

LA JERARQUÍA IMPLICACIONAL DE BERLIN Y KAY A PARTIR DE FRECUENCIAS

BERLIN AND KAY'S IMPLICATIONAL HIERARCHY BASED ON FREQUENCIES

Antoni Brosa Rodríguez
M. Dolores Jiménez López
Universitat Rovira i Virgili

RESUMEN

Este artículo se ocupa de la relación de las jerarquías implicacionales en semántica, en concreto estudia la propuesta de Berlin y Kay acerca de los colores, y analiza nuevas posibilidades de trabajo presentes gracias a la lingüística de corpus y la lingüística computacional. La hipótesis de partida es que un análisis cuantitativo de frecuencias basadas en el uso real de las lenguas coincide con la propuesta de Berlin y Kay, pensada en términos evolutivos. Esto lleva a que nos planteemos los siguientes objetivos: permitir la comprobación de si se respeta el orden de la jerarquía implicacional original a partir de frecuencias; ofrecer una mayor precisión en los niveles de la jerarquía original donde hay más de un color; posibilitar la muestra de la preferencia por los distintos sinónimos en relación con el color en inglés propuesto por Berlin y Kay (si los hubiere, en lenguas concretas). La metodología empleada combina el uso de PanLex para la obtención léxica y Kontext para el procesamiento de las ocurrencias en todos los textos reales de un modo universal y multilingüístico. Esto permite obtener unos resultados prometedores, donde se señalará cómo se van superando todos los problemas encontrados en una investigación de tal calibre. En primer lugar, se detalla la evolución de dicha jerarquía, sus posteriores implicaciones, su uso actual y sus ventajas y limitaciones. Posteriormente, se ofrece una propuesta de metodología de investigación de los colores desde una perspectiva cuantitativa y enfocada en la tipología lingüística. El protocolo que se ofrece está pensado para, en primer lugar, disponer de datos de distintas lenguas que permitan una revisión de la jerarquía a partir del concepto de saliencia. Este complemento cuantitativo permite enriquecer el modelo ya vigente. En segundo lugar, los datos extraídos a partir de la guía de actuación aquí presentada permiten una mayor precisión a la hora de jerarquizar colores que, en la propuesta canónica de Berlin y Kay, se hallan en el mismo nivel. Finalmente, también pretendemos reducir el grado de anglocentrismo de la propuesta mediante una expansión de los colores ingleses en las distintas divisiones propias de cada lengua (o sinónimos) con una ponderación de estas etiquetas en cuanto a su frecuencia.

PALABRAS CLAVE: Jerarquía implicacional, semántica, colores, corpus, cuantitativo, tipología lingüística

ABSTRACT

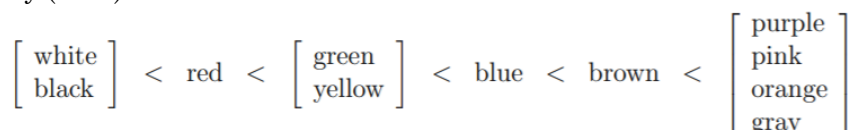
This study is concerned with the relationship between implicational hierarchies in semantics, in particular the one proposed by Berlin and Kay about colours, and the new possibilities offered by corpus linguistics and computational linguistics. The starting hypothesis is that a quantitative analysis of frequencies based on actual language use coincides with Berlin and Kay's proposal, thought of in evolutionary terms. This leads to the following objectives: to allow the verification of whether the order of the original implicational hierarchy is respected on the basis of frequencies; to provide greater precision in the levels of the original hierarchy where there is more than one color; to make it possible to show the preference for the different synonyms in relation to the color in English proposed by Berlin and Kay (if any, in specific languages). The methodology employed combines the use of PanLex for lexical retrieval and Kontext for processing occurrences in all real texts in a universal and multilingual way. This leads to promising results, where it will be shown how all the problems encountered in a research of this caliber are being overcome. First, the evolution of this hierarchy, its subsequent implications, its current use and its advantages and limitations are detailed. Subsequently, a proposal for a colour research methodology is offered from a quantitative perspective and focused on linguistic typology. The protocol offered is designed, first of all, to provide data from different languages that allow for a review of the hierarchy based on the concept of salience. This quantitative complement allows the existing model to be enriched. Secondly, the data extracted from the action guide presented here allow for greater precision when hierarchising colours which, in Berlin and Kay's canonical proposal, are on the same level. Finally, we also aim to reduce the degree of Anglocentrism of the proposal by expanding the English colours into the various language-specific divisions (or synonyms) with a weighting of these labels in terms of their frequency.

KEYWORDS: *Implicational hierarchy, semantics, colours, corpus, quantitative, linguistic typology*

1. INTRODUCCIÓN

El trabajo que presentamos se sitúa en el ámbito de los estudios sobre universales lingüísticos y pretende revisar, cuantitativamente y trabajando con nuevos datos extraídos de corpus de varias lenguas, el universal lingüístico conocido como “la jerarquía implicacional de los colores”, propuesto por Berlin y Kay (1969). Este universal presenta una clasificación de los colores básicos como se ve a continuación en la figura 1:

Figura 1. Jerarquía implicacional de los colores básicos por etapas, propuesta por Berlin y Kay (1969)



Según la teoría propuesta por Berlin y Kay, si una lengua posee alguno de estos colores, dispondrá también de los que se sitúan a su izquierda. Por ejemplo, si una lengua presenta la etiqueta azul, también contará con verde/amarillo, rojo y

blanco/negro. Esta clasificación de la universalidad de los colores básicos se realizó, inicialmente, a través de un experimento con hablantes de 20 lenguas. De ser cierta y universal, permitiría conocer la evolución de las lenguas del mundo (que solo podrían avanzar o retroceder en un nivel) e, incluso, de las distintas variedades de una misma lengua.

La propuesta de Berlin y Kay sigue suscitando interés en los investigadores. Se aplica a lenguas nuevas (García Ferrer, 2009) o se analizan aspectos morfológicos o semánticos que podrían contribuir a tal clasificación (McCarthy et al., 2020). No obstante, también ha sido ampliamente cuestionada, tanto por la metodología utilizada como por no cumplirse en determinadas lenguas (Levinson, 2000; Saunders, 2000). Además, también se ha comprobado dicho universal en su realización en lenguas específicas, ya que se puede ver la influencia de dicha teoría en lenguas con todos los términos, con subgrupos de colores básicos. El ruso, es una de las lenguas en las que más se ha trabajado este universal lingüístico, ya sea a través de formularios-experimentos (Paramei, Griber y Mylonas, 2018, Andrews 1994), o por medio del cálculo de su frecuencia en corpus (Zasorina 1977; Corbett y Morgan, 1988).

1.1. Hipótesis y objetivos

En este estudio pretendemos diseñar una metodología que permita una revisión cuantitativa de la propuesta de universal semántico de Berlin y Kay. La hipótesis de partida es que un análisis cuantitativo de frecuencias basadas en el uso real de las lenguas coincide con la propuesta de Berlin y Kay, pensada en términos evolutivos. No obstante, dada la dificultad de hacer un estudio así, no existe todavía ninguna metodología para ello (al menos que nosotros conozcamos). Por ello, los objetivos detrás de esta propuesta son:

- Permitir la comprobación de si se respeta el orden de la jerarquía implicacional original a partir de frecuencias.
- Ofrecer mayor precisión en los niveles de la jerarquía original donde hay más de un color.
- Posibilitar la muestra de la preferencia por los distintos sinónimos en relación con el color en inglés propuesto por Berlin y Kay (si los hubiere, en lenguas concretas).

El artículo se estructura de la siguiente manera. En primer lugar, revisaremos las jerarquías implicacionales, y en concreto la de colores, su evolución y sus implicaciones posteriores. Posteriormente, presentaremos nuestra propuesta de investigación mediante corpus, comentaremos los aspectos positivos y los inconvenientes o limitaciones. Concluiremos destacando algunas líneas de trabajo futuro.

2. ESTADO DE LA CUESTIÓN

Los *universales en el lenguaje*, idealmente, son generalizaciones que recogen propiedades del lenguaje o de las lenguas que son estables en todas las lenguas posibles del mundo (Siemund, 2011). El estudio de estos, ya desde las primeras propuestas de los años 60, se puede abordar desde dos metodologías distintas. Esto genera una primera diferenciación teórica entre lo que se conoce como *universales superficiales* y *universales profundos* (Newmeyer, 2008). Las principales diferencias

de estas dos perspectivas han sido perfectamente sintetizadas por Comrie (1989: 1–2):

The two approaches can be contrasted on a number of parameters, the most important of these being the following: the data base for research on language universals (a wide range of languages, or a highly restricted set of languages); the degree of abstractness of analysis that is required in order to state language universals (for instance, in terms of more concrete or more abstract syntactic analysis); and the kinds of explanations advanced for the existence of language universals. (...) On the one hand, some linguists have argued that in order to carry out research on language universals, it is necessary to have data from a wide range of languages; linguists advocating this approach have tended to concentrate on universals statable in terms of relatively concrete rather than very abstract analyses, and have tended to be open, or at least eclectic, in the kinds of explanations that may be advanced for the existence of language universals. On the other hand, some linguists have argued that the best way to learn about language universals is by the detailed study of a small number of languages; such linguists have also advocated stating language universals in terms of abstract structures and have tended to favor innateness as the explanation for language universals.

En el caso de los *universales profundos*, la búsqueda de invariancias se realiza a través de una o pocas lenguas y atiende a la lógica de la gramática universal. Es decir, que los universales formulados solo tendrán lógica dentro de esos mismos preceptos. Otra característica distintiva entre ambas vías es la recogida bajo la segunda distinción teórica: *universales absolutos* y *universales estadísticos* (Abraham, 1989). La diferencia, como indica Odden (2003: 38) radica en la presencia o aceptación de las excepciones:

The relation between absolute observational universal and statistical tendencies of the Greenbergian type is that these are both statements about observations, i.e., they are statements about languages and not about grammars. An absolute universal is one where 100% of human languages have a certain property, and a statistical tendency is one where a significant number, but not 100% have the property.

Los *universales profundos*, por su naturaleza específica, son presentados sin excepciones. Ello lo favorece, por un lado, el análisis reducido de lenguas y el alto nivel de abstracción de las características investigadas. Por otro lado, se entienden los universales como manifestaciones (o parámetros) en relación con un principio universal y, por lo tanto, la ausencia del universal en una lengua no se considera una excepción como tal. En cambio, los *universales superficiales* o *tipológico-funcionales* suelen manifestarse como *universales estadísticos*, es decir, con excepciones. Este hecho es controvertido y ha generado una gran cantidad de bibliografía y distintas etiquetas (como *generalización tipológica*, *tendencia universal*, *preferencia universal*, *patrones entre lenguas*, *generalización estadística*, *generalización entre lenguas*, *tendencia lingüística*, *casi universal*, etc.) que no han sido capaces de resolver la contradicción inherente al objeto de estudio (Jakobson, 1963; Greenberg, 1978; Dryer, 1998; Cysouw, 2003; Bickel, 2007; Guevara y Scalise, 2009; Whaley, 2010; Siemund, 2011; Pericliev, 2012).

La última de las distinciones, también en relación con las dos vías posibles para formular universales (Bickel, 2010) y, además, la más relevante para este estudio, es la que se establece entre *universales incondicionales* y *universales condicionales* o *implicacionales* (Greenberg, Osgood y Jenkins, 1963).

La distinción fundamental entre ambas radica en el nivel de afectación de la formulación. En el caso de los *universales incondicionales* (en todas las lenguas hay

y) su aplicación es en la totalidad del universo, es decir, no hay ninguna restricción de aplicación. Los *universales condicionales* (si en una lengua hay *x*, entonces hay *y*), se aplican a la totalidad del tipo lingüístico enunciado en la condición inicial. O, en palabras de Croft (2003: 53): “Implicational universals differ from unrestricted universals in that they do not assert that all languages belong to one type. Instead, they describe a restriction on logically possible language types that limits linguistic variation but does not eliminate it.”

Los universales dentro de la corriente innatista tendrán casi siempre la forma sin condiciones previas, ya que se intenta describir el lenguaje como tal, en su totalidad. Los universales dentro de la corriente tipológica tendrán casi siempre la forma con condiciones previas, pues se intenta entender el comportamiento de las lenguas de un determinado tipo lingüístico, de ahí su estrecha relación (Steedman, 2009; Grandi, 2014). O, dicho de otro modo, el objetivo último es predecir el comportamiento de una lengua a partir de sus características. Este hecho es también el que destaca el mismo Croft (2003) como diferencial: “What makes implicational universals more interesting than unrestricted universals above all, is that they state a dependency between two logically independent parameters”.

2.1. Jerarquías implicacionales

Uno de los aspectos más interesantes en la investigación de universales en las lenguas son las jerarquías implicacionales, una evolución o ampliación de los universales implicacionales o condicionales descritos anteriormente. De hecho, se podría definir una jerarquía como una suma de dos o más universales condicionales. Sin embargo, la complejidad de las jerarquías va más allá de la mera suma de condiciones. Uno de los aspectos más interesantes es el orden de los constituyentes dentro de dicha jerarquía, que no es arbitrario. Si se puede constatar la presencia de uno de los elementos a la jerarquía, se confirma automáticamente la presencia de todos los constituyentes a su izquierda, sin excepción.

La última de las características relevantes de las jerarquías implicacionales es su poder de predicción, ya sea en términos de ampliación o de reducción. En términos diacrónicos, de evolución de las lenguas, se puede aumentar o reducir uno de los elementos de la jerarquía en una lengua determinada con el paso del tiempo. Algunos ejemplos se han podido atestiguar en el cambio del latín a las lenguas romance, por ejemplo. De hecho, también es posible hablar de variación intralingüística en el nivel de adhesión a la jerarquía implicacional dentro de las variedades de una misma lengua (Corbett, 2010). Es decir, distintos dialectos de una lengua pueden hallarse en un nivel distinto de la jerarquía.

Se encuentran jerarquías implicacionales en todos los niveles de la lengua. De hecho, Corbett (2010) glosa las manifestaciones más relevantes. Según este lingüista, en sintaxis, se puede destacar la “*accessibility hierarchy*” (jerarquía de accesibilidad). En morfología, la “*agreement hierarchy*” (jerarquía de concordancia) y la “*animacy hierarchy*” (jerarquía de animacidad). En léxico-semántica, la “*colour hierarchy*” (jerarquía de los colores):

- La *jerarquía de accesibilidad* (Comrie y Keenan, 1979) se puede resumir en lo presentado en la tabla 1:

Tabla 1. Jerarquía implicacional de la relativización del nombre

| 1> | 2> | 3> | 4> | 5> | 6 |
|--------|------------|--------------|---------|----------|----------------|
| Sujeto | ObjDirecto | ObjIndirecto | Oblicuo | Genitivo | ObjComparativo |

Sin embargo, el orden de los dos primeros elementos (objeto y sujeto) ha sufrido distintas modificaciones en las últimas décadas. Este resumen se puede interpretar del siguiente modo: si una lengua puede relativizar un nombre como objeto indirecto, podrá también relativizar ese mismo nombre en función de objeto directo o sujeto. Es decir, si el español permite la construcción “El astronauta al que le di un libro”, también será posible construir “El astronauta que conocí ayer” o “El astronauta que me vio ayer”.

- La *jerarquía de concordancia* (Corbett, 1979) ordena los niveles en los que se puede producir concordancia en casos de palabras con dualidad singular/plural en su interpretación. Un ejemplo podría ser la palabra *committee* (formalmente singular, pero que incluye a un grupo de personas). Si nos basamos en la jerarquía propuesta en la tabla 2, es más fácil encontrar en las lenguas “ellos” como sustituto del *comité* que “quienes” como relativo de *comité*. Además, también se han hecho algunos estudios de corpus en los cuales se han apoyado estos datos con frecuencias de aparición en lenguas con más de una posibilidad.

Tabla 2. Jerarquía implicacional de la concordancia

| 1> | 2> | 3> | 4> |
|--------------------|--------------------|-----------|------------|
| Pronombre Personal | Pronombre Relativo | Predicado | Atributivo |

- La *jerarquía de animacidad* (Smith-Stark, 1974) detalla el nivel de precisión en la distinción de los intervinientes en las distintas lenguas. Su resumen es mostrado en la tabla 3:

Tabla 3. Jerarquía implicacional de la animacidad

| 1> | 2> | 3> | 4> | 5> | 6> | 7 |
|----------|--------------|-------------|--------------|--------|---------|-----------|
| Hablante | Destinatario | 3ra Persona | Familiaridad | Humano | Animado | Inanimado |

Dicha distinción puede producirse en distintos aspectos lingüísticos, en muchos o pocos, dependiendo de la lengua. Por ejemplo, es conocida la distinción existente en las lenguas eslavas como el ruso, en el uso del caso acusativo, entre animacidad y no animacidad. Es decir, en ruso, *kot* (gato) y *stol* (mesa), usadas en situación de acusativo, tendrán terminaciones distintas, a pesar de ser ambas masculinas (por su terminación consonántica), por el hecho de diferir en cuanto a animacidad. En el caso del español, por ejemplo, se puede destacar el uso de la preposición *a* con objeto directo de persona, algo que no ocurre con el resto de los seres vivos (huyendo del debate de la personificación de animales domésticos por ser, precisamente, ese peldaño más cercano). Así, pues, por ejemplo, si constatamos que en español existe la distinción entre humanos y animados, también existirán todas las distinciones anteriores. Si el ruso, en el futuro, pierde alguna de sus distinciones, será, precisamente, la mencionada, por estar en el límite de su

escala. Esta jerarquía, junto a la de número (*singular>plural>dual>paucal/trial*) ya aparece mencionada en Greenberg (1963).

Más allá de esta selección con las jerarquías más importantes, se pueden encontrar en otros ámbitos como la fonética o fonología, por ejemplo, la jerarquía de grupos consonánticos: *CV>CCV>CCCV* (Greenberg et al., 1963). También jerarquías que unen características de distintos módulos, como la jerarquía de caso (Primus, 2010) en la que se evalúa la incidencia del tono: *nominativo/absolutivo>acusativo/ergativo>dativo>otros casos oblicuos*.

Además, las jerarquías propuestas han servido para posteriores aplicaciones, como son revisiones cuantitativas en lenguas específicas con textos reales, como es el caso de Fox (1987); revisiones desde una metodología más generativa (Sigurd, 1989) o, incluso, experimentos psicolingüísticos en lenguas como el japonés con las cláusulas de relativo y su revisión de facilidad de relativización y procesamiento (Ozeki y Shirai, 2007).

En el caso de léxico-semántica, la jerarquía más relevante es la que se trata en este estudio, acerca de los colores, propuesta por Berlin y Kay (1969).

2.2. Jerarquía implicacional de los colores

A pesar de que ha habido propuestas de jerarquías implicacionales en semántica desde un punto de vista más formal (o abstracto) como las recogidas en Siewierska (1988) o Allan (1987), la jerarquía implicacional sobre los colores propuesta por Berlin y Kay es la que ha gozado de más éxito. Berlin y Kay (1969) han propuesto la clasificación de los colores básicos como presentamos en la tabla 4:

Tabla 4. Jerarquía implicacional de los colores básicos, según Berlin y Kay

| 1> | 2> | 3> | 4> | 5> | 6 |
|--------|------|----------|------|--------|---------|
| Blanco | Rojo | Verde | Azul | Marrón | Púrpura |
| Negro | | Amarillo | | | Rosa |
| | | | | | Naranja |
| | | | | | Gris |

En primer lugar, se debe hacer establecer una oposición entre todos los colores presentes en la jerarquía y los que no aparecen. Las etiquetas recogidas en la jerarquía anterior se entienden como *basic color terms* o palabras de colores básicos. En esta primera labor de clasificación, los dos lingüistas diferencian aquellos colores que se consideran básicos de los que no lo son. Para ello, utilizan, diversos criterios:

- Obligación de ser monolexémico: un color básico no puede estar compuesto por más de un lexema. Tampoco estar subyugado a un color más amplio. Por ello, un color como *azul marino* jamás podrá considerarse básico. Este criterio atiende a la lógica de que dicho color se basa en otro, que matiza y además son dos lexemas.
- La *salience* o prominencia, es decir, el uso de criterios de frecuencia y distribución para discernir qué colores son básicos.
- No disponer de un hiperónimo.
- Ser aplicable a cualquier objeto (no como *rubio*, por ejemplo).
- No estar modificado por morfemas.

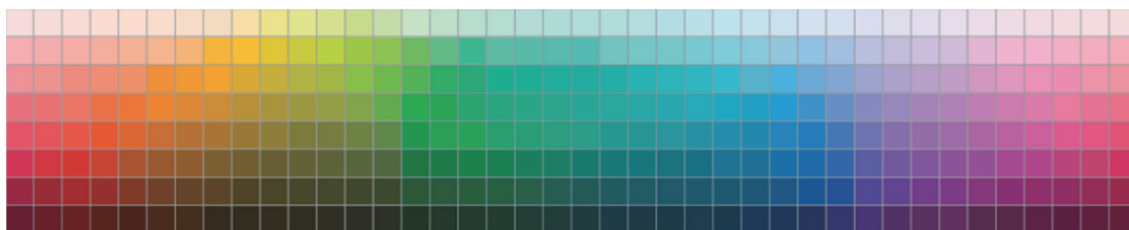
Aunque pueda creerse que el criterio de saliencia ha sido utilizado no solo para diferenciar colores básicos de los que no lo son, sino también para establecer los distintos niveles de la jerarquía, este no fue el proceso de la investigación de Berlin y Kay, al menos, en la primera de sus publicaciones. Los distintos elementos de la jerarquía se corresponden, en realidad, con distintas etapas evolutivas, según los autores. Es decir, todas las lenguas cuentan con los colores blanco y negro (o claro y oscuro), el primer nivel. Posteriormente, la mayoría de las lenguas optaron por añadir una oposición entre esos dos colores y el color rojo. Posteriormente, gran parte de las lenguas que añadieron alguna etiqueta similar a *rojo* a su repertorio, también decidieron añadir verde y amarillo, y, así, sucesivamente.

La propuesta de Berlin y Kay no está exenta de polémica, pues existe el riesgo de relacionar lenguas más evolucionadas o menos según su adhesión a la jerarquía y, por extensión, considerar que los hablantes que solo poseen *blanco*, *negro* y *rojo* en su lengua son menos evolucionados e inferiores a los hablantes que poseen todos los colores de la jerarquía, por ejemplo.

Esta jerarquía original ha sufrido distintas críticas y también alguna ligera modificación (producto de dichas críticas), por ejemplo, la unión o separación en un mismo nivel de los colores amarillo y verde. Saunders (1995) y Levinson (2000) son dos de los principales revisores de la propuesta de Berlin y Kay, que cuenta con problemas inherentes al estudio tipológico y muy similares a los encontrados en otras jerarquías implicacionales, como el caso de la jerarquía WH en Mackenzie (2009).

Después de reconocer todos los méritos y aciertos, Saunders (1995) se centra en aspectos más metodológicos y generales del diseño del estudio. Por ejemplo, es evidente que la propuesta parte de un claro anglocentrismo (un sesgo típico en los estudios tipológicos sobre universales). Hay algunas lenguas que tienen el aspecto del brillo como rasgo distintivo a la hora de tratar los colores. Como el inglés no tiene esta oposición a la hora de discernir colores, se obvia. O, en concreto, si se toma como ejemplo el japonés, se podrá apreciar cómo el modo de componer los colores es totalmente diferente, hecho que produce una amplia variedad de formas de un mismo color, sin que ninguna esté subordinada a otra. En otras palabras, la propuesta tiende a una simplificación que genera un difícil encaje o una difícil representación de los distintos mecanismos existentes en las distintas lenguas del mundo. La otra crítica principal de Saunders (1995) se centra en la metodología de recogida de datos. El estudio original se basa en los datos de 98 lenguas distintas, no obstante, hay una mezcla de las fuentes utilizadas. Por un lado, 20 de esas lenguas se han analizado mediante Munsell, es decir, mediante una paleta de colores en celdas diseñada por Munsell (figura 2) y presentada a hablantes de 20 lenguas distintas para que ubiquen las distintas etiquetas de los colores en los grupos de celdas. El tamaño de la muestra de hablantes también suele cuestionarse, así como otras influencias como son el bilingüismo de estos (inglés) y una misma zona geográfica compartida. Por otro lado, los colores de las otras 78 lenguas han sido obtenidos mediante etnografías, gramáticas o similares, es decir, recogidos en publicaciones con datos de segunda mano.

Figura 2. Muestra de paleta de celdas de colores en Munsell



Levinson (2000) se centra en aportar nuevos datos, de lenguas específicas que suponen excepciones a la propuesta de clasificación jerárquica de Berlin y Kay. El caso más llamativo aportado por este, seguramente, es la lengua Yéli que, según Levinson, no posee ni la palabra *color*. Por ello, los colores son reduplicaciones de objetos que contienen ese color, como el caso de mgîdî (noche) y mgîdîmgîdî (negro). Esta argumentación es un claro contrapunto a otro de los criterios propuestos por Berlin y Kay para diferenciar colores básicos de secundarios y que no hemos explicitado por estar ampliamente desmentida, el hecho de que los colores deben ser abstractos y no designar, además, objetos concretos del mundo.

En los últimos años, además, desde la lingüística computacional se ha tratado de probar el modelo de Berlin y Kay con distintas técnicas y nuevos datos. El ejemplo más representativo de ello aparece en McCarthy et al. (2020), que pretenden ofrecer nuevos datos con experimentos computacionales para confirmar si se cumplen los criterios de abstracción, formas monoléxicas y saliencia. No obstante, para revisar la saliencia, por ejemplo, solamente analizan corpus en lengua inglesa. Por lo tanto, se podría entender más como una prueba de concepto, de la metodología propuesta, que como una presentación de datos que permitan extraer conclusiones desde una perspectiva universal.

Muchos de los aspectos comentados anteriormente han generado una respuesta de los propios autores en posteriores publicaciones (Cook, Kay y Regier, 2005; Regier, Kay y Cook, 2005; Kay et al., 2009). Para dotar a la jerarquía de una mayor universalidad, se han analizado más lenguas (unas 110) y mediante una metodología unificada, conocida como World Color Survey. Se ha optado por perfeccionar la recogida de datos mediante Munsell. También se ha explicado de un modo más detallado el proceso de recogida de datos y los criterios empleados. Finalmente, se presentan los datos obtenidos de un modo más concreto y se justifican las elecciones determinadas en cada una de las lenguas.

En último lugar, también es necesario mencionar la amplia aceptación que sigue teniendo la jerarquía hoy en día, décadas después. A pesar de las críticas recibidas, son muchas las investigaciones realizadas en los últimos años que se basan en la propuesta de Berlin y Kay para poder entender mejor distintos aspectos específicos de las lenguas. Por ejemplo, Uusküla (2008b) revisa los Basic Color Terms en checo, en húngaro (Uusküla y Sutrop, 2010) y en lenguas eslavas en general (Uusküla, 2008a) desde un punto de vista cuantitativo, para disponer de datos más fiables y de naturaleza distinta. En este mismo sentido, Kandi et al. (2015) deciden aplicar la misma metodología empleada en la World Color Survey para poder documentar el nombramiento de colores en persa, lengua no documentada por Berlin y Kay en sus estudios. Además, desglosan los resultados por ciudades, ya que, aunque en la mayoría de los lugares se replican los resultados exactos de Berlin y Kay, en un par de ciudades persas, según estos, solo aparecen 10 colores básicos en vez de 11.

Otras lenguas específicas han sido analizadas, de nuevo, pero a partir de nuevas metodologías, como es el caso del italiano (Paggetti, Menegaz y Paramei, 2016) o el ruso (Paramei, Griber y Mylonas, 2018). A pesar de que se sigue utilizando la propuesta de Munsell, la selección de participantes es distinta a la propuesta por Berlin y Kay, así como los datos en los que específicos que se destacan como, por ejemplo, el encaje de los dos azules presentes en ruso (comparables a los italianos, precisamente). En el caso del ruso, también se ha producido, por ejemplo, una investigación microdiacrónica de semántica distribucional de dos etiquetas en competición por designar un color básico, en concreto, el *marrón* (Bochkarev et al., 2020). El estudio se ha realizado desde una perspectiva cuantitativa y de corpus, a través de n-gramas en Google Books. En el caso de Zaslavskiy et al. (2021), se realiza una revisión de los colores básicos (BCT) en la lengua Nafaanra desde una perspectiva diacrónica, para tratar de comprobar si hay cambio entre las etapas propuestas por Berlin y Kay. Los resultados obtenidos son muy relevantes para estudiar el contacto lingüístico y su incidencia en las lenguas. En el caso del Nafaanra, en 1978, en la World Color Survey aparecen 3 distinciones (es decir, 2 etapas): blanco/claro, negro/oscurito, rojo/cálido. En 2018, 40 años más tarde, en cambio, Zaslavsky et al. constatan la presencia de 10 colores, por influencia del inglés y un afán de una mayor precisión.

Además, también hay usos industriales derivados de la jerarquía analizada (Jameson y Webster, 2019). Por ejemplo, una confirmación de la universalidad de los 11 colores recogidos en Berlin y Kay puede permitir una clasificación de configuración cromática de los distintos dispositivos en esos términos. En caso contrario, pueden generarse complicaciones para los usuarios en distintos lugares del mundo.

3. METODOLOGÍA

Teniendo en cuenta lo dicho en el apartado anterior, consideramos que es necesario contar con más datos, de fuentes todavía no exploradas, para poder confirmar o perfilar la robustez del modelo de Berlin y Kay y, así, que las innovaciones expuestas puedan seguir creciendo y mejorando. Así, pues, proponemos una nueva línea de investigación cuantitativa para la jerarquía implicacional de los colores.

La propuesta que se desarrollará a continuación se basa en el uso de herramientas computacionales, de corpus anotados y digitalizados, para aportar datos cuantitativos de textos reales que sean un complemento o contrapunto a los datos aportados en los *Basic Color Terms* (BCT) o la *World Color Survey* (WCS).

3.1. Muestra de lenguas: Conveniencia

Dada la dificultad para obtener datos en las lenguas, en los estudios de tipología, se suele crear un *set* de lenguas equilibrado para que los resultados sean representativos (Mattiola, 2020). El equilibrio pasa por una consideración de distintas características condicionantes de la influencia en la variedad lingüística: herencia común, proximidad geográfica, influencia del tipo lingüístico (Miestamo, Bakker y Arppe, 2016). Cabe decir que las dos propuestas mencionadas, BCT y WCS, no han sido creadas teniendo en cuenta estos criterios. Desgraciadamente, dada la disponibilidad de corpus digitalizados anotados homogéneamente en distintas

lenguas en la actualidad, todavía no puede hacerse un *set* representativo, ya sea de variedad o de probabilidad. Por ello, a la espera de que crezca la disponibilidad de lenguas, se propone un trabajo basado en lo conocido como muestra de conveniencia, a partir de los datos disponibles, como una primera aproximación.

3.2. Herramienta de obtención de los colores: PanLex

Una vez se han fijado las lenguas que servirán para el estudio, se deben obtener todas las etiquetas que serán analizadas. En otras palabras, debemos obtener las formas de los colores en distintas lenguas. Puesto que se trabajará con lenguas poco conocidas para los investigadores, es necesario contar con un marco que ofrezca la seguridad de la obtención de etiquetas y sus posibles sinónimos (para evitar la parcialidad producto de una clasificación anglocentrista, como ya se ha comentado).

Lo más habitual sería ir directamente a las gramáticas (o listas léxicas y similares) de las lenguas que se analizarán para, así, obtener los distintos colores. No obstante, cabe recordar que pretendemos trabajar con corpus con textos reales y actuales. Por ello, como hemos mostrado en el caso de Zaslavsky et al. (2021), en una lengua, especialmente en los últimos años, pueden ocurrir muchas modificaciones. Por un lado, la forma de la palabra puede ser distinta. Por otro lado, la lengua se puede hallar en un estado de la jerarquía implicacional distinto. Siguiendo con el ejemplo de la lengua Nafaanra, que ha cambiado tanto de 1978 a 2018, si se utilizan los documentos canónicos de los años 80 para analizar textos de 2020, pueden existir inconsistencias considerables.

Por ello, proponemos aprovechar la herramienta utilizada en McCarthy et al. (2020) en un estudio similar con doble traducción de lenguas: *PanLex* (Kamholz, Pool y Colowick, 2014). Dicha herramienta tiene la ventaja de tener datos actualizados de un gran número de lenguas, también de lenguas poco estudiadas. Obviamente, las palabras que aparecen son lemas, hecho que facilita la forma que será utilizada posteriormente para recoger toda la variabilidad de los corpus.

3.3. Herramienta de análisis y corpus: Kontext

Actualmente, disponemos de distintas herramientas que permiten la comparación de corpus de distintas lenguas, gracias a su anotación bajo los mismos preceptos. Seguramente, la más relevante y utilizada en los últimos años es *Universal Dependencies* (Nivre et al., 2016), que contiene corpus anotados por dependencias y con etiquetas lo más universales posible. Su uso está especialmente pensado para la sintaxis, aunque también podrían hacerse búsquedas de lemas a partir de los datos recogidos.

No obstante, la herramienta que recomendamos para el estudio que proponemos es *Kontext* (Machálek, 2020). Este recurso ha sido diseñado bajo el mismo entorno y filosofía que *Universal Dependencies* (UD) y es un recurso integrado dentro de las múltiples herramientas en Lindat-CLARIN.

Las ventajas frente a UD son, fundamentalmente, dos. Por un lado, *Kontext* presenta una interfaz muy visual e intuitiva. Esto permite una búsqueda de elementos (con distintos parámetros) sin necesidad de dominio de Python o R. Así, pues, su uso es altamente recomendable para lingüistas cuyo dominio en programación no sea elevado. Por otro lado, *Kontext* engloba todos los *treebanks* de UD, pero no se limita solo a estos. Esto quiere decir, por lo tanto, que hay más corpus

anotados disponibles, de otras fuentes además de UD, como pueden ser los corpus paralelos ofrecidos por la Unión Europea, por ejemplo.

3.4. Tipología de corpus: Variada

Una vez hemos fijado el recurso que se utilizará y las lenguas en consideración, debemos decidir, en el caso de que existan varias, la tipología de corpus que sirve para la investigación que proponemos. Podemos resumir los tipos de corpus hallados en Kontext en 4 grupos distintos:

1. Noticias, reseñas y wiki
2. Legal, médico, no ficción
3. Actas parlamentarias
4. Blog, redes sociales, webs

Lo habitual en un estudio de este tipo es trabajar con el último de los grupos, pues se supone que hay una mayor variedad sociolingüística en los textos. Ello genera una mayor certidumbre en términos de saliencia. Es decir, distintos grupos sociales producen una determinada etiqueta de color y, por lo tanto, es menos probable que sea propia de un sociolecto determinado.

No obstante, recomendamos recolectar los datos de los cuatro grupos, si es posible. En estudios futuros, más detallados, se podría analizar la distinción de tipología de corpus como variable. Es decir, sería interesante tratar de capturar invariancias comunes a las lenguas en tipos de texto determinados. En análisis preliminares que hemos realizado, no hemos detectado diferencias significativas entre las distintas tipologías de corpus y ofrecen resultados similares. La única diferencia entre corpus radica en el tamaño de estos, obviamente.

Además, los datos se pueden clasificar y procesar de los dos modos. Dado que no hay diferencias entre los distintos corpus, es plausible ofrecer los resultados unificados de todos los corpus y, por otro lado, los datos pormenorizados para una comparación entre tipologías en las distintas lenguas.

4. RESULTADOS

Los datos que se obtienen a través de los distintos corpus no son comparables, dada la disparidad de tamaños. Por eso, otra de las facilidades que ofrece Kontext es, directamente, el índice por millón de tokens (IPM). Esta cifra es, simplemente, la homogenización de los resultados obtenidos de frecuencia (*hits*) si todos los corpus tuvieran 1 millón de tokens, es decir, el mismo tamaño. Así, pues, los resultados ya pueden compararse, sea entre tipologías, lenguas, o ambas opciones.

Además, también existe la posibilidad de trabajar con la frecuencia media reducida (*ARF*), que no es más que la fórmula de corrección de resultados para garantizar homogeneidad en la distribución del elemento de estudio, mostrada en la figura 3. En otras palabras, una vez aplicada la fórmula correctora, el resultado obtenido ofrece garantías de que la aparición de dicho color no se concentra únicamente en un texto del corpus o en una parte determinada del corpus y ofrece una media de este.

Figura 3. Fórmula de la Average Reduced Frequency

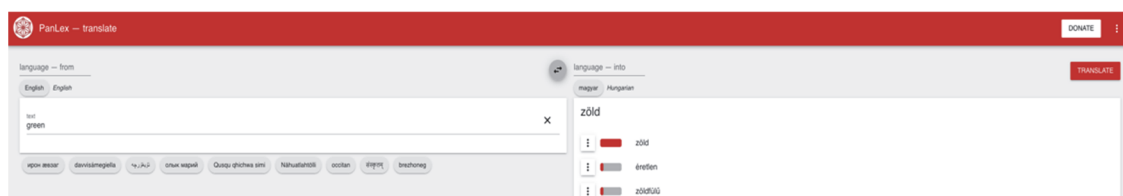
$$ARF = \frac{1}{v} \sum_{i=1}^f \min(d_i, v)$$

En relación con la poca disponibilidad de lenguas con corpus etiquetados homogéneamente (obviamente, de lenguas no europeas), también se debe mencionar el problema de hallar corpus disponibles, pero con un número de tokens insuficiente. En otras palabras, los corpus pequeños no permiten un estudio fiable del objeto de estudio. Por ello, se han realizado pruebas para valorar cuál es el umbral recomendable a la hora de llevar a cabo este estudio (o similares). Después de una revisión de distintas lenguas, se ha constatado que con un corpus de menos de 50.000 tokens no ofrece ningún tipo de resultado. No obstante, recomendamos considerar corpus con más de 100.000 tokens para llevar a cabo estudios cuantitativos sobre colores (y, por extensión, aspectos léxicos y semánticos similares).

Por lo tanto, como se ha comentado anteriormente, al no disponer de un número adecuado de corpus, y al no tener resultados concluyentes previos, recomendamos que la muestra sea de todas las lenguas disponibles hasta el momento, sin equilibrios, y sin tener en cuenta los corpus de menos de 100.000 tokens.

En relación con PanLex, se ofrece una lista de sinónimos (o posibles sinónimos) para la palabra buscada con un porcentaje de similitud en distintos contextos. Por ello, se recomienda recoger todos los lemas superiores a un 66% de similitud, que se corresponde con la proporción de dos tercios, un criterio bastante útil a la hora de trabajar con universales en tipología. Es posible utilizar el criterio de la doble traducción para verificar una correcta adecuación del término y la ausencia de homónimos. Así, se podrá mitigar la influencia de la lengua inglesa a la hora de fijar los estadios de colores. Un ejemplo de uso de PanLex sería el que se muestra en la figura 4.

Figura 4. Muestra de uso de la herramienta PanLex



Como es habitual en los estudios de tipología, al estudiar lenguas de tipos tan diversos, afloran dificultades a la hora de analizarlas. El primero que debe tenerse en cuenta, es el de los lemas. Obviamente, si se quiere ver la frecuencia del color *blanco* en los corpus del español, no nos interesará, solamente, todos los “blanco” que aparezcan, sino que debemos tener en cuenta los “blanca, blancas, blancos, blanquitos, blanquitas, ...”. Este hecho es más complejo, todavía, en las lenguas con declinaciones, con una variabilidad mucho mayor. Por ello, es necesario que la búsqueda se produzca siempre a través de lemas. En el caso de Kontext, permite hacer la búsqueda así directamente, uno de los motivos por los cuales se ha seleccionado.

La segunda de las dificultades que debemos solucionar es la distinción entre mayúsculas y minúsculas. Hay lenguas cuyas palabras empiezan por mayúscula. Por ello, es necesario considerar tanto mayúsculas como minúsculas para contabilizar

todas las palabras posibles. En el caso de Kontext se debe marcar la opción en *query type*.

La última dificultad que se puede producir en las lenguas es la homonimia. Por ejemplo, en euskera, *hori/horia* hace referencia al color *amarillo*. No obstante, dicha forma también se utiliza para el relativo (*que*). Es decir, si no se hace ninguna distinción, la frecuencia de *amarillo* en euskera será anormalmente alta. Por ello, debemos aplicar un *CQL (query)* en Kontext a todas las lenguas en las que haya aparecido un caso de homonimia a la hora de la construir la lista gracias a PanLex con la doble traducción. Hay dos posibilidades de configurar el *query*. Por un lado, podemos tener en cuenta, únicamente, los adjetivos, como en: [upos= “ADJ” & lemma= “hori”]. En el caso de que prefiramos incluir todas las categorías gramaticales excepto la que genera el caso de homonimia, simplemente se deben incluir en “upos” todas las categorías posibles para etiquetar excepto esa.

Seguramente, cuando haya más lenguas disponibles, será más razonable utilizar la segunda de las opciones, ya que hay algunas lenguas (poco comunes) en las que los colores pueden no ser adjetivos y estar etiquetados con otra categoría gramatical. A pesar de ello, se opte por la opción que se opte, naturalmente, se debe ser consecuente en el análisis de todos los corpus, mediante la misma elección y, además, se debería mencionar la vía utilizada y sus implicaciones.

5. CONCLUSIONES

En definitiva, creemos que, si se siguen las directrices que proponemos, será más fácil pasar de analizar cuantitativamente la jerarquía de colores de lenguas individuales a través de corpus, como ya ocurre, a hacerlo con múltiples lenguas. Disponer de datos cuantitativos en distintas lenguas a partir de textos reales sería una gran ayuda para el desarrollo y la mejora de la jerarquía implicacional de Berlin y Kay. Es altamente probable que haya una convergencia entre la motivación evolucionista de las etapas de estos investigadores y un reflejo en la frecuencia del uso de las distintas etiquetas, por causas funcionales, como comenta Cristofaro (2012) en relación con las jerarquías implicacionales y su explicación tipológica.

A partir de estos nuevos datos, extraídos de PanLex y Kontext, en primer lugar, podemos obtener el número total de ocurrencias de los colores en distintas lenguas. Esto permitirá una comparación entre el orden de colores mediante frecuencias y el orden evolucionista de BCT o WCS, como ya se ha mencionado. En segundo lugar, gracias a los datos obtenidos mediante la metodología propuesta, es posible explicitar la tendencia de las lenguas por los colores situados en una misma etapa por Berlin y Kay. Es decir, por ejemplo, es posible ordenar el grado de relevancia de los colores ubicados en el último nivel: violeta, rosa, naranja y gris. En tercer lugar, después de extraer las distintas etiquetas en PanLex para un mismo color y ponderarlas en Kontext, también logramos ofrecer una lista de etiquetas correspondientes a un único término en inglés. Además, también se valora cuál de ellas tiene un mayor uso y, por lo tanto, saliencia y cuál de ellas es más residual.

En resumen, creemos que es necesario aprovechar las ventajas que ofrecen los avances en lingüística de corpus y lingüística computacional para revisar, contrastar y mejorar los grandes avances ya hechos en lingüística. En este caso concreto, la jerarquía implicacional de colores, comparada con frecuencias de textos reales en distintas lenguas, puede ser aplicada a un terreno mucho más natural, lejos de datos simplificados y de segunda mano (gramáticas) o datos condicionados y elicitados

(Munsell). Esta unión de la teoría canónica de colores y textos reales puede tener aplicaciones futuras muy interesantes. Por un lado, desde un plano más teórico, puede ayudar a promover estudios de más jerarquías implicacionales en semántica, un ámbito poco explorado todavía. Las facilidades ofrecidas por las herramientas mostradas pueden ayudar al descubrimiento de nuevos marcos teóricos en tipología. Por otro lado, desde un plano más práctico, añadir a la jerarquía implicacional datos reales que la calibren, puede ayudar a adaptar o perfilar el uso de colores en productos internacionales (Jameson y Webster, 2019). Además, también sería útil e interesante aplicar dichos conocimientos a la enseñanza de lenguas extranjeras, por ejemplo, para facilitar el proceso de enseñanza y aprendizaje.

REFERENCIAS

- Abraham, W. 1989. Language universals: The Chomskyan approach vs. Greenberg's typological approach. *Belgian Journal of Linguistics* 4: 9–25.
- Allan, K. 1987. Hierarchies and the coin of left conjuncts (with particular attention to English). *Journal of Linguistics* 23(1): 51–77.
- Andrews, D. R. 1994. The Russian color categories Sinij and Goluboj: An experimental analysis of their interpretation in the standard and emigré Languages. *Journal of Slavic Linguistics* 2(1): 9–28.
- Berlin, B. y Kay, P. 1969. *Universality and evolution of basic color terms*. Berkeley: University of California Press.
- Bickel, B. 2007. Typology in the 21st century: Major current developments. *Linguistic Typology* 11(1): 239–251.
- Bickel, B. 2010. Absolute and statistical universals. En P. Colm Hogan ed. *The Cambridge encyclopedia of the language sciences*. Cambridge: Cambridge University Press, 77–79.
- Bochkarev, V., Shevlyakova, A., Paramei, G. y Rakhilina, E. 2020. A quantitative study of Russian colour terms buryj and koričnevij in the Google Books Ngram corpus. *CEUR Workshop Proceedings* 2852: 1–10.
- Comrie, B. 1989. *Language universals and linguistic typology*. Chicago: The University of Chicago Press.
- Comrie, B. y Keenan, E. L. 1979. Noun phrase accessibility revisited. *Language* 55(3): 649–664.
- Cook, R. S., Kay, P. y Regier, T. 2005. The World Color Survey Database. En H. Cohen y C. Lefebvre eds. *Handbook of categorization in cognitive science*. Amsterdam: Elsevier, 223–241.
- Corbett, G. 1979. The Agreement Hierarchy. *Journal of Linguistics* 15: 203–224.
- Corbett, G. 2010. Implicational hierarchies. En J. J. Song ed. *The Oxford handbook of linguistic typology*. Oxford: Oxford University Press, 1–13.
- Corbett, G. G. y Morgan, G. 1988. Colour terms in Russian: Reflections of typological constraints in a single language. *Journal of Linguistics* 24(1): 31–64.
- Cristofaro, S. 2012. Typological universals in synchrony and diachrony: The evolution of number marking. Presentado en SLE, Stockholm.
- Croft, W. 2003. *Typology and universals*. Cambridge: Cambridge University Press.
- Cysouw, M. 2003. Against implicational universals. *Linguistic Typology* 7(1): 89–101.
- Dryer, M. S. 1998. Why statistical universals are better than absolute universals. En *Papers from the 33rd Annual Meeting of the Chicago Linguistics Society*, 1–23.
- Fox, B. A. 1987. The noun phrase accessibility hierarchy reinterpreted: Subject primacy or the absolutive hypothesis? *Language* 63(4): 856–870.
- García Ferrer, D. J. 2009. La jerarquía de términos de colores básicos en pemón y yukpoa, lenguas caribes de Venezuela. *Revista de Artes y Humanidades UNICA* 10(3): 15–38.
- Grandi, N. 2014. *Fondamenti di tipologia linguistica*. Roma: Carocci.



- Greenberg, J. 1978. Typology and cross-linguistic generalizations. En J. Greenberg, C. Ferguson y E. Moravcsik eds. *Universals of human language*. Stanford: Stanford University Press, 33–59.
- Greenberg, J. H. 1963. *Universals of language*. Cambridge, Massachusetts: The M.I.T. Press.
- Greenberg, J., Osgood, C. y Jenkins, J. 1963. Memorandum concerning language universals. En J. Greenberg, C. Ferguson y E. Moravcsik eds. *Universals of language*. Stanford: Stanford University Press, 255–264.
- Guevara, E. y Scalise, S. 2009. Searching for universals in compounding. En S. Scalise, E. Magni y A. Bisetto eds. *Universals of language today*. Bologna: Springer, 101–128.
- Jakobson, R. 1963. Implications of language universals for linguistics. En J. Greenberg, C. Ferguson y E. Moravcsik eds. *Universals of language*. Cambridge, Massachusetts: The M.I.T. Press, 208–219.
- Jameson, K. A. y Webster, M. A. 2019. Color and culture: Innovations and insights since Basic color terms—Their universality and evolution (1969). *Color Research and Application* 44(6): 1034–1041.
- Kamholz, D., Pool, J. y Colowick, S. M. 2014. PanLex: Building a resource for panlingual lexical translation. *Proceedings of the 9th International Conference on Language Resources and Evaluation, LREC 2014*, 3145–3150.
- Kandi, S. G., Tehran, M. A., Hassani, N. y Jarrahi, A. 2015. Color naming for the Persian language. *Color Research and Application* 40(4): 352–360.
- Kay, P., Berlin, B., Maffi, L. y Merrifield, W. 2009. *Color naming across languages. Color categories in thought and language*. Cambridge: Cambridge University Press.
- Levinson, S. C. 2000. Yeli Dnye and the Theory of Basic Color Terms. *Journal of Linguistic Anthropology* 10(1): 3–55.
- Machálek, T. 2020. KonText: Advanced and flexible corpus query interface. En *Proceedings of LREC 2020*, 7005–7010.
- Mackenzie, J. L. 2009. Content interrogatives in a sample of 50 languages. *Lingua* 119(8): 1131–1163.
- Mattiola, S. 2020. Two language samples for maximizing linguistic variety. *AMS Acta*.
- McCarthy, A. D., Wu, W., Mueller, A., Watson, B. y Yarowsky, D. 2020. Modeling color terminology across thousands of languages. *EMNLP-IJCNLP 2019 – 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference*, 2241–2250.
- Miestamo, M., Bakker, D. y Arppe, A. 2016. Sampling for variety. *Linguistic Typology* 20(2): 233–296.
- Newmeyer, F. J. 2008. Universals in syntax. *Linguistic Review* 25(1–2): 35–82.
- Nivre, J., de Marneffe, M.-C., Ginter, F., Goldberg, Y., Hajič, J., Manning, C. D., ... y D. Zeman. 2016. Universal dependencies v1: A multilingual treebank collection. *Proceedings of the 10th International Conference on Language Resources and Evaluation, LREC 2016*, 1659–1666.
- Odden, D. 2003. Languages and universals. *Journal of Universal Language* 4(1): 33–74.
- Ozeki, H. y Shirai, Y. 2007. Does the noun phrase accessibility hierarchy predict the difficulty order in the acquisition of Japanese relative clauses? *Studies in Second Language Acquisition* 29(2): 169–196.
- Paggetti, G., Menegaz, G. y Paramei, G. 2016. Color naming in Italian language. *Color Research and Application* 41(4): 402–415.
- Paramei, G., Griber, Y. A. y Mylonas, D. 2018. An online color naming experiment in Russian using Munsell color samples. *Color Research and Application* 43(3): 358–374.
- Pericliev, V. 2012. Formulating and misformulating language universals. *Folia Linguistica* 46(1): 211–232.
- Primus, B. 2010. Case-marking typology. En J. J. Song ed. *The Oxford handbook of linguistic typology*. Oxford: Oxford University Press, 1–16.
- Regier, T., Kay, P. y Cook, R. S. 2005. Focal colors are universal after all. *Proceedings of the National Academy of Sciences of the United States of America* 102(23): 8386–8391.

- Saunders, B. 1995. Disinterring basic color terms: A study in the mystique of cognitivism. *History of the Human Sciences* 8(4): 19–38.
- Saunders, B. 2000. Revisiting basic color terms. *Journal of the Royal Anthropological Institute* 6(1): 81–99.
- Siemund, P. 2011. *Linguistic universals and language variation*. Berlin: De Gruyter Mouton.
- Siewierska, A. 1988. *Word order rules*. New York: Croom Helm.
- Sigurd, B. 1989. A supplementary relativization hierarchy based on the complexity of the relative phrase. *Studia Linguistica* 43(1): 33–46.
- Smith-Stark, T. C. 1974. The plurality split. En M. W. La Galy, R. A. Fox y A. Bruck eds. *Papers from the Tenth Regional Meeting, Chicago Linguistic Society, April 19–21*. Chicago: Chicago Linguistic Society, 657–671.
- Steedman, M. 2009. Foundations of Universal Grammar in planned action. En M. Christiansen, C. Collins y S. Edelman eds. *Language universals*. New York: Oxford University Press, 174–199.
- Uusküla, M. 2008a. Basic colour terms in Finno-Ugric and Slavonic languages: Myths and facts, 206.
- Uusküla, M. 2008b. The basic colour terms of Czech. *Trames* 12(1): 3–28.
- Uusküla, M. y Sutrop, U. 2010. The puzzle of two terms for red in Hungarian. En J. Wohlgemuth y M. Cysouw eds. *Rara & rarissima*. Berlin: De Gruyter Mouton, 359–376.
- Whaley, L. 2010. Syntactic typology. En J. J. Song ed. *The Oxford handbook of linguistic typology*. Oxford: Oxford University Press, 1–21.
- Zaslavsky, N., Garvin, K., Kemp, C., Tishby, N. y Regier, T. 2021. The evolution of color naming reflects pressure for efficiency: Evidence from the recent past. Publicado en *Journal of Language Evolution*, lzac001, 2022.

