpus). It features 58 Bulgarian and 69 Ukrainian lexemes (just over ½ of all) and has 141 edges. (The most frequent pair from the big corpus which is missing here is *сладкиш* : *тістечко*, followed by *хляб* : *опріснок* from the Bible.) In the centre is Uk *калач*, which is 8 edges away from the farthest nodes, Uk *скоринка* and *шкуринка* (at the right) and *торт* (in the lower left). The system is no longer fully interconnected; 13 Bulgarian and 10 Ukrainian terms are not linked to the rest. They are in the upper right corner of the figure, from right to left: Bg *кифла* : (*рогалик* : (*кифличка*, *красан*), *книш* : *книш*, *булочка*), Uk *крихта* : (*троха*, *трошица*), Uk *марципан* : (*бадемовка*, *марципан*), Bg *трохичка* : *кришка*, Bg *попара* : *тетеря*, Bg *порязаница* : *луста*, Bg *кекс* : *кекс*, Bg *сухар* : *сухар*.

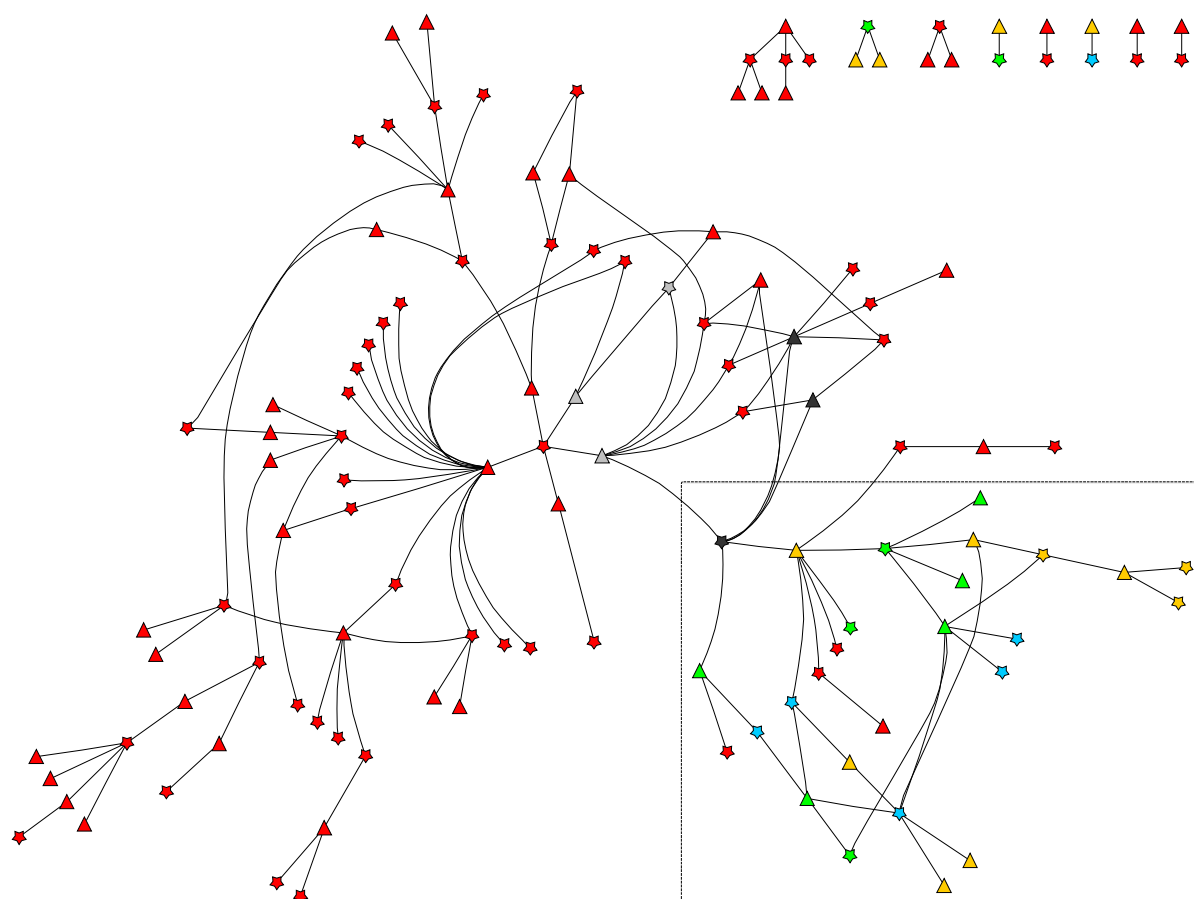


Figure 2: A network based on texts with Bulgarian or Ukrainian originals

Among the pairs in CUB some are extremely rare, such as *сандвич* : *хліб* or *питка* : *пиріг*, which only occur one time each and thus obviously have little to tell us about the highly frequent words that

compose them. Other pairs are more telling, e. g., *хамбургър*¹⁰ : *гамбургер*: it appears 6 times, accounting for all occurrences of the Bulgarian word and all but one (a sole *сандвич* : *гамбургер*, which ensures the contact of this pair to the rest of the system) of the Ukrainian. This suggests that a way of making the network more informative would involve labelling every edge by the number of occurrences of the corresponding word pair or another numeric value reflecting the relevance of the edge. Another way would be to make the edges directed by having them point from the less to the more frequent lexeme, e. g., from Uk *гамбургер* (7) to Bg *сандвич* (171) and from the latter to Uk *хліб* (1442); such edges will often be interpretable as indicating an ‘is a kind of’ relation.

The network can also be made more observable by excluding part of the correspondences, choosing them among the ones that are least well supported by corpus data. The one in Figure 3 has been made with the condition that if two word pairs share a word and one is more frequent than the other, we may not drop the first and keep the second.

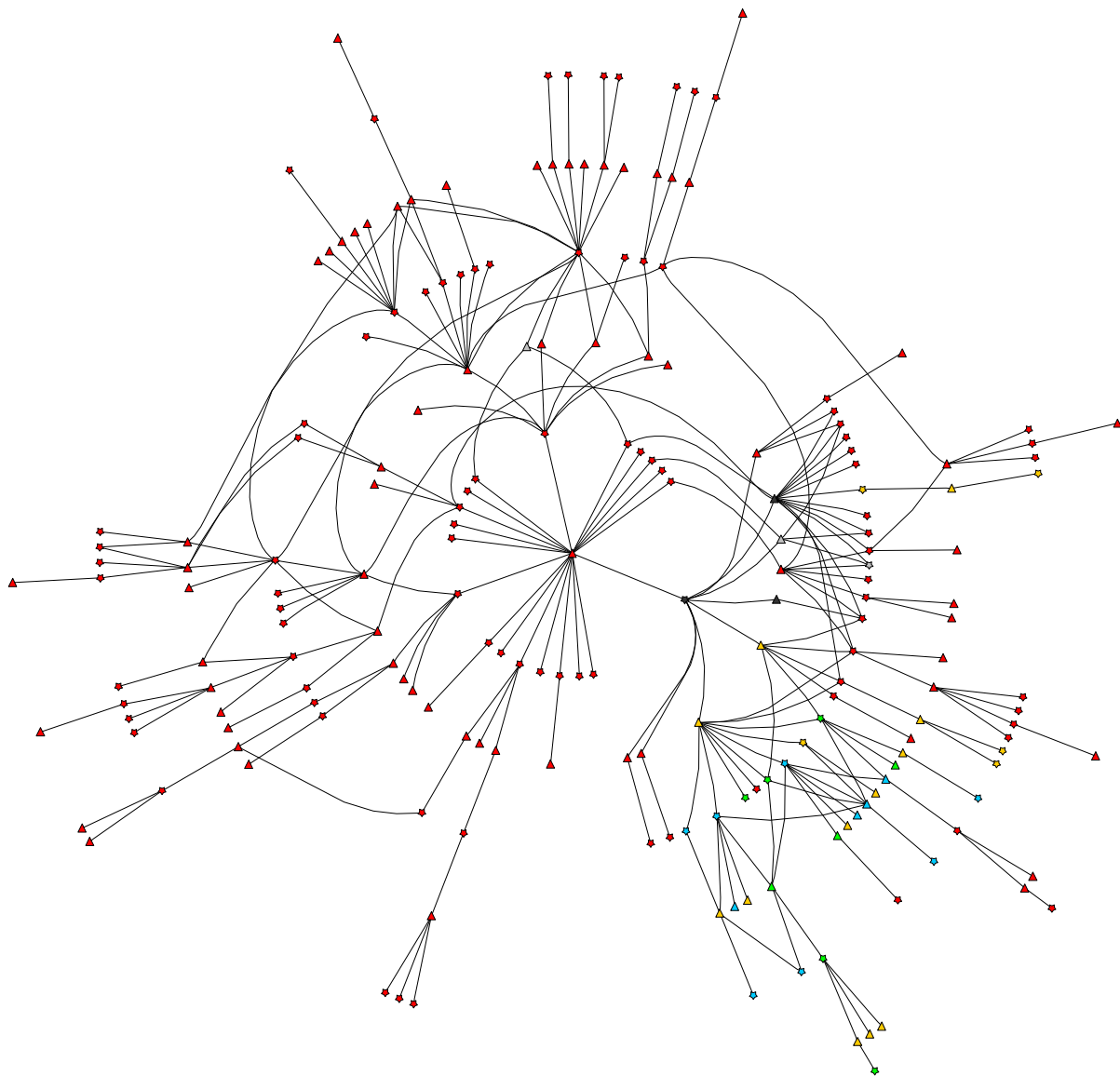


Figure 3: An abridged lexicosemantic network of bread

¹⁰ Thus in the source (a 1983 translation of E. M. Remarque’s novel *Shadows in Paradise*); remarkably, the contemporary language has settled for the form *хамбургер*, which is a less precise rendering of the English pronunciation.

The graph has 251 edges, and no further edge can be removed while preserving both this feature and the integrity of the system. Here, too, the central position is held by Bg *питка*, which is seven edges away from the most distant nodes, Uk *гостія* and *кришка*. (Another option would have been Uk *хліб*.) The greatest distance here is 13 edges, as between Uk *гостія* or *кришка* and Bg *вареник*, *геврек*, *кекс*, *равиоли* or *соленка*.

We can also relax the requirement that the edges from each node must be chosen among the most frequent ones, and take only as many edges as are needed to keep the system fully connected (namely 200, one fewer than the nodes). Such a minimal network is shown in Figure 4. The graph is drawn as a tree, but this should not be taken as implying a hierarchy, since no edges are directed; Bg *питка* is the top node not because it is a root, but because in this graph, as in the other two, it is central: from it all other nodes can be reached in no more than 14 moves, the most distant ones (lowermost in the picture) being Bg *вареник*, *равиоли* and *бухта*. (The other candidate for the central position is Uk *корж*, 14 edges away from Uk *гостія*. In the picture these are the ‘root’ of the right-hand subtree and the lowermost ‘leaf’ of the subtree on the left, respectively.)

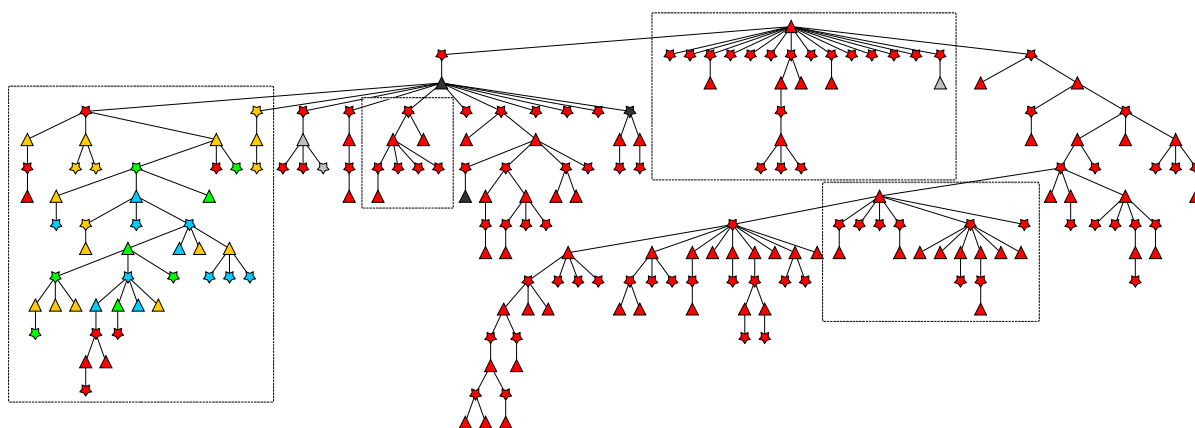


Figure 4: A minimised lexicosemantic network of bread

The largest distance, 27 edges, separates Bg *вареник*, *равиоли* and *бухта* from Uk *гостія*. The pair of cognate words that are farthest apart are Bg *кифла* ‘bun’ and *кифличка*; there are 16 edges between them. Between Uk *батон* and *батончик* there are 14 edges, as between Bg *сухар* ‘rusk’ and *сухарче*, and there are 12 between Uk *сухар* or *сухарець* on one hand and *сухарик* on the other.

What is especially intriguing about this tree is that its subtrees contain semantically well-formed subsets of the bread lexicon of the two languages. There are two big subtrees, one on the left headed by Uk *калач* (91 nodes) and one on the right headed by Uk *корж* (85 nodes), and as a general rule the words for types and quantities of bread are found in the former and for more complex products of bakery and confectionery in the latter. Here again there is a domain where the piece words are concentrated; another, including the centre, with words for types of bread by content (25 nodes); and in the subtrees smaller semantic areas can be recognised, for example sandwiches (in the left-hand tree, headed by Uk *бутерброд*) or desserts (in the right-hand one, headed by Bg *сладкиши*). In the figure these areas are highlighted by boxes.

A further refinement of the networks may involve a correction to the weight of an edge based on the number of sectors or texts in which it is encountered. This would reduce the impact of pairs which occur multiple times but are the handiwork of a single translator, such as Bg *курабия* : Uk *корж*, seen 7 times in Mykhailo Stelmakh’s novel *The Four Fords* and nowhere else.¹¹

5. Conclusions

We would highlight some valuable traits of the proposed approach:

¹¹ Such cases are rare. Their origin – oversight, *faux ami*, influence of a third language, pursuit of pragmatic rather than semantic equivalence, etc. – is of definite interest and merits special study; but in the context of constructing an adequate bilingual lexicosemantic network they demonstrate the desirability of manual edition involving expertise in both languages.

1. formalisation: the sequence of actions is precisely outlined, logically substantiated and carried out;
2. universality: the procedure can be applied to any field defined by a certain concept;
3. relevance: the research base is grounded in both the generalised translation experience embodied in the parallel texts collected in CUB and the fruits of the interpretative lexicography of the two languages embodied in the respective dictionaries;
4. objectivity: the exact rules of action (the formalisation and automation of the procedure) contribute to reducing the subjective component in the linguistic research as much as possible;
5. comprehensive coverage: the multilinguality of the sources (the presence of parallel translations from third languages in CUB) increases the diversity of detected entries;
6. ‘double hit’: the use of a parallel corpus allows building a network on the basis of two languages and for both languages simultaneously, transgressing the boundaries of translation from a source language to a target language.

It should be kept in mind that this method will not always be able to reveal all entries of a particular field on its own. For this task it is better to use a stepwise method of vocabulary identification with substantial use of dictionary interpretations.¹² But when it comes to finding new meanings or even entries not registered by explanatory dictionaries, the use of the parallel corpus method can give interesting results.

Further steps and directions in the development of the research can include an overlay of the obtained network of meanings on similar networks built by other (deductive and inductive) methods in order to compare their coverage, as well as combining the results, so as to obtain a more complete and structured overall system of meanings for each language.

References

- Bojkiv, I., Izjumov, O., Kalyshevs'kyj, H., and Trokhymenko, M. (1955). *Slovník chuzhomovnykh slov*. New York: M. Borets'kyj.
- Derzhanski, I. and Siruk, O. (2019). The Intensifying Prefix *pre-* in a Corpus of Bulgarian and Ukrainian Parallel Texts. In *Digital Presentation and Preservation of Cultural and Scientific Heritage. Conference Proceedings. Vol. 9*. Sofia: Institute of Mathematics and Informatics—BAS, pages 177–188. http://dipp.math.bas.bg/images/2019/177-188_12_2.10_fDiPP2019-67_f_v.1a.F_20190908.pdf.
- Fellbaum, C., Ed. (1998). *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.
- Hrinchenko, B. (1958). *Slovar' ukrajins'koji movy*. Kyjiv: Vyd-vo Akademiji nauk Ukrajins'koji RSR. <http://hrinchenko.com>.
- Panchev, T., Ed. (1904). *Rechnik na bǎlgarskij ezyk s tǎlkuvanie rechite na bǎlgarski i na ruski. Sǎbral i iztǎlkuval Najden Gerov. Chast 5: R–Ja*. Plovdiv: “Sǎglasie”.
- Panchev, T., Ed. (1908). *Dopǎlnenie na bǎlgarskija rechnik ot N. Gerov*. Plovdiv: “Trud”.
- RBE (1977—). *Rechnik na bǎlgarskija ezik*. Sofiya: Izdatelstvo na BAN “Prof. Marin Drinov”. <http://ibl.bas.bg/rbe/>.
- SUM (1970–1980). *Slovník ukrajins'koji movy v 11 tomach*. Kyjiv: Naukova dumka. <http://sum.in.ua>.
- SUM-20 (2010—). *Slovník ukrajins'koji movy u 20 tomach*. Kyjiv: Naukova dumka. <https://slovník.me/dict/newsum>.
- Turchyn, Je. (1985). Leksychna realizacija mikropolja “chastyny khlibyny” v ukrajins'kykh hovorakh, *Strukturni rivni ukrajins'kykh hovoriv*, pages 146–165. Kyjiv: Naukova dumka.

¹² The inclusion of dictionaries of dialects is especially advisable, as they are likely to present greater lexical variety than standard explanatory dictionaries, witness the wealth of words for pieces of bread in Ukrainian dialects summarised and classified in (Turchyn, 1985).

Appendix A. Bulgarian Wordlist

1. бадемовка 2	19. кифла 51	35. късче 10	56. попара 7	75. тиганица 9,
2. баница 60	20. кифличка 6	36. къшей 22	57. порязаница 2	тиганичка 1
3. баничка 15	21. кнедла 8	37. листо 2	58. просеник 1	76. торта 116
4. безквасник 2	22. книш 2	38. марципан 7,	59. просфора 3	77. тортичка 7
5. бисквита 33,	23. козунак 24,	марципанен 1	60. пудинг 16	78. точено 1
бисквитка 1	козуначен 1	39. маршал 1	61. пърженка 2	79. тригуна 1
6. блин 13	24. комат 10	40. маца 3, мац 1	62. равиоли 1	80. троха 41
7. бутерброд 3	25. коричка 16,	41. меденка 1	63. реване 1	81. трошица 2,
8. бухта 8,	кора ₁ 2,	42. мекица 2	64. резен 8,	трохица 1
бухтичка 1	корица 1	43. нафора 27	резенче 1	82. трохица 3,
9. вареник 25	26. кора ₂ 1	44. пай 3	65. самун 46,	трошичка 1
10. вафла 9	27. кравай 10	45. палачинка 17	самунче 1	83. филийка 32
11. геврек 25,	28. кравайче 15	46. парче 71	66. сандвич 171	84. филия 53
гевречен 1	29. крайче 2	47. парченце 9	67. симида 1	85. франзела 5,
12. геврече 8	30. краициник 6,	48. паста 16	68. сладка 19	франзелка 1
13. еклер 2	крайициник 2,	49. пастет 8	69. сладкиш 96	86. хамбургър 6
14. залък 33,	крайициник 2,	50. пирог 15	70. соленка 2	87. хапка 8
залче 1	краециник 1	51. пирожка 39	71. среда 6	88. хлебец 8
15. кадаиф 1	31. кроасан 3	52. пита ₁ 21	72. сухар 63	89. хлебец 49
16. канола 4	32. курабийка 14	53. пита ₂ 7	73. сухарче 3	90. хляб 1533
17. кейк 11	33. курабия 27	54. питка 139	74. тако 2	91. шрудел 9
18. кекс 2	34. къс 3	55. погача 14		

Appendix B. Ukrainian Wordlist

1. бабка 9	19. вівсяник 1	44. кришка 8	68. паска 19,	92. сухарець 3
2. балабуха 1	20. галета 6	45. кукурудзяник 1	пасочка 1	93. сухарик 4
3. баница 2,	21. гамбургер 7	46. кулеб'яка 1	69. паишет 12	94. тартинка 4
баниця 1	22. глевтяк 1	47. кулич 1	70. перепічка 13	95. тетеря 5,
4. батон 2	23. горохвяник 2	48. кусок 14, кус 1	71. печиво 56	тетерка 1
5. батончик 3	24. гостія 1	49. кусень 12	72. пиріг 116	96. тістечко 71,
6. бісквіт 13	25. грінка 43,	50. лигун 1	73. пиріжечок 5	тістечковий 2
7. бріюш 2	гріночка 1	51. луста 1	74. пиріжок 42,	97. товченик 1
8. бублик 26,	26. душеник 2	52. маківник 2	пиріжковий 1	98. торт 118
бубликовий 1,	27. житник 1	53. малай 1	75. підпалок 9	99. тортик 5
бубличний 1	28. завиванець 1	54. марципан 9,	76. пляцинда 1	100. тост 1
9. бубличок 3	29. кавалок 3	марципановий 277.	77. плячок 1	101. трубочка 4
10. булка 28	30. калач 28	55. маца 3	78. потапці 1	102. хліб 1440,
11. булочка 63,	31. калачик 11	56. медівник 1	79. прісне 3	хлібний 2
півбулочки 1	32. канাপка 2	57. млинець 35,	80. пряник 18,	103. хлібець 20
12. бутерброд 133,	33. картопляник 6	млиничик 2	пірник 1,	104. хлібина 37,
бутербродний 3,	34. квашене 2	58. м'якуш 5	попряничок 1	півхлібини 2,
бутербродик 1	35. кекс 4	59. м'якушка 4	81. пудинг 14	півхлібинки 1,
13. буханець 34,	36. книш 5	60. налисник 3	82. пундик 4	хлібинка 1
буханка 2,	37. колобок 4	61. облатка 28	83. рогалик 7	105. чурек 1
боханець 1,	38. корж 83	62. окраєць 22,	84. сандвіч 39,	106. икуруинка 6
півбухана 1	39. коржик 31	окрайчик 1	сендвіч 6	107. имат 6
14. вареник 28,	40. коровай 11	63. оладка 7	85. скиба 3	108. иматок 81
вареничок 1	41. крекер 2	64. опріснок 25	86. скибка 34	109. иматочок 30
15. варениця 1	42. крендель 4,	65. оціпок 1	87. скибочка 13	110. шуляк 1
16. ватрушка 4	крендельок 1	66. паляниця 13,	88. скоринка 11	
17. вафля 5,	43. крихта 39,	паляничка 1	89. слойка 1	
вафельний 2	крихтітка 1,	67. пампушка 9	90. струдель 9	
18. витічка 3	крихточка 1		91. сухар 60	