Electronic Romanian-English Contrastive Collection of Food Set Phrases. Methodological Aspects

Petronela SAVIN*

Key-words: *electronic collection, Romanian, English, food, set phrases, methodology*

1. Lexicography and new IT&C technologies. To a new e-collection

Lexicography benefits from the evolution of linguistic research using the new IT&C technologies, electronic collections having been made by using relational databases, mark up languages, professional dictionary writing system, wordnets (lexical databases) (Rothkegel 2007; Dziemianko 2010; Chen 2010). In spite of the emergence and development of electronic phraseological dictionaries, inspired by theoretical work on collocations and idioms (Mel'čuk, 2006), such as DAFLES[1] and ELDIT[2], there is no computerized electronic instrument to reflect the functionality of a particular onomasiological bilingual corpus of set phrases.

In this context, we propose an *Electronic Romanian-English Contrastive Collection of Food Set Phrases* which will be organized as a web-oriented database, which takes into account the new (digital) methodologies in modern lexicography. We consider that, due to the essential position that food set phrases hold in the reflection of the mechanisms of language expressivity, compiling a Romanian-English collection of set phrases whose resources could be enhanced by means of software instruments, is very productive. A collection of food idioms, as a complex database, will illustrate the main dimensions of the set phrases which reveal, on the one hand, conceptual structures and metaphorical modes of conceptualizing the world (Lakoff/Johnson 1980; Liu 2002), and, on the other hand, "cultural taxons" based on social interaction, phenomena of material culture, intertextual phenomena, fictive conceptual domains, cultural symbols (Dobrovol'skij 1998, Piirainen 2007, Mieder 2008).

The linguistic approach of this project enables us to answer more precisely various questions concerning potential explanations of cross-linguistic similarities of set phrases as well as illustrate elements related to cultural specificity. The informational approach to this project offers the possibility of using the results of linguistic research for the development of digital instruments for the processing of

* "Vasile Alecsandri" University of Bacau, Romania.

[1] http://ilt.kuleuven.be/blf/prod.php.

[2] http://eldit.eurac.edu/.

BDD-A24469 © 2016 Institutul de Filologie Română „A. Philippide"
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 20:49:32 UTC)

natural language. The development of informational applications which would use the generated database in automatic translations or corpus analysis (for example, the analysis of feelings) is only one of the many possible future outcomes of the project.

2. Technical features of the electronic collection

2.1. The corpus of this collection will be constituted of *set phrases* (collocations, idioms, quasi-idioms, catch phrases, routines and proverbs) excerpted from the most important Romanian and English lexicographical works[3] as well as from the World Wide Web. Despite the doubts regarding the use of a web corpus, we resort to it because the size of this corpus is so big that the probability of drawing wrong conclusions is very limited.

2.2. The corpus design will be organized as matrices of terminological information. The novelty lies in the development of this matrix composed of description-data for each food word-image and the related set phrases. The data will serve as a benchmark for various automated filtering-display reports, will correlate set phrases separately and simultaneously in the two languages, will ensure uniformity and will confer the corpus the nature of system. The matrix structure is as follows: data on ***word-entry*** reflecting a food image (the WE categories are: the *WE language*, the *WE etymology*, the *WE morphological category*, the *WE meaning*, the *WE realia classes*, the *WE set phrases*) and data on ***set phrases*** where they function (the SPH categories are: the *SPH type*, the *SPH literal translation*, the *SPH meaning*, the *WE paradigm signification in SPH*, the *SPH general concept*, the *SPH text-example*).

We provide below an example of the categories of the word-entry PÂINE.

Word-entry (WE)	**WE language**	**WE etymology**	**WE morph. category**	**WE meaning**	**WE realia classe**
PÂINE	Romanian	Lat. *panis*	noun	‘a common food made from flour, water, and usually yeast’	Dishes made from cereals

[3] The corpus of Romanian phraseological structures will be taken from Iuliu Zanne’s (1895–1912) monumental collection *Proverbele românilor din România, Bucovina, Ungaria, Istria şi Macedonia. Proverbe, zicători, povăţuiri, cuvinte adevărate, asemănări, idiotisme şi cimilituri cu un glosar româno-france*s, vols. I–X, Bucuresti, Editura Librariei Socecu & Comp., together with both the old and new series of the most important work of Romanian lexicography, DA = [Academia Română] *Dicţionarul limbii române*; I/I, letters A-B,1913; I/II, letter C, 1940; I/III, fascicle I, D -*de*, 1949; II/I, letters F–I, 1934; II/II, fasc. I, J -*lacustru*, 1937; II/III, *ladă* - *lojniţă* [etc.] and DLR = [Academia Română] *Dicţionarul limbii române* (new series), Editura Academiei, tome VI, letter M (1965–1968); VII/1, letter N (1971); VII/2, letter O (1969); VIII/1-5, letter P (1972–1984); IX, letter R (1975); X/l-5, letter S (1986–1994); XI/1, letter S (1978); XI/2-3, letter T (1982–1983); XII/1, letter T (1994); XIII/1-2, letter V (V- veni; 1997; *venial–vizurină*; 2003); IV, letter L (*L – lherzolită*; 2008); V, letter L (*Li-luzulă*, 2008).

The corpus of English set phrases will be taken taken from Manser, Martin, *The Facts on File Dictionary of Proverbs*, New York, Infobase Publishing, 2007; Siegfried, Judith, *Oxford Dictionary of Idioms*, OUP, 2nd Ed., 2004, and Speake, Jennifer, *Oxford Concise Dictionary of Proverbs*, OUP, 2003 and *The Free Dictionary by Farlex*, available online.

BDD-A24469 © 2016 Institutul de Filologie Română „A. Philippide”
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 20:49:32 UTC)

Set phrases (*SPH*)	**SPH type**	**SPH literal translation**	**SPH meaning**	**WE paradigm signification in SPH**	**SPH general concept**	**SPH text-example**
a lua (*cuiva*) *pâinea de la gură*	idiom	to take the bread from somebody's mouth	'to leave somebody without the possibility of making a living'	Food – image for the necessary means of subsistence	extortion	External link

Each category of information is important in studying set phrases and some of them represent a way to filter the collection's material.

The *WE language* category contains the options Romanian language and English language and therefore the database can be viewed as a single language collection or a bilingual collection.

The *WE morphological* category enables the filtering of the material after the lexical-semantic characteristics of the *WE*. Similarly, *WE meaning* category allows you to select the entry with the same meaning in both languages, allowing contrastive study. *WE etymology* enables illustration of etymological layers of *WE*, relevant in the evolution of meanings of *SPH*. *WE realia class* is especially important because the relationship with the real of the *WE* is defining in the construction of the meaning and exploitation of the phraseological level. The whole material will be filtered according to the reference classes or subclasses: Food (Fruits, Vegetables, Cereals, Cereal dishes, Milk, Milk products, Honey, Meat and meat dishes, Egg, Salt), Drinks (Water, Alcoholic beverages, Refreshments), Characteristics, Actions, States, Organs, Instruments (Containers, Tools, Facilities).

The category of the *WE set phrases* requires branching phraseological unit basis. Each *SPH* have the categories: the *SPH literal translation*, the *SPH meaning*, the *SPH text-example* (with an external link), which provides essential lexicographical information. The literal translation of each set phrase is important not only in the analysis of its internal structure, but also in the selection of similar form pairs in the two languages, Romanian and English.

The category *SPH type* is a way of filtering material allowing the organization depending on the types of set phrases (idioms, quasi-idioms, catch phrases, routines and proverbs).

The following two categories, that enable filtering of all the material, the *WE paradigm signification in SPH* and *SPH general concept*, are ways of semantic processing of the phraseological material.

2.3. The main semantic problem in organizing this corpus is to track the extent to which we can identify a relationship between the literal and phraseological meaning of the fixed structures based on food images. Previous research has led us to the conclusion that there are relevant features of the food image involved in the meaning of the phraseologism as a whole, labelled as the category *WE signification paradigm in SPH* (see Savin 2012)[4]. Thus, we have identified paradigms such as:

[4] In Petronela Savin, *Romanian Phraseological Dictionary. The Onomasiological Field of Human Nourishment*, Iași, Institutul European, 2012, the set phrases related to feeding were alphabetically

BDD-A24469 © 2016 Institutul de Filologie Română „A. Philippide"
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 20:49:32 UTC)

Food – image for the necessary means of subsistence, *Good taste – mark of the positive*, *Bad taste – mark of the negative*, *Lack of nutritional quality – image of non-value*, *Lack of edibility – mark of mischief*, *Accessibility – image of laziness*, *Inadequacy of context – image of the mismatch*, *Rarity – image of value* etc. This category enters a direct relationship with another matrix category *The SPH general concept*, reflecting the general meaning of the set phrase through a key-word: *hospitality*, *gratefulness*, *value*, *forgiveness*, *wisdom*, *altruism*, *power*, *pleasure*, *prosperity*, *honesty*, *loyalty*, *goodness*, *greed*, *misfortune*, *cheating*, *temptation*, *treachery*, *loss*, *ignorance*, *extortion*, *stupidity*, *poverty*, *selfishness* etc. Through digital tools, we can monitor the compatibility between *WE paradigm signification in SPH* and *SPH general concept*, each set phrase being integrated in complex systems of representation proving the functionality of some general principles of the metaphoric transfer. The very compatibility between the meaning of food imagery and the general meaning of the phraseological structure, reduced to categories of the good-bad opposition, could be exponential in future applications for the use of the database in the improvement of automatic translations or in corpus analysis of the feeling analysis type.

For example, by filtering the material by *WE meaning* 'a common food made from flour, water, and usually yeast' and *WE paradigm signification in SPH, Food – image for the necessary means of subsistence*, we could get on the English interface the following information:

Word-entry (*WE*)	**WE language**	**WE etymology**	**WE morph. category**	**WE realia classe**
PÂINE	Romanian	Lat. *panis*	noun	Dishes made from cereals
BREAD	English	Old English *bread*	noun	Dishes made from cereals

Set phrases (*SPH*)	**SPH type**	**SPH literal translation**	**SPH meaning**	**SPH general concept**	**SPH text-example**
bucată de pâine or *pâinea zilnică, pâinea cea de toate zilele*	idiom	piece of bread or daily bread	'someone's basic income'	value	External link
someone's bread and butter	idiom		'someone's basic income; someone's livelihood – the source of one's food'	value	External link
a lua (*cuiva*) *pâinea de la gură*	idiom	to take the bread from somebody's mouth	'to leave somebody without the possibility of making a living'	extortion	External link
take the bread out of people's	idiom		'deprive people of their livings,	extortion	External link

organized into nuclei, according to the central food image, as well as into series based based on the signification paradigm of the food image at the level of the phraseologism.

BDD-A24469 © 2016 Institutul de Filologie Română „A. Philippide"
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 20:49:32 UTC)

mouths			especially by competition or unfair working practices'		
a mânca pâinea şi sarea cuiva	idiom	to eat someone else's bread and salt	'to be received into somebody's house, to enjoy somebody's benevolence'	extortion	External link
a mânca pâine şi sare (*cu cineva*) *break bread with someone*	idiom idiom	to eat bread and salt with someone	'to live together' 'to eat a meal with someone'	cohabitation altruism	External link External link
Fie pâinea cât de rea, / Tot mai bună-n vatra mea *Dry bread at home is better than roast meat abroad*	proverb proverb	No matter how bad the bread, / It is still better on my hearth	'there's no other place like home' 'there is no place like home'	wisdom wisdom	External link External link
Whose bread I eat, his song I sing.	proverb		'people show loyalty to, or comply with the demands of, those who employ, pay, or feed them'	loyalty	External link

For long, in translation theory and, therefore, also in lexicography, the emphasis was placed on recording the phraseological meaning, the literal component of the phraseological structures being ignored. With the increasing interest in the development of natural language processing techniques, there was more acutely highlighted the issue of the relationship between the phraseological meaning, opaque to automatic recognition, and the literal component of the stable word structures, the only transparent one. From all the phraseologisms, the number of those with a full idiomatic nature, opaque in terms of the construction of meaning, is small. Therefore, an important question to be answered by analyzing the *Electronic Romanian-English Contrastive Collection of Food Set Phrases* is describing the link between the two levels of meaning, the phraseological and the literal meaning. One theory regarding the figurativeness of set phrases may start from the assumption that the *Signification paradigm of the WE in SPH* may be essential to ensure transparency of meaning to the entire phraseological structure. This hypothesis, once validated, would provide a new field for the studies dealing with figurativeness, highlighting for the first time, the role of semantic relations established at the level of an onomasiological field. The following stage should reveal the extent to which the features identifiable at the level of this field may be extrapolated to other fields, and also generalized to the whole domain of phraseology.

BDD-A24469 © 2016 Institutul de Filologie Română „A. Philippide"
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 20:49:32 UTC)

2.4. Therefore, the solutions to the different tasks of the requirements of this corpus reflect the state of the art concerning the modern achievements of collection editing and phraseologic research in the Digital Age.

We will build also an online and digital portal giving access to the collection that the project is aiming at. Furthermore, an online research infrastructure is taken into account to manage and analyze the data, as there is a big quantity and heterogeneity of data to be treated. This question and also that of the sustainability of the data were taken into consideration. In fact, the project aims to shape the data in such a way that it can be made fruitful for further research on European phraseology and interlinked with other phraseological material from other collections. This project will be fully developed in terms of a relational database, both as a work tool and as a way of presenting the results. Thus it could serve various corpus analyses through techniques for the automatic processing of natural language, by offering the possibility to find out the frequency with which each set phrase is used in the online medium, the type of texts in which it is used, its capacity to be modified in use etc. Therefore, such a data base enables, among other things, the basic conditions for a full exploitation of the material, while keeping the needs of maximum transparency of data.

3. Degree of originality

The main novelties, the original elements of the methodological model of the collection proposed are:

a) it has an innovative and integrative character, as it systemically articulates theoretical founding principles of lexicography with the most recent phraseological research and IT&C techniques;

b) it has an operational character, as it offers the possibility of further development of phraseological dictionaries for various domains and starting from miscellaneous types of corpora and it also contributes to the development of natural language processing techniques;

c) it has a complex character, because it is interdisciplinary, pluridisciplinary, and transdisciplinary, as it combines linguistic research methods with those used in computer science;

d) it has a formative character as it encourages the use of modern lexicographical tools, and develops lexicographers' competences and abilities.

4. Impact

The estimated impact of the project envisages several dimensions:

a) *scientific*: the development of innovative models for the formal and computerized representation of a *Romanian-English Contrastive Collection of Food Set Phrases*, which, through the intermediary of computer science means, capitalizes upon a linguistic material according to principles of both semasiological and onomasiological approaches; adopting the principles of the computational linguistic in realizing the contrastive collection mean a stimulation of the studies in phraseological research topics, such as semantic and semiotic aspects of phrasemes, cognitive aspects, contrastive phraseology and translation, phraseography;

BDD-A24469 © 2016 Institutul de Filologie Română „A. Philippide"
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 20:49:32 UTC)

b) *socio-economic*: the socio-professional insertion of philology and computer science graduates on the labor market at the current level of both its requirements and of the expectations of the knowledge-based society with reference to the use of new IT&C technologies;

c) *psycho-pedagogical*: enhancing lexicographers' motivation and receptivity towards new IT&C technologies, as well as improving the design and implementation of lexicographers' training; the competent utilization by beginner lexicographers of a modern instrument for collection compilation.

The target audiences are not only linguists, translators, anthropologists and specialists in informatics, but also the public with an interest in the expressiveness of both the Romanian language and the English language.

Bibliography

Chen 2010: Y. Chen, *Dictionary Use and EFL Learning. A Contrastive Study of Pocket Electronic Dictionaries and Paper Dictionaries*, "Int. J. Lexicography", no. 23(3), p. 275–306.

Coşeriu 1994: E. Coşeriu, *Lingvistica din perspectivă spaţială şi antropologică. Trei studii*, Chişinău, Editura Ştiinţa.

Dobrovol'skij 1998: D. Dobrovol'skij, *On cultural component in the semantic structure of idioms* in Ďurčo, P. (ed.), *Europhras 97: International Symposium. September 2-5, 1997, Liptovský Jan. Phraseology and Paremiology*, Bratislava, p. 55–61.

Dumistrăcel 1980: Stelian Dumistrăcel, *Lexic românesc. Cuvinte, metafore, expresii*, Bucureşti, Editura Ştiintifică şi Enciclopedică.

Dziemianko 2010: Anna Dziemianko, *Paper or Electronic? The Role of Dictionary Form in Language Reception, Production and the Retention of Meaning and Collocations*, in "Int. J. Lexicography", no. 23(3), p. 257–273.

Lakoff, Johnson 1980: G. Lakoff, M. Johnson, *Metaphors we live by*, Chicago, University of Chicago Press.

Liu 2002. D. Liu, *Metaphor, culture, and worldview. The case of American English and the Chinese language*, Lanham, Maryland.

Mel'čuk 2006: Igor Mel'čuk, *Explanatory Combinatorial Dictionary*, in Giandomenico Sica (ed.), *Open Problems in Linguistic and Lexicography*, Monza (Italy), Polimetrica.

Mieder 2008: W. Mieder, *Strategies of wisdom: Anglo-American and German proverb studies*, New York, Peter Lang Publishing.

Piirainen 2007: E. Piirainen, *Phrasemes from a cultural semiotic perspective*, in Burger, H. et al. (eds.). *Phraseologie/Phraseology: Ein internationales Handbuch zeitgenössischer Forschung/An International Handbook of Contemporary Research*, vol. 1, Berlin, Walter de Gruyter, p. 208–219.

Rothkegel 2007: Annely Rothkegel, *Computerlinguistische Aspekte der Phraseme*, H. et al. (eds.). *Phraseologie/Phraseology: Ein internationales Handbuch zeitgenössischer Forschung/An International Handbook of Contemporary Research*, vol. 2, Berlin-New York, Walter de Gruyter.

Savin 2012: Petronela Savin, *Romanian Phraseological Dictionary. The Onomasiological Field of Human Nourishment*, Iaşi, Institutul European.

Trandabăţ 2013; Diana Trandabăţ, *Towards extracting relations from unstructured data through natural language semantics*, in *Proceedings of the 2nd Workshop on New Frontiers in Mining Complex Patterns* (*NFMCP 2013*), Prague, *23–27 September 2013*, p. 225–236.

BDD-A24469 © 2016 Institutul de Filologie Română „A. Philippide"
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 20:49:32 UTC)

Sources

DA = [Academia Română] *Dicţionarul limbii române*; I/I, letters A-B,1913; I/II, letter C, 1940; I/III, fascicle I, D *-de*, 1949; II/I, letters F-I, 1934; II/II, fasc. I, J *-lacustru*, 1937; II/III, *ladă - lojniţă* [etc.].

DLR = [Academia Română] *Dicţionarul limbii române* (new series), Editura Academiei, tome VI, letter M (1965–1968); VII/1, letter N (1971); VII/2, letter O (1969); VIII/1-5, letter P (1972–1984); IX, letter R (1975); X/l-5, letter S (1986–1994); XI/1, letter S (1978); XI/2-3, letter T (1982-1983); XII/1, letter T (1994); XIII/1-2, letter V (V-veni; 1997; *venial–vizurină*; 2003); IV, letter L (*L – lherzolită*; 2008); V, letter L (*Li-luzulă*, 2008).

Manser 2007: Martin Manser, *The Facts on File Dictionary of Proverbs*, New York, Infobase Publishing, 2007.

Siegfried 2004: Judith Siegfried, *Oxford Dictionary of Idioms*, Oxford, OUP, 2nd Ed.

Speake 2003: Jennifer Speake, *Oxford Concise Dictionary of Proverbs*, Oxford, OUP.

The Free Dictionary by Farlex, available online: http://www.thefreedictionary.com/.

Zannne 1895–1912: Iuliu Zanne, *Proverbele românilor din România, Bucovina, Ungaria, Istria şi Macedonia. Proverbe, zicători, povăţuiri, cuvinte adevărate, asemănări, idiotisme şi cimilituri cu un glosar româno-france*s, vol. I–X, Bucureşti, Editura Librăriei Socecu & Comp.

Abstract

This project's main objective is the elaboration of an electronic Romanian-English contrastive collection of food set phrases. This research intends to make use of digital scientific lexicographical methods in the field of phraseology, providing an answer to the challenge regarding the research strategies of the language study from the perspective of cultural universals and cultural specificity. Even if the cross-linguistic similarities of set phrases according to their cultural foundations seem to gain ground in today's research, the electronic collections are less convincing. We believe that the contrastive research of the phraseological structures of two languages must overcome syntactical formalism by studying homologous onomasiological fields digitally developed. The study of the fixed structures of languages directly connected to the real, in our case, to nourishment, is an answer to Eugeniu Coşeriu's challenge regarding the various demands of linguistic research which, as compared with other subjects, entails the most numerous connections with man's way of being in general.

Bringing together language and reality in the digital space represents the most important challenge of the proposed research. An electronic Romanian-English contrastive collection of food set phrases, systemically organized according to digital markers, is, to our knowledge, the first attempt at an electronic revaluation of a bilingual onomasiological field of set phrases.

BDD-A24469 © 2016 Institutul de Filologie Română „A. Philippide"
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 20:49:32 UTC)