

EL COLOR ROJO EN LOS DICCIONARIOS Y EN EL CORPUS ELECTRÓNICO

Mária SPIŠIAKOVÁ, Nina MOCKOVÁ
Universidad de Economía de Bratislava

Abstract (En): The present paper deals with pluriverbal constructions containing the element of red colour in Spanish language (rojo). 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 (En): red; colour; Spanish; lexicalized constructions; dictionaries; electronic corpus

Palabras claves (Es): color; rojo; español; construcciones lexicalizadas; diccionarios; corpus electrónico

Introducción

Hoy en día existen diferentes métodos de búsqueda en el campo de la investigación lingüística. Gracias a las tecnologías modernas, los lingüistas tienen a su disposición nuevas posibilidades y se les abren nuevos horizontes en el campo del estudio de cualquier idioma. Hace unas décadas apenas podíamos imaginar esta amplia gama de posibilidades. Las investigaciones lingüísticas actuales son mucho más dinámicas y el trabajo de los lingüistas es cada vez más de carácter interdisciplinario. La recopilación de datos se realizaba, tradicionalmente, de los diccionarios académicos impresos, de la prensa, de los archivos, se recogían datos directamente en la calle, sobre el terreno, etc. Gracias al Internet y las tecnologías informáticas ahora podemos hacer toda esta investigación desde un ordenador y en un tiempo mucho más corto. En este sentido, uno de los logros más significativos para los lingüistas es la aparición del corpus lingüístico. Pero, ¿qué es lo que se encuentra en un corpus y qué hay en un diccionario? ¿Es lo mismo? ¿Hay diferencias? ¿Y si las hay, cuáles son?

La lingüística de corpus empezó a desarrollarse a finales de los años 60 y principios de los 70 del siglo XX cuando aparecieron los primeros corpus en la lengua inglesa. «Gracias al rápido e incesante desarrollo de las tecnologías de comunicación e informática podemos considerar la lingüística de corpus como una de las disciplinas lingüísticas más dinámicas en la actualidad, formando la parte principal de la lingüística aplicada de la informática»¹ (ČERMÁK, 2000 en ŠIMKOVÁ, 2008:11).

Algunos lingüistas (LINQUIST, 2009, HOFFMAN et al., 2008) tienden a definir la lingüística de corpus más bien como una metodología a través de la cual se puede

¹ Traducción nuestra

realizar la investigación lingüística en diversas áreas del lenguaje. Afirman que, en comparación con las disciplinas lingüísticas autónomas que describen o explican ciertas características del lenguaje, la lingüística de corpus no se centra en características particulares, sino que puede utilizarse para explorar cualquier área del uso del lenguaje. Esto significa que la metodología de la lingüística de corpus puede emplearse en la investigación lexicográfica, lexicológica, morfológica, sintáctica, estilística o semántica, en el análisis del discurso, variación del lenguaje, estudios contrastivos, etc.

En oposición a estas ideas, algunos otros lingüistas (ČERMÁK, 2000, ŠIMKOVÁ, 2008) afirman que la lingüística de corpus es una rama de la lingüística que tiene su propia metodología y objeto de estudio. ČERMÁK (2000) confirma que, gracias a la continua y rápida evolución y desarrollo de las tecnologías de información y comunicación, que desempeñan un papel esencial en la existencia de la lingüística de corpus, podemos decir que es una de las disciplinas lingüísticas modernas con el desarrollo más dinámico. Así, ŠIMKOVÁ (2008) describe el objeto de estudio de la lingüística de corpus: «Investiga y describe la lengua, o sea, su uso (parole) en el contexto real, el cual se busca en el corpus y se elabora por medio del método estadístico»² (ŠIMKOVÁ, 2008:11). En resumen, toda investigación basada en corpus es conocida como lingüística de corpus tanto si la consideramos como un conjunto de métodos o como un campo independiente de la lingüística.

La lingüística de corpus en el sentido moderno comenzó a desarrollarse a principios de la década de los años 80 del siglo pasado, en gran medida, gracias al desarrollo de la tecnología informática. «La base esencial de la existencia de la lingüística de corpus radica en la creación de un corpus en sí mismo, que puede caracterizarse como una base de datos electrónica de textos escritos o hablados que han pasado por un cierto procesamiento del lenguaje informático de acuerdo con las normas lingüísticas estándar válidas de un idioma determinado para el que se diseñó el corpus. Dicha base de datos de textos es objetiva en el sentido de que se ha recopilado de acuerdo con algunos criterios extralingüísticos lo que significa que cualquiera puede acceder a investigaciones previas basadas en el corpus» (AHOLA, 2011:6). Según SINCLAIR (1996), se trata de una muestra representativa del idioma. Por último, mencionamos la descripción del corpus de la Real Academia Española: «Poder trabajar con los corpus electrónicos juega un papel importante en la lingüística moderna ya que ofrece datos empíricos, esenciales para estudiar la escala completa de los fenómenos lingüísticos del idioma actual, al mismo tiempo que ofrece características globales de un idioma en un idioma específico en un período histórico» (RAE, 2014, <https://www.rae.es/recursos/banco-de-datos/crea>).

«Durante la última década, la creación de los corpus electrónicos ha sido reconocida como una herramienta efectiva para obtener datos lingüísticos en situaciones donde la recopilación de los corpus tradicionales resultaría demasiado costosa y lenta» (JAKUBÍČEK et al., 2013 en BENKO, ZAKHAROV, 2016:4245). Por lo tanto, la lingüística de corpus y los corpus electrónicos han reemplazado una gran cantidad de investigaciones individuales sobre el terreno o la investigación a base de los medios impresos, libros, etc. «Una de las principales ventajas de realizar una

² Traducción nuestra

investigación lingüística a través del corpus lingüístico es, sin duda, la posibilidad de una verificación fácil de los resultados obtenidos que pueden repetirse porque el mismo material está a disposición de todos los investigadores lingüísticos (así como de todas las demás personas interesadas) sin restricciones temporales o de área»³ (ŠIMKOVÁ, 2008:12). Con respecto a las definiciones mencionadas, podemos resumir los principales beneficios de la investigación basada en corpus: el procesamiento de datos por ordenador evita errores del factor humano en el análisis y, por lo tanto, dichos datos son fiables. Los datos se contextualizan (muestran el uso del lenguaje en contextos lingüísticos reales), por lo que proporcionan una gran cantidad de ejemplos en contextos reales. En relación con este punto, podemos decir que los datos del corpus son cuantitativos en la forma en que muestran los resultados desde la perspectiva de su frecuencia o estadísticas. Luego, todos los datos pueden procesarse rápidamente y a un costo mínimo (operaciones como la búsqueda, selección, clasificación, categorización, etc.). Finalmente, los grandes corpus se actualizan constantemente, lo que permite a los investigadores ver los cambios recientes en el idioma.

1. Objetivo, metodología, hipótesis

El presente artículo es un estudio parcial de las unidades léxicas. Nos enfocamos en el tema de los cromatismos del color rojo en el idioma español. Este estudio está relacionado con el proyecto VEGA titulado *Chromatizmus a jeho konotácie v kontexte slovanských a románskych jazykov (Cromatismo y sus connotaciones en el contexto de las lenguas eslavas y románicas)*. El objetivo del dicho proyecto es comparar y contrastar las formas, el uso, la apariencia, el significado, etc. de las unidades léxicas con un elemento cromático en su estructura en las lenguas romances y eslavas seleccionadas.

Los objetivos del presente estudio son los siguientes:

- a) encontrar en qué se diferencian las unidades léxicas con el componente cromático rojo en su estructura en el corpus de las unidades en los diccionarios;
- b) analizar las diferencias de uso, estructura, significado y las apariencias de estas unidades léxicas en los diccionarios y en el corpus;
- c) señalar qué unidades se encuentran tanto en los diccionarios como en el corpus, es decir, cuáles son iguales en las dos fuentes.

En nuestra investigación combinamos los métodos filológicos tradicionales como es la investigación lingüística sincrónica con la metodología básica de la investigación de corpus. «El objetivo de la investigación sincrónica de la lengua es la comunicación lingüística. Hay que tener en cuenta que la comunicación lingüística es un componente de la interacción social (las relaciones de las personas en sus actividades) y que la lengua como herramienta de comunicación está determinada por las necesidades de la comunidad lingüística. El estado actual de la investigación de la lengua tanto como su funcionamiento indican que la lingüística sincrónica implica la teoría de la comunicación»⁴ (DOLNÍK, 2013:10). Tenemos en

³ Traducción nuestra

⁴ Traducción nuestra

cuenta que el uso de los colores en el lenguaje es el producto de la sociedad, de su evolución y desarrollo. Por otra parte, según Dolník «La lingüística sincrónica no se interesa solamente por la estructura de la lengua, sino también por los cambios en el sistema. El objetivo de su estudio es también el dinamismo sincrónico de la lengua. Estudia los cambios lingüísticos, las relaciones entre ellos y busca qué tendencias se reflejan en el dinamismo de la lengua estudiada. Esta parte del estudio de la lengua es interesante también desde el punto de vista de la confrontación ya que la búsqueda de las similitudes o diferencias en el dinamismo sincrónico de las lenguas anima a buscar la respuesta a la cuestión qué causa los cambios en el sistema»⁵ (DOLNÍK, 2013:12).

En nuestro estudio nos hemos centrado en buscar las coincidencias y las diferencias en el uso del cromatismo rojo en los diccionarios tradicionales y en los recursos modernos como son los corpus electrónicos. Ya que, efectivamente, aquí se puede reflejar el dinamismo de la lengua y se pueden ver los cambios en el sistema. Partimos de la premisa de Dolník de que la lengua es un fenómeno cultural en el que se conservan y reflejan las interpretaciones del mundo y, a la vez, funciona como una herramienta de la interpretación.

El objetivo de nuestro estudio no es elaborar una teoría o concepto, sino recopilar, ver y comparar las diferencias de la presencia de las unidades léxicas con el elemento cromático rojo en los diccionarios clásicos y en el corpus electrónico. El material recopilado y los resultados del trabajo pueden servir de base y de punto de partida para las investigaciones siguientes sobre el dinamismo de la lengua actual y para formar un concepto teórico sobre las diferencias entre las metodologías tradicionales (diccionarios) y modernas (corpus).

Con respecto al uso de los colores en el lenguaje, y de acuerdo con nuestras investigaciones previas en esta área, podemos decir que el negro y el blanco son los colores más utilizados en el lenguaje. Según SPIŠIAKOVÁ (2016), el rojo, el verde y el azul también son colores frecuentes, aunque no tanto como el negro y el blanco. Estas afirmaciones coinciden con las de Vasilevič: «En las siguientes etapas del desarrollo de la lengua, a los colores negro y blanco se iban añadiendo los nombres para el color rojo, más tarde verde y azul [...]»⁶⁷ (en ULIČNÁ, 2019:51). Algo menos usados y frecuentes son los colores amarillo, marrón, violeta/púrpura, rosa y naranja. El color rojo resulta ser el cuarto color en cuanto a la frecuencia del uso de las unidades léxicas con un elemento cromático en el idioma español. Se han obtenido resultados comparables en el estudio centrado en las unidades fraseológicas con un elemento cromático en español realizado por MOCKOVÁ (2017). Este resultado se puede observar en la siguiente figura:

⁵ Traducción nuestra

⁶ Traducción nuestra

⁷ Traducción nuestra

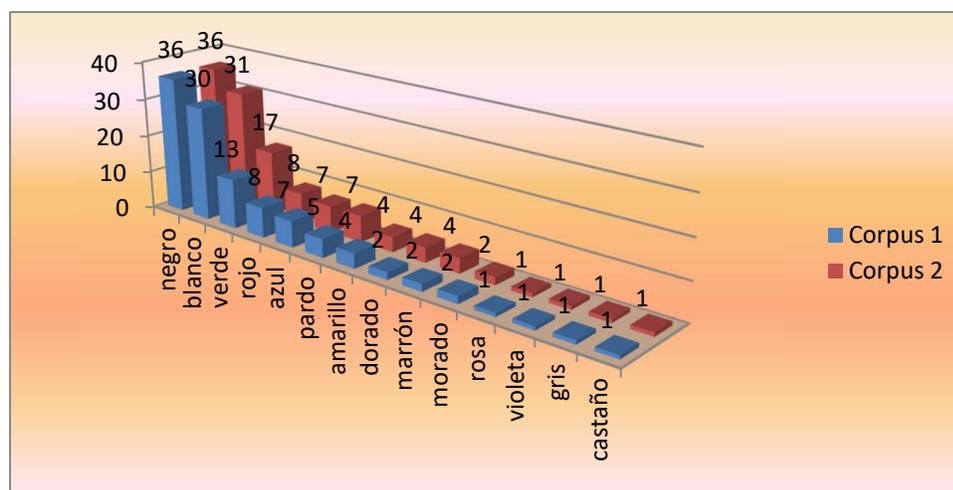


Figura 1 Distribución de colores (MOCKOVÁ, 2017:43)

En la presente investigación hemos incluido todas las unidades léxicas en cuanto a su estructura, es decir, las monoverbales, pluriverbales, sintagmáticas o construcciones de frases completas siempre que la unidad contenga el color rojo como uno de sus componentes. Nos interesa la apariencia y el uso de todas aquellas unidades léxicas con el color rojo en el caso de que cumplan al menos uno de los siguientes criterios:

- 1) la unidad sea figurativa, es decir, que sea evidente la motivación de la denominación;
- 2) sea lexicalizada;
- 3) el color haya perdido su significado original de “ser rojo“ o de un color como tal (por ejemplo, los apellidos);
- 4) sea idiomática y, por lo tanto, literalmente no traducible;
- 5) sea expresiva;
- 6) sean lexemas derivados del rojo (por ejemplo, *sonrojar*).

Las expresiones como *té rojo*, *pieles rojas* o *frutas rojas* también han sido incluidas ya que se trata de expresiones lexicalizadas aunque no sean figurativas. Hemos omitido solo las expresiones donde el color rojo designa el color real de un objeto, por ejemplo, *jersey rojo* o *coche rojo*. Tampoco hemos tomado en cuenta los títulos de las obras literarias, revistas, nombres de los clubs deportivos (p.e. *Tiburones rojos*) y nombres geográficos desconocidos.

En primer lugar, a través del método de selección y de análisis contextual recopilamos todas las unidades con el color rojo en su estructura de los diccionarios y del corpus *Araneum*. Las unidades léxicas encontradas en las dos fuentes (electrónica e impresa) las comparamos; seleccionamos las que se encontraban en los dos sitios, es decir, las que eran comunes en ambas partes y las que eran diferentes. Hemos utilizado los siguientes diccionarios: el *Diccionario de la Real Academia Española* (2014), *Diccionario de uso de María Moliner* (2007), *Diccionario fraseológico del español actual* (2004) *Španielsko-slovenský a slovensko-španielsky frazeologický slovník* (2017) (Diccionario fraseológico

español-eslovaco y eslovaco-español). El corpus recogido de los diccionarios lo comparamos con las expresiones encontradas en el corpus electrónico *Araneum*. Hemos optado por este corpus por varias razones:

- es el corpus que se utiliza en nuestro proyecto, se elabora en Eslovaquia y, a parte del eslovaco, contiene otros 31 idiomas incluyendo sus variantes;
- no distingue entre el español europeo y americano⁸;
- la búsqueda es muy rápida y sencilla, es decir, se escribe el lema *rojo* en masculino singular y el *Araneum* busca todas las acepciones en todas las formas gramaticales.

El autor, Benko, caracteriza el corpus de la siguiente manera: «La primera etapa del Proyecto Aranea se ha dirigido a la creación de una familia de un corpus electrónico [...] para los idiomas que se hablan en Eslovaquia y sus países vecinos, así como para los principales idiomas extranjeros que se enseñan en las universidades eslovacas. Esta etapa está a punto de completarse y la familia Aranea actualmente contiene corpus (generalmente) en dos tamaños, y algunos idiomas también tienen variantes específicas de la región» (BENKO, 2017:16). Los tamaños estándar son: el *Maius* y el *Minus*: el *Maius* contiene unos 1,2 billones de tokens y el *Minus* unos 120 millones. El *Minus* es una muestra aleatoria del 10% del *Maius* y es una herramienta diseñada no solo para los fines educativos universitarios, pero también para las investigaciones lingüísticas (estudios contrastivos) y lexicográficas (mono o bilingües). La composición de los textos abarcados es similar (específica para la web) en cuanto a su extensión, tipología, distribución por géneros y registros (BENKO, 2014:248). El *Araneum Hispanicum* cuenta con todos los textos que han aparecido en forma electrónica en español en el año 2013 que son de diferente índole: artículos periodísticos, revistas en línea, textos literarios en línea, documentos, blogs, etc. La obtención de los datos fuente ha sido realizada gracias a la herramienta SpiderLing, un rastreador web empleado para recolectar los datos de texto de las páginas web. SpiderLing contiene un mecanismo para la codificación de caracteres y otro para la detección del idioma (BENKO, 2014:249).

Bajo estas condiciones podemos decir que el corpus electrónico *Araneum Hispanicum Minus* es lo suficientemente equilibrado y representativo para nuestra investigación. Además, posibilita descargar las versiones fuente de los datos a través del mecanismo de acceso unificado. En la siguiente tabla se puede ver la distribución cuantitativa de los documentos en el corpus por idiomas:

⁸ Nos interesaban todas las unidades léxicas usadas en España o en América.

	Domains	Docs	Tokens	Docs per domain	Tokens per doc
de	80,722	2,332,921	1,200,000,087	28.9	514.4
en	23,968	1,163,007	1,200,048,075	48.5	1031.8
es	22,343	1,439,567	1,049,739,252	64.4	729.2
fr	48,398	1,780,315	1,233,336,202	36.8	692.8
pl	58,338	1,783,411	1,110,120,825	30.6	622.5
ru	37,200	1,034,734	1,216,800,424	27.8	1176.0
sk	33,037	1,724,512	1,200,003,757	52.2	695.9

Tabla 1 Datos descargados, filtrados y normalizados según Benko (2014)

En caso del español, el corpus no distingue, por ahora, entre la variante europea y la americana, lo que es importante y conveniente para nuestra investigación, ya que la mayoría de los diccionarios usados en nuestro estudio son de carácter panhispánico y nos interesaban todas las expresiones posibles con el cromatismo rojo en su estructura.

El *Araneum Hispanicum Minus* (No Sketch Engine versión) es accesible a través del registro gratuito y nuestra búsqueda fue del siguiente modo: buscamos la palabra clave *rojo* y se mostraron todas las combinaciones y usos posibles con este lexema. En esta fase pudimos determinar el significado de las unidades cromáticas léxicas en contextos concretos. Nuevamente, a través del método de selección manual, elegimos solo aquellas que eran adecuadas para nuestra investigación, es decir, las que cumplían al menos con uno de los criterios definidos anteriormente. Encontramos 13 393 ejemplos en *Minus* y 120 073 ejemplos en *Maius*. Utilizamos el método analítico en las siguientes etapas de la investigación. Trabajamos solo con la parte *Minus*, ya que después de casi 2000 palabras, los contextos del uso del rojo comenzaron a repetirse y apenas aparecían nuevos lexemas. A pesar de esta situación, encontramos 13 393 lexemas de los cuales seleccionamos solo un ejemplo de cada uso (un significado) para nuestro propio corpus del trabajo. Los lexemas los seleccionamos manualmente a base de los criterios mencionados anteriormente. No se han usado las herramientas estadísticas del corpus ni otras.

En total, obtuvimos 139 expresiones del corpus electrónico y 48 expresiones de los diccionarios. Luego separamos las expresiones que se encontraban tanto en el corpus como en los diccionarios, es decir, las iguales o comunes para las dos fuentes y comparamos las que eran diferentes. Para completar la búsqueda verificamos una vez más si las expresiones aparecidas solo en los diccionarios no aparecían en la versión *Maius* de forma que buscamos la expresión completa, por ejemplo, *rojo como la amapola*. Después de realizar el análisis necesario, se procedió al método sintético para poder hacer generalizaciones y sacar conclusiones.

Teniendo en cuenta las diferencias entre el corpus electrónico y los diccionarios tradicionales suponemos que:

- H1: En el corpus electrónico se encontrarán más expresiones con el color rojo en su estructura que en los diccionarios.

- H2: En el corpus electrónico se encontrarán solo las expresiones del uso actual y en los diccionarios se encontrarán también las expresiones arcaicas.
- H3: En el corpus electrónico se encontrarán solo las construcciones monoverbales o biverbales y en los diccionarios se encontrarán también expresiones más largas en forma de frases simples o compuestas como dichos, proverbios, comparaciones.

2. Uso y significado del color rojo

De acuerdo con las investigaciones anteriores realizadas en relación con el presente tema (HOFFMANN, 2008, KVAPIL 2012, SPIŠIAKOVÁ, 2016, MOCKOVÁ, 2017) en este capítulo intentamos resumir todas las posibilidades del significado y del uso del color rojo en el idioma español teniendo en cuenta todos los que se puedan encontrar en los diccionarios y en el corpus Araneum. Según SMIEŠNY (2010), el rojo es uno de los colores más populares y está presente en casi todas las culturas. El rojo es el color de la felicidad en gran parte de Asia, especialmente en China, ya que simboliza el éxito y buenas perspectivas para el futuro. «El rojo es el color del amor y la pasión, pero también es el símbolo de la sangre derramada de los mártires en la religión cristiana y por eso es el color de la liturgia.» (KVAPIL, 2012:337). Los profesores generalmente corrigen los trabajos de sus alumnos con tinta roja. El rojo se opone al verde cuando el verde representa una situación segura o la corrección de algo, mientras que el rojo simboliza peligro o prohibición. En la política (española) se opone al color azul ya que con el color rojo se identifican los revolucionarios o republicanos y los azules son los monarquistas, igual que en los señales de tráfico donde el rojo significa stop/prohibido y el color azul es informativo o significa permiso. En la mayoría de los casos, el color no se quedó en las realidades extralingüísticas, sino que también pasó al lenguaje. El uso lingüístico y el significado extralingüístico del color rojo se puede resumir en varios apartados:

- la sangre es roja, lo que dio origen a denominaciones como *glóbulos rojos*;
- el rojo simboliza la ira, una persona enfadada tiene la cara roja (*estar/ponerse rojo como un tomate*);
- en el cristianismo el rojo es el símbolo del derrame de la sangre de los mártires;
- históricamente, el rojo o el morado se consideraban colores de la realeza o de las ceremonias;
- simboliza amor, pasión, erotismo (*rosa roja, corazón rojo de San Valentín*);
- simboliza los partidos socialistas, en su mayoría comunistas (*partido rojo*);
- en la contabilidad, la pérdida se resalta con el color rojo (*números rojos*);
- la organización internacional voluntaria de salud se llama *Cruz Roja*;

- *los Jemeres Rojos* era, originalmente, la denominación de una unidad guerrillera rebelde en los campamentos de Camboya, más tarde, desde 1960, era el nombre del partido comunista en Camboya;
- *El Ejército Rojo* de la antigua Unión Soviética;
- la *bandera roja* tiene múltiples significados: 1. «en la historia – símbolo de desafío, 2. en la época de la Revolución Francesa – símbolo del patriotismo, 3. después de la Gran Revolución Socialista de Octubre, el gobierno soviético adoptó la bandera roja con el martillo y la hoz como la bandera nacional» (KVAPIL, 2012:337), 4. símbolo de peligro de nadar en el mar;
- «las unidades léxicas en español como *alerta roja* (emergencia), *teléfono rojo* (línea directa) o *números rojos* pertenecen al campo semántico común de alerta.» (MOCKOVÁ, 2017:102-103);
- «el rojo señala el cambio del color de la piel humana (especialmente de la cara o partes de ella) debido a la influencia de fuertes emociones negativas, principalmente ira, pero también vergüenza o frío, lo que condujo a numerosas construcciones comparativas: “rojo + como + sustantivo” o “más + rojo + que + sustantivo”» (MOCKOVÁ, 2018:36);
- el rojo significa *alto/pare* en las señales de tráfico;
- el color rojo describe los síntomas de muchas enfermedades, p. ej. *ojos rojos*, *garganta roja*, *manchas rojas en la piel*, etc.

En cuanto al adjetivo *rojo* en español, existe, excepto esta forma básica, toda la escala de otros adjetivos derivados de diferente base etimológica lo que demuestra una importante presencia de este color en el idioma español. A continuación presentamos los ejemplos del Diccionario de Uso de María MOLINER (2007):

- adjetivos con la base léxica roj-: *rojo*, *enrojecido*, *rojal*, *rojeante*, *rojeto*, *rojizo*;
- adjetivo con la base léxica carm-: *carmesí*, *carmin*, *carminoso*;
- adjetivos con la base léxica purpur-: *purpúreo*, *purpurino*, *empurpurado*;
- adjetivos con la base léxica ros-: *rosa*, *rosáceo*, *rosado*, *róseo*;
- adjetivos con la base léxica cobr-: *cobreño*, *cobrizo*, *encobrado*;
- otros adjetivos: *aborrachado*, *abrasilado*, *almagrado*, *aloque*, *ardiente*, *arrebolado*, *bermejo*, *brasilado*, *choco*, *coccíneo*, *colorado*, *coralino*, *encarnadino*, *encarnado*, *encendido*, *eritreo*, *granate*, *grancé*, *gules*, *ígneo*, *lacre*, *paco*, *rodeno*, *vino*.

Sin embargo, en nuestro trabajo tratamos exclusivamente las expresiones con el elemento cromático, *rojo* teniendo en cuenta que todas las formas con el mismo significado darían como resultado una investigación mucho más profunda y mucho más extensa.

3. Análisis

En el corpus electrónico *Araneum* hemos encontrado 139 contextos diferentes en los que se puede usar el color rojo, mientras que en los diccionarios hemos encontrado un número mucho más bajo, solo 48 contextos. En primer lugar, con respecto a las expresiones comunes en las dos fuentes, hay 26 expresiones que se podían encontrar tanto en los diccionarios, como en el corpus electrónico. En segundo lugar, en cuanto a las expresiones diferentes, en los diccionarios se han encontrado solo 22 unidades, mientras que en el corpus 113. Esta situación se muestra bien en la siguiente figura:

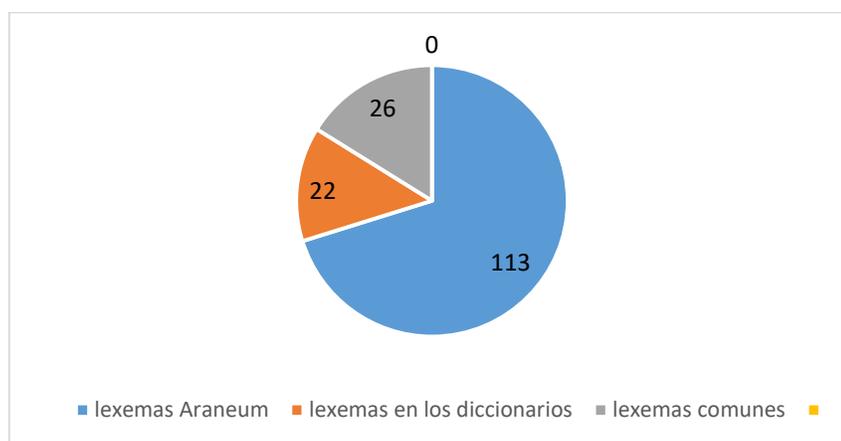


Figura 2 Distribución de las unidades léxicas en los diccionarios y en el corpus *Araneum*

Como podemos ver en la figura 2, las expresiones encontradas las hemos dividido en tres grupos básicos que vienen descritos y analizados a continuación.

El primer grupo está formado por los lexemas (22 en total) que solo se encuentran en los diccionarios tradicionales:

- ***poner rojo a alguien*** (avergonzar a alguien): en Araneum encontramos la misma expresión *poner rojo a alguien*, pero no en el contexto de estar avergonzado, sino en el contexto relacionado con la envidia;
- ***ser/estar/quedarse rojo como***: ninguno de estos verbos apareció en el Araneum;
- ***rojo como la acerola/la amapola/la grana/el guindo/un tomate/una brasa/un tomate***: de todas las construcciones comparativas encontradas en los diccionarios, las únicas comparaciones comunes en las dos fuentes fueron: *rojo como un tomate* y *rojo como un cáncer*;
- ***más rojo que el fuego/el sol/una grana/la lumbre/un pimiento/un tomate***: la construcción comparativa *más rojo que* la hemos encontrado en el Araneum con elementos diferentes en la parte llamada comparatum:
 - o *que aquella sangre*
 - o *que la gamba*

- que la antigua bandera de la U.R.S.S.;
- **rojo alambrado**
- **coco rojo, consuelda roja, cuervo de cabeza roja, gusano de sangre roja, oligisto rojo, pudrición roja**: son términos específicos de la botánica, zoología y geología formadas por la construcción «sustantivo + (preposición) + adjetivo»;
- **plata roja** (un mineral de color rubí): en Araneum hemos encontrado *oro rojo*.

El segundo grupo consta de los lexemas comunes (26 en total), es decir, las unidades que se encuentran tanto en los diccionarios como en el Araneum y las hemos dividido en 3 subcategorías:

- a) lexemas que tienen la misma forma y el mismo significado en las dos fuentes;
 - b) lexemas que tienen la misma forma, pero diferente significado;
 - c) lexemas que tienen la misma forma, pero uno de sus significados es diferente o ampliado.
- a) Lexemas que tienen la misma forma y el mismo significado:
- **alerta roja**: situación de peligro;
 - **piel roja**: nombre aplicado a los indios de América del Norte;
 - **teléfono rojo**: línea telefónica especial destinada al intercambio de informaciones urgentes entre las autoridades políticas de distintos países y entre otras personas con cargos de responsabilidad;
 - **glóbulos rojos**: tipo de célula sanguínea;
 - **libro rojo**: documentos diplomáticos publicados por el gobierno con el fin de informar a los órganos legislativos o al pueblo;
 - **ponerse rojo**: ponerse colorado por algún sentimiento (de vergüenza, ira...);
 - **al rojo blanco**: se aplica al estado de una materia, particularmente un metal, incandescente y a tan elevada temperatura que el color rojo se vuelve blanco;
 - **al rojo cereza**: se aplica al estado de la materia, particularmente del metal, incandescente, cuando tiene color rojo oscuro;
 - **agua roja**: agua muy caliente, hirviendo;
 - **farolillo rojo**: persona, equipo o grupo que ocupa el último puesto, especialmente en una competición;
 - **ocre rojo, plomo rojo, sándalo rojo, fósforo rojo, palo rojo, tuna roja**: términos específicos de diferentes áreas científicas;
 - **rojo como un tomate/un cangrejo**.

b) Lexemas que tienen la misma forma, pero diferente significado:

En esta subcategoría se encuentra solo una expresión: **lápiz rojo** que resultó ser interesante desde el punto de vista de su significado. El diccionario de la RAE explica esta unidad con un sinónimo que es almagre, mientras tanto, en el corpus electrónico podemos encontrar múltiples ejemplos de su significado como metáfora de censura o prohibición: *el lápiz rojo de la censura, aplicar el lápiz rojo a las políticas europeas, pasar el lápiz rojo a algo*, etc.

c) Lexemas que tienen la misma forma, pero uno de los significados es diferente o ampliado:

- **al rojo vivo**: los diccionarios mencionan tres significados de esta construcción igual que el corpus electrónico, pero solo dos de ellos son iguales. El tercer significado del diccionario (*color debido a la vergüenza*) no se encuentra en el Araneum. Por otro lado, el Araneum incluye un nuevo significado que no existe en los diccionarios y se trata de un famoso y exitoso programa de televisión llamado *Al rojo vivo* en el canal televisivo la Sexta.

- **números rojos**: en los diccionarios aparece solo el significado de que algo o alguien está en números rojos, es decir, tiene pérdidas. Mientras tanto, en el corpus esta construcción ha ampliado el horizonte colocacional, como podemos ver en los siguientes ejemplos: *rectas rojas de las estadísticas, cuentas rojas, facturas, contexto rojo* o se ha producido un cambio formal como en los siguientes ejemplos: *Apple en rojo, saldo en rojo*. Además, en el corpus electrónico esta construcción aparece también en el contexto de una mala situación no solo financiera, sino también en el contexto de la educación.

- **rojo como**: en los diccionarios aparecen 7 comparaciones mientras que en el Araneum 19, siempre con un comparatum diferente, por ejemplo: *rojo como la guindilla, el pimiento, la gamba, el fuego, la lava, la sangre*, etc. Hemos encontrado unas comparaciones nuevas, ocasionales como: *ponerse roja como Heidi, rojo como las banderas sindicalistas, como un hígado mal cocido*.

- **ponerse rojo**: en el sentido de volverse físicamente rojo (cara, mejillas, etc.) debido a la vergüenza. Este significado es común en los diccionarios y en el corpus. Además, en el corpus electrónico hemos encontrado también la expresión *ponerse rojo de rabia* en el sentido de volverse rojo debido a la ira.

El tercer grupo, el más numeroso (113 unidades en total), es el grupo formado por las expresiones encontradas solo en el corpus Araneum:

- la mayoría de las unidades léxicas del Araneum son colocaciones. La construcción más frecuente es «sustantivo + adjetivo». Muchas de ellas son términos de la botánica o zoología, como, por ejemplo: *frutos rojos, pimiento rojo, alubias rojas, maíz rojo, atún rojo, araña roja, ciervo rojo, gamba roja, lince rojo*, etc.

- también son frecuentes los términos geológicos o nombres de diferentes materias, como, por ejemplo: *ladrillo rojo, oro rojo, madera roja, granito rojo, mármol rojo, tierra roja, marea roja (flotador rojo)*, etc.

- las unidades lexicalizadas que tienen la forma de «sustantivo + adjetivo»: *crónica roja, páginas/notas rojas, la lista roja, el planeta rojo, Jemeres Rojos, tarjeta roja, alfombra roja*, etc.
- dentro de la estructura de las unidades formadas por un «sustantivo + adjetivo» hemos encontrado denominaciones nuevas u ocasionales interesantes como: *Gorila roja* (Hugo Chávez), *escudería/escuadra roja* (Ferrari), *la operadora roja* (Vodafone), *biotecnología roja*.
- se han encontrado dos colocaciones que se refieren a la sangre o muerte: *muerte roja, rastro rojo*;
- en el corpus Araneum también aparecen expresiones que se refieren a los comunistas o republicanos: *los rojos, los rojillos* (irónicamente), *Internacional Sindical Roja, estrella roja* (la estrella comunista rusa o la estrella de Che Guevara), *brigadas rojas, Ejército Rojo, camisas rojas, ponchos rojos*, etc.
- *la Roja*: el equipo de fútbol de la representación española. Aparece también en una construcción ampliada: *la furia/la locura roja*.
- el apellido *Rojas*: es un apellido bastante frecuente en español.

Conclusiones

Para concluir nuestro trabajo podemos decir que, en primer lugar, se ha encontrado un número de expresiones con el elemento cromático del color rojo más elevado en el corpus electrónico *Araneum* que en los diccionarios tradicionales. En total, hemos analizado 187 expresiones que contienen el color rojo en su estructura, de las cuales en el corpus *Araneum* se han encontrado 139 construcciones, mientras que en los diccionarios tan solo 48.

En segundo lugar, desde el punto de vista formal, las construcciones más frecuentes y más productivas han sido las comparativas de tipo: «rojo + como + sustantivo», eventualmente, «más + rojo + que», de las cuales se encontraban más en el corpus electrónico *Araneum* que en los diccionarios estudiados (22 vs. 7). Al mismo tiempo, se ha visto un resultado muy similar con las construcciones de «sustantivo + rojo», de las cuales en el corpus electrónico se ha encontrado un número más elevado que en los diccionarios (43 vs. 9). Con respecto a los términos específicos, estos se encontraban más en los diccionarios que en el corpus, principalmente teniendo el carácter de términos técnicos y generalmente desconocidos (*coco rojo, consuelda roja, fósforo rojo*, etc.), mientras que en el corpus eran más frecuentes las expresiones comunes del lenguaje cotidiano junto con los términos generalmente conocidos como *té rojo, frutas rojas, pimiento rojo*, etc.

En tercer lugar, desde el punto de vista semántico, podemos decir que en el corpus electrónico se hallan muchas unidades léxicas con el significado ampliado o modificado; lo que señala que su significado (pero a veces también su forma) han sufrido cambios como resultado natural del proceso constante y dinámico de la evolución del lenguaje, causado por su uso constante y repetido por parte de los miembros de la comunidad lingüística hispanohablante. En el corpus también se han

visto algunas expresiones nuevas con un significado nuevo que podrían caracterizarse como neologismos o neologismos estilísticos. Por otro lado, los diccionarios tradicionales no reflejan el estado actual del idioma español contemporáneo, no pueden reaccionar y procesar todos los cambios de manera tan dinámica y rápida. Aunque hemos mencionado anteriormente que el color rojo se relacionaba con los símbolos reales, ni en los diccionarios, ni en el corpus, se han encontrado expresiones que evidencien este significado.

Gracias a los resultados mencionados arriba podemos concluir que el corpus electrónico *Araneum* contiene más expresiones con el color rojo en su estructura que las fuentes lexicográficas tradicionales. La diferencia ha sido de 113 expresiones. Esto significa que nuestra primera hipótesis, H1 ha sido confirmada. Al mismo tiempo, en el corpus electrónico solo se encuentran las unidades léxicas actuales, mientras que en los diccionarios aparecen también algunas de las expresiones arcaicas, lo que significa que también nuestra segunda hipótesis, H2 ha sido confirmada. Por último, tanto en el *Araneum*, como en los diccionarios, la construcción comparativa fue la más abundante. En el *Araneum* se presentan con abundancia las construcciones simples (como «sustantivo + rojo» que era la más frecuente) tal y como habíamos supuesto, pero, por otro lado, en las fuentes lexicográficas no aparecían las construcciones en forma de frases completas o construcciones paremiológicas como habíamos supuesto. Se han encontrado solo algunas construcciones verbales (como *ponerse rojo*). Así que nuestra tercera hipótesis, H3 no ha podido ser confirmada.

Nuestra investigación también demuestra que en el corpus electrónico se refleja el dinamismo del idioma, el cual se puede ver en los cambios o ampliación del significado, en los cambios formales, en la aparición de nuevos lexemas ya lexicalizados y otros que son ocasionales, etc.

La aportación de nuestro estudio parcial de las unidades léxicas con el cromatismo rojo en su estructura la vemos, sobre todo, en la recopilación del material lingüístico del corpus electrónico *Araneum*, que puede servir para la elaboración de los diccionarios generales bilingües, monolingües, fraseológicos, especializados o temáticos y también para estudiar más profundamente los cambios en el sistema y el dinamismo de la lengua. Con más estudios similares se podría llegar a una conclusión más compleja sobre los cambios sistemáticos en el español actual.

Este artículo es parte del proyecto VEGA/Chromatizmus a jeho konotácie v kontexte slovanských a románských jazykov 1/0107/18 (Cromatismo y sus connotaciones en el contexto de las lenguas eslavas y románicas)

BIBLIOGRAFÍA

AHOLA Jussi (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 Vladimír (2014), Aranea: Yet Another Family of (Comparable) Web Corpora, in: In SOJKA Petr, BENKO Vladimír (2016), Two Years of Aranea: Increasing Counts and Tuning the Pipeline, in: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2016)*, Portorož, European Language Resources Association (ELRA), 2016, pp. 4245-4248.
- BENKO Vladimír, BUTAŠOVÁ Anna (2017), Teaching corpus linguistics with Aranea web corpora, in: *Trudy meždunarodnoj konferencii Korpusnaja lingvistika – 2017*, Sankt-Peterburg, Sankt-Peterburgskij gosudarstvennyj universitet – Institut lingvističeskich issledovanij RAN – Rossijskij gosudarstvennyj pedagogičeskij universitet im. A. I. Gercena, p. 16-21.
- ČERMÁK František (2000), Jazykový korpus: prostředek a zdroj poznání, in: *Studie z korpusové lingvistiky. Acta Universitatis Carolinae Philologica 3 – 4, 1997*, Praha, Nakladatelství Karolinum, p. 15-37.
- DOLNÍK Juraj (2013), *Všeobecná jazykoveda*, Veda, Bratislava.
- HOFFMANN Sebastian et al. (2008), *Corpus Linguistics with BNC web – a Practical Guide*. Frankfurt am Main, Peter Lang.
- HORÁK Aleš, KOPEČEK Ivan, PALA Karel (eds.): *Text, Speech and Dialogue. 17th International Conference, TSD 2014*, Brno, Czech Republic, September 8-12, 2014. Proceedings. LNCS 8655. Springer International Publishing Switzerland, p. 257-264.
- KORPUS Araneum: [Internet]. Available at <http://aranea.juls.savba.sk/aranea>
- KVAPIL Roman (2012), Farba v komunikačnom procese, in: *FORLANG – Cudzie jazyky v akademickom prostredí*, Košice, Technická univerzita v Košiciach, p. 333-340.
- LINDQUIST Hans (2009), *Corpus Linguistics and the Description of English*, Edinburgh, Edinburgh University Press.
- MOCKOVÁ Nina (2017), *Unidades fraseológicas con nombres de colores en su estructura* (dizertačná práca), Univerzita Komenského v Bratislave, Bratislava.
- MOCKOVÁ Nina (2018), Chromatizmy červenej farby v slovenčine a španielčine, in: *Lingua et vita 14, VII/2*, Bratislava, Ekonóm, p. 31-41.
- MOLINER María (2007), *Diccionario de uso del español*, 3^a edición, Madrid, Gredos.
- Real Academia Española (2014), *Diccionario de la lengua española*. [En línea]. 23.^a ed. Madrid: Espasa Calpe, S. A.
- SMIEŠNY Jozef (2010), Farba – komunikačný signál nielen pre dopravné križovatky, in: *Finančný manažment a controlling v praxi*, 3. Bratislava, Iura Edition, p. 13-19.
- SPIŠIAKOVÁ Mária (2016), The contrastive analysis of the use, meaning and connotations of the colors in the Slovak and Spanish language [El análisis contrastivo del uso, significado y connotaciones de los colores en la lengua eslovaca y española], in: *XLinguae Journal, Volume 9 Issue 3, June 2016*, p. 104-128.
- ŠIMKOVÁ Mária (2008), Korpusová lingvistika na Slovensku, in: *Jazykovedný časopis*, 2008, Vol. 59, No. 1-2, Bratislava, Slovak Academic Press, p. 11-24.

- SECO Manuel, ANDRÉS Olimpia, RAMOS Gabino (2004), *Diccionario fraseológico documentado del español actual. Locuciones y modismos españoles*, Madrid, Aguilar.
- TRUP Ladislav, BAKYTOVÁ Jana, (2017), *Španielsko-slovenský a slovensko-španielsky frazeologický slovník*, Bratislava, Mikula.
- ULIČNÁ Martina (2019), Paralely a asymetria výskytu chromatismu biely vo vicaslovných pomenovaniach a frazeologizmoch v ruskom a slovenskom jazyku, in: *Lingua et vita, ročník VIII/1, č. 15*. Bratislava, Ekonóm, p. 50-66.