Chromatisms of red colour in academic dictionaries and in the web corpus (Spanish, French)

Maria Spisiakova – Nina Mockova

Abstract
The present paper deals with pluriverbal constructions containing the element of red colour. The red colour is after the black and the white, together with the green, one of the most productive colours able to form lexicalized constructions. This kind of expressions can belong to everyday language, to specific scientific areas, or to the idiomatic field of language. This article focuses on the types and number of these expressions through the analysis of two main sources: the traditional academic dictionaries and the electronic web corpus *Araneum*.
Key words: language, red, colour, dictionary, corpus, linguistics

Introduction
In present days, there are plenty of methods how to do a research and to look for data when studying practically any area of a language. Thanks to modern technology linguists are given new possibilities and new horizons are opened in the field of linguistic research. However, maybe only few decades ago, we could hardly imagine this rich scale of possibilities. In fact, it means that the present day linguistic researches are much more dynamic and more importance (and opportunity) is given to researches of interdisciplinary character. Therefore, some linguists (Lindqvist, 2009, Hoffmann et al., 2008) tend to define corpus linguistics rather as a methodology throughout which linguistic research in various language areas can be realized. They state that in comparison with the autonomous linguistic disciplines, which describe or explain certain language features, corpus linguistics does not centre on particular features, but can be utilized to explore any area of language use. It means that the corpus linguistic methodology can be used in lexicographical, lexicological, morphological, syntactical, stylistic or semantic research, in discourse analysis, language variation, contrastive studies, etc.
In opposition to this ideas, some other linguists (Cermak, 2000, Simkova, 2008) claim that corpus linguistics is a branch of linguistics, having its own methodology and object. Thanks to the continuous and fast evolution, development and improving of the information and communication technologies playing an essential role in the existence of the corpus linguistics, we can speak now about it as about one of the most dynamically developing modern linguistic disciplines, being at the same time considered as the main component of the applied computer linguistics (Cermak, 2000). It investigates and describes language, or, let´s say, its usage (parole) in real contexts, that are looked for in corpora and are processed with the use of statistic methods" (Simkova, 2008). Altogether, all research based on corpora has become known as corpus linguistics and it does not make any change whether we contemplate about it as a set of methods or as an independent field of linguistics.
Corpus linguistics in the modern sense began to develop in the early 1980s of the past 20th century, to a large extend thanks to the real development of computer technology, and the first electronic corpora were created for English language. The essential base of the corpus linguistics existence lays in the creation of a corpus itself, which can be characterized as an electronic database of written or spoken texts, which in their users´ form had passed through a certain computer language processing according to the valid standard linguistic norms of a given language for which the corpus had been

designed. Such text database is objective in the sense that it has been collected according to some extralinguistic criteria, what means that anyone can access to previous researches based on the corpus (Ahola, 2011). Such selected and processed collection of pieces of language is then considered to be a representative sample of the language (Sinclair, 1996 in Ahola, 2011). Working with the electronic corpora plays an important role in the modern linguistics, because it offers empirical data, essential for studying the whole scale of linguistic phenomena of the present day language, offering, at the same time, global characteristics of a language in a specific historical period (RAE, 2014).

As we said, obtaining data through individual excerption or background researching was done traditionally from printed dictionaries, press, field survey, card registers, etc. All this work can be done now thanks to the Internet and the information technologies, we can say, at one computer and, what is more important, within a comparably shorter time horizon. Moreover, the corpus linguistics and the electronic corpora have replaced a great amount of individual terrene research or researching of the printed media, books, etc. One of the main advantages of realizing a linguistic research through the corpus linguistic is for sure the possibility of an easy verification of the obtained results that can be repeated, because the same material is at the disposition to all linguistic researchers (as well as to all other people interested in the matter) without any areal or temporal restrictions (Simkova, 2008). Regarding the mentioned definition, we can assume the main benefits of corpus data: computer processing of data avoids human factor errors in analysis and therefore such data is reliable. The data is contextualized (it means, it shows language use in real linguistic contexts), so it provides a large number of examples occurred in real contexts. In connection to this point, we can say that corpus data is quantitative in the way that it shows results from the perspective of their frequency or statistics, etc. Then, all the data can be processed quickly and at minimal cost (operations like searching, selecting, sorting out, categorizing, etc.). Finally, large corpora are constantly updated, what enables researchers to capture even recent changes in the language. In conclusion, we could consider the modern large electronic corpora as a kind of authority for linguists, in the same way as printed dictionaries.

**Objective, methodology, hypothesis**

For the purposes of our present research we decided to work with the electronic corpus *Araneum*. "The first stage of the *Aranea* Project (Benko 2014, 2016, Benko, Zakharov 2016) has been targeted to creation of a family of dozen+ Giga-word web corpora for languages spoken in Slovakia and its neighbouring countries, as well as for the main foreign languages taught at Slovak universities. This stage is next to completed and the *Aranea* family currently contains corpora for 18 languages in (usually) two sizes, with some languages having also region-specific variants" (Benko, 2017).

In our study we focus on the topic of the chromatisms with the presence of the red colour in Spanish language (*rojo* in Spanish). This intention stems from our recent project we are working on called VEGA *Chromatism and its connotations in the context of Slavonic and Romance languages*. In this project we compare and contrast forms, usage, appearance, etc. of naming units with a chromatic element in their structure in selected Romance and Slavonic languages. We investigate the appearance of these units in the contemporary stage of the given languages; it means, of the lexemes that are actual and are used by the contemporary language users. Similarly, as in our present paper, also in our main project we work with the already mentioned electronic corpus Araneum, together with *Slovak National Corpus* and other electronic corpora of other countries representing the languages that take part in the project. Of course, for the purpose of selection of the given lexical units we use all the accessible dictionaries, whether in their printed or electronic version.

Therefore, at this point we are interested and would like to verify:

differences in appearance of these lexical units in dictionaries and in corpora,
differences in the nature of those units regarding the source difference,
which of those units are commonly used both in dictionaries and corpora.

Regarding the usage of the individual colours in language, and according our previous researches in this area, we can say that black and the white are the most commonly used colours in our lives as reflected also in the language, so their frequency of appearance and of usage is really high. According to M. Spišiaková, the red, the green and the blue are quite frequent as well, although not at such a high level as the black and the white. Something less used and frequent are: yellow, brown, violet/purple, pink/rose and orange. The red colour results as the fourth colour used for the naming units with a chromatic element in the Spanish language (Spisiakova, 2016). Comparable results had been obtained also in the study done by Mockova (2017), focused on idioms with a chromatic element in Spanish language. This result is well displayed in the following diagram[1]:
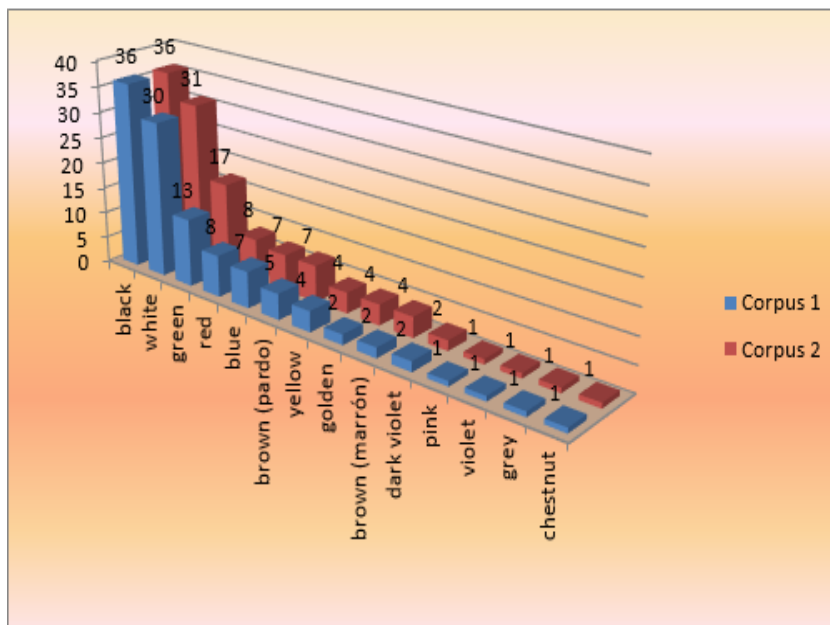
Figure 1. Distribution of the chromatic idioms according to colours

The object of our research are all lexical units according to their structure, what means, monoverbal, pluriverbal, syntagmatic naming units or whole phrase constructions, always when there is *red* as one of their constituents. We are interested in the appearance and usage of all those lexical units containing the red colour under the condition that they meet at least one of the following criteria:

---

[1] The terms *Corpus 1* and *Corpus 2* are related to the study of Spanish chromatic idioms, in which we worked with another corpora, not all of them mentioned in our present study.

there is a at least minimal presence of the figurativeness, so that the motivation of the naming unit is present,

they are lexicalized,

the colour is present in naming units in which it already had lost its original meaning of "being red" or of a colour as such (e.g. surnames),

they are idiomatic and therefore not literally translatable,

there is a stylistic expressivity,

the colour appears also in other lexemes derived from *red* (e.g. *sonrojar*).

In our research we include constructions like *té rojo* (red tea), *piel roja* (red skin) or *frutas rojas* (red fruits), because although there is missing the figurativeness in some of them, they are considered to be fixed lexicalized units with a unique meaning. We excluded only such naming units in which the red colour had the sole meaning of some red object, for instance, *red sweater, red car*, etc. and the names of literary or artistic pieces, names of newspaper, books, etc. Simultaneously, we did not regard the names of some sport clubs or other companies, like, for example, *Tiburones rojos* (Red Sharks) – a football club of Veracruz (Mexico), or, at the same time, some of the not globally known toponyms.

In our present paper, and, especially in the initial part of our research, we use the method of excerption, when the following lexicographic sources in Spanish language have been consulted: monolingual explanatory dictionaries *Diccionario de la Real Academia Española* (2014) and *Diccionario de uso de María Moliner* (2007), a bilingual translating dictionary *Španielsko-slovenský a slovensko-španielsky frazeologický slovník* (2017) and an idiomatic dictionary *Diccionario fraseológico del español actual* (2004)[2] in their printed or electronic versions. The excerpted dictionary units are in the next phase compared and contrasted through the comparative method with the results obtained from the excerption from the electronic corpus Araneum. We decided for the Araneum corpus because of various reasons:

it is a corpus created and processed in Slovakia and apart from the Slovak language it contains other 31 languages, including varieties,

it does not distinguish between the European and the American standard form of the Spanish language[3],

the process of searching is easy in that way that it is enough with the lemma form of the word *rojo* (red) – masculine form, singular – and immediately all used forms of the basic form are shown in Araneum, including also the derived lexemes with the same root (*-roj-*)[4],

it is the corpus we work with also in our main VEGA project.

The corpus Araneum counts with 32 languages actually, but the number of the languages available in the corpus increases each year, as new languages appear. It contains also variants of English, for example, the Asian or African English, or the African or Canadian French. In the case of Spanish, as we already mentioned above, it

[2] *Diccionario de la Real Academia Española - The Spanish Royal Academy Dictionary*
Diccionario de uso de María Moliner - Dictionary of Usage of María Moliner
Španielsko-slovenský a slovensko-španielsky frazeologický slovník - Spanish-Slovak and Slovak-Spanish Idiomatic Dictionary

[3] We were interested in all naming units with the red colour in Spanish, without taking into account whether some of them are limited to the European standard or the American one, which would be even more complicated according to the indivudual countries in some cases.

[4] Comparably, dictionaries usually list pluriverbal expressions only in the infinitive form and in the active voice, with the open subject position which can be filled mostly by an agent (Fellbaum, 2019: 737).

does not distinguish between the European and the American variants yet, what we consider an advantage for us for our purposes. Similarly, the majority of the Spanish language dictionaries has also the panhispanic character. The basic size of a corpus for each language is *Maius* (1.2 milliards of tokens, about 1 milliard of words). We were working with the version *Minus*, what is a 10% random sample of the Maius version. They are more or less identical in content, texts are "put" in there through the process of the "general crowling"; it means the corpora contain genres typical for the Internet.

Taking into account the mentioned differences between the electronic corpora and the traditional dictionaries, we were led to suppose and therefore to build the following hypothesis:

H1: In the electronic corpus Aranea there are considerably more naming units than in the traditional dictionaries.

H2: In the electronic corpus appear only the current naming units, meanwhile the traditional dictionaries contain also archaic and paremiological units.

H3: In the electronic corpus appear more mono-, bi- or triverbal constructions, meanwhile in the dictionaries there are more whole phrase constructions.

Firstly, through the method of selection and the contextual analysis we retrieved all units with the red colour from the dictionaries, then from the corpus Araneum and the results were crosschecked and compared and contrasted to find the common and the different ones. In the Araneum corpus we were searching by the base form *rojo* and all possible combinations and usages with this lexeme were shown to us. In this phase we were able to determine the character of the given lexical chromatic units in concrete contexts and, again through the method of selection, select only those which were suitable for our research, meeting at least one of the above predefined criteria. Regarding our case, we found 13 393 examples in the part Minus and even 120 073 examples in the part Maius. We were utilizing the analytical method in the next stages of our research. We were working only with the Minus part, as by the frequency of 2000 the contexts of the usage of the word *rojo* started to repeat and only hardly we crossed with a new context. In spite of this situation, we listed all the 13 393 displayed results and chose only one example of each usage for our research data sample. Altogether, we obtained 139 expressions from the electronic corpus and 48 expressions from the dictionaries. Then we put together the common appearances and usages from both sources and compared the different ones. Within an additional step we checked also those units that we had found in the dictionaries but not in Araneum in the way that we taped the complete expression, for instance, *rojo como la amapola* (literally *red as the red poppy*) in the Maius part. After the necessary analysis were done, we proceeded to the synthetic method to be able to make generalizations and conclusions.

Usage and symbolism of the red colour
According to our previous researches realized in relation to our present topic (Kvapil, 2012, Spisiakova, 2016, Mockova, 2017, 2018), in this chapter we try to summarize all possible meanings and usage of the red colour in the Spanish language, taking into account how many of them can be found in dictionaries and how many in Araneum corpus.

According to J. Smiesny (2010), the red is one of the most popular colours, present in almost every culture. The red is considered to be the colour of happiness in the main part of Asia, especially in China, symbolizing success and good perspective for the future. The red is also the colour of love and passion. Teachers usually correct the written works of their students with a red ink. The red stands in opposition to the green, representing safe situation or correctness of something, whereas the red

symbolizes danger or prohibition (Kvapil, 2012). The following situations testify the rich usage and a pretty high frequency of the red colour. In the majority of the cases, the colour did not stayed in the extralinguistic reality, but did passed also to the language since "all phraseologically related meanings are connotative, that is, they reflect images of the real world indirectly. They verbalize images with lexemes, adding some personal emotions and assessments and including figurative representations" (Kuprieva et al, 2019: 36):

the blood is red what gave origin to naming units like *cuerpo rojo* (red corpuscle)
red symbolizes anger – an angry person has a red face, e. g. *estar rojo como un tomate* (to be red as a tomato)
in Christianity the red is the symbol of the bloodshed of the martyrs
historically, red or purple were considered royal or ceremonial colours
red symbolizes love, passion, erotics (*red rose, red St. Valentine´s heart*)
it symbolizes the socialistic parties, mostly communistic *(red party)*
in book-keeping, depth or loss were/are highlighted with red colour (*red numbers*)
the *Red Cross* – international voluntary health organization
the *red Khmeres* – originally a rebellious guerrilla units in the Cambodia camps, later, since 1960, name for the communistic party in Cambodia
the *Red Army* of the former Soviet Union
red flag with multiple meanings: 1. in the History as a symbol of defiance, 2. in the time of the French Revolution symbol of patriotism, 3. after the Great October Socialist Revolution the Soviet government adopted the red flag with hammer and grass-hook as their national flag (Kvapil, 2012), 4. the red flag as a warranty symbol of swimming in a sea (Spisiakova, 2016)
the lexical units in Spanish as *alerta roja* (emergency), *teléfono rojo* (hot line) or *números rojos* (red numbers) belong to the common semantic field of EMERGENCY (Mockova, 2017)
red signalizes the change of the colour of the human skin (especially of the face or parts of it) due to the influence of strong negative emotions, mostly anger, but also shame or cold, what led to numerous comparative constructions like: "rojo + como + noun" or "más + rojo + que + noun" (Mocková, 2017, 2018)
red means "stop" in traffic signals
red colouring as symptom of a plenty of diseases, e.g. *red eyes, red throat, red pickles or spots on the skin*, etc.

There are various articles and studies in French language as well, dealing with the usage of the red colour in the current French language. There are several expressions which have the same form and meaning in French, in Spanish and even in Slovak. An example can be: **être rouge comme une tomate/estar rojo como un tomate/byť červený ako paradajka** (**to be as red as a tomato**). On the other hand, there are expressions existing only in the French language, like: **être la lantern rouge** with the meaning of being somebody the last one; or **tirer á boulet rouges** meaning to criticise roughly somebody. Some of these multiword units would have similar equivalents in the Spanish language, also containing a chromatic element, although not the red one (**poner verde a alguien** in Spanish, where the same meaning is built by the green colour). Some of the French-speaking linguists dealing with this topic, that means, the usage of the red colour in language, we can mention, for instance, Claude Thomasset (2006) alebo Fritz Blakolmer (2017).

In relation to the sole adjective – *red*, it means *rojo* in Spanish – except for this basic form there exist the whole scale of other adjectives derived from etymologically other base, declaring the rich presence of this colour in the Spanish language. To illustrate

the situation, we present the examples from the dictionary *Diccionario de uso de María Moliner*:

adjectives with the lexical base -roj-: *rojo, enrojecido, rojal, rojeante, rojeto, rojizo*
adjective with the lexical base carm-: *carmesí, carmín, carminoso*
adjectives with the lexical base -purpur-: *purpúreo, purpurino, empurpurado*
adjectives with the lexical base ros-: *rosa, rosáceo, rosado, róseo*
adjectives with the lexical base -cobr-: *cobreño, cobrizo, encobrado*
other adjectives: *aborrachado, abrasilado, almagrado, aloque, ardiente, arrebolado, bermejo, brasilado, choco, coccíneo, colorado, coralino, encarnadino, encarnado, encendido, eritreo, granate, grancé, gules, ígneo, lacre, paco, rodeno, vino*.

The same happens also in French: in every dictionary we can find the basic lexeme *rouge*, but, beside it, there are lexemes like *sanglant* or *sanguine*, which derive originally from Latin, referring to the red colour of the blood.

However, in our present paper we deal with the expressions having as the chromatic element the word *rojo* exclusively, not to complicate the matter, as taking into account all forms with the same meaning would result into a much deeper and much more extensive research.

Analysis

In the electronic corpus Araneum we found together 139 contexts in which the red colour can be used, meanwhile in dictionaries we could find a comparably lower number – only 48. Firstly, regarding the common naming units, there were 26 expressions which could be found as in dictionaries as in the electronic corpus. Secondly, speaking of the different ones, 22 units could be found only in dictionaries and 113 only in the corpus. This situation is well displayed in the diagram below:

■ Araneum   ■ dictionaries   ■ common units   ■
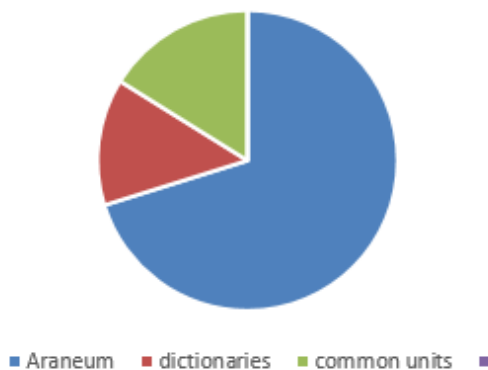
Figure 2. Distribution of the units in dictionaries and in *Araneum*

As we can see in the diagram above, we were able to divide our gathered data into three main categories. First, we introduce lexical units existing only in the traditional dictionaries:

poner rojo a alguien – avengonzarle (lit. make sb. ashamed): in Araneum we found the same expression *poner rojo a alguien*, but not in the context of being ashamed but rather in the context of envy.

ser/estar/quedarse rojo como (lit. to be[5]/become red as): in Araneum none of these verbs appeared in the meaning of comparing something to the red colour.

rojo como la acerola[6]/la amapola/la grana/el guindo/un tomate/una brasa/un tomate (lit. red as the acerola/the red poppy/the carmine/the sour cherry/a tomato/an amber/a cancer): from all these comparative constructions found in dictionaries, in Araneum the only comparison in common were *rojo como un tomate* and *rojo como un cáncer* (red as a tomato/a cancer).

más rojo que el fuego/el sol/una grana/la lumbre/un pimiento/un tomate (more red than the fire/the sun/the carmine/ the flame/a pepper/a tomato): in Araneum we did not find the above mentioned dictionary examples, however, we found the comparative construction "more red than + noun" with new, other nouns, as seen in the following expressions:

que aquella sangre (lit. than that blood)

que la gamba (lit. than prawn)

que la antigua bandera de la U.R.S.S. (lit. than the former Soviet Union flag)

rojo alambrado (lit. flaming red)

coco rojo, consuelda roja, cuervo de cabeza roja, gusano de sangre roja, oligisto rojo, pudrición roja: constructions "noun + (preposition) + adjective" belong to the area of specific terms from Botany, Zoology and Geology.

plata roja (lit. red silver): actually, it means a ruby-colour mineral; in Araneum we crossed with the unit *oro rojo* (lit. red gold).

Secondly, within the category of naming units existing indifferently in dictionaries and in the Araneum corpus, we devided them into 3 subcategories:

the same form and meaning of lexemes,

the same form of lexemes but with different meaning,

the same form of lexemes but one of their meanings is different or extended.

Hereinafter, we can see the complete result as it continues:

a) The same form and meaning of the lexemes:

alerta roja (the highest level of emergency)

piel roja (red-skinned man – North American Indian)

teléfojo rojo (hotline)

glóbulos rojos (red corpuscle)

libro rojo (red book[7])

ponerse rojo (to become red)

al rojo blanco (very bright red, almost white, under the influence of a very high temperature)

al rojo cereza (dark red colour under the influence of a very high temperature)

agua roja (very hot water)

farolillo rojo (lit. red lighthouse – the last person in a competition)

ocre rojo, plomo rojo, sándalo rojo, fósforo rojo, palo rojo, tuna roja (specific terms of various scientific disciplines)

rojo como un tomate/un cangrejo (red as a tomato/cancer)

b) The same form of the lexemes but with different meaning:

We came across only with one single expression – **lápiz rojo**, which turned out to be interesant from the point of view of its definition, varying its definition in the DRAE

---

[5] In Spanish there are two verbs expressing the English meaning of *to be*, which are *ser* and *estar* and their usage is bent to grammatical rules.

[6] A tropical fruit called also *Barbados cherry*.

[7] In Spanish it reffers to diplomatic documents published by the government to inform the legislative organs or the public.

dictionary from its usage in the Araneum corpus. Meanwhile the DRAE dictionary explains this unit leterally under another lemma *almagre* as "red earth colour", in the electronic corpus we can find multiple examples of its meaning as metaphor of censorship or prohibition: *el lápiz rojo de la censura* – the red pencil of censorship, *aplicar el lápiz rojo a políticas europeas* – to apply the red pencil to the European politics, *pasar el lápiz rojo* – to deface something with a red pencil, etc.

c) The same form of the lexemes but one of their meanings is different or extended:
al rojo vivo (aching, flaming): dictionaries mention three meanings of this constructions, the same as the electronic corpus, but, interesantly, only two of them are the same. The other dictionary meaning (aching due to shame) we did not discovered in Araneum. On the other hand, Araneum includes a new meaning of this unit, not existing in dictionaries, which is the name of a famous and successfull TV program called *Al rojo vivo*.
números rojos (red numbers): in dictionaries there we can check only one meaning, which is that something/somebody is in red numbers, meaning "to be in loss". On the other side, in the corpus we discovered that this construction has extended the part linguistically called *comparandum*, as seen in the following resulting constructions: for instance, *rectas rojas de las estadísticas* (red finishing point), *cuentas rojas* (red accounts), *facturas rojas* (red invoices), *contexto rojo* (red context). Or, an another observation we made at this point was that certain formal change had taken place, as in the following examples: *Apple en rojo* (Apple in red), *saldo en rojo* (balance in red). Moreover, in the electronic corpus this construction appears also in the meaning af a bad situation, not just finantial, for example, bad educational results.
rojo como (red as): in dictionaries there existed 7 comparisons, in the Araneum corpus there were 19 similar comparisons, always with another *comparatum*; as, for instance: *rojo como la guindilla* (hot pepper), *el pimiento* (pepper), *la gamba* (prawn), *el fuego* (fire), *la lava* (lava), *la sangre* (blood), atd'. But, on the other side, new and interesting comparisons arose: *rojo como las banderas sindicalistas* (red as the union flags), *como un hígado mal cocido* (as a badly cooked liver), *ponerse roja como Heidi* (to become red as Heidi). In the last example, from the grammatical point of view, only the feminine gender of the adjective exists (*roja*), as it reffers to a feminine literary character (*Heidi*).
ponerse rojo (to turn red): in the sence of becoming physically red (face, cheeks, etc.), overhelmingly because of being ashamed. This meaning was common to dictionaries and to the corpus. Moreover, in the electronic corpus we found also the expression *ponerse rojo de rabia* in the sense of becoming red due to anger.

The third main group consists of naming units which showed to be used in the Araneum corpus exclusively:
the majority of the naming units discovered in the Araneum corpus were collocations. The most frequent construction was "noun + adjective", where the majority of them were terms from Botany and Zoology, as, for instance: **frutos rojos** (red fruits), **pimiento rojo** (red pepper), **alubias rojas** (red bean) **maíz rojo** (red corn), **atún rojo** (red tuna), **araña roja** (red spider), **ciervo rojo** (red deer), **gamba roja** (red prawn), **lince rojo** (bay lynx), etc.
similarly frequent were geological terms or names of different materials, as, for example: **ladrillo rojo** (red brick), **oro rojo** (red gold), **madera roja** (red wood), **granito rojo** (red granite), **mármol rojo** (red marble), **tierra roja** (red earth), **marea roja** (red float), etc.
lexicalized units having the form "noun + adjective" not restricted to a specific scientific terms: **crónica roja** (red chronicle), **páginas/notas rojas** (red pages/notes),

la lista roja (red list), **el planeta rojo** (red planet), **Jemeres rojos** (Red Khmers), **tarjeta roja** (red card), **alfombra roja** (red carpet), etc.

among the mentioned scheme "noun + red" we could also discover some interesting, new or ocassional naming units, like: **Gorila roja** (Red gorilla → Hugo Chávez), **scudería/escuadra roja** (red racing stable → Ferrari), **la operadora roja** (red operator → Vodafone), **biotecnología roja** (red biotechnology).

we discovered two collocations reffering to the meaning of blood or death: **muerte roja** (red death), **rastro rojo** (red trail → blood).

in the corpus Araneum, there appeared expressions reffering to the red colour in connection with communists in various forms: **los rojos** (the red [ones]) → **los rojillos** (ironically – the diminutive of it), **Internacional Sindical Roja** (International Red Unions), **estrella roja** (red star → the Russian communistic star ~ the Che Guevarra´s star), **brigadas rojas** (red brigades), **Ejército Rojo** (Red Army), **camisas rojas** (Red Shirts), **ponchos rojos** (red ponchos), etc..

la Roja as the naming unit for the Spanish football representation; the same in a complex construction like*: la furia/la locura roja* (the red craziness/folly).

in the Spanish language the surname "**Roja**s" derived from the colour is very frequent.

Conclusions

By the time of resuming our work, we can assume that, firstly, a comparably higher number of expressions with the element of red colour can be found in the electronic corpus than in the traditional dictionaries; a fact that showed us in relation 139 : 48 units. That means, that we investigated alltogether 187 expressions containing the red colour in their structure.

Secondly, from the formal point of view, from the most frequent types of constructions, one of the most productive was "rojo + como + noun", eventually, "más + rojo + que", where, once again, more of them were present in the electronic corpus *Araneum* than in all the studied dictionaries (22 : 7). At the same time, a very similar result was shown with the lexicalized units with the formal pattern "noun + red", where a much more higher number of them could be found again in the corpus, in comparison to the dictionaries (43 : 9). At this point we did not take into account the biological, geographical, medical, etc. terms, only the commonly used expressions like *alfombra roja, piel roja, alerta roja*, etc. Regarding the mentioned terminology, more of those expressions were present in dictionaries, mainly having the character of some technical and not generally known terms (*coco rojo, consuelda roja, fósforo rojo*, etc.), whereas in the corpus the common expressions of everyday language were more frequent, paralely with the generally known terms like *té rojo, red fruits, red pepper*, etc.

Thirdly, from the semantic point of view, we can say that in the electronic corpus we discovered many cases of extention of the meaning of certain units; what means, that their meaning (but sometimes also their form) suffers changes, as a natural result of the constant and dynamic process of the language evolution caused by its usage by the members of this linguistic community. In other words, we came across few new forms with new meaning, which could be characterized as neologisms or ocassionalisms. On the other side, the traditional dictionaries do not reflect the very recent stage of the contemporary Spanish language, they are not able to react and to process all the changes so dynamically and quickly. Although we mentioned within our work that one of the meanings of the red colour was its connection to the royal symbols, surprisingly, neighter in the dictionaries, nor in the corpus we found an example of this accurate meaning.

Thanks to the above described results we can now conclude at the end of our paper that in the electronic corpus *Araneum* more expressions can be found than in the traditional lexicographic sources. The difference was 113 expressions. That means

that **our first hypothesis H1 has been confirmed.** Paralely, in the electronic corpus only the current units were found, whereas the dictionaries cited also some of the archaic expressions, what means that also **our second hypothesis H2 has been confirmed.** Lastly, as in *Araneum* as in the dictionaries, the comparative construction were abundant. Simultaneously, *Araneum* presented overwhelmingly the simple constructions (like "noun + red", which was the most frequent one) – as we had supposed – but, on the other side, the lexicographic sources did not present the whole phrased constructions, nor any paremiological expressions – as we had supposed as well – only some verbal constructions (like *ponerse rojo*). That is why we cannot speak about confirmation of **our third hypothesis H3**, but we rather conclude that it **has not been confirmed.**

Bibliographic references
AHOLA, J. 2011. A Corpus-Driven Approach to the Infinitival Complementation Patterns of the Verb HELP in Recent Centuries [Pro Gradu Thesis]. Tampere: University of Tampere.
BENKO, V. – BUTASOVA, A. 2017. Teaching corpus linguistics with Aranea web corpora. In: Trudy mezdunarodnoj konferencii "Korpusnaja lingvistika - 2017", pp. 16-21. ISSN 2412-9623.
BENKO, V. 2014. Aranea: Yet Another Family of (Comparable) Web Corpora. In: Petr Sojka, Ales Horak, Ivan Kopecek and Karel Pala (Eds.): Text, Speech and Dialogue. 17th International Conference, TSD 2014, pp. 257-264. LNCS 8655. ISBN 978-3030007935.
BLAKOLMER, F. 2017. La couleur rouge dans la symbolique, l'art et le langage du monde egeen de l'age du Bronze. In : Trivium, Revue franco-allemand de sciences humianes et sociales, 27, Available online: http://journals.openedition.org/trivium/5572. ISSN 1963-1820
CERMAK, F. 2000. Jazykovy korpus: prostredek a zdroj poznani. In: Studie z korpusove lingvistiky. Acta Universitatis Carolinae Philologica 3 – 4, 1997, pp. 15-37. ISSN 0567-8269.
FELLBAUM, C. 2019. How flexible are idioms? A corpus-based study. In: Linguistics, vol. 57, n. 4, pp. 735-767. ISSN 1613-396X.
HOFFMANN, S. et al. 2008. Corpus Linguistics with BNC web - a Practical Guide. Frankfurt am Main: Peter Lang. ISBN 978-3-631-56315-1.
KUPRIEVA, I., A. 2019. et al. Semantic features of the phraseological units with the component light. In: Humanities and Social Reviews, vol. 7, n. 4, pp. 33-38. ISSN 2395-6518.
KVAPIL, R. 2012. Farba v komunikacnom procese, In: FORLANG - Cudzie jazyky v akademickom prostredi, pp.333-340. ISBN 978–80–553–0982–8.
LINDQVIST, H. 2009. Corpus Linguistics and the Description of English. Edinburgh: Edinburgh University Press. ISBN 978-14-744-2172-0.
MOCKOVA, N. 2018. Chromatizmy cervenej farby v slovencine a spanielcine. In: Lingua et vita 14, VII/2, Ekonom, Bratislava. pp. 31-41.
MOCKOVA, N. 2017. Unidades fraseologicas con nombres de colores en su estructura [dizertacna praca]. Bratislava: Univerzita Komenskeho v Bratislave.
MOLINER, M. 2007. Diccionario de uso del español, 3ª edicion. Madrid: Gredos. ISBN 978-84-249-2886-5.
REAL ACADEMIA ESPAÑOLA. 2014. Diccionario de la lengua española. 23.ª ed. [En linea]. Madrid: Espasa Calpe, S. A.

SIMKOVA, M. 2008. Korpusova lingvistika na Slovensku. In: Jazykovedny casopis, vol. 59, n. 1-2, pp. 11-24. ISSN 0021-5597.

SMIESNY, J. 2010. Farba – komunikacny signal nielen pre dopravne krizovatky. In: Financny manazment a controling v praxi, 3. Bratislava: Iura Edition. ISSN 1337-7574.

SPISIAKOVA, M. 2016. The contrastive analysis of the use, meaning and connotations of the colors in the Slovak and Spanish language. [El analisis contrastivo del uso, significado y connotaciones de los colores en la lengua eslovaca y española]. In: XLinguae, vol. 9, n. 3, pp. 104-128. ISSN 1337-8384.

TRUP, L. – BAKYTOVA, J. 2017. Spanielsko-slovensky a slovensko-spanielsky frazeologicky slovnik. Bratislava: Mikula. ISBN 978-80-88814-95-5.

THOMASSET, C. 2006. L'ecriture du texte scientifique. l′Universite Parise-Sorbonne, Paris. ISBN: 2-84050-421-9

VV.AA. ARANEA. A, Family of Comparable Gigaword Web Corpora, Available online: http://aranea.juls.savba.sk/aranea

Words: 6123
Characters: 38 599 (21,44 standard pages)

doc. Mgr. Mária Spišiaková, PhD.
Department of Romance and Slavic Languages
Faculty of Applied Languages
University of Economics in Bratislava
Dolnozemská cesta 1
852 35 Bratislava
Slovakia
maria.spisiakova@euba.sk

Mgr. Nina Mocková, PhD.
Department of Romance and Slavic Languages
Faculty of Applied Languages
University of Economics in Bratislava
Dolnozemská cesta 1
852 35 Bratislava
Slovakia
nina.mockova@euba.sk