

Los índices de relatividad, densidad y eficiencia informativa en las lenguas: estudio de las correlaciones matemáticas entre palabras y fonemas

Indices of relativity, density and informative efficiency of languages: a study of mathematical correlations between words and phonemes

ENRIQUE J. VERCHER GARCÍA
Universidad Complutense de Madrid, España
evercher@ucm.es
<https://orcid.org/0000-0002-3263-6199>

MANUEL BULLEJOS LORENZO
Universidad de Granada, España
bullejos@ugr.es
<https://orcid.org/0000-0001-6914-8052>

Resumen

El presente artículo recoge y analiza en 459 lenguas del mundo el número de palabras (*tokens*) y el número de sonidos y fonemas (*unidades fónicas convencionales de token* o UFCT) que emplean dichas lenguas para expresar una misma información (en el análisis central de nuestro estudio el texto fuente empleado, en concreto, son los 10 primeros artículos de la Declaración Universal de los Derechos Humanos). Asimismo, estudia las correlaciones

Abstract

This article compiles and analyses, in 459 world languages, the number of words (*tokens*) and the number of sounds and phonemes (*token conventional phonemic units* or TCPU) that these languages use to express the same information (the source text used in the main analysis of our study is, specifically, the first 10 articles of the Universal Declaration of Human Rights). Additionally, it studies the mathematical correlations existing between

Para citar este artículo: Vercher García, E. J.; Bullejos Lorenzo, M. (2022). Los índices de relatividad, densidad y eficiencia informativa en las lenguas: estudio de las correlaciones matemáticas entre palabras y fonemas. *ELUA*, (37), 23-66. <https://doi.org/10.14198/ELUA.18583>

Recibido: 05/01/2021, Aceptado: 06/05/2021

© 2022 Enrique J. Vercher García, Manuel Bullejos Lorenzo



Este trabajo está sujeto a una licencia de Reconocimiento 4.0 Internacional de Creative Commons (CC BY 4.0)

matemáticas existentes entre tokens, unidades fónicas convencionales de token y contenido informativo; correlaciones que dan lugar a los denominados índice de relatividad informativa (coeficiente resultante de dividir el número de tokens entre el número de UFCT), índice de densidad informativa (coeficiente resultante de dividir el número de UFCT entre el número de tokens), índice de eficiencia informativa léxica (coeficiente resultante de dividir 100 entre el número de tokens) e índice de eficiencia informativa fónica (coeficiente resultante de dividir 100 entre el número de UFCT). El objetivo del análisis es aportar algo más de luz a los principios matemáticos del lenguaje, a la denominada economía del lenguaje y al conocimiento sobre las características de las distintas lenguas según su tipología morfológica.

La investigación se centra en analizar la relación existente entre estos índices y la tipología morfológica predominante de cada lengua.

El resultado es una gran cantidad de datos y estadísticas de las que extraemos una serie de conclusiones referidas al uso de recursos léxicos y fónicos en las lenguas: 1) los índices de relatividad informativa y de densidad informativa tienen una relación directa con el tipo morfológico; 2) el número total de UFCT empleadas para expresar un mismo contenido de información no depende del tipo morfológico; 3) la diferencia entre el número total de UFCT empleadas por distintas lenguas puede ser muy elevada; 4) la correlación entre número de tokens y valor medio de UFCT por token muestra una relación lineal negativa (es decir, a mayor número de palabras, las longitudes de las palabras disminuyen); 5) a mayor número de tokens totales empleado para expresar una misma información corresponde por regla general un mayor número de UFCT totales (a mayor número de palabras totales, mayor número de fonemas totales); 6) la denominada economía del lenguaje no parece funcionar de igual manera en todas las lenguas, al menos en lo que a uso de recursos (fonemas) se refiere.

El presente trabajo introduce como nuevos factores de estudio y descripción de las lenguas los citados índices, así como una serie de herramientas de estudio que podrían ser aplicadas en el futuro a otras investigaciones lingüísticas.

tokens, token conventional phonemic units and informative content; correlations which result in the so-called *index of informative relativity* (coefficient resulting from dividing the number of tokens by the number of TCPU's), *index of informative density* (coefficient resulting from dividing the number of TCPU's by the number of tokens), *lexical informative efficiency index* (coefficient resulting from dividing 100 by the number of tokens) and *phonetic informative efficiency index* (coefficient resulting from dividing 100 by the number of TCPU's). The objective of the analysis is to shed more light on the mathematical principles of language, the so-called economy of language and knowledge of the characteristics of different languages based on their morphological typology.

The research is focused on analysing the relationship existing between these indices and the predominant morphological typology of each language.

The result is a large quantity of data and statistics from which we draw a series of conclusions regarding the use of lexical and phonetic resources in languages: 1) the index of informative relativity and the index of informative density depend on the morphological type; 2) the total number of TCPUs used to express the same information does not depend on the morphological type; 3) the difference between the total number of TCPUs used by different languages may be very large; 4) the correlation between the number of tokens and average TCPU value per token has a negative linear relationship (that is, the larger the number of words, the shorter the lengths of words); 5) a larger number of total tokens used to express the same information corresponds to a larger number of total TCPUs (the greater the total number of words, the greater the total number of phonemes); 6) the so-called economy of language does not seem to work in the same way in all languages, at least with regard to use of resources (phonemes).

The present article introduces the aforementioned indices as new factors of study and description of languages, as well as a series of study tools that could be applied in the future to other linguistic research.

PALABRAS CLAVE: índice de relatividad informativa; índice de densidad informativa; índice de eficiencia informativa léxica; índice de eficiencia informativa fónica; tokens.

KEYWORDS: index of informative relativity; index of informative density; lexical informative efficiency index; phonetic informative efficiency index; tokens.

1. INTRODUCCIÓN

La idea para la elaboración del presente estudio surgió de nuestra experiencia en el mundo de la traducción, en el que se constataba la gran diferencia en el número de palabras que empleaban dos lenguas distintas para expresar una misma información (texto). Comenzó, así, un arduo trabajo de búsqueda y descripción de varios cientos de idiomas con el fin de registrar el número de palabras y el número de fonemas que empleaba cada uno para expresar un mismo contenido semántico. El presente artículo es el fruto del análisis de la correlación entre el número de palabras y el número de fonemas empleados en cada una de las lenguas estudiadas, así como de la relevancia que pudiera tener en estas correlaciones el tipo morfológico.

El objetivo del análisis es aportar algo más de luz a los principios matemáticos del lenguaje, a la denominada economía del lenguaje y al conocimiento sobre las características de las distintas lenguas según su tipología morfológica. Estos datos aportarán un mayor conocimiento sobre la estructuración y funcionamiento de las lenguas.

Así pues, lo que realizamos en el presente estudio es un análisis de textos en el que cuantificamos la proporción entre el número de palabras y el número de fonemas empleados, pero también analizamos la correlación entre el número de fonemas que posee una lengua y el número de fonemas que necesita usar para transmitir una misma información.

El concepto mismo de *palabra* es polémico y carece de una definición unánime. Ha sido ampliamente debatido en la lingüística por autores como Jespersen, Admoni, Brøndal, Šahmatov, Vinogradov, González Calvo, Ušakova, Mauro, Krivososov, Ščerba, Steblin-Kamenskij, Serebrennikov, Suprun, Sunik, Migirin, Reichenbach, Potebnja, Hockett y un largo etcétera y ya lo tratamos también nosotros en otros trabajos (Vercher García, 2011: 54 y ss.; Vercher García, 2019)¹. Sea como fuere, la división de palabras también forma parte de la “lengua natural”, en tanto en cuanto los hablantes de una lengua son conscientes y coinciden en la identificación de las palabras (y aun así, existirían algunas vacilaciones). Para este estudio nos atenemos, pues, a las convenciones y normas de cada lengua en la identificación y división de sus palabras. Creemos, además, que esta decisión queda justificada por ser este un trabajo cuantitativo-estadístico, no un trabajo de análisis morfo-sintáctico o semántico, en cuyo caso sería más apropiado el uso de conceptos como los de *sintema* y *silema*, introducidos por A. Martinet (1985; 1989).

Nosotros para el presente trabajo nos vamos a valer del concepto de *token*, empleado en estudios relacionados con la carga de la información, y que en nuestro caso cuantitativamente va a equivaler a la *palabra* según la división convencional que cada lengua hace de sus propias unidades léxicas.

¹ Aprovechamos para comentar que, dado que para un trabajo de esta naturaleza se ha tenido que manejar un elevado número de obras bibliográficas, no vamos a saturar el texto con referencias, remitimos desde un primer momento a las descripciones de las distintas lenguas que se realizan en las obras que se recogen en la bibliografía.

Por otro lado, como unidades fónicas que constituyen cada uno de los token emplearemos el concepto de *unidad fónica convencional de token (UFCT)*. Las UFCT en ciertas lenguas de nuestro estudio (aquellas con un mayor alejamiento entre escritura y fonética o de las que disponemos como lingüistas de un mayor conocimiento y herramientas de estudio) están tomadas de transcripciones fonéticas y, por tanto, corresponden exactamente a sonidos; es el caso del alemán, catalán, danés, español, francés, hebreo, inglés, inuktitut, irlandés, italiano, japonés, latín, luxemburgués, portugués, tamil, telugu y turco. En el resto de casos, cuando por cuestiones técnicas no ha sido posible contar con transcripciones fonéticas fiables, y tratándose, además, de lenguas con sistemas de escritura más recientes y más fonéticas, las UFCT se corresponden con las grafías del texto escrito. Para el presente tipo de estudio (datos macroestadísticos y matemáticos) son equivalentes cuantitativamente los sonidos y los fonemas; así, por ejemplo, /'dweNde/ (transcripción fonológica) y [ˈdwɛ̃ndɛ] (transcripción fonética) tienen el mismo número de UFCT: seis².

Las diferentes proporciones entre tokens y UFCT nos dan una serie de índices sobre el número de palabras necesario para expresar un contenido semántico y la cantidad de recursos fónicos (cantidad de fonemas o sonidos) necesarios para expresar ese mismo contenido semántico.

Así, por índice de relatividad informativa entenderemos el coeficiente resultante de dividir el número de tokens entre el número de UFCT, ya que estamos hablando de la proporción entre el número de tokens y el número de UFCT empleados en un mismo texto. El índice de relatividad informativa tiene como potencial valor máximo 1 (si un idioma tuviera tantas palabras como número de UFCT).

Por su parte, denominamos índice de densidad informativa al coeficiente resultante de dividir el número de UFCT entre el número de tokens, ya que indica proporcionalmente la cantidad de información contenida en cada UFCT (evidentemente, hablamos de una proporción matemática, de uso de recursos, no de que la información —la carga semántica— pueda ser fragmentada en fonemas). Este índice viene a indicar *de facto* la longitud media de las palabras en una lengua.

Estos dos índices reflejan valores resultantes del análisis dentro de una misma lengua dada (proporción entre tokens y UFCT empleados por esa lengua).

Sin embargo, también sería útil comparar los recursos lingüísticos en forma de unidades léxicas y unidades fónicas que necesitan distintas lenguas para expresar una misma información. Así pues, si aplicamos el valor numérico de 100 a la información contenida en un texto dado, el coeficiente resultante de dividir 100 entre el número de tokens o entre el número de UFCT empleados para expresar esa misma información en distintas lenguas nos dará como resultado lo que denominaremos índice de eficiencia informativa léxica e índice de eficiencia informativa fónica respectivamente³.

2 Para el conteo de UFCT hay que precisar también que las africadas se contabilizan como una única unidad y que rasgos como el tono, la aspiración o la longitud no contabilizan como una unidad más, sino que se consideran rasgos propios de un mismo fonema y sonido y por tanto de la misma UFCT.

3 Para ser estrictos, al analizar los índices informativos a partir de un texto dado lo que obtenemos son los datos con respecto a ese texto concreto. Los índices informativos podrían variar en función del tipo de texto (literario o coloquial, técnico o poético, etc.) dentro de una misma lengua. No obstante, el presente estudio nos va a dar una idea general de los índices de relatividad, densidad y eficiencia en los que se mueven las distintas lenguas analizadas.

Los índices de eficiencia informativa léxica y fónica son valores relativos resultantes de comparar los coeficientes obtenidos en distintas lenguas sobre la base de un mismo contenido informativo (un mismo texto traducido a distintas lenguas). Evidentemente, somos conscientes de que no existe una única traducción posible exacta, pero, no obstante, creemos que el uso de ciertos textos *minoriza* este problema. Es por ello que en nuestro estudio nos valemos principalmente de la Declaración Universal de los Derechos Humanos, ya que se trata de un texto con un contenido preciso que no da lugar a muchas interpretaciones ni variaciones, además de ser un texto accesible en gran cantidad de lenguas y de ser un texto de redacción reciente y con traducciones a las distintas lenguas en un periodo bastante reducido en términos históricos y de evolución de las lenguas.

Para el presente estudio hemos analizado un total de 459 lenguas pertenecientes a familias, macrofamilias y filos (o *phyla*) lingüísticos diferentes (incluyendo lenguas aisladas) y de características lingüísticas y tipología morfológica variada. Como curiosidad, entre las lenguas analizadas se incluyen algunas lenguas artificiales, lo que nos permitirá ver los valores y correlación que tienen estas lenguas con respecto a lenguas naturales.

2. EXPOSICIÓN Y ANÁLISIS ESTADÍSTICO DE LOS ÍNDICES DE RELATIVIDAD, DENSIDAD Y EFICIENCIA INFORMATIVA DE LAS LENGUAS

En el presente trabajo nos valemos de un doble método: *estadística descriptiva* para mostrar los datos cuantitativos obtenidos durante la recogida de muestras de las distintas lenguas (se muestra en las distintas tablas y figuras insertadas) y *estadística inferencial* para extraer conclusiones basándonos en los datos obtenidos (es lo que se desarrolla en el cuerpo del texto de los apartados 3 –*Análisis de datos*– y 4 –*Conclusiones*–).

En la tabla 2 insertada como Apéndice al final del artículo exponemos los datos obtenidos en las lenguas analizadas indicando tipología morfológica, n.º de tokens, n.º de UFCT, índice de relatividad informativa, índice de densidad informativa, índice de eficiencia informativa léxica e índice de eficiencia informativa fónica. Pasamos, previamente, a aclarar algunas cuestiones sobre la misma (vid. Tabla 2 en el Apéndice).

En la columna de *tipología morfológica* indicamos el tipo morfológico predominante de cada lengua: A/I (analítica-aislante), SA (sintética aglutinante), SF (sintética fusionante) y SP (sintética polisintética). Entendemos que la moderna tipología lingüística a partir de J. Greenberg no habla de “tipos morfológicos puros”, sino de índices graduales como el índice de síntesis (que varía de lenguas más analíticas a lenguas más sintéticas) o el índice de aglutinación (entre lenguas más aglutinantes –mayor segmentación de las marcas morfológicas– y lenguas más fusionantes), además de los índices de *flexión*, *composición*, *prefijación*, *sufijación*, *aislamiento*, *flexión pura* y *concordancia* (vid. Greenberg, 1954). No obstante esta puntualización, lo cierto es que en las descripciones de las lenguas se suele hablar de predominio de un tipo morfológico y, dado el carácter de estudio macroestadístico que aquí presentamos, nos valdremos de esta clasificación en tipos morfológicos predominantes, pues nos vale para los objetivos del presente estudio. Asimismo, vamos a considerar en nuestro estudio como un mismo grupo las lenguas aislantes y analíticas, ya que la diferencia entre estas dos nociones no afecta a los resultados cuantitativos con los que trabajamos.

También incluimos una columna con observaciones sobre el texto fuente utilizado. En la mayor parte de los casos se trata de la traducción completa y correcta de los 10 primeros ar-

tículos de la Declaración Universal de los Derechos Humanos (DUDH), como se ha señalado. En estos casos se han usado las lenguas para el análisis de los cuatro índices estudiados.

En algunos casos la traducción de esos 10 primeros artículos es incompleta, hecho que hemos indicado con la letra “i”. En otros casos no nos hemos podido valer de la traducción de los 10 primeros artículos de la DUDH y hemos usado otros textos fuentes. En estos dos últimos casos hemos usado las lenguas para el análisis del índice de relatividad informativa y del índice de densidad informativa por ser valores intralingüísticos, pero no los hemos usado para el análisis del índice de eficiencia informativa léxica ni del índice de eficiencia fónica por ser valores relativos que se basan, precisamente, en la comparación con otras lenguas.

Puntualizamos igualmente que nos valemos del sistema lingüístico tal y como aparece en los textos fuentes, y que normalmente será la lengua estandarizada literaria, cuando exista, o una variante lingüística concreta en el caso de lenguas con un *continuum* de hablas (por ejemplo, el aymara). El sistema de análisis que proponemos permitiría, no obstante, hacer un estudio de una variedad lingüística dada, si así se quisiera. Por ejemplo, en ciertas zonas de Andalucía la sílaba terminada en vocal + /s/ pierde el fonema /s/ abriendo la vocal, con lo que numéricamente sería un fonema en lugar de dos en esas variedades lingüísticas.

Una vez expuestos los datos obtenidos, pasaremos a analizarlos y extraer las posibles conclusiones (estadística inferencial). Puntualizamos previamente que trabajamos con tres conjuntos de lenguas. El primer conjunto sería el conformado por todas las lenguas consultadas (459), conjunto que puede servir para el estudio de ciertas cuestiones como los índices de relatividad informativa y densidad informativa, ya que, como apuntábamos anteriormente, son coeficientes intralingüísticos independientes para cada lengua (en la Tabla 1 se muestran todas estas lenguas). Sin embargo, ciertos datos de lenguas para los que no hemos usado un mismo texto fuente (los 10 primeros artículos de la DUDH) podrían distorsionar los resultados, con lo que en nuestra investigación y análisis de datos hemos descartado aquellas lenguas que no usaban dicho texto como texto fuente y hemos trabajado también, por tanto, con un segundo conjunto de lenguas formado solo por aquellas cuyo texto fuente han sido los diez primeros artículos completos y fiables de la DUDH (376 lenguas). Finalmente, como ocurre en cualquier análisis estadístico, los valores extremos pueden distorsionar resultados, por lo que paralelamente también realizaremos un análisis valiéndonos de un tercer conjunto constituido solo por las lenguas resultantes tras depurar las que presentan valores extremos; en este caso el número de lenguas con las que hemos trabajado es de 296.

De este modo, el número de lenguas de cada tipo morfológico predominante y su porcentaje con respecto al total son los que exponemos en los cuadros 1 y 2:

Cuadro 1. Número de lenguas por tipo morfológico predominante (análisis bruto de 376 lenguas)

<i>Tipos morfológicos</i>	Frecuencia	Porcentaje
<i>A/I</i>	65	17.29 %
<i>SA</i>	172	45.74 %
<i>SF</i>	94	25 %
<i>SP</i>	45	11.97 %
<i>Todos</i>	376	100 %

Cuadro 2. Número de lenguas por tipo morfológico predominante
(análisis depurado de 296 lenguas)

<i>Tipos morfológicos</i>	Frecuencia	Porcentaje
<i>A/I</i>	55	18.58 %
<i>SA</i>	150	50.67 %
<i>SF</i>	67	22.63 %
<i>SP</i>	24	8.10 %
<i>Todos</i>	296	100 %

3. ANÁLISIS DE DATOS

De los datos obtenidos en el conjunto de todas las lenguas podemos concluir, en primer lugar, que los índices de relatividad informativa y de densidad informativa tienen una relación directa con el tipo morfológico. Así pues, las lenguas analíticas-aislantes van a emplear mayor número de palabras para expresar un mismo contenido y las lenguas polisintéticas un menor número de palabras. Por otro lado, las lenguas analíticas-aislantes presentan las palabras más breves, mientras que las lenguas polisintéticas presentan las palabras más largas. Y estos datos se consiguen tanto analizando todas las 376 lenguas que tienen los diez primeros artículos de la DUDH como texto fuente, como depurando el análisis a 296 lenguas eliminando los valores extremos. Estas conclusiones, que podrían ya *a priori* parecer lógicas, se ven confirmadas y expuestas con exactitud cuantitativamente en nuestro estudio. Es lo que mostramos en los cuadros 3 y 4, en los que incluimos las *medias* (el promedio del conjunto de valores), las *medianas* (el valor que ocupa el lugar central de todos los datos cuando estos están ordenados de menor a mayor) y las *desviaciones típicas* (medida de la dispersión de los datos respecto a la media) para tokens.

Cuadro 3. Medias, Medianas y Desviaciones típicas para tokens
(análisis bruto de 376 lenguas)

<i>Tipos morfológicos</i>	Medias para tokens	Medianas para tokens	Desviaciones típicas para tokens
<i>A/I</i>	373.5538	361.0	81.92261
<i>SA</i>	281.5116	266.0	75.50543
<i>SF</i>	291.5000	288.5	46.77026
<i>SP</i>	276.4889	247.0	81.81287

Cuadro 4. Medias, Medianas y Desviaciones típicas para tokens
(análisis depurado de 296 lenguas)

<i>Tipos morfológicos</i>	Medias para tokens	Medianas para tokens	Desviaciones típicas para tokens
<i>A/I</i>	382.1091	373.0	77.33230
<i>SA</i>	271.0333	264.0	57.81217
<i>SF</i>	298.8209	294.0	31.84381
<i>SP</i>	239.7917	231.5	43.34491

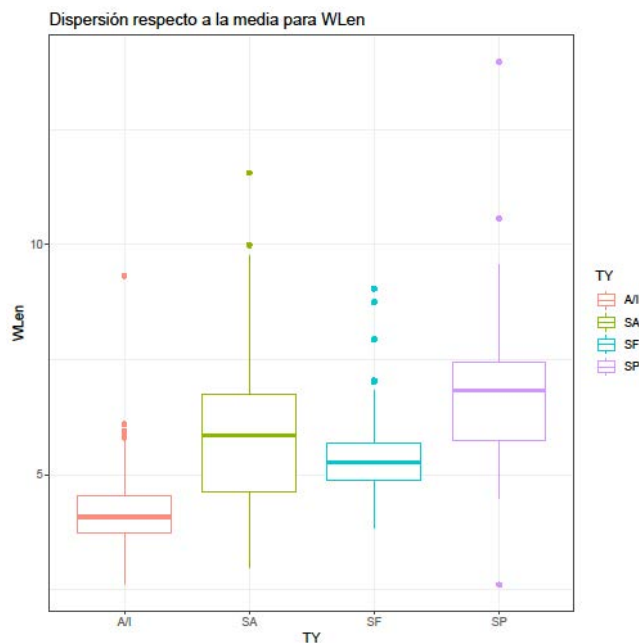
El dato de las desviaciones típicas para tokens nos indica que las lenguas sintéticas fusionantes tienden a estar más agrupadas en el valor medio del número de tokens empleado para expresar una misma información, tienen menos dispersión en términos estadísticos. Es decir, para expresar el contenido de los diez primeros artículos de la Declaración Universal de los Derechos Humanos las lenguas fusionantes han usado una media de 291 tokens y en general todas las lenguas fusionantes se mueven en torno a esa cifra (298 tokens en el análisis depurado). Sin embargo, vemos que las lenguas analíticas-aislantes han usado una media de 373 tokens (382 en el análisis depurado) y las polisintéticas, de 276 (239 en el análisis depurado), pero que hay muchas lenguas de estos dos tipos morfológicos que se alejan de esa media, cosa que ocurre tanto en el análisis del conjunto de 376 lenguas como en el depurado de solo 296 lenguas. En nuestra opinión, esta mayor coherencia de las lenguas fusionantes podría estar relacionada con el hecho de pertenecer la mayoría de ellas a un mismo filo lingüístico (el indoeuropeo), frente a la mayor dispersión genética de las lenguas analíticas-aislantes y las polisintéticas. Las lenguas aglutinantes, por su parte, emplean una media de 281 tokens (271 en el análisis depurado).

El índice de relatividad informativa, como dato que muestra la potencial capacidad semántica de cada unidad fónica con respecto al número total de tokens utilizado por esa lengua, oscila entre los 0.3840 del toba y los 0.06993 del inuktitut de Canadá. O dicho de manera sencilla, el toba va a necesitar de media apenas algo más de 2.5 fonemas para formar una palabra, mientras que el inuktitut va a necesitar más de 14 fonemas de media para formar una palabra.

Por otro lado, los datos registrados también muestran que las lenguas analíticas-aislantes presentan unos valores de índice de densidad informativa claramente menores que el resto de tipos morfológicos, con una media de 4.3035 (o dicho de manera sencilla, las palabras de las lenguas analíticas-aislantes tienen una media de 4.3 fonemas) y una dispersión muy pequeña (prácticamente todas las lenguas analíticas-aislantes se van a mover en torno a un valor muy similar a ese). En el otro extremo estarían las lenguas polisintéticas con un índice de densidad informativa de 6.8032 (es decir, las palabras de las lenguas polisintéticas tienen una media de 6.8 fonemas). Por su parte, las lenguas aglutinantes tendrían un índice de densidad informativa de 5.7769 y las fusionantes, de 5.3644⁴ (vid. Figura 1):

4 El análisis depurado con solo 296 lenguas viene a aumentar todavía más la diferencia entre grupos morfológicos (en torno al 4 de las A/I, frente a las SP que superarían el 7 de media).

Figura 1. Dispersión respecto a la media para índice de densidad informativa (WLen)



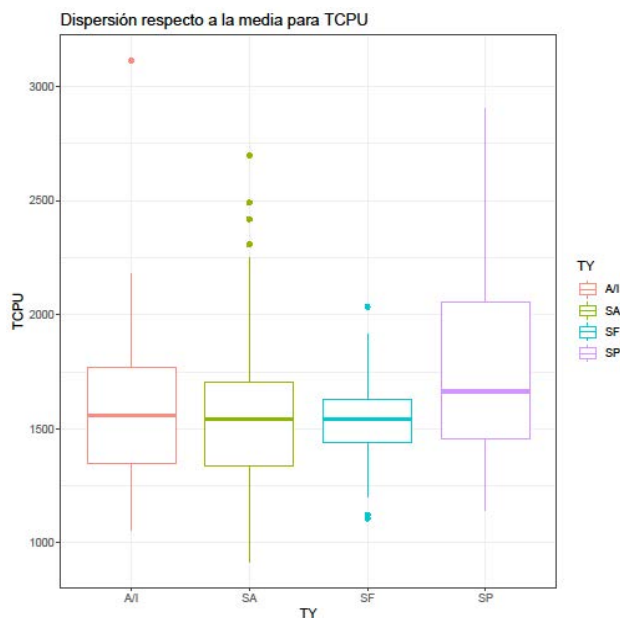
En segundo lugar, podemos concluir que el número total de UFCT empleadas para expresar un mismo contenido de información no depende del tipo morfológico. Todos los tipos morfológicos han empleado un número medio de UFCT estadísticamente similar que se mueve entre los 1,533 de media de las lenguas SF, los 1,546 de las SA, los 1,570 de las A/I y los 1,809 de las SP (y teniendo en cuenta las medianas, incluso valores más próximos: 1,558.0 las A/I, 1,543.5 las SA, 1,540.5 las SF y 1,664.0 las SP)⁵. La caja de dispersión de la media de UFCT empleadas, además, es muy pequeña, lo que quiere decir que, en general, todas las lenguas de todos los tipos morfológicos se van a mover en valores muy cercanos a la media (quizá con la excepción de las lenguas SP, que vemos que presentan una mayor dispersión en valores por encima de la media). Es lo que mostramos gráficamente en el Cuadro 5 y la Figura 2:

Cuadro 5. Medias, Medianas y Desviaciones típicas para UFCT
(análisis bruto de 376 lenguas)

<i>Tipos morfológicos</i>	Medias para UFCT	Medianas para UFCT	Desviaciones típicas para UFCT
<i>A/I</i>	1,570.431	1,558.0	337.0922
<i>SA</i>	1,546.506	1,543.5	289.2658
<i>SF</i>	1,533.691	1,540.5	166.6511
<i>SP</i>	1,809.533	1,664.0	495.1565

⁵ El análisis depurado con solo 296 lenguas arroja valores muy similares.

Figura 2. Dispersión respecto a la media para UFCT

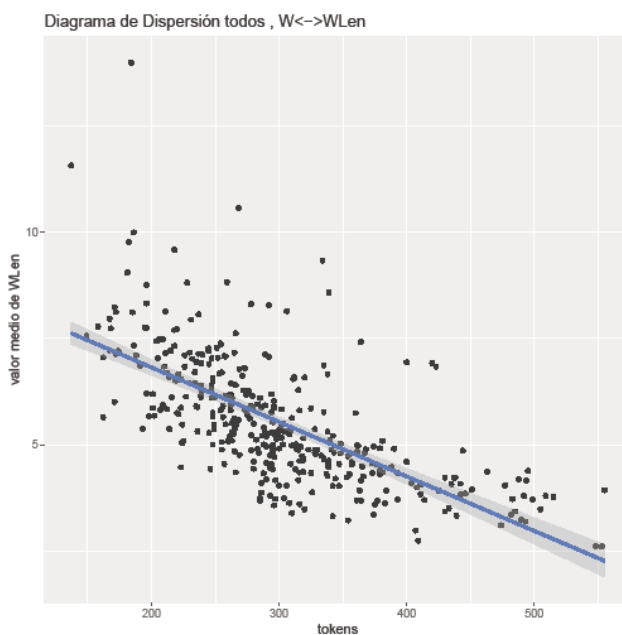


Estas dos primeras conclusiones no son incompatibles entre sí: el tipo morfológico de una lengua determina que esta emplee mayor o menor número de palabras para expresar una misma información y que sus palabras de media sean más largas o más cortas, pero también significa que el número total de fonemas o sonidos que necesita para expresar esa información no depende de su tipo morfológico. Sin embargo, nuestro estudio también revela que, aunque el tipo morfológico no determine el número total de UFCT empleadas, la diferencia entre el número total de UFCT empleadas por distintas lenguas puede ser muy elevada. Así, entre el arabela (con 2,299 UFCT empleadas y un índice de eficiencia informativa fónica de 0.0435) y el diola (914 UFCT, índice de eficiencia informativa fónica 0.1094) hay una diferencia de 2.5 veces. Y eso tomando en consideración solo las lenguas del análisis depurado, en el no depurado las diferencias son mucho mayores. Es decir, lenguas como el arabela, el amahuaca o el aguaruna emplean más del doble de fonemas que el diola, el kabyé, el kaonde o el twi para transmitir un mismo contenido informativo. Esto se ve corroborado si los comparamos con otros textos fuentes: mientras que el arabela necesita 520 UFCT para expresar el Padrenuestro, lenguas como el diola (163), el twi (170) o el wolof (183) emplean muchas menos.

Dicho esto, comprobamos, asimismo, que la correlación entre número de tokens y valor medio de UFCT por token muestra una relación lineal negativa (es decir, a mayor número de palabras, las longitudes de las palabras disminuyen). La ratio de esa relación lineal no es la misma, no obstante, en todos los grupos morfológicos. El coeficiente medio teniendo en cuenta todas las lenguas es de -0.6781 ; y teniendo en cuenta los tipos morfológicos oscilaría entre el -0.7253 de coeficiente medio que tienen las lenguas aglutinantes

y el -0.4726 de coeficiente medio que tienen las lenguas analíticas-aislantes⁶. O dicho de otra manera, en las lenguas aglutinantes tiende a reducirse muy rápidamente la longitud de las palabras si se aumenta el número de fonemas que contienen de media cada una; mientras que en el caso de las lenguas analíticas-aislantes también se tiende a reducir la longitud de las palabras si se aumenta el número de fonemas de media de cada una, pero menos bruscamente. Como vemos, los grupos morfológicos que presentan mayores valores en el número de tokens / longitud de los tokens (las SF que son las que tienen tokens más largos y menor número de tokens; y las A/I que son las que tienen tokens más cortos pero mayor número de tokens) son los que tienen una relación lineal más suave, por debajo de la media. Las lenguas SA y SF son las que presentan una relación lineal más fuerte, por encima de la media (al ir aumentando el número de tokens que necesitan, va aumentando rápidamente el número medio de UFCT que contienen los tokens de ese idioma). Vid. Figuras 3 y 4:

Figuras 3 y 4. Diagramas de dispersión (relación lineal entre n.º de tokens y valor medio de UFCT) para el conjunto de todas las lenguas y por tipo morfológico.



6 En el análisis depurado con solo 296 lenguas los datos son: -0.4157 (A/I), -0.2269 (SP), -0.7446 (SF), -0.7156 (SA), -0.7345 (media usando todos los tipos morfológicos).



Esta conclusión que acabamos de analizar no es incompatible con otra tendencia estadística presente en el estudio realizado: la de que a mayor número de tokens totales empleado para expresar una misma información corresponde por regla general un mayor número de UFCT totales (a mayor número de palabras totales, mayor número de fonemas totales), independientemente de su tipo morfológico. O ejemplificándolo con dos casos concretos: tanto el wayuu como el shipibo-conibo son lenguas polisintéticas y de media sus palabras son más largas que las de lenguas analíticas-aislantes, pero, mientras que el wayuu ha empleado 167 tokens y 1,206 UFCT para expresar los diez primeros artículos de la DUDH, el shipibo-conibo ha empleado 423 tokens y 2,892 UFCT, es decir, muchas más palabras, pero también muchos más fonemas. Un estudio como el que hemos realizado permite nuevamente, no obstante, afinar esta tendencia general: el coeficiente de correlación entre número de palabras y número de letras presenta una relación lineal positiva en torno al 0.33 (si tenemos en cuenta todas las lenguas) y es independiente del tipo morfológico. Es decir, las lenguas van a aumentar su número de fonemas si aumenta su número de palabras en una proporción similar, en torno al 0.33, siendo las lenguas A/I las que presentan una relación lineal más brusca (0.5112). En el análisis depurado con solo 296 lenguas los resultados son los siguientes: 0.8691 (A/I), 0.8293 (SP), 0.5294 (SF), 0.3866 (SA), 0.4115 (media usando todos los tipos morfológicos).

No podemos dejar de mencionar el caso de las lenguas artificiales analizadas en el presente estudio (ido, esperanto e interlingua). Los datos concretos para estas lenguas son los siguientes:

Tabla 1. Lenguas artificiales, tipología morfológica, n.º de tokens, n.º de UFCT, índice de relatividad informativa, índice de densidad informativa, índice de eficiencia informativa léxica, índice de eficiencia informativa fónica.

LENGUA	TIPOLOGÍA MORFO-LÓGICA	OBSERVACIONES SOBRE EL TEXTO FUENTE	N.º DE TOKENS	N.º DE UFCT	ÍNDICE DE RELATIVIDAD INFORMATIVA (TOKENS/UFCT)	ÍNDICE DE DENSIDAD INFORMATIVA (UFCT/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA LÉXICA (100/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA FÓNICA (100/UFCT)
Ido	SA	A	248	1,240	0.2	5	0.403225806	0.080645161
Esperanto	SA	A	314	1,538	0.204161248	4.898089172	0.318471338	0.065019506
Interlingua	SF	A	313	1,567	0.199744735	5.006389776	0.319488818	0.063816209

De estos datos extraemos la conclusión de que las lenguas artificiales no presentan valores extremos, sino que están acordes con la media general de las lenguas naturales. Sus números de tokens y de UFCT totales empleados están dentro de la media general. Sus valores para los índices estudiados y las conclusiones que podríamos extraer también son coherentes con los de las lenguas naturales. De hecho, no se ha eliminado ninguna de ellas en el análisis matemáticamente depurado. Entendemos que esto es así porque las lenguas artificiales analizadas (ido, esperanto e interlingua), a pesar de ser lenguas construidas, se basan en gran medida (por ejemplo, su vocabulario) en lenguas naturales.

A este respecto también sería interesante comprobar los datos que arrojan lenguas artificiales no basadas en lenguas naturales (por ejemplo, quenya, sindarin o klingon). Si tomamos como texto fuente el Padrenuestro vemos que los valores de los índices de relatividad informativa (quenya: 0.1990; sindarin: 0.2378; klingon: 0.1189) y de densidad informativa (quenya: 5.0244; sindarin: 4.2045; klingon: 8.4074) no sobrepasan los valores extremos ni por arriba ni por debajo de algunas lenguas naturales que hemos analizado.

4. CONCLUSIONES

Los datos arrojados por el estudio nos llevan a una conclusión sorprendente: la denominada economía del lenguaje no parece funcionar de igual manera en todas las lenguas, al menos en lo que a uso de recursos (fonemas) se refiere. Si bien es verdad, como hemos visto anteriormente, que todos los tipos morfológicos se mueven en unos valores medios similares (es decir, el número total de fonemas empleados no está en correlación con el tipo morfológico), lo cierto es que hay una gran diferencia entre el número de UFCT empleados por las lenguas que presentan los valores más bajos de número de UFCT y las lenguas con valores más altos.

Pero nuestro estudio también aporta otros datos interesantes para el conocimiento de ciertas características de las lenguas que sí van a depender de aspectos como la tipología morfológica. Datos que, en nuestra opinión, abren la puerta a nuevas categorizaciones lingüísticas más allá de la tipología morfológica tradicional o de los índices de síntesis, aglutinación, flexión, composición, prefijación, sufijación, aislamiento, flexión pura y concordancia. De este modo, entrarían como factor de estudio y descripción de las lenguas los índices de *relatividad informativa*, *densidad informativa*, *eficiencia informativa léxica* y *eficiencia informativa fónica*. En nuestro estudio nos hemos centrado en el aspecto de la tipología morfológica, pero sería aplicable a otros campos como la tipología genética, los sistemas fonético-fonológicos de las lenguas, etc.

Los índices de relatividad y densidad informativa suponen nuevos indicadores para el estudio de la tipología morfológica. Además, existe una cierta correlación entre estos índices y el índice de síntesis, ya que, por regla general, cuanto más analítica sea una lengua mayor será su relatividad informativa (y menor será su densidad informativa).

El índice de eficiencia informativa léxica puede tener aplicaciones en el campo de la traductología, ya que establece de manera precisa la variación en número de palabras que tendrá un texto en la lengua de partida y en la lengua meta (hablamos de valores generales extraídos a partir del estudio realizado, evidentemente en cada caso concreto de una traducción el número de palabras puede variar, además de que por regla general en la lengua meta de una traducción se suele aumentar el número de palabras). Esto puede ser importante en el mercado de la traducción profesional, puesto que hoy en día se suele tarifar por palabra.

Estos índices también pueden ayudar a identificar o descifrar lenguas. A modo de ejemplo curioso, el voynichés, la lengua del manuscrito Voynich (con 37,919 tokens y unas 170,000 UFCT) tiene un índice de densidad informativa inferior a 4.5, lo que descartaría, en cualquier caso, que se trate de una lengua polisintética o de un idioma con un alto índice de síntesis.

El presente trabajo supone en realidad la elaboración de una serie de herramientas de estudio que podrían ser aplicadas en el futuro a otras investigaciones (análisis de un corpus mayor, comparación entre textos fuente de diversa naturaleza dentro de una misma lengua, comparación bilateral más profunda entre dos lenguas o dos familias, incorporación de los índices presentados en este trabajo a estudios de tipología lingüística, aplicación de estos índices en el caso de lengua hablada y de otros registros lingüísticos, etc.). Queda abierto todo un campo de investigación en materia de tipología y de lingüística matemática.

REFERENCIAS BIBLIOGRÁFICAS

- Adone, D., Brück, M. A. y Gabel, A. (2018). Kot nou vire tourne nou tand li. *Serial Verb Constructions at the Interface between Grammar and Culture: Case-Study Kreol Seselwa (Seychelles Creole)*. *Quaderni di Linguistica e Studi Orientali*, 4, 15-46. <http://dx.doi.org/10.13128/QULSO-2421-7220-23838>
- Aikhenvald, A. Y. (2000). *Classifiers: A typology of noun categorization devices*. Oxford University Press.
- Ainsworth, Z. (2018). The Veps Illative: The Applicability of an Abstractive Approach to an Agglutinative Language. *Transactions of the Philological Society*. <https://doi.org/10.1111/1467-968X.12142>
- Anderson, J. D. (1895). A Collection of Kachári Folk-Tales and Rhymes. *Assam Secretariat Printing Office* (online en The Project Gutenberg <https://www.gutenberg.org/files/53506/53506-h/53506-h.htm>).
- Anderson, S. C. (2015). *A Phonological Sketch of Lamso'*. SIL. <https://www.sil.org/system/files/reapdata/16/87/00/16870014076648162723540523929249624288/LamsoPhonologySketch2015A4.pdf>.
- Aronoff, M. y Fudeman, K. (2005). *What is morphology?* Blackwell.
- Berg, A., Pretorius, R. y Pretorius, L. (2012). Exploring the Treatment of Selected Typological characteristics of Tswana in LFG. En M. Butt y T. Holloway King (eds.). *Proceedings of the LFG12 Conference*. <http://web.stanford.edu/group/cslipublications/cslipublications/LFG/17/papers/lfg12bergetal.pdf>.
- Beshears, A. (2017). *The Demonstrative Nature of the Hindi/Marwari Correlative* [tesis doctoral, Queen Mary University of London]. https://www.sil.org/system/files/reapdata/13/53/97/135397489734230802427407731441106180531/Beshears_A_PhD_131017.pdf.

- Bhosale, G., Kembhavi, S., Amberkar, A., Mhatre, S. y Popale, L. (2011). Processing of Kridanta (Participle) in Marathi. En *Proceedings of ICON-2011: 9th International Conference on Natural Language Processing*. Macmillan Publishers. <https://www.cse.iitb.ac.in/~pb/papers/icon11-marathi-kridant.pdf>.
- Biggs, B. (1960). Morphology-Syntax in a Polynesian Language. *The Journal of the Polynesian Society*, 69 (4), 376-379. https://www.jstor.org/stable/20703854?seq=1#page_scan_tab_contents
- Borchers, D. (2008). *A Grammar of Sunwar: Descriptive Grammar, Paradigms, Texts and Glossary*. BRILL.
- Brassett, C., Brassett, P. y Lu, M. (2006). *The Tujia language*. Lincom Europa.
- Cope, A. T. (1966). *Zulu Phonology, Tonology and Tonal Grammar* [tesis doctoral, University of Natal] <https://researchspace.ukzn.ac.za/handle/10413/3320>
- Coupe, A. R. (2007). *A Grammar of Mongsen Ao*. De Gruyter Mouton. <https://doi.org/10.1515/9783110198522>
- Creissels, D. (2019). Morphology in Niger-Congo Languages. En *The Oxford Encyclopedia of Morphology*. <http://www.denisceissels.fr/public/Creissels-morphNC.pdf>. <https://doi.org/10.1093/acrefore/9780199384655.013.535>
- Damonte, F. (2005). The Mirror Principle and the Order of Verbal Extensions: Evidence from Pular. En G. Booij *et al.* (eds.). *On-line Proceedings of the Fifth Mediterranean Morphology Meeting* (pp. 337-358). University of Bologna.
- DeLancey, S. (2011). On the Origins of Sinitic. En Z. Jing-Schmidt (ed.). *Proceedings of the 23rd North American Conference on Chinese Linguistics*, (vol.1., pp. 51-64). University of Oregon.
- Driem, G. van (1987). *A Grammar of Limbu*. Mouton De Gruyter. <https://doi.org/10.1515/9783110846812>
- Ferguson, C. A. (1982). Simplified registers and linguistic theory. En L. K. Obler y L. Menn (ed.). *Exceptional Language and Linguistics* (pp. 49-66). Academic Press.
- Gelderen, E. van (2016). *Cyclical Change Continued*. John Benjamins. <https://doi.org/10.1075/la.227>
- Genetti, C. (2004). *Tibeto-Burman languages of Nepal: Manange and Sherpa*. Pacific Linguistics-The Australian National University.
- Gerner, M. (2013). *A Grammar of Nuosu*. Mouton De Gruyter. <https://doi.org/10.1515/9783110308679>
- Gilley, L. (2004). *Morphophonemic Orthographies in Fusional Languages: The Cases of Dinka and Shilluk*. University of Khartoum- SIL International.
- Good, J. (2012). How to become a “Kwa” noun. *Morphology*, 22(2), 1-37. <https://doi.org/10.1007/s11525-011-9197-2>
- Greenberg, J. H. (1954). A Quantitative Approach to the Morphological Typology of Languages. En R. F. Spencer (ed.). *Method and Perspective in Anthropology: Paper in Honor of Wilson D. Wallis* (pp. 192-220). University of Minnesota Press.
- Greksáková, Z. (2018). *Tetun in Timor-Leste: The Role of Language Contact in its Development*. [Tesis doctoral, Universidade de Coimbra]. <https://estudogeral.uc.pt/handle/10316/80665>
- Hale, A. y Sresthacharya, I. (1972). *Toward a revision of Hale's roman Newari orthography*. Kathmandu: Summer Institute of Linguistics (disponible online <https://www.sil.org/resources/archives/36851>).
- Hardman, M. J., Vásquez, J., Yapita, J. de D., *et al.* (2001). *Aymara: Compendio de estructura fonológica y gramatical*. Instituto de Lengua y Cultura Aymara. <https://fhcevirtual.umsa.bo/bteavirtual/?q=node/166>
- Hari, A. M. (2010). *Yohlmo Grammar Sketch*. Tribhuvan University.
- Haspelmath, M., Dryer, M., Gil, D. y Comrie, B. (2005). *The World Atlas of Language Structures*. Oxford University Press.
- Haspelmath, M. y Michaelis, S. M. (2017). Analytic and synthetic: Typological change in varieties of European languages. En I. Buchstaller y B. Siebenhaar (eds.). *Language Variation - European Perspectives VI: Selected papers from the 8th International Conference on Language Variation in Europe (ICLaVE 8), Leipzig 2015*. John Benjamins. <https://doi.org/10.1075/silv.19.01has>
- Heine, B. y Nurse, D. (2008). *A linguistic geography of Africa*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511486272>

- Jitwiriyanon, S. (2019). Ban Pa La-U Sgaw Karen Tones: an Analysis of Semitones, Quadratic Trendlines and Coefficients. En *RGJ Seminar Series LXXXII on Southeast Asian Linguistics* (pp. 60-77). Research Institute for Languages and Cultures of Asia at Mahidol University.
- Kilarski, M. (2013). *Nominal Classification: A History of its Study from the Classical Period to the Present*. John Benjamins Publishing Company. <https://doi.org/10.1075/sihols.121>
- King, D. (2010). *Voice and Valence-Altering Operations in Falam Chin: A Role and Reference Grammar Approach* [Tesis doctoral, The University of Texas at Arlington] http://www.tibetoburman.net/files/King_2010_Voice_and_valence-altering_operations_in_Falam_Chin.pdf.
- King, J. T. (2008). *A grammar of Dhimal* [Tesis doctoral, Leiden University]. <https://openaccess.leidenuniv.nl/bitstream/handle/1887/13072/Kopie%20van%20Dhimal%20Grammar.pdf?sequence=1>
- Konnerth, L., Morey, S., Sarmah, P. y Amos, T. (2015). *North East Indian Linguistics 7*. The Australian National University. <https://core.ac.uk/download/pdf/156698443.pdf>.
- Kornai, A. (2001). *Mathematical Linguistics*. Springer-Verlag https://eprints.sztaki.hu/7913/1/Kornai_1762289_ny.pdf
- Krishan, S. (2001). A Sketch of Raji Grammar. 国立民族学博物館調査報告, 19, 449-497. <http://doi.org/10.15021/00002153>
- Laboratorio de Lengua y Cultura Víctor Franco (2019). *Jñatjo (mazahua)*. Centro de Investigaciones y Estudios Superiores en Antropología Social.
- Launey, M. (1992). *Introducción a la lengua y a la literatura Náhuatl*. Universidad nacional autónoma de México.
- Lee, S.-W. (2011). *Eastern Tamang Grammar Sketch* [Trabajo de fin de máster, Graduate Institute of Applied Linguistics]. <https://www.sil.org/resources/archives/42567>.
- Leufkens, S. (2011). Kharia: A transparent language. *Linguistics in Amsterdam*, 4, 75-95. https://www.researchgate.net/publication/241888149_Kharia_a_transparent_language.
- Mandelbrot, B. (1961). On the theory of word frequencies and on related markovian models of discourse. En R. Jakobson (ed.). *Structure of language and its mathematical aspects* (pp. 190-219). American Mathematical Society. <https://doi.org/10.1090/psapm/012/9970>
- Manson, K. (2003). *Karenic Language Relationships: A Lexical and Phonological Analysis*. Payap University. https://www.sil.org/system/files/reapdata/76/92/97/76929773418782758038670312229278361244/Manson_2003_Karen_lg_relationships.pdf. <https://doi.org/10.1024/0369-8394.92.3.97>
- Martinet, A. (1955). *Économie des changements phonétiques. Traité de phonologie diachronique*. Éditions A. Francke.
- Martinet, A. (1987). *Sintaxis general*. Gredos.
- Martinet, A. (1993). *Función y dinámica de las lenguas*. Gredos.
- Mchombo, S. (2004). *The Syntax of Chichewa*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511486302>
- Mensah, E. O. (2010). On Efik Prefixing Morphology. *The Buckingham Journal of Language and Linguistics*, 3, 187-203. <https://doi.org/10.5750/bjll.v3i0.31>
- Michaud, A. (2012). The Complex Tones of East/Southeast Asian Languages: Current Challenges for Typology and Modelling. *Third International Symposium on Tonal Aspects of Languages* (TAL2012), Nanjing, China, 2012, pp. 1-7. HAL Id: halshs-00676251. <https://halshs.archives-ouvertes.fr/halshs-00676251>.
- Micheli, I. (2018). *Grammatical Sketch and Short Vocabulary of the Ogiek Language of Mariashoni*. Università di Trieste.
- Mihas, E. (2010). *Essentials of Ashéninka Perené Grammar* [Tesis doctoral, University of Wisconsin-Milwaukee]. https://www.researchgate.net/publication/50198629_ESSENTIALS_OF_ASHENINKA_PERENE_GRAMMAR.
- Mnguhenen Sokpo, R. (2016). An Autosegmental Analysis of Tiv Phonology [Tesis doctoral, Benue State University] <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.424.286&rep=rep1&type=pdf>

- Moran, S. y McCloy, D. (2019). *PHOIBLE 2.0*. Max Planck Institute for the Science of Human History. <http://phoible.org> (21-10-2019).
- Moreno Cabrera, J. C. (2003). Síntesis y análisis en las lenguas. Crítica de la morfología clásica y de algunas de sus aplicaciones sincrónicas y diacrónicas. *Estudios de Lingüística de la Universidad de Alicante*, 17, 465-504. <https://core.ac.uk/download/pdf/16359747.pdf>
- Moreno Cabrera, J. C. (1997). *Introducción a la Lingüística. Enfoque tipológico y universalista*. Síntesis.
- Musumali, G. (2014). *Zambezi Lunda and Zambezi Luvale. Phonological and Lexical Similarities. Series 1. Phonemic contrasts*. Autoedición. https://www.academia.edu/39625196/series_1_phonemic_contrasts_of_lunda_and_luvale.
- Mutaka, N. M. (2007). *Kinande: A Grammar Sketch* (Version 1.0). Rutgers, The State University of New Jersey-Afranaph Project.
- Naula Guacho, J. y Burns, D. D. (1975). *Bosquejo gramatical del quichua de Chimborazo*. Universidad Central del Ecuador.
- Nchare, A. L. (2012). *The Grammar of Shupamem*. New York University. ling.auf.net/lingbuzz/001441/current.pdf.
- Ndze, W. F. (2015). *A Comparative Morphological Analysis of Lamnsó and the English Compound Words*. Ahmadu Bello University.
- Nurse, D. (2007). Did the Proto-Bantu verb have a Synthetic or an Analytic Structure? *SOAS Working Papers in Linguistics*, 15, 239-256. <https://www.soas.ac.uk/linguistics/research/workingpapers/volume-15/file37812.pdf>.
- Osada, T. (2008). Mundari. En D.S. Anderson (ed.). *The Munda Languages* (pp. 99-164). Routledge.
- Overall, S. E. (2007). *A Grammar of Aguaruna*. Research Centre for Linguistic Typology-La Trobe University.
- Owen-Smith, T. (2014). *Grammatical Relations in Tamang, a Tibeto-Burman Language of Nepal* [Tesis doctoral, University of London] https://eprints.soas.ac.uk/23664/1/Owen-Smith_4167.pdf.
- Paster, M. (2010). The Verbal Morphology and Phonology of Asante Twi. *Studies in African Linguistics*, 39(1), 77-120. <https://doi.org/10.32473/sal.v39i1.107285>
- Paudyal, K. P. (2013). *A Grammar of Chitoniya Tharu* [Tesis doctoral, Tribhuvan University]. <http://202.45.147.21:8080/jspui/bitstream/123456789/297/1/75.%20Krishna%20Prasad%20Paudyal.pdf>.
- Passy, P. (1890). *Etude sur les changements phonétiques et leurs caractères généraux*. Firmin-Didot.
- Payne, D. L. (1974). *Nasality in Aguaruna*. The University of Texas at Arlington.
- Peirce, C. S. S. (1906). Prolegomena to an Apology for Pragmaticism. *The Monist*, 16, 492-546. <https://archive.org/details/jstor-27899680>. <https://doi.org/10.5840/monist190616436>
- Peterson, J. (2010). *Kharia Texts. Glossed, translated and annotated by John Peterson*. Autoedición. https://pure.mpg.de/rest/items/item_402190_2/component/file_679587/content.
- Phillips, A. (1999). *Western Thailand Pwo Karen Text Collection*. Thammasat University-Summer Institut of Linguistics.
- Plaisier, H. (1968). *A Grammar of Lepcha* [Tesis doctoral, Universidad de Leiden]. <https://scholarlypublications.universiteitleiden.nl/handle/1887/4379>
- Poeta, T. (2016). *Reference to objects in Makhuwa and Swahili discourse*. University of London. https://eprints.soas.ac.uk/26161/1/4437_Poeta.pdf.
- Pokharel, M., Donohue, M. y Gautam, B. (AÑO). Kusunda linguistics. The Australian National University. <http://kusunda.linguistics.anu.edu.au/index.php>
- Poletto, R. (1998). *Topics in Runyankore Phonology* [Tesis doctoral, The Ohio State University].
- Provencher, C. (2012). *A Morpho-Syntactic Analysis of Marshallese Determiner Phrases* [Tesis doctoral, Université du Québec à Montréal]. <https://archipel.uqam.ca/4685/1/M12379.pdf>.
- Rai, T. M. (2017). *A Sociolinguistic Survey of Chamling: A Tibeto-Burman Language*. Tribhuvan University. <http://cdltu.edu.np/site/images/linsun/chamling.pdf>.

- Ramos, P. A. (2015). *La fonología y morfología del zapoteco de San Pedro Mixtepec* [Tesis doctoral, Centro de Investigaciones y Estudios Superiores en Antropología Social]. <http://repositorio.ciesas.edu.mx/bitstream/handle/123456789/399/D278.pdf?sequence=1&isAllowed=y>
- Regmi, D. R. y Regmi, A. (2011). *A Sociolinguistic Survey of Bhujel: A Tibeto-Burman Language*. Tribhuvan University. <http://cdltu.edu.np/site/images/linsun/bhujel.pdf>.
- Rich, F. (1963). Arabela phonemes and high-level phonology. En Benjamin F. Elson (ed.). *Studies in Peruvian Indian Languages: I* (pp. 193-206). Summer Institute of Linguistics of the University of Oklahoma.
- Ringmacher, M. (1989). Los modelos tipológicos y la descripción del Guaraní. *Amerindia*, 14. https://www.researchgate.net/publication/268000650_Los_modelos_tipologicos_y_la_descripcion_del_Guarani.
- Romano, A. (2018). Preliminary Analysis of Ditammari Tonal Patterns. *Studi e Ricerche*, 3, 251-261.
- Sapir, J. D. (1969). *A Grammar of Diola-Fogny*. Cambridge University Press.
- Sayce, A. H. (1880 [2015]). *Introduction to the Science of Language. Volume 2*. Cambridge University Press.
- Schackow, D. (2015). *A grammar of Yakkha* (Studies in Diversity Linguistics 7). Language Science Press-Freie Universität. https://doi.org/10.26530/OAPEN_603340
- Schadeberg, T. C. (1982). Nasalization in Umbundu. *Journal of African Languages and Linguistics*, 4, 109-132. https://openaccess.leidenuniv.nl/bitstream/handle/1887/8809/5_1234891_024.pdf. <https://doi.org/10.1515/jall.1982.4.2.109>
- Sibanda, G. (2004). *Verbal Phonology and Morphology of Ndebele* [Tesis doctoral, University of California]. <https://escholarship.org/uc/item/6cf9w3j2>.
- Štekauer, P., Valera, S. y Kórtvélyessy, L. (2012). *Word-Formation in the World's Languages: A Typological Survey*. Cambridge University Press. <https://doi.org/10.1515/lity-2013-0022>
- Sweet, H. (1888). *A history of English sounds from the earliest period*. Clarendon Press.
- Thurgood, G. y LaPolla, R. J. (2003). *The Sino-Tibetan Languages*. Routledge.
- Turin, M. (2012). *A grammar of the Thangmi language: with an ethnolinguistic introduction to the speakers and their culture*. BRILL.
- UNITED NATIONS, *Universal Declaration of Human Rights*. https://readtiger.com/www.unicode.org/udhr/assemblies/full_all.html and <https://www.ohchr.org/SP/UDHR/Pages/SearchByLang.aspx>.
- Vendryes, J. (1939). Parler par Économie. En *Mélanges de linguistique offerts à Charles Bally* (pp. 49-62). Université de Genève-Georg & Cie.
- Vercher García, E. J. (2011). *Estudio comparado de partículas modales en ruso y español*. EAE Publishing.
- Vercher García, E. J. (2019). El modelo estructural-funcional del sistema de clases de palabras. El caso de las partículas modales en la lengua española. *Analecta Malacitana* (AnMal Electrónica), 46, 111-130.
- Vesalainen, O. (2016). *A Grammar Sketch of Lhomi*. SIL International. https://www.sil.org/system/files/reapdata/27/32/78/27327800011828341569489257185488473676/sillcdd_34.pdf.
- Vuillermet, M. (2013). *The grammar of fear in Ese'ejá. Syntax Circle*. University of California.
- Vydrin, V. (2016). Tonal inflection in Mande languages: The cases of Bamana and Dan-Gwtaa. En E. Palancar y L. Jean Léo (eds.). *Tone and Inflection: New facts and new perspectives* (pp. 83-105). De Gruyter. <https://doi.org/10.1515/9783110452754-005>
- Walls, T. (2015). *Tsafiki: A Linguistic Analysis of the Ecuadorian Language*. [Trabajo de fin de grado, University of Central Arkansas] https://www.academia.edu/12048805/Tsafiki_A_Linguistic_Analysis_of_the_Ecuadorian_Language.
- Welmers, W. E. (1952). Notes on the Structure of Bariba. *Language*, 28 (1), 82-103. <https://doi.org/10.2307/409991>
- Wetzel, L. (2018). Types and Tokens. En E. N. Zalta (ed.). *The Stanford Encyclopedia of Philosophy* (Fall 2018 Edition). The Metaphysics Research Lab.

- Wilde, C. P. (2008). *A Sketch of the Phonology and Grammar of Rājibanshi*. University of Helsinki <https://helda.helsinki.fi/bitstream/handle/10138/19290/asketcho.pdf?sequence=1&isAllowed=y>.
- Wojtylak, K. (2017). Verbal classifiers in Murui (Witotoan) - What are they? En International Workshop “Genders and classifiers in Amazonia and beyond”, 10 de agosto de 2017. https://www.academia.edu/34231601/Verbal_Classifiers.
- Wozny, D. y Cassin, B. (dirs.) (2014). *Les intraduisibles du patrimoine en Afrique subsaharienne*. Demopolis. <https://doi.org/10.4000/books.demopolis.515>
- Zariquiey Biondi, R. (2011). *A grammar of Kashibo-Kakataibo* [Tesis doctoral, La Trobe University]. http://etnolinguistica.wdfiles.com/local--files/tese%3Azariquiey-2011/zariquiey_biondi_2011.pdf.
- Zerbian, S. (2007). A First Approach to Information Structuring in Xitsonga/Xichangana. *SOAS Working Papers in Linguistics*, 15, 65-78. <https://www.soas.ac.uk/linguistics/research/workingpapers/volume-15/file37800.pdf>.
- Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort*. Addison-Wesley Press.

Expresamos nuestro agradecimiento a Tania García Arévalo, Taidor Lam, Vicente García Rojas, Elena Erickson de Hollenbach, Klaus Zimmermann, Nicholas R. Rolle, Faust Noam, Justin B. Hopkins, John Smith, Ronald Asher, Maria Keet, Joan Byamugisha, Victor Tapuyo Pianchiche, Huia Jahnke, Charles Marfo, Regina Caesar y Hona Black.

APÉNDICE

Tabla 2. Lenguas, tipología morfológica, n.º de tokens, n.º de UFCT, índice de relatividad informativa, índice de densidad informativa, índice de eficiencia informativa léxica, índice de eficiencia informativa fonética.

LENGUA	TIPOLOGÍA MORFOLÓGICA	OBSERVACIONES SOBRE EL TEXTO FUENTE	N.º DE TOKENS	N.º DE UFCT	ÍNDICE DE RELATIVIDAD INFORMATIVA (TOKENS/UFCT)	ÍNDICE DE DENSIDAD INFORMATIVA (UFCT/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA LÉXICA (100/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA FONÉTICA (100/UFCT)
Abkhaz	SA		211	1,508	0.139920424	7.146919431	0.473933649	0.066312997
Achehese	SA		325	1,901	0.170962651	5.849230769	0.307692308	0.052603893
Achuar-Shiwiar	SA		289	2,060	0.140291262	7.128027682	0.346020761	0.048543689
Adja	A/I		411	1,575	0.260952381	3.832116788	0.243309002	0.063492063
Adyge	SA		251	1,826	0.137458927	7.274900398	0.398406375	0.054764513
Ajaan Oromo (Oromiffa)	SA		239	1,502	0.159121172	6.284518828	0.418410042	0.066577896
Afrikaans	A/I		295	1,473	0.200271555	4.993220339	0.338983051	0.067888663
Aguaruna	SA		338	2,250	0.150222222	6.656804734	0.295857988	0.044444444
A'ingae	SP		206	1,542	0.133592737	7.485436893	0.485436893	0.064850843
Albanian	SF		317	1,569	0.202039516	4.949526814	0.315457413	0.063734863
Altay	SA		296	1,735	0.170605187	5.861486486	0.337837838	0.057636888
Amahuaca	SP		259	2,284	0.113397548	8.818532819	0.386100386	0.043782837
Amarakaeri	SP		292	2,062	0.141610087	7.061643836	0.342465753	0.048496605
Amazigh / Tamazight, Central Atlas	SA		319	1,490	0.21409396	4.670846395	0.313479624	0.067114094
Amis	SA		265	1,283	0.206547155	4.841509434	0.377358491	0.077942323
Amesha-Yanesha	SP		420	2,903	0.144677919	6.911904762	0.238095238	0.034447124

LENGUA	TIPOLOGÍA MORFOLÓGICA	OBSERVACIONES SOBRE EL TEXTO FUENTE	N.º DE TOKENS	N.º DE UFCT	ÍNDICE DE RELATIVIDAD INFORMATIVA (TOKENS/UFCT)	ÍNDICE DE DENSIDAD INFORMATIVA (UFCT/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA LÉXICA (100/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA FÓNICA (100/UFCT)
Ao	SA		214	1,148	0.18641115	5.364485981	0.46728972	0.087108014
Arabela	SP		335	2,299	0.145715528	6.862686567	0.298507463	0.043497173
Arabic	SF		278	1,645	0.16899696	5.917266187	0.35971223	0.060790274
Armenian	SA		261	1,609	0.162212554	6.164750958	0.383141762	0.062150404
Asante	A/I		316	1,138	0.277680141	3.601265823	0.316455696	0.087873462
Asháninka	SP		218	2,088	0.10440613	9.577981651	0.458715596	0.04789272
Assamese	SF	*	23	138	0.166666667	6	n/a	n/a
Assyrian (Aoraya)	SF	*	185	920	0.201086957	4.972972973	n/a	n/a
Asturian (Bable)	SF		340	1,740	0.195402299	5.117647059	0.294117647	0.057471264
Awadhi	SF	*	66	280	0.235714286	4.242424242	n/a	n/a
Awapit	SA		195	1,512	0.128968254	7.753846154	0.512820513	0.066137566
Aymara	SA		158	1,228	0.128664495	7.772151899	0.632911392	0.081433225
Ayoreo	SF	i	185	796	0.23241206	4.302702703	n/a	n/a
Azeri/Azerbaijani	SA		272	1,707	0.159343878	6.275735294	0.367647059	0.058582308
Bahasa Indonesia	SA		296	1,785	0.165826331	6.030405405	0.337837838	0.056022409
Bahasa Melayu (Malay)	SA		271	1,793	0.151143335	6.616236162	0.36900369	0.055772448
Bai	A/I		283	1,684	0.168052257	5.950530035	0.35335689	0.059382423
Belanda Viri	SA		342	1,133	0.301853486	3.312865497	0.292397661	0.088261253
Balinese	SA		294	1,819	0.161627268	6.18707483	0.340136054	0.054975261
Balochi	SF	*	291	1,307	0.222647284	4.491408935	n/a	n/a

LENGUA	TIPOLOGÍA MORFOLÓGICA	OBSERVACIONES SOBRE EL TEXTO FUENTE	N.º DE TOKENS	N.º DE UFCT	ÍNDICE DE RELATIVIDAD INFORMATIVA (TOKENS/UFCT)	ÍNDICE DE DENSIDAD INFORMATIVA (UFCT/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA LÉXICA (100/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA FÓNICA (100/UFCT)
Bambara	A/I		290	1,215	0.238683128	4.189655172	0.344827586	0.082304527
Bamum	SA		433	1,515	0.285808581	3.498845266	0.230946882	0.066006601
Baoule/Baule	A/I		490	1,586	0.308953342	3.236734694	0.204081633	0.063051702
Bari	SA		347	1,396	0.248567335	4.023054755	0.288184438	0.071633238
Basque (Euskara)	SA		253	1,688	0.149881517	6.671936759	0.395256917	0.059241706
Batonu (Bariba)	SA		310	1,052	0.294676806	3.393548387	0.322580645	0.095057034
Belorus (Belaruski)	SF		286	1,679	0.170339488	5.870629371	0.34965035	0.059559261
Bemba	SA		256	1,512	0.169312169	5.90625	0.390625	0.066137566
Bengali	SF		234	1,569	0.149139579	6.705128205	0.427350427	0.063734863
Béti	SA	i	209	778	0.268637532	3.722488038	n/a	n/a
Bhojpuri	SF		294	1,556	0.188946015	5.292517007	0.340136054	0.064267352
Bhujel (Gharti)	SA	*	211	1,202	0.175540765	5.696682464	n/a	n/a
Bichelamar	A/I		430	1,816	0.236784141	4.223255814	0.23255814	0.055066079
Bikol/Bicolano	SA		342	1,781	0.192026951	5.207602339	0.292397661	0.056148231
Bizisa, Bizika, Tujia	A/I		273	1,590	0.171698113	5.824175824	0.366300366	0.062893082
Bodo (Kachari)	SA	*	556	3,313	0.167823725	5.958633094	n/a	n/a
Bora	SP		228	2,008	0.113545817	8.807017544	0.438596491	0.049800797
Bosnian	SF		263	1,391	0.18907261	5.288973384	0.380228137	0.071890726
Breton	SF		293	1,302	0.225038402	4.443686007	0.341296928	0.076804916
Bugisnese	SA		275	1,726	0.159327926	6.276363636	0.363636364	0.057937428

LENGUA	TIPOLOGÍA MORFOLÓGICA	OBSERVACIONES SOBRE EL TEXTO FUENTE	N.º DE TOKENS	N.º DE UFCT	ÍNDICE DE RELATIVIDAD INFORMATIVA (TOKENS/UFCT)	ÍNDICE DE DENSIDAD INFORMATIVA (UFCT/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA LÉXICA (100/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA FÓNICA (100/UFCT)
Bulgarian (Balgarski)	SF		299	1,588	0.188287154	5.3111036789	0.334448161	0.062972292
Burmese/Myanmar	A/I		334	3,115	0.107223114	9.326347305	0.299401198	0.032102729
Cakchiquel	SA		444	2,157	0.205841446	4.858108108	0.225225225	0.046360686
Candoshi-Shapra	SP		237	1,910	0.12408377	8.05907173	0.421940928	0.052356021
Cantones	A/I		313	1,382	0.226483357	4.415335463	0.319488818	0.07223589
Caquite	SP		268	2,832	0.094632768	10.56716418	0.373134328	0.035310734
Cashibo-Cacetaibo	SP		379	1,862	0.203544576	4.91292876	0.263852243	0.053705693
Catalan (Català)	SF		295	1,433	0.205861828	4.857627119	0.338983051	0.069783671
Cebuano	SA		304	1,617	0.188002474	5.319078947	0.328947368	0.061842919
Chaa'pala (Cayapa)	SA	*	574	3,215	0.178538103	5.601045296	n/a	n/a
Chamorro	SA		319	1,642	0.194275274	5.147335423	0.313479624	0.06090134
Chayahuita	SP		185	1,501	0.123251166	8.113513514	0.540540541	0.066622252
Checheva (Nyanja)	SA		239	1,528	0.156413613	6.393305439	0.418410042	0.065445026
Chickasaw	SA	i	85	769	0.11053316	9.047058824	n/a	n/a
Chin Falam	SA		483	1,798	0.268631813	3.722567288	0.207039337	0.055617353
Chin Hakha	SA		495	2,169	0.228215768	4.381818182	0.202020202	0.046104195
Chin Matu (Nga La)	A/I	*	363	1,603	0.226450405	4.415977961	n/a	n/a
Chin Tiddim	SA		354	1,673	0.211595935	4.725988701	0.282485876	0.059772863
Chinanteco	A/I		361	1,332	0.271021021	3.689750693	0.27700831	0.075075075

LENGUA	TIPOLOGÍA MORFOLÓGICA	OBSERVACIONES SOBRE EL TEXTO FUENTE	N.º DE TOKENS	N.º DE UFCT	ÍNDICE DE RELATIVIDAD INFORMATIVA (TOKENS/UFCT)	ÍNDICE DE DENSIDAD INFORMATIVA (UFCT/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA LÉXICA (100/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA FÓNICA (100/UFCT)
Chimanteco, Ajiñán	A/I	i	269	798	0.337092732	2.966542751	n/a	n/a
Chinese (Mandarin)	A/I		294	1,319	0.222896133	4.486394558	0.340136054	0.075815011
Chiquitano	SP	*	227	1,502	0.151131824	6.616740088	n/a	n/a
Chuuk (Trukese)	SA		388	1,735	0.223631124	4.471649485	0.257731959	0.057636888
Chuvash	SA		316	1,834	0.172300981	5.803797468	0.316455696	0.054525627
Cokwe	SA		224	1,225	0.182857143	5.46875	0.446428571	0.081632653
Corsican	SF		318	1,501	0.211858761	4.720125786	0.314465409	0.066622252
Crioulo (Cabo Verde)	A/I		437	1,843	0.237113402	4.217391304	0.228832952	0.05425936
Crioulo da Guiné-Bissau (Guinea Bissau Creole)	SF		316	1,458	0.216735254	4.613924051	0.316455696	0.068587106
Croatian	SF		258	1,413	0.182590234	5.476744186	0.387596899	0.070771408
Cymraeg (Welsh)	SF		290	1,436	0.201949861	4.951724138	0.344827586	0.069637883
Czech (Cesky)	SF		248	1,406	0.176386913	5.669354839	0.403225806	0.071123755
Dagaare	SA		430	1,474	0.291723202	3.427906977	0.23255814	0.067842605
Daghani	SA		393	1,456	0.269917582	3.704834606	0.254452926	0.068681319
Dangme	A/I		548	1,431	0.382948987	2.611313869	0.182481752	0.069881202
Danish (Dansk)	SF		283	1,384	0.204479769	4.890459364	0.353356689	0.072254335
Dendi	SA	i	241	868	0.27764977	3.601659751	n/a	n/a
Dhanuk (Magahi)	SF		255	1,245	0.204819277	4.882352941	0.392156863	0.080321285

LENGUA	TIPOLOGÍA MORFOLÓGICA	OBSERVACIONES SOBRE EL TEXTO FUENTE	N.º DE TOKENS	N.º DE UFCT	ÍNDICE DE RELATIVIDAD INFORMATIVA (TOKENS/UFCT)	ÍNDICE DE DENSIDAD INFORMATIVA (UFCT/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA LÉXICA (100/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA FÓNICA (100/UFCT)
Dhimal (Dhimal)	SA	*	329	1,927	0.170731707	5.857142857	n/a	n/a
Dine. Navajo (Navaho)	SP		273	1,857	0.147011309	6.802197802	0.366300366	0.053850296
Dinka	SF	i	301	1,077	0.279480037	3.57807309	n/a	n/a
Diola (Jola-Fogny)	SA		162	914	0.177242888	5.641975309	0.617283951	0.10940919
Dioula	A/I		482	1,620	0.297530864	3.360995851	0.20746888	0.061728395
Ditammari	SA		285	1,049	0.271687321	3.680701754	0.350877193	0.095328885
Drung (Dulong, Tyrung)	SA	i	181	1,035	0.174879227	5.718232044	n/a	n/a
Dutch (Nederlands)	SF		294	1,694	0.173553719	5.761904762	0.340136054	0.059031877
Dzongkha/Bhutanese	SA	*	207	1,017	0.203539823	4.913043478	n/a	n/a
Edo	A/I		414	1,530	0.270588235	3.695652174	0.241545894	0.065359477
Efik (Ibibio)	SA		393	1,700	0.231176471	4.325699746	0.254452926	0.058823529
Ellinika (Greek)	SF		310	1,744	0.177752294	5.625806452	0.322580645	0.05733945
English	A/I		280	1,289	0.217222653	4.603571429	0.357142857	0.077579519
Ese Eja	SP		186	1,364	0.136363636	7.333333333	0.537634409	0.073313783
Esperanto	SA	A	314	1,538	0.204161248	4.898089172	0.318471338	0.065019506
Estonian (Eesti)	SA		239	1,652	0.144673123	6.912133891	0.418410042	0.060532688
Even	SA	i	146	871	0.167623421	5.965753425	n/a	n/a
Evenki	SA	i	124	941	0.131774708	7.588709677	n/a	n/a
Ewe/Eve	A/I		290	1,326	0.218702866	4.572413793	0.344827586	0.075414781

LENGUA	TIPOLOGÍA MORFOLÓGICA	OBSERVACIONES SOBRE EL TEXTO FUENTE	N.º DE TOKENS	N.º DE UFCT	ÍNDICE DE RELATIVIDAD INFORMATIVA (TOKENS/UFCT)	ÍNDICE DE DENSIDAD INFORMATIVA (UFCT/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA LÉXICA (100/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA FÓNICA (100/UFCT)
Fante	A/I		342	1,376	0.248546512	4.023391813	0.292397661	0.072674419
Faroese	SF		261	1,355	0.192619926	5.191570881	0.383141762	0.073800738
Farsi/Persian	SA		316	1,373	0.23015295	4.344936709	0.316455696	0.072833212
Fijian	A/I		319	1,390	0.229496403	4.357366771	0.313479624	0.071942446
Filipino (Tagalog)	SA		302	1,652	0.182808717	5.470198675	0.331125828	0.060532688
Finnish	SA		218	1,678	0.129916567	7.697247706	0.458715596	0.059594756
Fiole/Kongo/ Kikongo (Angola)	SA		298	1,531	0.194644024	5.137583893	0.33557047	0.065316786
Fon	A/I		409	1,120	0.365178571	2.738386308	0.244498778	0.089285714
Francoprovençal / Occitan (Fribourg)	SF		337	1,452	0.232093664	4.308605341	0.296735905	0.068870523
Francoprovençal / Occitan (Savoie)	SF		344	1,445	0.238062284	4.200581395	0.290697674	0.069204152
Francoprovençal / Occitan (Valais)	SF		329	1,573	0.209154482	4.781155015	0.303951368	0.063572791
Francoprovençal / Occitan (Vaud)	SF		381	1,657	0.229933615	4.349081365	0.262467192	0.060350003
French (Français)	SF		325	1,408	0.230823864	4.332307692	0.307692308	0.071022727
Frisian	SF		289	1,529	0.189012426	5.290657439	0.346020761	0.065402224
Friulian (Friulano)	SF		335	1,557	0.215157354	4.647761194	0.298507463	0.064226076
Fulfulde, Fula (Nigerian)	SA		245	1,291	0.189775368	5.269387755	0.408163265	0.077459334
Ga	A/I		377	1,386	0.272005772	3.676392573	0.265251989	0.072150072
Gaelge (Irish Gaelic)	SF		326	1,264	0.257911392	3.877300613	0.306748466	0.079113924
Gagauz	SA		229	1,478	0.154939107	6.454148472	0.436681223	0.067658999

LENGUA	TIPOLOGÍA MORFOLÓGICA	OBSERVACIONES SOBRE EL TEXTO FUENTE	N.º DE TOKENS	N.º DE UFCT	ÍNDICE DE RELATIVIDAD INFORMATIVA (TOKENS/UFCT)	ÍNDICE DE DENSIDAD INFORMATIVA (UFCT/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA LÉXICA (100/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA FÓNICA (100/UFCT)
Gàidhlig Albanach (Scottish Gaelic)	SF		323	1,577	0.204819277	4.882352941	0.309597523	0.063411541
Galician (Galego)	SF		288	1,553	0.185447521	5.392361111	0.347222222	0.0643915
Gan	A/I	*	281	1,338	0.210014948	4.761565836	n/a	n/a
Ganda	SA		214	1,411	0.151665485	6.593457944	0.46728972	0.070871722
Garhwali	SF	*	61	322	0.189440994	5.278688525	n/a	n/a
Garifuna	SP		220	1,264	0.174050633	5.745454545	0.454545455	0.079113924
Georgian	SA		244	1,768	0.13800905	7.245901639	0.409836066	0.056561086
German (Deutsch)	SF		268	1,497	0.179024716	5.585820896	0.373134328	0.066800267
Gondi, Northern	SA	*	271	1,113	0.243486074	4.10701107	n/a	n/a
Gonja	A/I		320	1,113	0.287511231	3.478125	0.3125	0.08984726
Greenlandic	SP		184	2,573	0.071511854	13.98369565	0.543478261	0.038865138
Guarani	SP		235	1,632	0.143995098	6.944680851	0.425531915	0.06127451
Guarayo	SA		239	1,462	0.163474692	6.117154812	0.418410042	0.068399453
Guen (Gen, Mina)	A/I		425	1,664	0.255408654	3.915294118	0.235294118	0.060096154
Gujarati	SF		269	1,823	0.147558969	6.776951673	0.371747212	0.054854635
Haitian Creole (Kreyol)	A/I		384	1,506	0.25498008	3.921875	0.260416667	0.066401062
Haitian Creole (popular)	A/I		509	1,932	0.263457557	3.7956778	0.196463654	0.051759834
Hani	A/I		357	1,639	0.217815741	4.591036415	0.280112045	0.061012813
Hankuko (Korean)	SA		297	1,237	0.240097009	4.164983165	0.336700337	0.080840744
Hausa	A/I		442	1,803	0.245146977	4.07918552	0.226244344	0.055463117

LENGUA	TIPOLOGÍA MORFOLÓGICA	OBSERVACIONES SOBRE EL TEXTO FUENTE	N.º DE TOKENS	N.º DE UFCT	ÍNDICE DE RELATIVIDAD INFORMATIVA (TOKENS/UFCT)	ÍNDICE DE DENSIDAD INFORMATIVA (UFCT/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA LÉXICA (100/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA FÓNICA (100/UFCT)
Hausa/Haoussa	SF		442	1,803	0.245146977	4.07918552	0.226244344	0.055463117
Hawaiian	A/I		375	1,348	0.278189911	3.594666667	0.266666667	0.074183976
Hayu (Wayu)	SF	*	178	1,237	0.143896524	6.949438202	n/a	n/a
Hebrew	SF		236	1,254	0.188197767	5.313559322	0.423728814	0.079744817
Hiligaynon	SA		338	1,792	0.188616071	5.301775148	0.295857988	0.055803571
Hindi	SF		328	1,759	0.186469585	5.362804878	0.304878049	0.056850483
Hmong (Miao) N'jua	A/I		515	1,940	0.265463918	3.766990291	0.194174757	0.051546392
Hmong (Miao) Northern Qiangdong	A/I		499	1,851	0.269584009	3.709418838	0.200400802	0.054024851
Hmong (Miao) Southern Qiangdong	A/I		442	1,691	0.261383797	3.825791855	0.226244344	0.059136606
Ho	SA	*	346	1,560	0.221794872	4.50867052	n/a	n/a
Hokkien	A/I	*	53	227	0.233480176	4.283018868	n/a	n/a
Huastec (San Luis Potosí)	SA		294	1,336	0.22005988	4.544217687	0.340136054	0.074850299
Huastec (Sierra de Otontepec)	SA		379	1,662	0.228038508	4.385224274	0.263852243	0.060168472
Huastec (Veracruz)	SA		224	1,129	0.198405669	5.040178571	0.446428571	0.088573959
Huitoto Murui	SP		201	1,244	0.161575563	6.189054726	0.497512438	0.080385852
Hungarian	SA		245	1,689	0.145056246	6.893877551	0.408163265	0.059206631
HyoImo	SA	*	227	1,131	0.200707339	4.982378855	n/a	n/a
Icelandic (Íslenska)	SF		290	1,505	0.19269103	5.189655172	0.344827586	0.066445183

LENGUA	TIPOLOGÍA MORFOLÓGICA	OBSERVACIONES SOBRE EL TEXTO FUENTE	N.º DE TOKENS	N.º DE UFCT	ÍNDICE DE RELATIVIDAD INFORMATIVA (TOKENS/UFCT)	ÍNDICE DE DENSIDAD INFORMATIVA (UFCT/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA LÉXICA (100/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA FÓNICA (100/UFCT)
Ido	SA	A	248	1,240	0.2	5	0.403225806	0.080645161
Idoma	SA		363	1,424	0.25491573	3.922865014	0.275482094	0.070224719
Igbo	A/I		373	1,614	0.23110285	4.327077748	0.268096515	0.061957869
Ijaw	SA	i	196	793	0.247162673	4.045918367	n/a	n/a
Iloko/Ilocano	SA		335	1,931	0.173485241	5.764179104	0.298507463	0.051786639
Indonesian / Bahasa Indonesia	SA		276	1,724	0.160092807	6.246376812	0.362318841	0.05800464
Inerlingua	SF	A	313	1,567	0.199744735	5.006389776	0.319488818	0.063816209
Inuktitut (Canada)	SP	*	201	2,874	0.06993737	14.29850746	n/a	n/a
Italian	SF		311	2,034	0.152900688	6.540192926	0.321543408	0.049164208
Japanese	SA		357	1,619	0.220506485	4.535014006	0.280112045	0.061766523
Javanese	SA		315	1,979	0.159171299	6.282539683	0.317460317	0.050530571
Jnatjo (Mazahua)	A/I	i	206	890	0.231460674	4.32038835	n/a	n/a
Kabardian	SA		247	1,736	0.142281106	7.028340081	0.4048583	0.057603687
Kabuverdianu	A/I		438	1,850	0.236756757	4.223744292	0.228310502	0.054054054
Kabyè	SA		223	996	0.223895582	4.466367713	0.448430493	0.100401606
Kanauji	SF	*	360	1,643	0.219111382	4.563888889	n/a	n/a
Kannada	SA		182	1,777	0.102419809	9.763736264	0.549450549	0.05627462
Kanuri Yerwa	SA		216	1,321	0.163512491	6.115740741	0.462962963	0.075700227
Kaonde	SA		171	1,027	0.166504382	6.005847953	0.584795322	0.097370983
Kapampangan (Pampangan)	SA		303	1,667	0.181763647	5.501650165	0.330033003	0.059988002

LENGUA	TIPOLOGÍA MORFOLÓGICA	OBSERVACIONES SOBRE EL TEXTO FUENTE	N.º DE TOKENS	N.º DE UFCT	ÍNDICE DE RELATIVIDAD INFORMATIVA (TOKENS/UFCT)	ÍNDICE DE DENSIDAD INFORMATIVA (UFCT/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA LÉXICA (100/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA FÓNICA (100/UFCT)
Karakalpak	SA		262	1,619	0.161828289	6.179389313	0.381679389	0.061766523
Karelian	SA		245	1,765	0.138810198	7.204081633	0.408163265	0.056657224
Karen (S'gaw)	A/I	*	145	1,630	0.088957055	11.24137931	n/a	n/a
Kasem	SA		374	1,255	0.298007968	3.355614973	0.267379679	0.079681275
Kashmiri	SF	*	158	631	0.250396197	3.993670886	n/a	n/a
Kazakh	SA		248	1,630	0.152147239	6.572580645	0.403225806	0.061349693
Khakas	SA		262	1,687	0.155305276	6.438931298	0.381679389	0.059276823
Kharia	SA	*	313	1,692	0.18498818	5.405750799	n/a	n/a
Khasi	A/I		451	1,778	0.253655793	3.942350333	0.22172949	0.05624297
Khmer	A/I		360	2,067	0.174165457	5.741666667	0.277777778	0.048379294
Kibushi (Shibushi, Kibushi)	SA		289	1,712	0.168808411	5.923875433	0.346020761	0.058411215
K'iche' (Quiché)	SP		341	1,664	0.204927885	4.879765396	0.293255132	0.060096154
Kimbundu	SA	*	316	1,395	0.226523297	4.414556962	n/a	n/a
Kinyamwezi (Nyamwezi)	SA		193	1,037	0.18611379	5.375056995	0.518134715	0.096432015
Kinyarwanda	SA		196	1,215	0.161316872	6.198979592	0.510204082	0.082304527
Kirundi	SA		234	1,566	0.149425287	6.692307692	0.427350427	0.06385696
Klau	SA		278	1,452	0.191460055	5.223021583	0.35971223	0.068870523
Koits-Sunuwar	SF	*	113	563	0.20071048	4.982300885	n/a	n/a
Komi-Permian	SF		207	1,204	0.17192691	5.816425121	0.483091787	0.083056478
Koulango	A/I	i	208	727	0.28610729	3.495192308	n/a	n/a

LENGUA	TIPOLOGÍA MORFOLÓGICA	OBSERVACIONES SOBRE EL TEXTO FUENTE	N.º DE TOKENS	N.º DE UFCT	ÍNDICE DE RELATIVIDAD INFORMATIVA (TOKENS/UFCT)	ÍNDICE DE DENSIDAD INFORMATIVA (UFCT/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA LÉXICA (100/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA FÓNICA (100/UFCT)
Kpellewo (Kpelle)	A/I	i	263	940	0.279787234	3.574144487	n/a	n/a
Krio	A/I	i	308	1,034	0.29787234	3.357142857	n/a	n/a
Kurdish, Central	SF		288	1,245	0.231325301	4.322916667	0.347222222	0.080321285
Kurug, Kurukh	SA	*	324	1,481	0.218771101	4.570987654	n/a	n/a
Kusunda (Kusanda)	SA	*	231	937	0.246531483	4.056277056	n/a	n/a
Kven	SA		255	1,708	0.149297424	6.698039216	0.392156863	0.058548009
Kyrgyz	SA		266	1,703	0.15619495	6.402255639	0.37593985	0.058719906
Ladin	SF		339	1,567	0.21633695	4.622418879	0.294985251	0.063816209
Ladino	SF		293	1,543	0.189889825	5.266211604	0.341296928	0.064808814
Lamiso '(Lám nso')	SA		354	1,139	0.310798946	3.217514124	0.282485876	0.087796313
Lao	A/I		285	1,737	0.164075993	6.094736842	0.350877193	0.057570524
Latin (Latina)	SF		232	1,628	0.142506143	7.017241379	0.431034483	0.061425061
Latvian	SF		248	1,611	0.153941651	6.495967742	0.403225806	0.062073246
Lepcha (Ntumpa, Rong, Rongpa)	SA	*	368	2,085	0.176498801	5.66576087	n/a	n/a
Lhomi	SA	*	333	1,581	0.210626186	4.747747748	n/a	n/a
Ligurian	SF		302	1,561	0.193465727	5.168874172	0.331125828	0.064061499
Limba	SA	i	234	888	0.263513514	3.794871795	n/a	n/a
Limbu	SF	*	61	442	0.13800905	7.245901639	n/a	n/a
Lingala	SA		319	1,480	0.215540541	4.639498433	0.313479624	0.067567568
Lithuanian (Lietuviskai)	SF		247	1,506	0.164010624	6.097165992	0.4048583	0.066401062

LENGUA	TIPOLOGÍA MORFOLÓGICA	OBSERVACIONES SOBRE EL TEXTO FUENTE	N.º DE TOKENS	N.º DE UFCT	ÍNDICE DE RELATIVIDAD INFORMATIVA (TOKENS/UFCT)	ÍNDICE DE DENSIDAD INFORMATIVA (UFCT/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA LÉXICA (100/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA FÓNICA (100/UFCT)
Lobiri	SA		485	1,662	0.291817088	3.426804124	0.206185567	0.060168472
Lozi	SA		300	1,211	0.247729149	4.036666667	0.333333333	0.082576383
Luba-Kasai (Tshituba)	SA		308	1,613	0.190948543	5.237012987	0.324675325	0.061996628
Luganda/Ganda	SA		214	1,411	0.151665485	6.593457944	0.46728972	0.070871722
Lunda/Chokwe-lunda	SA		172	1,227	0.140179299	7.13372093	0.581395349	0.081499593
Luvale	SA		196	1,519	0.129032258	7.75	0.510204082	0.065832785
Luxembourgish (Lëtzebuergesch)	SF		310	1,543	0.200907323	4.977419355	0.322580645	0.064808814
Macedonian	SF		288	1,441	0.199861207	5.003472222	0.347222222	0.069396253
Madurese	SA		250	1,531	0.163291966	6.124	0.4	0.065316786
Magahi	SF		255	1,290	0.197674419	5.058823529	0.392156863	0.07751938
Magar (Dhut)	SA	*	37	152	0.243421053	4.108108108	n/a	n/a
Maijithi	SF		234	1,509	0.155069583	6.448717949	0.427350427	0.066269052
Maunani, Maonan	A/I		223	1,308	0.170489297	5.865470852	0.448430493	0.076452599
Makonde	SA	i	136	904	0.150442478	6.647058824	n/a	n/a
Makua, Makhuwa (Mozambique)	SA		191	1,308	0.146024465	6.848167539	0.523560209	0.076452599
Malagasy	SA		262	1,521	0.172255095	5.805343511	0.381679389	0.06574622
Malayalam	SA		137	1,584	0.086489899	11.5620438	0.729927007	0.063131313
Maldivian (Dhivehi)	SF	*	28	229	0.122270742	8.178571429	n/a	n/a
Maltese	SF		263	1,607	0.163658992	6.110266616	0.380228137	0.062227754

LENGUA	TIPOLOGÍA MORFOLÓGICA	OBSERVACIONES SOBRE EL TEXTO FUENTE	N.º DE TOKENS	N.º DE UFCT	ÍNDICE DE RELATIVIDAD INFORMATIVA (TOKENS/UFCT)	ÍNDICE DE DENSIDAD INFORMATIVA (UFCT/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA LÉXICA (100/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA FÓNICA (100/UFCT)
Mam	SA		301	1,618	0.186032138	5.375415282	0.33225914	0.061804697
Maninka	A/I		285	1,064	0.267857143	3.733333333	0.350877193	0.093984962
Manipuri	SA	*	26	200	0.13	7.692307692	n/a	n/a
Maori	A/I		555	2,181	0.254470426	3.92972973	0.18018018	0.045850527
Maori (Cook Islands) (Rarotongan)	A/I		477	1,929	0.247278383	4.044025157	0.209643606	0.051840332
Mapudungun (Mapuzgün)	SP		225	1,142	0.197022767	5.075555556	0.444444444	0.087565674
Marathi	SA		265	1,898	0.139620653	7.162264151	0.377358491	0.052687039
Marshallese	SF		446	1,717	0.259755387	3.849775785	0.224215247	0.058241118
Marwari	SF	*	223	800	0.27875	3.587443946	n/a	n/a
Matsés	SA		364	2,698	0.134914752	7.412087912	0.274725275	0.037064492
Mayan (Yucateco)	SP		381	1,938	0.196594427	5.086614173	0.262467192	0.051599587
Mazateco	SP		239	1,477	0.161814489	6.179916318	0.418410042	0.067704807
Mende	A/I		308	1,118	0.27549195	3.62987013	0.324675325	0.089445438
Mijisa	SA		306	1,235	0.247773279	4.035947712	0.326797386	0.08097166
Mikmaq/Micmac	SP		218	1,596	0.136591479	7.321100917	0.458715996	0.062656642
Minangkabau	SA		279	1,640	0.170121951	5.878136201	0.358422939	0.06097561
Minjiang (spoken)	A/I		295	1,302	0.226574501	4.413559322	0.338983051	0.076804916
Minjiang (written)	A/I		280	1,241	0.225624496	4.432142857	0.357142857	0.080580177
Miskito	SA		347	1,577	0.220038047	4.544668588	0.288184438	0.063411541

LENGUA	TIPOLOGÍA MORFOLÓGICA	OBSERVACIONES SOBRE EL TEXTO FUENTE	N.º DE TOKENS	N.º DE UFCT	ÍNDICE DE RELATIVIDAD INFORMATIVA (TOKENS/UFCT)	ÍNDICE DE DENSIDAD INFORMATIVA (UFCT/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA LÉXICA (100/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA FÓNICA (100/UFCT)
Mixe	SP		221	1,263	0.174980206	5.714932127	0.452488688	0.079176564
Mixteco	A/I	i	235	1,012	0.232213439	4.306382979	n/a	n/a
Mizo	SA		355	1,528	0.232329843	4.304225352	0.281690141	0.065445026
Moba	SA		407	1,213	0.335531739	2.98034398	0.245700246	0.082440231
Mon	A/I	*	26	86	0.302325581	3.307692308	n/a	n/a
Mongolian (Halh, Khalkha)	SA		277	1,556	0.178020566	5.61732852	0.36101083	0.064267352
Montenegrin	SF		265	1,487	0.178211163	5.611320755	0.377358491	0.067249496
Mooré/More	SA		439	1,460	0.300684932	3.325740319	0.227790433	0.068493151
Moxeño Trinitario	SP		226	1,605	0.140809969	7.101769912	0.442477876	0.062305296
Mozarabic (Ajami)	SF		302	1,761	0.17149347	5.831125828	0.331125828	0.056785917
Mundari	SP	*	220	1,583	0.138976627	7.195454545	n/a	n/a
Muzzi	A/I		315	1,536	0.205078125	4.876190476	0.317460317	0.065104167
Nahuatl	SP		320	2,105	0.152019002	6.578125	0.3125	0.047505938
Nanai	SA		215	1,549	0.138799225	7.204651163	0.465116279	0.064557779
Ndebele (Northern)	SA		167	1,329	0.12565839	7.958083832	0.598802395	0.075244545
Nenets	SF		248	1,426	0.173913043	5.75	0.403225806	0.070126227
Nepali	SF		226	1,544	0.146373057	6.831858407	0.442477876	0.064766839
Newar (Nepal Bhasa)	SF	*	72	325	0.221538462	4.513888889	n/a	n/a
Ngasan	SA		230	1,648	0.139563107	7.165217391	0.434782609	0.060679612
Ngangela, Nyemba	SA	*	355	1,754	0.202394527	4.94084507	n/a	n/a

LENGUA	TIPOLOGÍA MORFOLÓGICA	OBSERVACIONES SOBRE EL TEXTO FUENTE	N.º DE TOKENS	N.º DE UFCT	ÍNDICE DE RELATIVIDAD INFORMATIVA (TOKENS/UFCT)	ÍNDICE DE DENSIDAD INFORMATIVA (UFCT/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA LÉXICA (100/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA FÓNICA (100/UFCT)
Nigerian Pidgin English	A/I		505	1,756	0.287585421	3.477227723	0.198019802	0.056947608
Niue	A/I		435	1,771	0.245623941	4.071264368	0.229885057	0.056465274
Nivkh	SA		239	1,471	0.162474507	6.154811715	0.418410042	0.067980965
Nomatsiguenga	SP		211	1,715	0.12303207	8.127962085	0.473933649	0.058309038
Norwegian (Bokmål)	SF		265	1,424	0.186095506	5.373584906	0.377358491	0.070224719
Norwegian (Nynorsk)	SF		293	1,393	0.210337401	4.754266212	0.341296928	0.071787509
Nuer	SF	*	262	921	0.284473398	3.515267176	n/a	n/a
Nuosu/Yi	A/I	*	271	783	0.346104725	2.889298893	n/a	n/a
Nyanja/Chinyanja (Chechewa, Chichewa)	SA		239	1,528	0.156413613	6.393305439	0.418410042	0.065445026
Nzema	A/I		366	1,558	0.23491656	4.256830601	0.273224044	0.064184852
Ñahú (Otomi)	SF	i	277	916	0.302401747	3.306859206	n/a	n/a
Occitan	SF		282	1,443	0.195426195	5.117021277	0.354609929	0.069300069
Occitan Auvergnat	SF		287	1,495	0.191973244	5.209059233	0.348432056	0.066889632
Occitan Languedocien	SF		328	1,597	0.205385097	4.868902439	0.304878049	0.062617408
Ogiek	SA		162	1,142	0.141856392	7.049382716	0.617283951	0.087565674
Ojibwa / Ojibway / Ojibwe	SP	*	18	182	0.098901099	10.111111111	n/a	n/a
Oriya	SF	*	59	374	0.157754011	6.338983051	n/a	n/a
Orok	SA		220	1,698	0.129564193	7.718181818	0.454545455	0.058892815

LENGUA	TIPOLOGÍA MORFOLÓGICA	OBSERVACIONES SOBRE EL TEXTO FUENTE	N.º DE TOKENS	N.º DE UFCT	ÍNDICE DE RELATIVIDAD INFORMATIVA (TOKENS/UFCT)	ÍNDICE DE DENSIDAD INFORMATIVA (UFCT/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA LÉXICA (100/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA FÓNICA (100/UFCT)
Oromo, Borana-Arsi-Guji	SA		219	1,425	0.153684211	6.506849315	0.456621005	0.070175439
Oroqen	SA		188	1,335	0.14082397	7.10106383	0.531914894	0.074906367
Osetin (Ossetian)	SA		202	1,147	0.176111595	5.678217822	0.495049505	0.087183958
Oshiwambo (Ndonga)	SA		236	1,389	0.169906407	5.88559322	0.423728814	0.07199424
Otuho	SA		260	1,321	0.19682059	5.080769231	0.384615385	0.075700227
Paez	SP		204	1,517	0.134475939	7.43627451	0.490196078	0.065919578
Palauan	A/I		338	1,441	0.234559334	4.263313609	0.295857988	0.069396253
Panjabi (Punjabi)	SF		390	1,913	0.20386827	4.905128205	0.256410256	0.052273915
Papiamentu	A/I		330	1,611	0.204841713	4.881818182	0.303030303	0.062073246
Pashito/Pakhto	SF	*	38	148	0.256756757	3.894736842	n/a	n/a
Picard	SF		338	1,523	0.221930401	4.50591716	0.295857988	0.065659882
Pintupi-Luritja	SP		339	2,906	0.116655196	8.572271386	0.294985251	0.034411562
Pipil	SA		266	1,156	0.230103806	4.345864662	0.37593985	0.08650519
Polish (Polski)	SF		274	1,690	0.162130178	6.167883212	0.364963504	0.059171598
Ponapean, Pohnpeian	A/I		284	1,366	0.207906296	4.809859155	0.352112676	0.073206442
Portuguese	SF		297	1,550	0.191612903	5.218855219	0.336700337	0.064516129
Prouvençau	SF		282	1,443	0.195426195	5.117021277	0.354609929	0.069300069
Pulaar	SA		263	1,340	0.196268657	5.095057034	0.380228137	0.074626866
Pular	SA	i	183	904	0.202433628	4.93989071	n/a	n/a
Purhépecha	SA		240	1,708	0.140515222	7.116666667	0.416666667	0.058548009

LENGUA	TIPOLOGÍA MORFOLÓGICA	OBSERVACIONES SOBRE EL TEXTO FUENTE	N.º DE TOKENS	N.º DE UFCT	ÍNDICE DE RELATIVIDAD INFORMATIVA (TOKENS/UFCT)	ÍNDICE DE DENSIDAD INFORMATIVA (UFCT/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA LÉXICA (100/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA FÓNICA (100/UFCT)
Q'echi/Kekchi	SP		249	1,362	0.182819383	5.469879518	0.401606426	0.073421439
Quechua	SA		306	2,491	0.122842232	8.140522876	0.326797386	0.04014452
Quichua	SA		174	1,253	0.138866672	7.201149425	0.574712644	0.07980846
Rajasthani	SF	*	31	140	0.221428571	4.516129032	n/a	n/a
Rajbanshi / Rangpuri / Kamtapuri	SF	*	300	1,293	0.232018561	4.31	n/a	n/a
Rajji	A/I	*	101	348	0.290229885	3.445544554	n/a	n/a
Rhaeto-Romance (Rumanisch)	SF		292	1,538	0.189856957	5.267123288	0.342465753	0.065019506
Romagnolo (Sammarinese)	SF		348	1,668	0.208633094	4.793103448	0.287356322	0.059952038
Romani	SF		264	1,386	0.19047619	5.25	0.378787879	0.072150072
Romanian (Romána)	SF		309	1,648	0.1875	5.333333333	0.323624595	0.060679612
Rukonzo (Konjo)	SA		278	2,308	0.120450607	8.302158273	0.35971223	0.043327556
Rundi	SA		221	1,472	0.15013587	6.660633484	0.452488688	0.067934783
Runyankore-rukiga/Nkore-ki-ga	SA		149	1,126	0.132326821	7.55704698	0.67114094	0.088809947
Russian (Rusky)	SF		279	1,713	0.162872154	6.139784946	0.358422939	0.058377116
Saami/Lappish	SA		254	1,875	0.135466667	7.381889764	0.393700787	0.053333333
Salar	SA		201	1,178	0.170628183	5.860696517	0.497512438	0.084889643
Samoan	A/I		492	1,865	0.263806971	3.790650407	0.203252033	0.053619303
Sango (Sangho)	A/I		383	1,385	0.276534296	3.61618799	0.261096606	0.072202166
Sanskrit	SF		181	1,637	0.110568112	9.044198895	0.552486188	0.061087355

LENGUA	TIPOLOGÍA MORFOLÓGICA	OBSERVACIONES SOBRE EL TEXTO FUENTE	N.º DE TOKENS	N.º DE UFCT	ÍNDICE DE RELATIVIDAD INFORMATIVA (TOKENS/UFCT)	ÍNDICE DE DENSIDAD INFORMATIVA (UFCT/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA LÉXICA (100/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA FÓNICA (100/UFCT)
Santhali	SA	*	327	1,533	0.213307241	4.688073394	n/a	n/a
Säotomense	SF		286	1,107	0.258355917	3.870629371	0.34965035	0.090334237
Sardinian, Logudorese	SF		348	1,738	0.20023015	4.994252874	0.287356322	0.057537399
Sarnámi Hindustani	SF		379	1,628	0.232800983	4.295514512	0.263852243	0.061425061
Saxon, Low	SF		333	1,579	0.21089297	4.741741742	0.3003003	0.063331222
Scotts	SF		278	1,354	0.205317578	4.870503597	0.35971223	0.073855244
Secoya	SP		227	1,391	0.163191948	6.127753304	0.440528634	0.071890726
Seereer	SA		293	1,112	0.263489209	3.795221843	0.341296928	0.089928058
Serbian (Srpski)	SF		249	1,337	0.186237846	5.369477912	0.401606426	0.074794316
Seselwa Creole French	SA		328	1,578	0.207858048	4.81097561	0.304878049	0.063371356
Shangani	SA		361	1,800	0.200555556	4.986149584	0.27700831	0.055555556
Sharanahua	SP	i	138	1,106	0.12477396	8.014492754	n/a	n/a
Sherpa	A/I	*	143	566	0.252650177	3.958041958	n/a	n/a
Shilluk	SF	i	224	830	0.269879518	3.705357143	n/a	n/a
Shimaore	SA		367	1,765	0.207932011	4.809264305	0.272479564	0.056657224
Shingazidja	SA		295	1,631	0.180870632	5.528813559	0.338983051	0.061312078
Shipibo-Conibo	SP		423	2,892	0.14626556	6.836879433	0.236406619	0.034578147
Shona	SA		245	1,779	0.137717819	7.26122449	0.408163265	0.056211355
Shor	SA	i	150	865	0.173410405	5.766666667	n/a	n/a
Sia Pedee	SP		247	1,551	0.159252095	6.279352227	0.4048583	0.064474533

LENGUA	TIPOLOGÍA MORFOLÓGICA	OBSERVACIONES SOBRE EL TEXTO FUENTE	N.º DE TOKENS	N.º DE UFCT	ÍNDICE DE RELATIVIDAD INFORMATIVA (TOKENS/UFCT)	ÍNDICE DE DENSIDAD INFORMATIVA (UFCT/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA LÉXICA (100/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA FÓNICA (100/UFCT)
Sindhi	SF	*	32	157	0.203821656	4.90625	n/a	n/a
Sinhala	SF		252	1,775	0.141971831	7.043650794	0.396825397	0.056338028
Siona	SP		223	1,458	0.152949246	6.538116592	0.448430493	0.068587106
Stirionó	SA		246	1,089	0.225895317	4.426829268	0.406504065	0.091827365
siSwati (Swazi)	SA		292	2,419	0.120711038	8.284246575	0.342465753	0.041339396
Slovak	SF		248	1,423	0.174279691	5.737903226	0.403225806	0.070274069
Slovenian (Slovenscina)	SF		256	1,446	0.177040111	5.6484375	0.390625	0.069156293
Somali	SA		300	1,472	0.203804348	4.906666667	0.333333333	0.067934783
Soninké (Soninkaxaane)	A/I		297	1,106	0.268535262	3.723905724	0.336700337	0.090415913
Sorbian	SF		262	1,552	0.168814433	5.923664122	0.381679389	0.06443299
Sotho, Northern/ Pedi/ Sepedi	SA		404	1,657	0.243814122	4.101485149	0.247524752	0.06035003
Sotho, Southern/ Sotho / Sesotho / Sutu / Sesutu	SA		315	1,312	0.240091463	4.165079365	0.317460317	0.076219512
Sotho, Western/ Tswana/ Setswana	SA		373	1,699	0.219540906	4.554959786	0.268096515	0.058858152
Spanish	SF		299	1,580	0.189240506	5.284280936	0.334448161	0.063291139
Sukuma	SA		212	1,237	0.171382377	5.83490566	0.471698113	0.080840744
Sundanese	A/I		296	1,748	0.169336384	5.905405405	0.337837838	0.057208238
Surel (Sunwar)	SA	*	61	333	0.183183183	5.459016393	n/a	n/a
Sussu / Sousou / Sosso / Soso / Susu	A/I	i	199	927	0.214670982	4.658291457	n/a	n/a

LENGUA	TIPOLOGÍA MORFOLÓGICA	OBSERVACIONES SOBRE EL TEXTO FUENTE	N.º DE TOKENS	N.º DE UFCT	ÍNDICE DE RELATIVIDAD INFORMATIVA (TOKENS/UFCT)	ÍNDICE DE DENSIDAD INFORMATIVA (UFCT/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA LÉXICA (100/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA FÓNICA (100/UFCT)
Swahili/Kiswahili	SA		269	1,280	0.21015625	4.758364312	0.371747212	0.078125
Swampy Cree	SA	*	23	218	0.105504587	9.47826087	n/a	n/a
Swedish (Svenska)	SF		264	1,473	0.179226069	5.579545455	0.378787879	0.067888663
T'Simane	SA	i	188	1,091	0.172318973	5.803191489	n/a	n/a
Tacana	SP	i	188	968	0.194214876	5.14893617	n/a	n/a
Tagalog	SA		302	1,652	0.182808717	5.470198675	0.331125828	0.060532688
Tahitian	A/I		488	2,043	0.238864415	4.18647541	0.204918033	0.048947626
Tajik	SF		303	1,516	0.199868074	5.00330033	0.330033003	0.065963061
Talysh	SF		263	1,447	0.181755356	5.501901141	0.380228137	0.0691085
Tamang / Tam (Eastern)	SA	*	208	938	0.221748401	4.509615385	n/a	n/a
Tamashaq	SA		286	1,169	0.24465355	4.087412587	0.34965035	0.085543199
Tamazight	SA		296	1,394	0.212338594	4.709459459	0.337837838	0.071736011
Tamil	SA	*	21	224	0.09375	10.666666667	n/a	n/a
Tatar	SA		258	1,586	0.162673392	6.147286822	0.387596899	0.063051702
Telugu	SA		186	1,858	0.100107643	9.989247312	0.537634409	0.053821313
Tem	SA		266	1,453	0.183069511	5.462406015	0.375939985	0.068823125
Tetun	SA		257	1,168	0.220034247	4.544747082	0.389105058	0.085616438
Tetun Dili	A/I	i	229	944	0.242584746	4.122270742	n/a	n/a
Thai	A/I		368	1,798	0.204671858	4.885869565	0.27173913	0.055617353
Thangmi	SA	*	142	739	0.192151556	5.204225352	n/a	n/a

LENGUA	TIPOLOGÍA MORFOLÓGICA	OBSERVACIONES SOBRE EL TEXTO FUENTE	N.º DE TOKENS	N.º DE UFCT	ÍNDICE DE RELATIVIDAD INFORMATIVA (TOKENS/UFCT)	ÍNDICE DE DENSIDAD INFORMATIVA (UFCT/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA LÉXICA (100/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA FÓNICA (100/UFCT)
Tharu-Dangaura	SF	*	205	1,102	0.186025408	5.375609756	n/a	n/a
Themne	SA		364	1,339	0.27184466	3.678571429	0.274725275	0.074682599
Tibetan	SA		494	2,053	0.240623478	4.155870445	0.20242915	0.048709206
Ticuna	SP		348	1,747	0.199198626	5.020114943	0.287356322	0.057240985
Tigrinya (Tigrigna)	SF		231	1,834	0.125954198	7.939393939	0.432900433	0.054525627
Tiv	SA		493	1,571	0.313812858	3.186612576	0.202839757	0.063653724
Toba	SP		553	1,440	0.384027778	2.6039783	0.180831826	0.069444444
Tojol-a'b'al	SP		400	1,837	0.217746326	4.5925	0.25	0.054436581
Tok Pisin	A/I		353	1,493	0.236436705	4.229461756	0.283286119	0.066979236
Tonga (Zambia)	SA		209	1,562	0.133802817	7.473684211	0.4784689	0.064020487
Tongan (Tonga)	A/I		463	2,020	0.229207921	4.362850972	0.215982721	0.04950495
Totonaco	SP		266	2,026	0.131293189	7.616541353	0.375939985	0.049358342
Tsafiki	SF		198	1,122	0.176470588	5.666666667	0.505050505	0.08912656
Tsonga (Mozambique)	SA		377	1,707	0.220855302	4.527851459	0.265251989	0.058582308
Tsonga (Zimbabwe)	SA		362	1,800	0.201111111	4.972375691	0.276243094	0.055555556
Tswana	SA		373	1,699	0.219540906	4.554959786	0.268096515	0.058858152
Turkish (Türkçe)	SA		236	1,505	0.156810631	6.377118644	0.423728814	0.066445183
Turkmen	SA		212	1,418	0.149506347	6.688679245	0.471698113	0.070521862
Tuvan	SA		237	1,691	0.140153755	7.135021097	0.421940928	0.059136606
Twí (Akan Kasa) [Akuapem]	A/I		304	1,135	0.26784141	3.733552632	0.328947368	0.088105727

LENGUA	TIPOLOGÍA MORFOLÓGICA	OBSERVACIONES SOBRE EL TEXTO FUENTE	N.º DE TOKENS	N.º DE UFCT	ÍNDICE DE RELATIVIDAD INFORMATIVA (TOKENS/UFCT)	ÍNDICE DE DENSIDAD INFORMATIVA (UFCT/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA LÉXICA (100/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA FÓNICA (100/UFCT)
Twí (Akan Kasa) [Asante, ashanti]	A/I		296	1,057	0.280037843	3.570945946	0.337837838	0.094607379
Tzeltal, Oxchuc	SP		335	1,568	0.213647959	4.680597015	0.298507463	0.06377551
Tzotzil (Chamula)	SA		361	1,744	0.206995413	4.831024931	0.27700831	0.05733945
Uighur	SA		257	1,820	0.141208791	7.081712062	0.389105058	0.054945055
Ukrainian	SF		275	1,568	0.175382653	5.701818182	0.363636364	0.06377551
Umbundu	SA		220	1,433	0.153524075	6.513636364	0.454545455	0.069783671
Urarina	SP		258	1,824	0.141447368	7.069767442	0.387596899	0.054824561
Urdu	SF		364	1,511	0.240900066	4.151098901	0.274725275	0.066181337
Uzbek	SA		265	1,701	0.155790711	6.418867925	0.377358491	0.058788948
Vai	SA		474	1,472	0.32201087	3.105485232	0.210970464	0.067934783
Venda	SA		365	1,637	0.222968845	4.484931507	0.273972603	0.061087355
Venetian	SF		295	1,457	0.20247083	4.938983051	0.338983051	0.06863418
Veps	SA		247	1,581	0.156230234	6.400809717	0.4048583	0.063251107
Vietnamese	A/I		411	1,668	0.246402878	4.058394161	0.243309002	0.059952038
Vlach	SF		196	1,715	0.114285714	8.75	0.510204082	0.058309038
Waama	SA	i	157	545	0.288073394	3.47133758	n/a	n/a
Walloon/Wallon	SF		374	1,852	0.201943844	4.951871658	0.267379679	0.05399568
Worani/Wao Tededo/Huorani	SP		235	1,387	0.169430425	5.90212766	0.425531915	0.072098053
Waray/Waray-Waray	SA		343	1,652	0.207627119	4.816326531	0.29154519	0.060532688

LENGUA	TIPOLOGÍA MORFOLÓGICA	OBSERVACIONES SOBRE EL TEXTO FUENTE	N.º DE TOKENS	N.º DE UFCT	ÍNDICE DE RELATIVIDAD INFORMATIVA (TOKENS/UFCT)	ÍNDICE DE DENSIDAD INFORMATIVA (UFCT/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA LÉXICA (100/TOKENS)	ÍNDICE DE EFICIENCIA INFORMATIVA FÓNICA (100/UFCT)
Wayuu	SP		167	1,206	0.138474295	7.221556886	0.598802395	0.08291874
Weenhayek	SP		312	2,054	0.151898734	6.583333333	0.320512821	0.048685492
Wolof	SA		293	1,110	0.263963964	3.788395904	0.341296928	0.090090009
Xhosa	SA		196	1,631	0.120171674	8.321428571	0.510204082	0.061312078
Yagua	SA		172	1,397	0.123120974	8.122093023	0.581395349	0.071581961
Yakkha	SF	*	157	912	0.172149123	5.808917197	n/a	n/a
Yakut	SA		254	1,866	0.136120043	7.346456693	0.393700787	0.053590568
Yanomamö	SP		339	1,518	0.223320158	4.477876106	0.294985251	0.065876153
Yao	SA		210	1,436	0.146239554	6.838095238	0.476190476	0.069637883
Yapese	A/I		309	1,406	0.219772404	4.550161812	0.323624595	0.071123755
Yiddish	SF		280	1,537	0.182173064	5.489285714	0.357142857	0.065061809
Yoruba (Yorüba)	A/I		408	1,628	0.250614251	3.990196078	0.245098039	0.061425061
Yukagir	SA		168	1,299	0.129330254	7.732142857	0.595238095	0.076982294
Yuracaré	SA		205	1,442	0.142163662	7.034146341	0.487804878	0.069348128
Záparo	SP		209	1,241	0.168412571	5.937799043	0.4784689	0.080580177
Zapoteco	A/I	i	248	876	0.283105023	3.532258065	n/a	n/a
Zhuang	A/I		254	1,475	0.17220339	5.807086614	0.393700787	0.06779661
Zulu	SA		171	1,406	0.121621622	8.222222222	0.584795322	0.071123755

OBSERVACIONES SOBRE EL TEXTO FUENTE:

(sin marca): 10 primeros artículos de la Declaración Universal de los Derechos Humanos.

i: 10 primeros artículos de la Declaración Universal de los Derechos Humanos incompletos.

*: otros textos.

A: lengua artificial.

CLAVES DE LA TIPOLOGÍA MORFOLÓGICA:

A/I: lengua predominantemente analítica-aislante.

SA: lengua predominantemente sintética aglutinante.

SF: lengua predominantemente sintética fusionante.

SP: lengua predominantemente sintética polisintética.