

Coloma, Germán

**Working Paper**

## Efectos de compensación entre indicadores de la complejidad de los idiomas

Serie Documentos de Trabajo, No. 569

**Provided in Cooperation with:**

University of CEMA, Buenos Aires

*Suggested Citation:* Coloma, Germán (2015) : Efectos de compensación entre indicadores de la complejidad de los idiomas, Serie Documentos de Trabajo, No. 569, Universidad del Centro de Estudios Macroeconómicos de Argentina (UCEMA), Buenos Aires

This Version is available at:

<https://hdl.handle.net/10419/130820>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

**UNIVERSIDAD DEL CEMA  
Buenos Aires  
Argentina**

Serie  
**DOCUMENTOS DE TRABAJO**

**Área: Lingüística y Estadística**

**EFFECTOS DE COMPENSACIÓN  
ENTRE INDICADORES DE LA  
COMPLEJIDAD DE LOS IDIOMAS**

**Germán Coloma**

**Octubre 2015  
Nro. 569**

**[www.cema.edu.ar/publicaciones/doc\\_trabajo.html](http://www.cema.edu.ar/publicaciones/doc_trabajo.html)  
UCEMA: Av. Córdoba 374, C1054AAP Buenos Aires, Argentina  
ISSN 1668-4575 (impreso), ISSN 1668-4583 (en línea)  
Editor: Jorge M. Streb; asistente editorial: Valeria Dowding <jae@cema.edu.ar>**



# Efectos de compensación entre indicadores de la complejidad de los idiomas

Germán Coloma\*

## Resumen

En este trabajo utilizamos datos del Atlas Mundial de Estructuras Lingüísticas para una muestra de 100 idiomas y 60 características. Los valores de tales características son interpretados como variables binarias de complejidad, que luego se agrupan en cinco categorías (fonología, morfología, sintaxis, sistema verbal y vocabulario) para las cuales se calculan índices de complejidad. Tanto las variables como los índices son sometidos a un análisis de correlación estadística, buscando detectar posibles efectos de compensación. Para ello utilizamos coeficientes de correlación estándar y de correlación parcial, y controlamos por otros factores lingüísticos y extralingüísticos (ubicación geográfica, filiación, tamaño de la población). Detectamos así varios efectos de compensación importantes, que suelen quedar ocultos por otros elementos. Sus señales más evidentes son que las correlaciones entre variables de complejidad se incrementan cuando controlamos por otros factores, y que cualquier idioma es más complejo que cualquier otro de los incluidos en la muestra, al menos para una característica lingüística.

**Palabras clave:** complejidad, efectos de compensación, variables binarias, correlación parcial.

## 1. Introducción

El Atlas Mundial de Estructuras Lingüísticas (WALS, por su sigla en inglés) es una publicación que compila información sobre características estructurales de los distintos idiomas que se hablan en el mundo. En su versión electrónica vigente (Dryer & Haspelmath, 2013) contiene datos de 2679 idiomas y dialectos, que corresponden a 192 características pertenecientes a distintos componentes de la estructura gramatical de dichos idiomas.

Ninguna de las lenguas incluidas en el WALS tiene datos para todas las características que aparecen en el atlas, y tampoco existe ninguna característica que

---

\* Universidad del CEMA; Av. Córdoba 374, Buenos Aires, C1054AAP, Argentina. Teléfono: 6314-3000. Correo electrónico: [gcoloma@cema.edu.ar](mailto:gcoloma@cema.edu.ar). Agradezco los comentarios de Damián Blasi, Guiomar Ciapuscio, Verónica Nercesián y Frans Plank a una versión anterior del presente trabajo. Las opiniones expresadas en esta publicación son las del autor y no necesariamente las de la Universidad del CEMA.

aparezca informada para todos los idiomas. La lengua con mayor cobertura es el inglés (159 características informadas), pero hay muchos idiomas (por ejemplo, el asturiano o el taiwanés) para los cuales solo se presenta información acerca de una única característica. La característica para la cual hay más información es la que lleva el código 83A (orden de objeto y verbo en los enunciados), ya que para ella hay datos referidos a 1519 idiomas. En el otro extremo, la característica 141A (sistemas de escritura) solo reporta información referida a 6 idiomas.

Si bien el número de lenguas que aparecen en el WALSL es muy grande, los editores de dicho atlas han elegido una muestra de 100 lenguas cuya inclusión en los distintos capítulos es requerida “siempre que sea posible”. Se supone que esas lenguas forman una muestra relativamente balanceada en cuanto a su diversidad lingüística y geográfica, si bien en su elección pesaron también algunos otros factores relacionados con la importancia poblacional o cultural, y con la disponibilidad de descripciones gramaticales detalladas.<sup>1</sup>

Aprovechando que el WALSL concentra una cantidad muy importante de información, y que ha privilegiado la compilación de datos sobre una muestra particular de idiomas, en este trabajo usaremos dicha muestra para llevar a cabo una serie de análisis estadísticos que buscan detectar posibles relaciones entre los valores de 60 características diferentes. Dichas características provienen de distintas categorías gramaticales, y han sido codificadas por medio de variables que pueden interpretarse como medidas de complejidad.<sup>2</sup>

Con las variables así construidas procedimos a calcular correlaciones, usando tanto los coeficientes estándar (de Pearson) como los llamados “coeficientes de correlación parcial”. Después de eso computamos cinco índices de complejidad, surgidos de agrupar las variables en categorías, y obtuvimos un índice de complejidad total (que es el promedio de los índices parciales). Esos índices también fueron analizados en términos de su posible correlación, y todo esto generó una serie de conclusiones relacionadas con la existencia de efectos de compensación.

---

<sup>1</sup> La lista completa de idiomas está reproducida en el apéndice 1.

<sup>2</sup> La lista completa de variables está reproducida en el apéndice 2.

Un efecto de compensación (*trade-off*) entre variables que miden la complejidad de los idiomas representa una situación en la cual un nivel de complejidad mayor para cierto componente de la lengua aparece en correspondencia con un nivel de complejidad menor para otro componente. Este tipo de relación puede ocurrir entre características que pertenezcan a la misma categoría gramatical (por ejemplo, fonología, morfología, sintaxis), entre características que pertenezcan a distintas categorías, o entre medidas que representen a tales categorías.

El presente trabajo está organizado del siguiente modo. En la sección 2 reseñamos la literatura sobre *trade-offs* de complejidad, especialmente la que ha procurado hallar correlaciones negativas entre variables lingüísticas. En la sección 3 describimos la base de datos que hemos construido, y en la sección 4 analizamos los principales estadísticos calculados para dicha base de datos, en términos de la correlación entre las variables. En la sección 5 hacemos lo mismo para los índices de complejidad, y en la sección 6 llevamos a cabo algunos cálculos adicionales acerca de posibles comparaciones entre idiomas. Finalmente, la sección 7 está dedicada a resumir las principales conclusiones de todo el trabajo.

## **2. Reseña de la literatura**

La literatura sobre comparaciones interlingüísticas de complejidad es relativamente extensa y variada. Uno de sus trabajos más citados es el de McWhorter (2001), que propone la construcción de medidas de complejidad para los diferentes subsistemas de la lengua (por ejemplo, fonología, morfología, sintaxis, vocabulario) basándose en la idea de que un idioma es más complejo que otro si posee “más distinciones y/o reglas explícitas”.<sup>3</sup> Aplicando dicha definición, McWhorter concluye que los “idiomas criollos” (*creole languages*) son siempre más simples que los idiomas no criollos, ya que carecen de una serie de “sobrespecificaciones” que son típicas de las

---

<sup>3</sup> Esta definición se relaciona con lo que algunos autores denominan “complejidad absoluta”, que se define como “el número de partes de un sistema”. Sobre este punto, véase Miestamo (2008).

lenguas más tradicionales.<sup>4</sup>

La hipótesis de McWhorter se opone a una larga tradición lingüística que considera que “todos los idiomas son igualmente complejos”. Dicha tradición no se basa en una medición sistemática de indicadores de complejidad, sino en distintos enfoques teóricos que suponen tal igualdad.<sup>5</sup> Desde que la lingüística cuantitativa comenzó a desarrollar medidas de complejidad de los idiomas y a analizar la posible correlación entre dichas medidas, sin embargo, la hipótesis de igual complejidad ha sido vista en general como falsa.

Uno de los artículos más importantes que intentó medir la correlación entre indicadores de complejidad en un entorno interlingüístico es el de Shosted (2006), que estudia la posible existencia de correlación negativa entre complejidad fonológica y morfológica en una muestra de 32 idiomas. En ese artículo, que termina concluyendo que no existe una correlación estadísticamente significativa entre esas medidas, el autor mide la complejidad fonológica usando el número teórico de posibles tipos de sílabas en cada idioma, en tanto que su medida de complejidad morfológica viene dada por el número de posibles tipos de inflexión verbal. En la misma línea de investigación, Nichols (2009) mide la complejidad interlingüística en cinco categorías (fonología, morfología, clases de palabras, sintaxis y vocabulario) usando distintos índices, que calcula como promedios de varias medidas de complejidad en una muestra de 68 idiomas. Si bien no encuentra correlaciones negativas significativas entre dichos índices, sí halla una correlación positiva significativa entre su medida de complejidad morfológica y su medida de complejidad sintáctica.

Otro enfoque alternativo para medir la complejidad es usar medidas “empíricas” de dicho fenómeno. Fenk-Oczlon & Fenk (2008), por ejemplo, calculan una serie de cocientes (fonemas por sílaba, sílabas por palabra, fonemas por palabra, palabras por enunciado, etc.) que provienen de traducciones de varias oraciones simples (escritas originalmente en alemán) a 33 lenguas distintas. Luego computan coeficientes de

---

<sup>4</sup> Se denomina “idioma criollo” a una lengua natural que ha surgido como una versión simplificada de otra lengua natural, pero que luego de varias generaciones ha ganado autonomía respecto de aquella.

<sup>5</sup> Para una reseña de esta literatura, véase Sampson (2009) y Joseph & Newmeyer (2012).

correlación entre estos cocientes, y llegan a una serie de conclusiones que implican la existencia de correlaciones negativas significativas entre los mismos.<sup>6</sup>

Otra rama de la literatura es la que se inició con el trabajo de Nettle (1995), que relaciona la extensión de las palabras (medida usando una muestra de palabras para distintos idiomas) y la complejidad fonológica (medida usando indicadores teóricos tales como el número de fonemas distintos que tiene cada idioma). En dicho estudio, que usó una muestra de solo 10 idiomas, su autor encontró una correlación negativa muy fuerte entre talas medidas, lo cual generó toda una literatura posterior con muchas más observaciones y con metodologías de cómputo más sofisticadas (por ejemplo, Wichmann, Rama & Holman, 2011; y Moran & Blasi, 2014).

Otros enfoques del problema de medición de la complejidad se restringen a alguna categoría lingüística en particular, como la fonología, la morfología o la sintaxis. Maddieson (2007), por ejemplo, mide la complejidad fonológica usando tres indicadores distintos: inventario de fonemas (vocálicos y consonánticos), sistemas tonales y estructura silábica. Al buscar correlaciones entre dichas medidas, solo encuentra un coeficiente positivo y significativo entre el inventario de consonantes y la estructura silábica (es decir, cuanto más consonantes tiene un idioma, más compleja parece ser la estructura de las sílabas que utiliza), y un coeficiente negativo (menos significativo) entre complejidad del sistema tonal y estructura silábica.<sup>7</sup>

Otro ejemplo de búsqueda de efectos de compensación dentro de una categoría lingüística particular es el que lleva a cabo Sinnemäki (2008), que se enfoca en tres medidas alternativas de la complejidad (sintáctica) de la marcación de los argumentos en la frase (que son la marcación de núcleos, la marcación de complementos, y el uso de un orden fijo para los distintos tipos de palabras). Después de definir esas medidas en base a su capacidad para discriminar argumentos, Sinnemäki halla una correlación negativa significativa entre la marcación de complementos y el orden fijo de palabras, así como correlaciones también negativas pero no significativas entre marcación de núcleos y marcación de complementos, y entre marcación de núcleos y orden fijo de palabras.

---

<sup>6</sup> Sobre este tema, véase también Coloma (2014).

<sup>7</sup> Sobre este tema, véase también Coloma (2013).

Varios de los trabajos sobre efectos de compensación entre indicadores de complejidad han utilizado datos tomados del WALS. Esto ocurre con los artículos ya citados de Maddieson (2007), Sinnemäki (2008) y Moran & Blasi (2014), y con numerosos artículos relacionados con temas de complejidad lingüística y correlación, tales como los de Parkvall (2008), Dahl (2011) y Matasovic (2014).

En este artículo intentaremos encontrar efectos de compensación entre medidas de complejidad que ocurren tanto a nivel general (o sea, entre distintas categorías lingüísticas) como a nivel particular (o sea, dentro de una misma categoría). Esto se debe a que la base de datos que hemos construido se refiere a un número relativamente grande de características lingüísticas, que pueden a su vez agruparse en categorías. Todas nuestras medidas de complejidad, sin embargo, serán de carácter teórico o tipológico, ya que no provienen de textos específicos sino de las gramáticas de los diferentes idiomas, compiladas por los autores de los distintos capítulos del WALS.

### **3. Descripción de los datos**

Tal como mencionamos en la introducción, nuestra base de datos contiene información sobre 60 variables, y se refiere a las lenguas incluidas en la muestra de 100 idiomas del WALS. Esas variables se armaron usando información sobre 60 características idiomáticas, y en todos los casos son “variables binarias”, es decir, variables que toman un valor igual a uno en ciertos casos, e igual a cero en otros.

El criterio usado para definir las variables fue asignar un valor igual a uno a los idiomas que fueran más complejos en términos de cierta característica, y un valor igual a cero a los idiomas que fueran más simples en términos de dicha característica. Para hacerlo, seleccionamos todas las características para las cuales la información del WALS nos permitiera clasificar a los idiomas de acuerdo con alguna medida de complejidad. Por ejemplo, la característica 1A (inventario de fonemas consonánticos) es la base para definir una variable según la cual un idioma es complejo si tiene 26 o más consonantes, y simple si tiene 25 o menos. Del mismo modo, la variable basada en la característica 66A (tiempo pasado) le asigna un valor igual a uno (complejo) a los idiomas que hacen algún

tipo de distinción gramatical entre formas verbales del pasado y del presente, y un valor igual a cero (simple) a los que no hacen ningún tipo de distinción de esa clase.

Muchas características informadas en el WALS fueron excluidas del análisis porque no sirven para generar variables que puedan ser fácilmente interpretadas como medidas de complejidad. La característica 3A (cociente entre consonantes y vocales) es un ejemplo de esa situación, ya que no está claro que un idioma que tenga un cociente elevado entre consonantes y vocales sea más complejo o más simple que otro que tenga pocas consonantes por vocal. Lo mismo ocurre con la característica 31A (sistemas de asignación de género), ya que un sistema en el cual las palabras son masculinas o femeninas no es ni más simple ni más complejo que otro en el cual las clases de palabras estén basadas en otro tipo de consideración. Esto contrasta con el uso que hacemos de la característica 30A (número de géneros), que es la base de una variable según la cual un idioma es complejo si los sustantivos están marcados por algún tipo de distinción de clase (basada o no en el sexo) y simple si no hace ninguna distinción gramatical de género.

La construcción de nuestra base de datos también está condicionada por el hecho de que hay muchas observaciones para las cuales el WALS no nos da una información completa. Esto se debe a que solamente cinco idiomas de la muestra (inglés, finlandés, francés, indonesio y turco) tienen datos para todas las características seleccionadas, y solo siete de tales características (48A, 100A, 104A, 107A, 108A, 109A y 113A) tienen datos para los 100 idiomas. El número de datos faltantes es igual a 532, lo cual es relativamente poco si consideramos la cantidad total de datos que tiene la base que hemos construido (igual a 6000). La reconstrucción de dichos datos, sin embargo, requirió un esfuerzo considerable, y para eso tuvimos que usar distintas estrategias.

Varios datos que faltaban fueron relativamente fáciles de inferir usando información correspondiente a dialectos de la misma lengua (por ejemplo, árabe saudita en vez de árabe egipcio, o quechua norperuano en vez de quechua ecuatoriano). En otros casos en los cuales eso no fue posible, buscamos idiomas que pertenecieran a la misma subfamilia (por ejemplo, holandés en vez de alemán, o cebuano en vez de tagalo). En otros casos tuvimos que recurrir a una reconstrucción más indirecta, usando información

de idiomas que no estuvieran filogenéticamente tan cerca pero que estuvieran ligados por una relación de tipo geográfico (por ejemplo, yurok en vez de karok, o igbo en vez de yoruba). Dichos procedimientos, obviamente, tienen cierto grado de imprecisión y están sujetos a posibles errores de medición, pero de cualquier modo entendemos que la probabilidad de tales errores es baja porque en todos los casos los idiomas están clasificados en grupos binarios (simples o complejos), y esos grupos están distribuidos de un modo relativamente regular entre familias lingüísticas y zonas geográficas.

Las 60 características usadas para construir nuestra base de datos pertenecen a distintas áreas del WALSL, que son: fonología (11 características), morfología (6 características), categorías nominales (14 características), sintaxis nominal (4 características), categorías verbales (10 características), orden de las palabras (5 características), enunciados simples (8 características), y vocabulario (2 características). Para llevar a cabo algunos análisis agregados de complejidad, sin embargo, procedimos a reclasificar esas características usando un esquema diferente, y terminamos con una clasificación en cinco categorías: fonología (11 variables), morfología (11 variables), sintaxis (12 variables), sistema verbal (13 variables) y vocabulario (13 variables).<sup>8</sup>

Cuando la información que aparece en el WALSL para una característica en particular permite dividir la muestra de distintas maneras, hicimos la división entre idiomas simples y complejos de modo de que el número de observaciones en cada conjunto fuera relativamente parejo. Por ejemplo, la característica 41A (contrastes de distancia en demostrativos) clasifica a las lenguas en cinco grupos, de acuerdo con el número de contrastes marcados por distintas palabras o afijos. Basándonos en dicha distribución, consideramos que los idiomas que no tenían contrastes de este tipo y los que tenían solo un contraste (por ejemplo, el inglés, que contrasta “this” con “that”) eran simples, y que, en cambio, los idiomas con más de un contraste (por ejemplo, el español, que contrasta “este”, “ese” y “aquel”) eran complejos. Con este criterio, nos quedaron 48 idiomas complejos y 52 idiomas simples. Si hubiéramos usado otra forma de dividir la muestra, podríamos habernos quedado con 96 idiomas complejos y solo 4 idiomas

---

<sup>8</sup> Para saber qué característica pertenece a qué categoría, véase el apéndice 2.

simples (si hubiéramos considerado que solamente eran simples los idiomas sin ningún contraste entre demostrativos) o con 91 idiomas simples y solo 9 complejos (si hubiéramos exigido más de tres contrastes para que un idioma fuera catalogado como complejo).

Tomando en cuenta los 100 idiomas y las 60 características como un todo, nuestra clasificación generó 2637 datos complejos (44%) y 3363 datos simples (56%). La variable con un porcentaje mayor de observaciones complejas es la que corresponde a la característica 130A (dedo y mano), con 91 idiomas que usan distintas palabras para los conceptos de “dedo” y “mano”, y solo 9 idiomas que usan la misma palabra. La variable con un menor porcentaje de observaciones complejas, en cambio, es la que corresponde a la característica 11A (vocales anteriores redondeadas), con 7 idiomas que poseen al menos una vocal anterior redondeada como fonema,<sup>9</sup> y 93 idiomas que no tienen ninguna de esas vocales. Si observamos las distintas lenguas, vemos que la que tiene un porcentaje mayor de observaciones complejas es el idioma abjasio (que pertenece a la familia caucásica noroccidental y se habla en un zona cerca al Mar Negro), el cual tiene 41 características complejas y 19 simples. Los idiomas con más observaciones simples, en cambio, son el tailandés y el hmong njua (que pertenece a la familia hmong mien, y se habla en China), que tienen 14 características complejas y 46 simples.

#### **4. Análisis de correlación entre variables binarias**

El primer paso para estudiar la posible existencia de efectos de compensación entre medidas de la complejidad de los 100 idiomas de la muestra tomada del WALs consiste en computar coeficientes de correlación entre las 60 variables seleccionadas. Como todas las variables son binarias, sus correlaciones pueden inferirse utilizando los números que aparecen en tablas de contingencia tales como la del cuadro 1, que ejemplifica el caso de las variables derivadas de la característica 12A (estructura silábica) y 13A (tono). En él puede verse el número de observaciones que pertenecen a las cuatro situaciones posibles que ocurren en la muestra: lenguas no tonales con estructura silábica

---

<sup>9</sup> Uno de dichos idiomas es el francés, que usa ese tipo de vocales en palabras como “*sur*”, “*peu*” y “*soeur*”.

simple (39 observaciones), lenguas no tonales con estructura silábica compleja (29 observaciones), lenguas tonales con estructura silábica simple (29 observaciones), y lenguas tonales con estructura silábica compleja (3 observaciones).

**Cuadro 1: Tabla de contingencia entre estructura silábica y tono**

<b>Estructura silábica / Tono</b>	No Tonal	Tonal	<b>Total</b>
Simple	39	29	68
Compleja	29	3	32
<b>Total</b>	68	32	100

Para calcular el coeficiente de correlación estándar ( $r$ ) entre estas dos variables, también llamado “coeficiente de Pearson”, basta utilizar esta fórmula:

$$r = \frac{N(S, NT) \cdot N(C, T) - N(S, T) \cdot N(C, NT)}{\sqrt{N(S) \cdot N(C) \cdot N(NT) \cdot N(T)}} = \frac{39 \cdot 3 - 29 \cdot 29}{\sqrt{68 \cdot 32 \cdot 68 \cdot 32}} = -0.3327 \quad ;$$

donde  $N(S, NT)$ ,  $N(C, T)$ ,  $N(S, T)$  y  $N(C, NT)$  son los números en cada una de las cuatro celdas de la tabla de contingencia, y  $N(S)$ ,  $N(C)$ ,  $N(NT)$  y  $N(T)$  son las cantidades totales de observaciones referidas a estructuras silábicas simples, estructuras silábicas complejas, lenguas no tonales y lenguas tonales.

En este caso, el resultado de aplicar esta fórmula indica la existencia de una correlación negativa, que resulta ser estadísticamente significativa.<sup>10</sup> La misma implica que es más probable que los idiomas que tienen una estructura silábica relativamente simple utilicen el tono como un modo de distinguir entre distintas palabras (o entre distintas formas gramaticales de las palabras), en tanto que es raro que los idiomas con una estructura silábica más compleja empleen el tono en esa función. Esto puede verse como un ejemplo de *trade-off* entre fenómenos de complejidad idiomática, ya que mayor complejidad en una dimensión (estructura silábica) se relaciona con mayor simplicidad

<sup>10</sup> La significación estadística quiere decir que no es probable que el verdadero valor de un coeficiente sea cero. Para este coeficiente de correlación en particular, la probabilidad de que el mismo sea cero puede estimarse como igual a 0,07%. Para dos variables cualesquiera, cuya correlación se calcule usando 100 observaciones, el coeficiente de correlación es estadísticamente significativo a un nivel de probabilidad del 5% (es decir, la probabilidad de que la verdadera correlación sea cero es menor que 5%) si el coeficiente es mayor que 0,2 en valor absoluto.

en otra (tono) y viceversa.

Si computamos los coeficientes de correlación estándar para los 1770 pares de variables que pueden formarse utilizando las 60 características seleccionadas, vemos que hay 795 que son negativos (44,9%) y 975 que son positivos (55,1%). La mayoría de ellos, sin embargo, son estadísticamente insignificantes, y solo 85 coeficientes (4,8%) son a la vez negativos y significativos a un 5% de probabilidad (es decir, negativos y mayores que 0,2 en valor absoluto).<sup>11</sup> Entre ellos podemos mencionar algunos casos tales como el coeficiente que relaciona las características 100A (alineamiento morfosintáctico en frases verbales) y 119A (predicación nominal y espacial), cuya correlación ( $r = -0,3021$ ) indica que los idiomas que marcan el objeto en las frases verbales son más proclives a usar el mismo verbo para predicados nominales (por ejemplo, “Juan es un hombre”) y espaciales (por ejemplo, “Juan está en su casa”).<sup>12</sup> Otro caso de correlación negativa, que tiene lugar entre variables léxicas y cuyo coeficiente es igual a  $-0,2359$ , es el que indica que, cuando un idioma usa la misma palabra para “mano” y “brazo” (característica 129A), entonces tiende a usar una palabra diferente para “dedo” (característica 130A).

El uso de coeficientes de correlación estándar para detectar la existencia de efectos de compensación es relativamente simple de calcular, pero tiene la desventaja de que no toma en cuenta la posible interacción con otras variables relevantes.<sup>13</sup> Para resolver este problema pueden calcularse coeficientes de correlación parcial, que son medidas de la dependencia lineal entre dos variables que surgen cuando se elimina la influencia de otras variables. Para computar esos coeficientes, puede comenzarse armando una matriz de correlación de todas las variables bajo análisis, y luego invertirla. Una vez hecho eso, puede utilizarse la siguiente fórmula:

$$r = -\frac{P_{xy}}{\sqrt{P_{xx} \cdot P_{yy}}} \quad ;$$

---

<sup>11</sup> El número de coeficientes de correlación positivos y significativos también es relativamente bajo (186 casos, o sea el 10,5% de todos los coeficientes calculados).

<sup>12</sup> Ese sería, por ejemplo, el caso del inglés, que usa el mismo verbo (*to be*) para ambas situaciones.

<sup>13</sup> Una buena explicación de esa desventaja, en el contexto de las correlaciones que se obtienen usando variables tipológicas tomadas del WALs, es la que aparece en Dryer (2009).

donde  $p_{xy}$  es el coeficiente que corresponde al par de variables  $x$  e  $y$  en la matriz de correlación inversa, y  $p_{xx}$  y  $p_{yy}$  son los coeficientes que corresponden a dichas variables en la diagonal principal de dicha matriz inversa.<sup>14</sup>

Si llevamos a cabo ese procedimiento en nuestra muestra de 100 idiomas y 60 variables, terminaremos con otra matriz cuyos 1770 coeficientes de correlación para todos los posibles pares de variables son negativos (878 casos, 49,6%) o positivos (892 casos, 50,4%). El número de coeficientes negativos y estadísticamente significativos es ahora mayor que el que aparece en la matriz de correlación estándar, ya que en esta matriz de correlaciones parciales encontramos 162 coeficientes negativos y significativos (9,15%). Entre ellos podemos mencionar el caso de las características 1A (inventario de fonemas consonánticos) y 2A (inventario de tipos vocálicos), cuya correlación negativa ( $r = -0,3163$ ) indica que los idiomas con más consonantes tienden a tener relativamente menos fonemas vocálicos y viceversa. Medida a través de un coeficiente estándar, dicha correlación es negativa pero no significativa ( $r = -0,1622$ ).

Otro coeficiente que puede interpretarse como indicativo de un efecto de compensación, y que es estadísticamente significativo si lo computamos como una correlación parcial (y no si lo calculamos como un coeficiente de correlación estándar), es el que corresponde a las características 23A (marcación del objeto directo) y 30A (número de géneros). Dicho coeficiente (igual a  $-0,3437$  usando correlación parcial, y a  $-0,0750$  usando correlación estándar) indica que los idiomas que no marcan el objeto directo con un afijo determinado tienen una mayor tendencia a diferenciar gramaticalmente las palabras por género. Lo mismo ocurre con otras correlaciones tales como las que pueden establecerse entre las características 68A (tiempo perfecto) y 70A (modo imperativo), entre las características 84A (orden de objeto, verbo y complemento oblicuo) y 92A (partículas interrogativas polares), etc.

Un instrumento adicional que puede incluirse, a fin de mejorar la medición de los posibles efectos de compensación entre distintas características del WALs, es el uso de variables extralingüísticas. Una posibilidad relativamente directa consiste en adicionar

---

<sup>14</sup> Para una explicación más completa del concepto de correlación parcial, véase Prokhorov (2002).

variables que representen características de tipo geográfico, tales como variables binarias que tomen un valor igual a uno cuando un idioma pertenece a determinada región e igual a cero en caso contrario. Como el WALs divide a los idiomas en seis macro-áreas, hemos construido cinco variables que corresponden a las lenguas incluidas en cada una de dichas áreas, que son Africa (16 idiomas), Australia (7 idiomas), Papunesia (17 idiomas), Norteamérica (18 idiomas) y Sudamérica (13 idiomas).<sup>15</sup>

Otras tres variables binarias adicionales se refieren a características filogenéticas o de parentesco idiomático. Tales son las que corresponden a idiomas que pertenecen a tres familias con un número relativamente grande de observaciones en la muestra de 100 idiomas del WALs, que son las de los grupos indoeuropeo (8 idiomas), austronesio (8 idiomas) y Niger-Congo (7 idiomas). También introdujimos otras tres variables que toman un valor igual a uno cuando una lengua pertenece a una de las siguientes subregiones incluidas dentro de alguna de las seis macro-áreas: Asia sudoriental (birmano, hmong njua, mandarín, meithei, tailandés y vietnamita), Mesoamérica (jacalteco, mixteco, otomí, rama, yaqui y zoque) y el Amazonas (apurina, barasano, canela-kraho, hixkaryana, pirahá, sanuma, wari y yagua).

La última variable extralingüística que hemos construido tiene que ver con diferencias entre idiomas utilizados por un gran número de hablantes e idiomas hablados por un número más pequeño de personas.<sup>16</sup> Esta variable le asigna un valor igual a uno a las 33 lenguas cuyo número total de hablantes supera los 5 millones (mandarín, inglés, español, hindi, árabe, ruso, japonés, alemán, francés, indonesio, coreano, turco, vietnamita, persa, canarés, hausa, birmano, tagalo, yoruba, suajili, oromo, tailandés, malgache, griego, zulú, quechua, bereber, hebreo, mongol, finlandés, guaraní, georgiano y hmong njua), y un valor igual a cero a los 67 idiomas restantes.

Utilizando las 12 nuevas variables creadas, resulta posible calcular nuevos coeficientes de correlación parcial. Los mismos nos señalan 881 casos de correlación negativa (49,8%) y 889 casos de correlación positiva (50,2%). El número de coeficientes

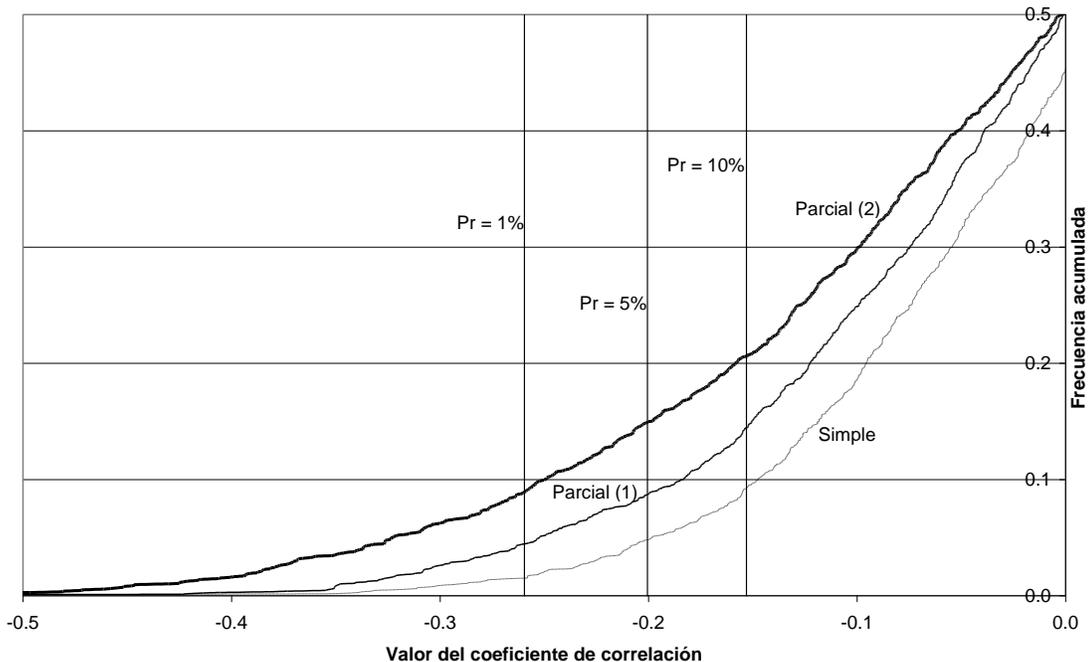
---

<sup>15</sup> La sexta macro-área definida en el WALs es Eurasia (29 idiomas), la cual funciona en nuestro análisis estadístico como “grupo de control” (*default group*).

<sup>16</sup> Sobre este tema, véase Dahl (2011).

negativos y significativos es ahora igual a 243 (13,7%), lo cual implica un incremento adicional respecto de nuestros dos cálculos previos.<sup>17</sup>

**Gráfico 1: Distribuciones acumuladas de los coeficientes de correlación**



Las diferencias entre los resultados de los tres cálculos de coeficientes de correlación pueden apreciarse a través de las funciones de distribución acumulada representadas en el gráfico 1. Las mismas surgen de ordenar los 1770 coeficientes generados por cada estimación del menor al mayor, y evaluar las frecuencias que corresponden a los distintos valores de tales coeficientes. Lo representado en el gráfico 1 se refiere exclusivamente a la parte de la distribución que corresponde a los valores negativos de los respectivos coeficientes calculados por los tres diferentes métodos (“simple”, “parcial (1)” y “parcial (2)”), y el último de ellos (es decir, el que incluye la interacción con las 12 variables extralingüísticas) es el que genera las mayores frecuencias de coeficientes negativos y significativos para los tres niveles de probabilidad

<sup>17</sup> Estos números son los que corresponden a los coeficientes de correlación parcial entre las 60 variables lingüísticas originales. El procedimiento de cálculo genera también otros 786 coeficientes (que surgen de incluir las 12 nuevas variables), cuyos valores no pueden ser comparados con los resultados previos.

generalmente utilizados en estadística (10%, 5% y 1%).

**Cuadro 2: Coeficientes de correlación entre variables fonológicas**

Característica	1A	2A	12A	13A
<b>Correlación estándar</b>				
1A: Consonantes	1.0000			
2A: Tipos vocálicos	-0.1622	1.0000		
12A: Estructura silábica	0.1485	0.0161	1.0000	
13A: Tono	0.0990	0.1666	<b>-0.3327</b>	1.0000
<b>Correlación parcial (caso 1)</b>				
1A: Consonantes	1.0000			
2A: Tipos vocálicos	<b>-0.3163</b>	1.0000		
12A: Estructura silábica	-0.0025	0.2122	1.0000	
13A: Tono	-0.0944	-0.0372	<b>-0.3180</b>	1.0000
<b>Correlación parcial (caso 2)</b>				
1A: Consonantes	1.0000			
2A: Tipos vocálicos	<b>-0.4108</b>	1.0000		
12A: Estructura silábica	-0.0019	0.1693	1.0000	
13A: Tono	<b>-0.3424</b>	-0.1763	<b>-0.2653</b>	1.0000

Los efectos del uso de correlaciones parciales y de la inclusión de variables extralingüísticas en el cálculo de los coeficientes pueden verse también si uno observa un conjunto reducido de casos como el representado en el cuadro 2. En dicho cuadro aparecen los coeficientes de correlación estándar y de correlación parcial para cuatro variables fonológicas (las correspondientes a las características 1A, 2A, 12A y 13A), con y sin la inclusión de las variables extralingüísticas. Puede verse así que, mientras el único coeficiente de correlación estándar que resulta negativo y significativo es el que corresponde a las características 12A (estructura silábica) y 13A (tono), el uso de técnicas de correlación parcial (caso 1) genera que también el coeficiente de correlación negativo entre las características 1A (consonantes) y 2A (tipos vocálicos) sea estadísticamente significativo. A su vez, el uso de variables extralingüísticas en el cómputo de los coeficientes de correlación parcial (caso 2) hace que el coeficiente que surge de relacionar las características 1A y 13A sea negativo y significativo. Esta última relación parece ser positiva y no significativa si la calculamos mediante un coeficiente de correlación estándar, y negativa y no significativa si la calculamos sin tener en cuenta su posible interacción con las 12 variables extralingüísticas.

Nótese además que el número de coeficientes negativos informados en el cuadro 2 (tanto significativos como no significativos) también se incrementa cuando pasamos de las correlaciones estándar a las correlaciones parciales. En la primera de dichas situaciones solo hay 2 coeficientes negativos, pero ese número aumenta a 5 cuando los coeficientes se calculan utilizando métodos de correlación parcial.

## 5. Análisis de complejidad agregada

Otro modo de buscar efectos de compensación entre indicadores de complejidad lingüística es el que surge de relacionar distintos grupos de características. Para hallar dichas relaciones, hemos agrupado las 60 variables en cinco categorías (fonología, morfología, sintaxis, sistema verbal y vocabulario), y con ellas hemos creado índices de complejidad. Dichos índices son promedios de variables binarias, y sus valores, por lo tanto, son números racionales que se encuentran en el espacio entre cero y uno. Esos números, a su vez, pueden promediarse y generar un “índice de complejidad total”, cuyo valor para cada idioma se ubica también entre cero y uno.<sup>18</sup>

Los índices de complejidad permiten ordenar a los distintos idiomas del más complejo al más simple, y producen *rankings* que difieren según la categoría de que se trate. La lengua con un índice más alto de complejidad total, dentro de las incluidas en la muestra de 100 idiomas del WALS, es el abjasio, que también se ubica primero en los *rankings* parciales de complejidad fonológica, verbal y léxica.<sup>19</sup> En la categoría morfológica, en cambio, el *ranking* está encabezado por el burushaski (que es una lengua aislada, hablada en Pakistán), en tanto que el *ranking* de complejidad sintáctica tiene en primer lugar al idioma karok (que se habla en Estados Unidos, y que también es una lengua aislada). Entre los idiomas con menores índices de complejidad, por su parte, están el rapanui (fonología), el cree (fonología), el sango (sistema verbal), el mixteco

---

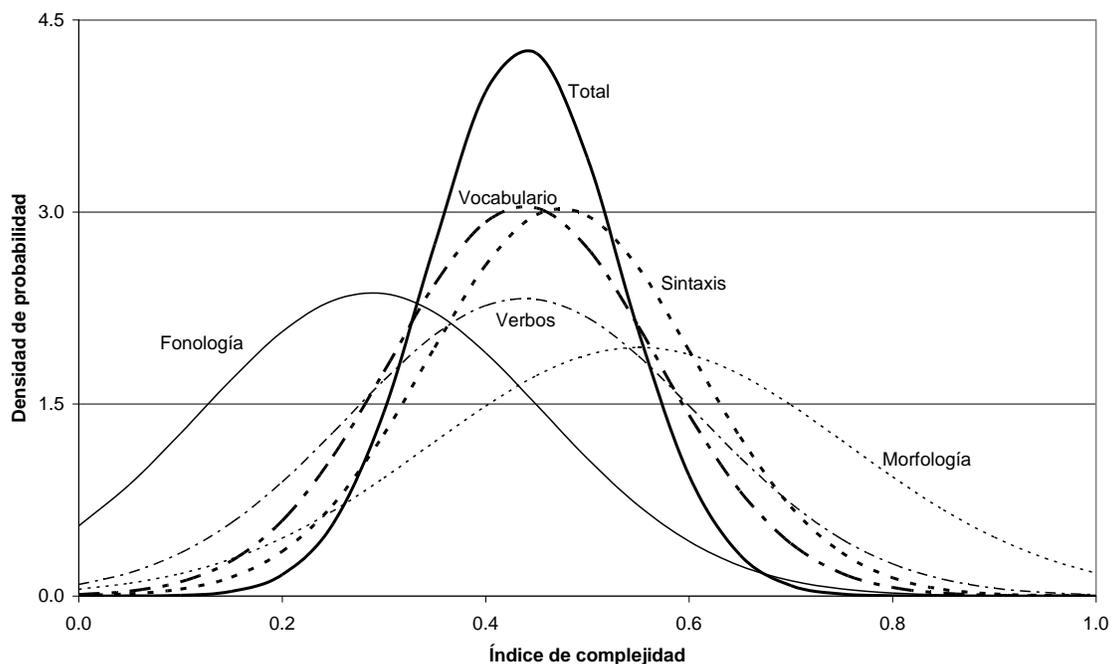
<sup>18</sup> Esto último es parecido a lo que hace Parkvall (2008), quien también utiliza mayoritariamente información del WALS (en su caso, de una edición anterior) y la complementa con otras fuentes. El índice creado por Parkvall es un promedio de 47 variables, y también se mueve en el rango entre 0 y 1.

<sup>19</sup> En esas categorías, el abjasio comparte el primer puesto con otros idiomas, tales como el francés, el lezguiano y el sango (fonología), el burushaski, el georgiano y el vasco (sistema verbal), y el barasano, el español y el zoque (vocabulario).

(sintaxis y vocabulario) y el tailandés (morfología y sintaxis). El tailandés es también el idioma con un menor índice de complejidad total.<sup>20</sup>

Cada índice tiene una distribución de frecuencias entre idiomas cuyos parámetros principales son la media ( $\mu$ ) y el desvío estándar ( $\sigma$ ). El índice con menor media es el de complejidad fonológica ( $\mu = 0,2882$ ), y el de mayor media es el de complejidad morfológica ( $\mu = 0,5573$ ). El índice de complejidad total tiene un valor promedio que está entre esos dos extremos ( $\mu = 0,4390$ ), y que además es muy cercano a la media que exhiben los índices de complejidad correspondientes a la sintaxis, el sistema verbal y el vocabulario. El índice de complejidad total, sin embargo, tiene un desvío estándar menor ( $\sigma = 0,0942$ ), que contrasta con los valores que pueden calcularse para los distintos índices parciales. Estos últimos van desde un mínimo de 0,1311 (vocabulario) a un máximo de 0,2078 (morfología).

**Gráfico 2: Distribuciones aproximadas de los índices de complejidad**



El efecto de estas diferencias en las frecuencias estimadas para los distintos

<sup>20</sup> La lista completa de índices está reproducida en el apéndice 3.

índices de complejidad puede apreciarse en el gráfico 2, en el cual hemos representado sendas aproximaciones de las mismas utilizando la llamada “distribución normal”. Nótese que los valores del índice de complejidad total están relativamente concentrados alrededor de la media, en tanto que las distribuciones de los índices parciales (en especial los referidos a la fonología, la morfología y el sistema verbal) están mucho más dispersas a lo largo de todo el espacio de probabilidad. Esta diferencia en cuanto a la dispersión de los índices puede verse como una evidencia en favor de la existencia de efectos de compensación entre medidas de complejidad, ya que indica que los idiomas son más desiguales cuando se los evalúa usando una única categoría lingüística (por ejemplo, fonología, morfología o sistema verbal), y menos desiguales cuando se utiliza una medida de su complejidad total.

Una posibilidad adicional para analizar la existencia de efectos de compensación entre distintos índices de complejidad es computar los coeficientes de correlación entre ellos. Como en este caso las variables no son binarias, la fórmula que debemos usar para calcular los coeficientes de correlación estándar es la siguiente:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \cdot \sum(y_i - \bar{y})^2}} \quad ;$$

donde  $x_i$  e  $y_i$  son los valores de dos índices distintos para el mismo idioma, y  $\bar{x}$  e  $\bar{y}$  son los valores promedio para dichos índices en la muestra de 100 idiomas.

Tal como vimos en la sección anterior, los coeficientes estándar pueden transformarse en coeficientes de correlación parcial a través de un proceso de inversión matricial (caso 1), el cual puede ser a su vez mejorado mediante la inclusión de las 12 variables extralingüísticas ya descritas (caso 2). Los resultados de dichos procedimientos, aplicados a los cinco índices parciales de complejidad definidos más arriba, aparecen en el cuadro 3.

Los valores de los coeficientes de correlación expuestos brindan cierto apoyo a la existencia de efectos de compensación entre los índices de complejidad, pero dicho apoyo es bastante más débil que el encontrado en la sección anterior para algunos casos

individuales. Nótese que, cuando computamos correlaciones estándar, ningún coeficiente negativo es significativamente distinto de cero. Más aún, los únicos valores negativos corresponden a la relación entre complejidad fonológica y sintáctica ( $r = -0,1370$ ), y a la relación entre complejidad fonológica y léxica ( $r = -0,1079$ ).

**Cuadro 3: Coeficientes de correlación entre índices de complejidad**

Índice de complejidad	Fonolog	Morfolog	Sintaxis	Verbos	Vocabul
<b>Correlación estándar</b>					
Fonología	1.0000				
Morfología	0.1688	1.0000			
Sintaxis	-0.1370	0.2908	1.0000		
Sistema verbal	0.1618	0.5721	0.1271	1.0000	
Vocabulario	-0.1079	0.0857	0.1365	0.0940	1.0000
<b>Correlación parcial (caso 1)</b>					
Fonología	1.0000				
Morfología	0.1458	1.0000			
Sintaxis	-0.1814	0.2848	1.0000		
Sistema verbal	0.0793	0.5432	-0.0415	1.0000	
Vocabulario	-0.1092	0.0213	0.0976	0.0695	1.0000
<b>Correlación parcial (caso 2)</b>					
Fonología	1.0000				
Morfología	0.0169	1.0000			
Sintaxis	-0.0142	0.2431	1.0000		
Sistema verbal	0.0575	0.4112	-0.0905	1.0000	
Vocabulario	<b>-0.2026</b>	0.0687	0.2433	0.0908	1.0000

Cuando introducimos técnicas de correlación parcial (caso 1), aparece un tercer coeficiente de correlación negativo (no significativo): el referido a la relación entre complejidad sintáctica y complejidad del sistema verbal. Un cuarto coeficiente negativo no significativo (fonología vs. morfología) surge cuando incluimos variables extralingüísticas en el cálculo de los coeficientes de correlación parcial (caso 2), pero esa inclusión hace que el valor absoluto del coeficiente negativo que era originalmente el más alto (fonología vs. sintaxis) se reduzca, pasando de “ $r = -0,1814$ ” a “ $r = -0,0142$ ”. También genera, sin embargo, el único coeficiente negativo que resulta ser estadísticamente significativo en todo el cuadro, que es el que relaciona la complejidad fonológica con la complejidad del vocabulario ( $r = -0,2026$ ).

Un ejercicio adicional que podemos llevar a cabo cuando clasificamos a las

características lingüísticas en categorías es ver si los coeficientes negativos que hallamos en la sección 4 pertenecen a variables incluidas en la misma categoría o a variables incluidas en categorías distintas. El cuadro 4 nos muestra la distribución de tales coeficientes, calculada a través de las proporciones entre el número de coeficientes negativos y significativos, y el total de coeficientes incluidos en cada celda de la matriz.

**Cuadro 4: Proporción de coeficientes de correlación negativos y significativos**

<b>Categoría</b>	<b>Fonolog</b>	<b>Morfolog</b>	<b>Sintaxis</b>	<b>Verbos</b>	<b>Vocabul</b>
<b>Correlación estándar</b>					
Fonología	3.6%				
Morfología	1.7%	1.8%			
Sintaxis	8.3%	6.1%	<b>6.1%</b>		
Sistema verbal	2.8%	1.4%	5.1%	2.6%	
Vocabulario	<b>9.8%</b>	3.5%	5.1%	5.9%	5.1%
<b>Correlación parcial (caso 1)</b>					
Fonología	10.9%				
Morfología	7.4%	9.1%			
Sintaxis	8.3%	5.3%	<b>12.1%</b>		
Sistema verbal	7.0%	10.5%	8.3%	10.3%	
Vocabulario	10.5%	10.5%	5.8%	11.2%	7.7%
<b>Correlación parcial (caso 2)</b>					
Fonología	16.4%				
Morfología	10.7%	16.4%			
Sintaxis	11.4%	12.9%	<b>21.2%</b>		
Sistema verbal	11.2%	16.8%	16.7%	12.8%	
Vocabulario	<b>19.6%</b>	7.7%	12.8%	13.6%	14.1%

De los valores informados en el cuadro 4 surge que la celda con el mayor porcentaje es siempre la correspondiente a variables fonológicas versus variables léxicas (vocabulario), con valores de 9,8%, 10,5% y 19,6%. Esto es consistente con el hecho de que tales categorías son las que exhiben los mayores coeficientes de correlación entre sus respectivos índices de complejidad. También puede observarse que la categoría que tiene un mayor porcentaje de coeficientes de correlación negativos y significativos entre sus propias variables es la sintaxis (6,1%, 12,1% y 21,2%), y esto vale tanto para los coeficientes de correlación estándar como para los dos tipos calculados de coeficientes de correlación parcial.

## 6. Comparación entre idiomas

Tal como mencionamos en la sección 2, la literatura sobre efectos de compensación entre medidas de complejidad interlingüística está relacionada con la hipótesis de que todos los idiomas son igualmente complejos. Los datos compilados para este trabajo son una fuente útil para estudiar la razonabilidad de dicha hipótesis, la cual puede ser evaluada utilizando distintos tipos de medición.

La respuesta más simple que puede dársele a la pregunta sobre la igualdad de la complejidad entre idiomas es rotundamente negativa, ya que en nuestra muestra hay cuatro lenguas (abjasio, burushaski, alemán y koasati) cuyos índices de complejidad total están por encima de 0,6 puntos, y al mismo tiempo hay otras tres lenguas (tailandés, vietnamita y hmong njua) cuyos índices de complejidad están por debajo de 0,25 puntos. Más aún, si dividimos la muestra en dos partes (los 50 idiomas con mayores índices de complejidad total, versus los 50 idiomas con menores índices) y calculamos sus respectivas medias, vemos que el primer grupo tiene un índice promedio igual a 0,5139, en tanto que el segundo tiene un índice promedio de 0,3641. La diferencia entre dichas medidas es por lo tanto igual a 0,1498, y ese número es estadísticamente distinto de cero para cualquier nivel razonable de probabilidad.

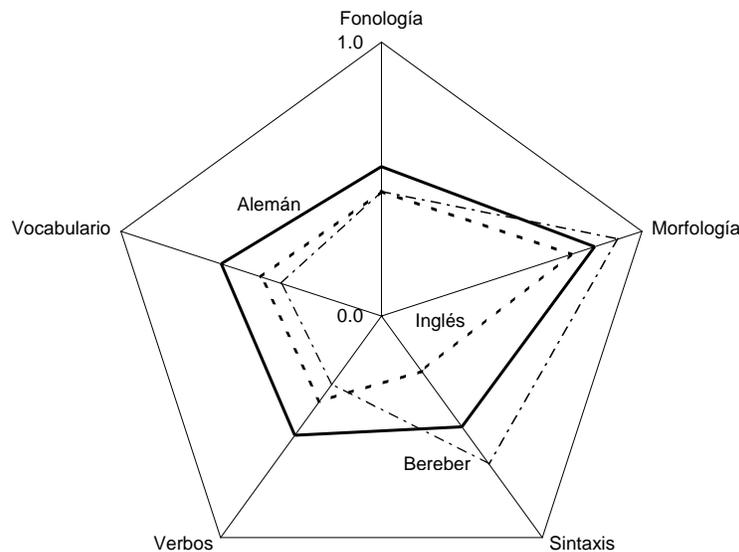
La hipótesis de igual complejidad, sin embargo, también se puede analizar de una manera más moderada. Supongamos, por ejemplo, que encontramos que un idioma es más complejo que otro en cierta categoría, pero que el segundo es más complejo que el primero en otra categoría. Podemos entonces decir que esos dos idiomas son igualmente complejos en cierto sentido, ya que la mayor complejidad que uno de ellos exhibe en una dimensión está compensada por la mayor complejidad del otro en una dimensión diferente.

En nuestra muestra de 100 lenguas, hay 4950 pares posibles de idiomas. En 3951 de ellos (79,82%) hay algún *trade-off* como el descrito en el párrafo anterior (es decir, el primer idioma tiene un índice de complejidad más elevado en una categoría, y el segundo tiene un índice de complejidad más elevado en otra categoría). Existen sin embargo 307 casos (6,02%) en los cuales un idioma es más complejo que otro en las cinco categorías

en las cuales hemos agrupado las características bajo estudio, y otros 692 casos más (13,98%) en los cuales hay un idioma que es más complejo que otro en alguna categoría e igualmente complejo en las restantes.

Para algunos idiomas, inclusive, este tipo de situaciones son particularmente comunes. El abjasio, por ejemplo, tiene índices de complejidad más altos en las cinco categorías que otros 38 idiomas incluidos en la muestra del WALS, y hay 23 idiomas más que tampoco lo superan en ninguna categoría. En el extremo opuesto, el warao (que es una lengua amerindia aislada que se habla en Venezuela) tiene índices de complejidad menores que otros 34 idiomas en las cinco categorías, y existe además un grupo adicional de 23 lenguas a las cuales no supera en complejidad en ninguna categoría.

**Gráfico 3: Diagrama de telaraña**



La dominación en términos de complejidad de una lengua sobre otra puede apreciarse utilizando diagramas de telaraña como el que aparece en el gráfico 3. En él hemos dibujado tres polígonos irregulares que representan distintos idiomas (alemán, inglés y bereber) en un espacio pentadimensional que captura la complejidad en cada una de nuestras categorías lingüísticas. El área de cada polígono puede ser vista como una

medida de la complejidad total del respectivo idioma, y el hecho de que un polígono esté totalmente incluido en otro representa una señal de dominación en términos de complejidad. Por ejemplo, el pentágono correspondiente al idioma inglés está incluido dentro del que corresponde al idioma alemán, puesto que el alemán es más complejo que el inglés en las cinco categorías que estamos analizando. Por el contrario, el pentágono que representa al bereber no está completamente incluido en ninguno de los otros dos pentágonos (a los cuales tampoco él incluye). Esto se debe a que el bereber es menos complejo que el alemán y que el inglés en lo que se refiere a su sistema verbal y a su vocabulario, pero más complejo que ellos en su morfología y en su sintaxis.

El uso de este tipo de comparaciones multidimensionales también se puede extender a todo el conjunto de variables definidas en la sección 3. Si hacemos eso, hallaremos una interpretación particular bajo la cual nuestra base de datos de 100 idiomas genera un apoyo relativamente fuerte a la hipótesis de igual complejidad. La misma tiene que ver con el siguiente hecho: no hay ninguna lengua en la muestra que sea tan simple como para no tener un valor de complejidad mayor que cualquier otra lengua en por lo menos una característica. De hecho, si chequeamos los 4950 pares de idiomas, veremos que en todos ellos siempre hay al menos una característica para la cual el primer idioma es más complejo que el segundo, y al menos otra característica para la cual el segundo idioma es más complejo que el primero.

Consideremos, por ejemplo, uno de los casos más extremos, que es el que relaciona los idiomas zulú y suajili. El zulú tiene un índice de complejidad total igual a 0,5336, el cual se encuentra bien por encima de la media (ocupa el puesto 17 en el *ranking* de 100 idiomas). El suajili, en cambio, tiene un índice de complejidad total de 0,4344, algo por debajo de la media (ocupa el puesto 52). El zulú sobrepasa en complejidad al suajili en cuatro de las cinco categorías que hemos definido en el presente trabajo, y en la restante (morfología) ambos son igualmente complejos. El zulú, además, es más complejo que el suajili en siete características diferentes (7A, 13A, 41A, 64A, 65A, 129A y 143F), pero el suajili lo supera en una característica (37A: artículos definidos). Por ende, si pensamos que, por tener artículos definidos, el suajili es capaz de

compensar los diferentes aspectos de tipo fonológico, sintáctico, verbal y léxico en los cuales el zulú es más complejo, entonces podemos decir que, en cierto sentido, el suajili y el zulú son igualmente complejos (o, por lo menos, que ninguno de los dos es más complejo que el otro en todas las posibles dimensiones lingüísticas analizadas).<sup>21</sup>

Otro ejemplo particularmente extremo es el que aparece cuando comparamos el idioma abjasio con el hmong njua. El abjasio es más complejo que el hmong njua en 33 características, y ese es el máximo número de características para las cuales una lengua supera a otra en nuestra base de datos. Más aún, los índices de complejidad del abjasio son mayores que los del hmong njua en las cinco categorías en las cuales hemos dividido la muestra. El hmong njua, sin embargo, es más complejo que el abjasio en 6 características (13A, 41A, 55A, 84A, 92A y 107A). Si esto alcanza para compensar las múltiples variables para las cuales el abjasio es más complejo que el hmong njua (que incluyen estructura silábica, consonantes no habituales, morfología flexiva, pronombres de cortesía, número de géneros, número de casos, alineamiento morfosintáctico, negación, casi todas las categorías verbales, etc.) es por lo tanto un asunto relacionado con la importancia relativa que uno le dé a esas seis características.

## **7. Consideraciones finales**

Las conclusiones que pueden extraerse de los distintos tipos de análisis realizados, utilizando la información de la muestra de 100 idiomas del WALs, dependen mucho del aspecto de la comparación en el cual uno esté más interesado. Por ejemplo, si solo observamos los coeficientes de correlación estándar de las 60 variables binarias que hemos construido, una conclusión posible es que los efectos de compensación entre medidas de complejidad tienen poca importancia, ya que solamente unos pocos pares de variables (4,8% del total) generan coeficientes estadísticamente significativos cuyo signo

---

<sup>21</sup> El suajili y el zulú son también idiomas que pertenecen a la misma sub-familia (bantú) y, quizás por eso, comparten un gran número de características: tienen el mismo valor en 52 de las 60 variables binarias definidas en el presente trabajo. El caso con mayor superposición de este tipo es el que relaciona al vietnamita y el tailandés, que no tienen un origen común pero sí pertenecen a la misma área lingüística. Estas dos lenguas solo difieren en las características 7A y 37A (en las cuales el vietnamita es el más complejo) y en la característica 129A (en la cual el tailandés es más complejo).

es negativo. La misma conclusión puede obtenerse si miramos los coeficientes de correlación estándar entre los cinco índices de complejidad calculados. En dicho conjunto hay solamente dos valores negativos (fonología vs. sintaxis, y fonología vs. vocabulario) y ninguno de ellos es estadísticamente significativo al 5% de probabilidad.

La importancia relativa de los *trade-offs* de complejidad, sin embargo, se incrementa considerablemente cuando tomamos en cuenta la posible interdependencia entre las distintas variables lingüísticas, y más aún cuando controlamos por factores de tipo geográfico, filogenético y de tamaño poblacional. Utilizando técnicas de correlación parcial, que permiten incluir todos esos elementos, hallamos que algunas correlaciones negativas importantes aparecen con cierta claridad, tales como las que relacionan consonantes con vocales, estructura silábica con tono, número de géneros con marcación del objeto directo, marcación de la persona verbal con predicación nominal y espacial, contrastes de distancia en los demostrativos con conjunción de la frase nominal, etc. También encontramos una correlación negativa y significativa entre los índices de complejidad fonológica y léxica, y una proporción relativamente elevada de correlaciones negativas y significativas entre variables de tipo sintáctico.

La existencia de efectos de compensación entre indicadores de complejidad, sin embargo, no implica necesariamente que todos los idiomas sean igualmente complejos en ningún sentido específico.<sup>22</sup> En la muestra de 100 idiomas del WALs, por ejemplo, hay algunas lenguas que son complejas en una gran cantidad de características, en tanto que hay otras que son simples en la mayor parte de ellas. Más aún, existen muchos pares de idiomas para los cuales una lengua es más compleja que la otra en las cinco categorías utilizadas para clasificar las distintas características idiomáticas (fonología, morfología, sintaxis, sistema verbal y vocabulario). Sin embargo, lo que no existe en la muestra bajo estudio es ningún idioma que domine a ningún otro en todas las características, y esto implica que cualquiera de dichos idiomas es más complejo que cualquiera de los 99 restantes en al menos una variable.

De cualquier modo, si quisiéramos encontrar un caso, fuera de nuestra muestra de

---

<sup>22</sup> Sobre este tema, véase Fenk-Oczlon & Fenk (2011).

100 idiomas, para el cual un idioma o dialecto estuviera completamente dominado por otro, la búsqueda no debería ser demasiado difícil. Por ejemplo, es muy probable que un idioma criollo sea más simple que su lengua madre en varias características lingüísticas, y que no sea más complejo que ella en el resto de las características. Lo que es probablemente imposible, sin embargo, es encontrar un idioma (criollo o no) que sea tan simple que no supere en complejidad a ningún otro idioma en ninguna característica. El único caso de este tipo sería una lengua que fuera simple en las 60 características bajo análisis, ya que para cada una de dichas características existen varios ejemplos de idiomas simples. Dicha lengua, por cierto, no existe en la muestra de 100 idiomas del WALS (en la cual los ejemplos más simples tienen al menos 14 características complejas), y tampoco es probable que exista fuera de dicha muestra.<sup>23</sup>

Si analizamos el problema de esta manera, los efectos de compensación entre indicadores de complejidad idiomática deben necesariamente existir, al menos en cierto grado. Una manera de ver eso es pensar en términos del enfoque propuesto por la “lingüística sinérgica”, cuya idea básica es que el lenguaje es un sistema organizado y autorregulado cuyas propiedades provienen de la interacción de distintos requisitos.<sup>24</sup> Entre ellos pueden mencionarse los requisitos de codificación (necesidad de proveer expresiones para distintos significados), economía (necesidad de utilizar el menor número posible de elementos para producir las expresiones deseadas) y estabilidad (necesidad de modificar el idioma lo menos posible para que pueda ser comprendido).

Si utilizamos este enfoque analizar el fenómeno de la complejidad idiomática, podemos pensar que la misma es buena para satisfacer el requisito de codificación, pero mala para satisfacer el requisito de economía. Como todo idioma necesita balancear el cumplimiento de ambos requisitos, a fin de funcionar como un medio de comunicación eficiente, es improbable que haya idiomas demasiado simples o demasiado complejos (o, más aún, demasiado simples o demasiado complejos en todos sus componentes). Esta sería la principal base teórica para justificar la existencia de efectos de compensación

---

<sup>23</sup> El extremo opuesto (es decir, una lengua que sea compleja en las 60 características analizadas) es también igualmente improbable.

<sup>24</sup> Para una buena reseña de esta literatura, véase Köhler (2005).

entre indicadores de complejidad. Sin embargo, la necesidad de estabilidad puede hacer que algunos idiomas se vuelvan más complejos que otros por razones históricas, ya que los cambios que se producen en un nivel de complejidad no necesariamente generan cambios opuestos en otro nivel (en virtud del costo que los mismos pueden tener en términos de la comprensión de la lengua por parte de personas que vienen utilizándola en su forma previa). Como los idiomas están sujetos a distintas presiones para ir modificándose (ligadas con su evolución cultural, tamaño de la población, contacto con otros idiomas, etc.), y esto puede implicar procesos de simplificación o complejización, el requisito de estabilidad hace que algunas lenguas se vuelvan más simples o más complejas que otras en un número variable de características.<sup>25</sup> Esta fuerza, sin embargo, nunca tendrá la capacidad suficiente como para eliminar completamente los efectos de compensación, ya que eso implicaría una violación de las reglas que ayudan a satisfacer simultáneamente los requisitos de codificación y de economía.

Como conclusión general de todo el análisis, por lo tanto, podemos afirmar que nuestro estudio permite visualizar la existencia de efectos de compensación entre indicadores de complejidad idiomática, pero también permite comprobar que dichos efectos suelen estar ocultos, debido a la interferencia de otros factores lingüísticos y extralingüísticos. Sus señales más evidentes, sin embargo, surgen de considerar los siguientes hechos estilizados:

- a) Las correlaciones negativas entre variables de complejidad se incrementan cuando controlamos por el efecto de otras variables;
- b) La distribución de la complejidad total entre idiomas tiene un desvío estándar menor que la distribución de la complejidad en cualquier componente particular del lenguaje (fonología, morfología, sintaxis, sistema verbal o vocabulario); y,
- c) Si tomamos una lista lo suficientemente larga de características lingüísticas, cuya presencia o ausencia pueda interpretarse como un indicador de complejidad, no hay ningún idioma que domine a ningún otro en todas esas características (especialmente si comparamos lenguas que no sean demasiado parecidas entre sí, como es el caso de las

---

<sup>25</sup> Este es el argumento básico de una rama de la literatura que relaciona la complejidad idiomática con factores sociolingüísticos. Véase, por ejemplo, Trudgill (2009).

incluidas en la muestra de 100 idiomas del WALs).

### Referencias bibliográficas

- Coloma, Germán (2013). Un modelo estadístico de ecuaciones simultáneas sobre la interacción de variables fonológicas, Documento de Trabajo Nro 519. Buenos Aires, Universidad del CEMA.
- Coloma, Germán (2014). La existencia de correlación negativa entre distintos aspectos de la complejidad de los idiomas, Documento de Trabajo Nro 536. Buenos Aires, Universidad del CEMA.
- Dahl, Osten (2011). Are Small Languages More or Less Complex than Big Ones? *Linguistic Typology* 15: 171-175.
- Dryer, Matthew (2009). Problems Testing Typological Correlations with the Online WALs. *Linguistic Typology* 13: 121-135.
- Dryer, Matthew & Martin Haspelmath (2013). *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Fenk-Oczlon, Gertraud & August Fenk (2008). Complexity Trade-Offs Between the Subsystems of Language. En M. Miestamo, K. Sinnemäki & F. Karlsson (eds.), *Language Complexity: Typology, Contact and Change*, 43-65. Amsterdam: John Benjamins.
- Fenk-Oczlon, Gertraud & August Fenk (2011). Complexity Trade-Offs in Language Do Not Imply an Equal Overall Complexity. En V. Solovyev & V. Polyakov (eds.), *Text Processing and Cognitive Technologies*, 145-148. Kazan: Kazan State University Press.
- Joseph, John & Frederick Newmeyer (2012). ‘All Languages Are Equally Complex’: The Rise and Fall of a Consensus. *Historiographia Linguistica* 39: 341-368.
- Köhler, Reinhard (2005). Synergetic Linguistics. En G. Altmann, R. Köhler & R. Piotrowski (eds.), *Quantitative Linguistics: An International Handbook*, 760-774. Berlín: De Gruyter.
- Maddieson, Ian (2007). Issues of Phonological Complexity: Statistical Analysis of the Relationship Between Syllable Structures, Segment Inventories and Tone Contrasts. En M. Solé, P. Beddor & M. Ohala (eds.), *Experimental Approaches to Phonology*, 93-103. Nueva York: Oxford University Press.
- Matasovic, Ranko (2014). Verbal and Adnominal Agreement: Areal Distribution and Typological Correlations. *Linguistic Typology* 18: 171-214.
- McWhorter, John (2001). The World’s Simplest Grammars Are Creole Grammars. *Linguistic Typology* 5: 125-166.
- Miestamo, Matti (2008). Grammatical Complexity in a Cross-Linguistic Perspective. En

- M. Miestamo, K. Sinnemäki & F. Karlsson (eds.), *Language Complexity: Typology, Contact and Change*, 23-41. Amsterdam: John Benjamins.
- Moran, Steven & Damián Blasi (2014). Cross-Linguistic Comparison of Complexity Measures in Phonological Systems. En F. Newmayer & L. Preston (eds.), *Measuring Grammatical Complexity*, 217-240. Nueva York: Oxford University Press.
- Nettle, Daniel (1995). Segmental Inventory Size, Word Length and Communicative Efficiency. *Linguistics* 33: 359-367.
- Nichols, Johanna (2009). Linguistic Complexity: A Comprehensive Definition and Survey. En G. Sampson, D. Gil & P. Trudgill (eds.), *Language Complexity as an Evolving Variable*, 110-125. Oxford: Oxford University Press.
- Parkvall, Mikael (2008). The Simplicity of Creoles in a Cross-Linguistic Perspective. En M. Miestamo, K. Sinnemäki & F. Karlsson (eds.), *Language Complexity: Typology, Contact and Change*, 265-285. Amsterdam: John Benjamins.
- Prokhorov, A. V. (2002). Partial Correlation Coefficient. En M. Hazewinkel (ed.), *Encyclopedia of Mathematics*. Nueva York: Springer.
- Sampson, Geoffrey (2009). A Linguistic Axiom Challenged. En G. Sampson, D. Gil & P. Trudgill (eds.), *Language Complexity as an Evolving Variable*, 1-18. Oxford: Oxford University Press.
- Shosted, Ryan (2006). Correlating Complexity: A Typological Approach. *Linguistic Typology* 10: 1-40.
- Sinnemäki, Kaius (2008). Complexity Trade-Offs in Core Argument Marking. En M. Miestamo, K. Sinnemäki & F. Karlsson (eds.), *Language Complexity: Typology, Contact and Change*, 67-88. Amsterdam: John Benjamins.
- Trudgill, Peter (2009). Sociolinguistic Typology and Complexification. En G. Sampson, D. Gil & P. Trudgill (eds.), *Language Complexity as an Evolving Variable*, 98-109. Oxford: Oxford University Press.
- Wichmann, Soren, Taraka Rama & Eric Holman (2011). Phonological Diversity, Word Length and Population Sizes Across Languages: The ASJP Evidence. *Linguistic Typology* 15: 157-177.

## Apéndice 1: Lista de idiomas en la muestra del WALS

Nro.	Idioma	Macro-área	Familia
1	Abjasio	Eurasia	Caucásica Noroccidental
2	Acoma	Norteamérica	Keresana
3	Alamblak	Papunesia	Sepik
4	Alemán	Eurasia	Indo-Europea
5	Amele	Papunesia	Trans-Neoguineana
6	Apurina	Sudamérica	Arahuaca
7	Árabe (Egipcio)	Eurasia	Afro-Asiática
8	Arapesh (Montañés)	Papunesia	Kombio
9	Asmat	Papunesia	Trans-Neoguineana
10	Bagirmi	Africa	Nilo-Sahara
11	Barasano	South America	Tucanoana
12	Bereber (Marroquí)	Africa	Afro-Asiática
13	Birmano	Eurasia	Sino-Tibetana
14	Burushaski	Eurasia	Burushaskiana
15	Canarés	Eurasia	Dravídica
16	Canela-Kraho	Sudamérica	Macro-Ge
17	Chamorro	Papunesia	Austronesia
18	Chukchi	Eurasia	Chukotkana
19	Coreano	Eurasia	Coreánica
20	Cree (Praderas)	Norteamérica	Algica
21	Daga	Papunesia	Dagana
22	Dani (Valle Inferior)	Papunesia	Trans-Neoguineana
23	Español	Eurasia	Indo-Europea
24	Finlandés	Eurasia	Urálica
25	Fiyiano	Papunesia	Austronesia
26	Francés	Eurasia	Indo-Europea
27	Georgiano	Eurasia	Kartveliana
28	Gooniyandi	Australia	Bunubana
29	Grebo	Africa	Niger-Congo
30	Griego (Moderno)	Eurasia	Indo-Europea
31	Groenlandés (Oeste)	Eurasia	Eskimo-Aleutiana
32	Guaraní	Sudamérica	Tupiana
33	Hausa	Africa	Afro-Asiática
34	Hebreo (Moderno)	Eurasia	Afro-Asiática
35	Hindi	Eurasia	Indo-Europea
36	Hixkaryana	Sudamérica	Caribana
37	Hmong Njua	Eurasia	Hmong-Mien
38	Imonda	Papunesia	Fronteriza
39	Indonesio	Papunesia	Austronesia
40	Inglés	Eurasia	Indo-Europea
41	Jacalteco	Norteamérica	Maya
42	Japonés	Eurasia	Japónica
43	Karok	Norteamérica	Karokiana

44	Kayardild	Australia	Tangkica
45	Kewa	Papunesia	Trans-Neoguineana
46	Khoekhoe	Africa	Khoisan
47	Kiowa	Norteamérica	Tanoana
48	Koasati	Norteamérica	Muskogeana
49	Koyraboro Senni	Africa	Nilo-Sahara
50	Krongo	Africa	Kadugliana
51	Kutenai	Norteamérica	Salish
52	Lakhota	Norteamérica	Siouan
53	Lango	Africa	Nilo-Sahara
54	Lavukaleve	Papunesia	Papú Oriental
55	Lezguiano	Eurasia	Daguestaniana
56	Luvale	Africa	Niger-Congo
57	Makah	Norteamérica	Wakashana
58	Malgache	Africa	Austronesia
59	Mandarín	Eurasia	Sino-Tibetana
60	Mangarrayi	Australia	Mangarrayiana
61	Mapuche	Sudamérica	Araucana
62	Maricopa	Norteamérica	Hokan
63	Martuthunira	Australia	Pama-Nyungan
64	Maung	Australia	Iwaidjana
65	Maybrat	Papunesia	Papú Occidental
66	Meithei	Eurasia	Sino-Tibetana
67	Mixteco (Chalcatongo)	Norteamérica	Oto-Mangueana
68	Mongol (Khalkha)	Eurasia	Altaica
69	Ngiyambaa	Australia	Pama-Nyungan
70	Oneida	Norteamérica	Irocuoiana
71	Oromo (Harar)	Africa	Afro-Asiática
72	Otomí (Mezquital)	Norteamérica	Oto-Mangueana
73	Paiwan	Papunesia	Austronesia
74	Persa	Eurasia	Indo-Europea
75	Pirahá	Sudamérica	Mura
76	Quechua (Ecuatoriano)	Sudamérica	Quechuana
77	Rama	Norteamérica	Chibchana
78	Rapanui	Papunesia	Austronesia
79	Ruso	Eurasia	Indo-Europea
80	Sango	Africa	Niger-Congo
81	Sanuma	Sudamérica	Yanomana
82	Slave	Norteamérica	Na-Dene
83	Suajili	Africa	Niger-Congo
84	Supyire	Africa	Niger-Congo
85	Tagalo	Papunesia	Austronesia
86	Tailandés	Eurasia	Tai-Kadai
87	Tiwi	Australia	Tiwiana
88	Tukang Besi	Papunesia	Austronesia
89	Turco	Eurasia	Altaica

90	Vasco	Eurasia	Vascónica
91	Vietnamita	Eurasia	Austro-Asiática
92	Warao	Sudamérica	Waraoana
93	Wari	Sudamérica	Chapacurana
94	Wichita	Norteamérica	Caddoana
95	Wichí	Sudamérica	Mataca
96	Yagua	Sudamérica	Peba-Yaguana
97	Yaqui	Norteamérica	Uto-Azteca
98	Yoruba	Africa	Niger-Congo
99	Zoque (Copainala)	Norteamérica	Mixe-Zoque
100	Zulú	Africa	Niger-Congo

## Apéndice 2: Lista de variables binarias de complejidad

Nro.	Característica	Compleja (variable = 1) si:
	<b>Fonología</b>	
1A	Inventario de fonemas consonánticos	Inventario de consonantes > 25.
2A	Inventario de tipos vocálicos	Tipos de vocales > 6.
4A	Sonoridad de oclusivas y fricativas	La sonoridad es distintiva en al menos un fonema oclusivo o fricativo.
6A	Consonantes uvulares	Existe al menos un fonema consonántico uvular.
7A	Consonantes glotalizadas	Existe al menos un fonema consonántico glotalizado.
10A	Vocales nasalizadas	La nasalización es distintiva en al menos un fonema vocálico.
11A	Vocales anteriores redondeadas	Existe al menos un fonema vocálico anterior redondeado.
12A	Estructura silábica	La estructura silábica es compleja.
13A	Tono	El tono es distintivo.
14A	Posición del acento	La posición del acento no está fija en una sílaba determinada.
19A	Presencia de consonantes no habituales	Hay al menos una consonante “no habitual” (labial-velar, clic, faríngea, /θ/).
	<b>Morfología</b>	
20A	Fusión de formantes flexivos	El idioma no es “aislante”.
22A	Síntesis en la flexión de los verbos	Categorías por palabra > 5.
23A	Marcación del objeto directo	Existe alguna marcación del objeto.
26A	Prefijación y sufijación en la morfología flexiva	Existe morfología flexiva (prefijación, sufijación o ambas).
28A	Sincretismo de casos	Existe marcación de casos.
29A	Sincretismo en la marcación de la persona o el número en la conjugación verbal	Existe marcación verbal de persona o número (sincrética o no sincrética).
30A	Número de géneros	Número de géneros > 1.
34A	Ocurrencia del plural nominal	Todos los sustantivos tienen forma plural, y su uso es obligatorio.

48A	Marcación de la persona en adposiciones	Hay algún tipo de marcación de persona.
49A	Número de casos	Número de casos > 1.
51A	Posición de clíticos y afijos de caso	Existen afijos de caso o clíticos adposicionales.
	<b>Sintaxis</b>	
58A	Flexión posesiva obligatoria	Existe flexión posesiva.
59A	Clasificación de los posesivos	Existen dos o más clases de clasificadores posesivos.
81A	Orden de sujeto, verbo y objeto	No hay un orden dominante.
84A	Orden de objeto, verbo y complemento oblicuo	No hay un orden dominante.
92A	Partículas interrogativas polares	Existen partículas interrogativas polares.
98A	Alineamiento morfosintáctico en frases nominales	El alineamiento no es neutral.
100A	Alineamiento morfosintáctico en frases verbales	El alineamiento no es neutral.
104A	Orden de los marcadores de persona en los verbos	Hay marcadores para el agente y para el paciente de los verbos transitivos.
113A	Negación simétrica y asimétrica	Hay estructuras distintas para las formas afirmativas y negativas.
119A	Predicación nominal y espacial	Hay distintos verbos para la predicación nominal y espacial.
143E	Morfemas negativos preverbales	Existen morfemas negativos preverbales.
143F	Morfemas negativos posverbales	Existen morfemas negativos posverbales.
	<b>Sistema verbal</b>	
65A	Aspecto perfecto e imperfecto	Hay marcación gramatical del aspecto perfecto e imperfecto.
66A	Tiempo pasado	Hay marcación gramatical del tiempo pasado.
67A	Tiempo futuro	Hay una forma flexiva del tiempo futuro.
68A	Tiempos perfectos	Existen tiempos perfectos diferenciados.
69A	Posición de los afijos de tiempo y aspecto	Hay formas flexivas para marcar el tiempo o el aspecto.
70A	Modo imperativo	Hay al menos una forma para el imperativo de segunda persona.
73A	Modo optativo	Hay una forma flexiva del modo optativo.
77A	Distinciones semánticas de evidencialidad	Existen marcadores gramaticales de evidencialidad.
79A	Irregularidades relacionadas con el tiempo o el aspecto verbal	Hay irregularidades relacionadas con el tiempo, el aspecto o ambos.
80A	Irregularidades relacionadas con el número	Los verbos son distintos si la persona es singular o plural.
107A	Construcciones pasivas	Existen construcciones pasivas.
108A	Construcciones antipasivas	Existen construcciones antipasivas (paciente implícito u oblicuo).
109A	Construcciones aplicativas	Existen construcciones aplicativas (benefactivas o no benefactivas).

	<b>Vocabulario</b>	
37A	Artículos definidos	Hay artículos definidos.
38A	Artículos indefinidos	Hay artículos indefinidos.
39A	Distinción inclusiva/exclusiva en pronombres independientes	Hay distintos pronombres personales plurales inclusivos y exclusivos.
40A	Distinción inclusiva/exclusiva en formas verbales	Hay distintas formas verbales plurales inclusivas y exclusivas.
41A	Contrastes de distancia en demostrativos	Hay más de un contraste de distancia.
44A	Distinción de género en pronombres personales independientes	Existe algún tipo de distinción de género entre pronombres personales.
45A	Pronombres personales de cortesía	Existe algún tipo de distinción de cortesía en pronombres de segunda persona.
47A	Pronombres reflexivos e intensificadores	Los intensificadores son distintos de los pronombres reflexivos.
55A	Clasificadores numerales	Existen clasificadores numerales.
63A	Conjunción de la frase nominal	Las conjunciones para frases nominales y comitativas son diferentes.
64A	Conjunción nominal y verbal	Las conjunciones para frases nominales y verbales son diferentes.
129A	Mano y brazo	Existen palabras distintas para “mano” y “brazo”.
130A	Dedo y mano	Existen palabras distintas para “dedo” y “mano”.

### Apéndice 3: Índices de complejidad

<b>Idioma</b>	<b>Fonología</b>	<b>Morfolog</b>	<b>Sintaxis</b>	<b>Verbos</b>	<b>Vocabul</b>	<b>Total</b>
Abjasio	0.6364	0.8182	0.5000	0.7692	0.6923	0.6832
Acoma	0.4545	0.4545	0.8333	0.3846	0.2308	0.4716
Alamblak	0.3636	0.8182	0.5000	0.5385	0.3846	0.5210
Alemán	0.5455	0.8182	0.5000	0.5385	0.6154	0.6035
Amele	0.2727	0.4545	0.6667	0.4615	0.3077	0.4326
Apurina	0.1818	0.7273	0.4167	0.3846	0.3077	0.4036
Árabe	0.5455	0.6364	0.4167	0.4615	0.3077	0.4735
Arapesh	0.1818	0.5455	0.5833	0.3077	0.3846	0.4006
Asmat	0.0000	0.3636	0.5833	0.5385	0.3846	0.3740
Bagirmi	0.5455	0.6364	0.5000	0.1538	0.3077	0.4287
Barasano	0.3636	0.7273	0.5000	0.3846	0.6923	0.5336
Bereber	0.4545	0.9091	0.6667	0.3077	0.3846	0.5445
Birmanio	0.4545	0.4545	0.5000	0.1538	0.4615	0.4049
Burushaski	0.5455	1.0000	0.5000	0.7692	0.4615	0.6552
Canarés	0.3636	0.7273	0.4167	0.6923	0.5385	0.5477
Canela	0.1818	0.3636	0.4167	0.3077	0.3077	0.3155
Chamorro	0.1818	0.3636	0.3333	0.4615	0.3846	0.3450
Chukchi	0.1818	0.7273	0.6667	0.3077	0.3077	0.4382
Coreano	0.2727	0.5455	0.3333	0.4615	0.6154	0.4457

Cree	0.0000	0.6364	0.6667	0.3077	0.6923	0.4606
Daga	0.0909	0.5455	0.4167	0.3077	0.2308	0.3183
Dani	0.0909	0.7273	0.5833	0.5385	0.3077	0.4495
Español	0.2727	0.6364	0.4167	0.6154	0.6923	0.5267
Finlandés	0.2727	0.8182	0.4167	0.4615	0.3077	0.4554
Fiyiano	0.1818	0.4545	0.5833	0.3077	0.5385	0.4132
Francés	0.6364	0.6364	0.3333	0.6923	0.6154	0.5828
Georgiano	0.4545	0.8182	0.3333	0.7692	0.3077	0.5366
Gooniyandi	0.0000	0.5455	0.6667	0.3846	0.3846	0.3963
Grebo	0.5455	0.6364	0.3333	0.6154	0.2308	0.4723
Griego	0.2727	0.8182	0.5000	0.5385	0.6923	0.5643
Groenlandés	0.1818	0.8182	0.5000	0.5385	0.3077	0.4692
Guaraní	0.1818	0.4545	0.5833	0.3846	0.6154	0.4439
Hausa	0.4545	0.5455	0.4167	0.2308	0.4615	0.4218
Hebreo	0.2727	0.7273	0.4167	0.4615	0.4615	0.4679
Hindi	0.4545	0.8182	0.4167	0.6154	0.2308	0.5071
Hixkaryana	0.1818	0.5455	0.4167	0.5385	0.5385	0.4442
Hmong Njua	0.3636	0.0000	0.3333	0.0769	0.3846	0.2317
Imonda	0.2727	0.6364	0.5000	0.6154	0.3077	0.4664
Indonesio	0.1818	0.0909	0.3333	0.3077	0.4615	0.2751
Inglés	0.4545	0.7273	0.2500	0.3846	0.4615	0.4556
Jacalteco	0.2727	0.4545	0.4167	0.5385	0.3846	0.4134
Japonés	0.3636	0.4545	0.4167	0.3846	0.5385	0.4316
Karok	0.3636	0.6364	0.8333	0.6154	0.3846	0.5667
Kayardild	0.0000	0.5455	0.5000	0.3846	0.3846	0.3629
Kewa	0.1818	0.6364	0.4167	0.6154	0.3077	0.4316
Khoekhoe	0.2727	0.6364	0.6667	0.4615	0.6154	0.5305
Kiowa	0.4545	0.6364	0.6667	0.5385	0.3846	0.5361
Koasati	0.3636	0.8182	0.7500	0.5385	0.5385	0.6017
Koyraboro	0.1818	0.1818	0.4167	0.3846	0.3077	0.2945
Krongo	0.3636	0.6364	0.4167	0.6154	0.4615	0.4987
Kutenai	0.3636	0.3636	0.5833	0.3077	0.3846	0.4006
Lakhota	0.4545	0.5455	0.5833	0.2308	0.5385	0.4705
Lango	0.3636	0.4545	0.4167	0.3846	0.3077	0.3854
Lavukaleve	0.0909	0.7273	0.5000	0.3077	0.6154	0.4483
Lezguiano	0.6364	0.6364	0.4167	0.6923	0.3077	0.5379
Luvale	0.1818	0.7273	0.5000	0.4615	0.5385	0.4818
Makah	0.5455	0.2727	0.6667	0.3077	0.5385	0.4662
Malgache	0.1818	0.4545	0.3333	0.3846	0.4615	0.3632
Mandarín	0.3636	0.1818	0.3333	0.2308	0.6154	0.3450
Mangarrayi	0.0909	0.8182	0.6667	0.3846	0.6923	0.5305
Mapuche	0.0909	0.6364	0.4167	0.3077	0.4615	0.3826
Maricopa	0.2727	0.7273	0.5000	0.6154	0.3077	0.4846
Martuthunira	0.0000	0.5455	0.3333	0.4615	0.3846	0.3450
Maung	0.0909	0.4545	0.5833	0.4615	0.4615	0.4104
Maybrat	0.0000	0.6364	0.4167	0.0769	0.3846	0.3029

Meithei	0.1818	0.5455	0.4167	0.4615	0.5385	0.4288
Mixteco	0.4545	0.4545	0.2500	0.3846	0.1538	0.3395
Mongol	0.3636	0.5455	0.3333	0.5385	0.2308	0.4023
Ngiyambaa	0.1818	0.5455	0.5833	0.3846	0.3077	0.4006
Oneida	0.2727	0.7273	0.6667	0.5385	0.4615	0.5333
Oromo	0.4545	0.7273	0.4167	0.4615	0.4615	0.5043
Otomí	0.5455	0.4545	0.5000	0.5385	0.3077	0.4692
Paiwan	0.1818	0.3636	0.3333	0.4615	0.4615	0.3604
Persa	0.2727	0.6364	0.4167	0.5385	0.5385	0.4805
Pirahá	0.2727	0.4545	0.4167	0.2308	0.2308	0.3211
Quechua	0.4545	0.8182	0.5000	0.4615	0.3846	0.5238
Rama	0.1818	0.4545	0.5833	0.3846	0.5385	0.4286
Rapanui	0.0000	0.1818	0.4167	0.1538	0.5385	0.2582
Ruso	0.3636	0.8182	0.4167	0.4615	0.3846	0.4889
Sango	0.6364	0.2727	0.2500	0.0000	0.2308	0.2780
Sanuma	0.1818	0.4545	0.3333	0.3077	0.5385	0.3632
Slave	0.5455	0.6364	0.5000	0.6923	0.4615	0.5671
Suajili	0.2727	0.6364	0.4167	0.5385	0.3077	0.4344
Supyire	0.3636	0.2727	0.4167	0.4615	0.4615	0.3952
Tagalo	0.0909	0.2727	0.2500	0.5385	0.6154	0.3535
Tailandés	0.2727	0.0000	0.2500	0.1538	0.4615	0.2276
Tiwi	0.0000	0.4545	0.6667	0.3846	0.3846	0.3781
Tukang Besi	0.0909	0.2727	0.5833	0.4615	0.3846	0.3586
Turco	0.3636	0.8182	0.4167	0.6154	0.4615	0.5351
Vasco	0.2727	0.7273	0.5833	0.7692	0.6154	0.5936
Vietnamita	0.3636	0.0000	0.2500	0.1538	0.4615	0.2458
Warao	0.0909	0.4545	0.3333	0.3846	0.2308	0.2988
Wari	0.2727	0.3636	0.4167	0.1538	0.5385	0.3491
Wichita	0.1818	0.5455	0.5000	0.6923	0.4615	0.4762
Wichí	0.1818	0.3636	0.7500	0.3077	0.5385	0.4283
Yagua	0.1818	0.5455	0.5000	0.3077	0.4615	0.3993
Yaqui	0.1818	0.6364	0.4167	0.6923	0.4615	0.4777
Yoruba	0.4545	0.2727	0.2500	0.0769	0.3846	0.2878
Zoque	0.0909	0.8182	0.5833	0.6923	0.6923	0.5754
Zulú	0.4545	0.6364	0.5000	0.6154	0.4615	0.5336
<b>Máximo</b>	<b>0.6364</b>	<b>1.0000</b>	<b>0.8333</b>	<b>0.7692</b>	<b>0.6923</b>	<b>0.6832</b>
<b>Mínimo</b>	<b>0.0000</b>	<b>0.0000</b>	<b>0.2500</b>	<b>0.0000</b>	<b>0.1538</b>	<b>0.2276</b>
<b>Promedio</b>	<b>0.2882</b>	<b>0.5573</b>	<b>0.4742</b>	<b>0.4377</b>	<b>0.4377</b>	<b>0.4390</b>
<b>Desvío Std.</b>	<b>0.1685</b>	<b>0.2078</b>	<b>0.1317</b>	<b>0.1717</b>	<b>0.1311</b>	<b>0.0942</b>