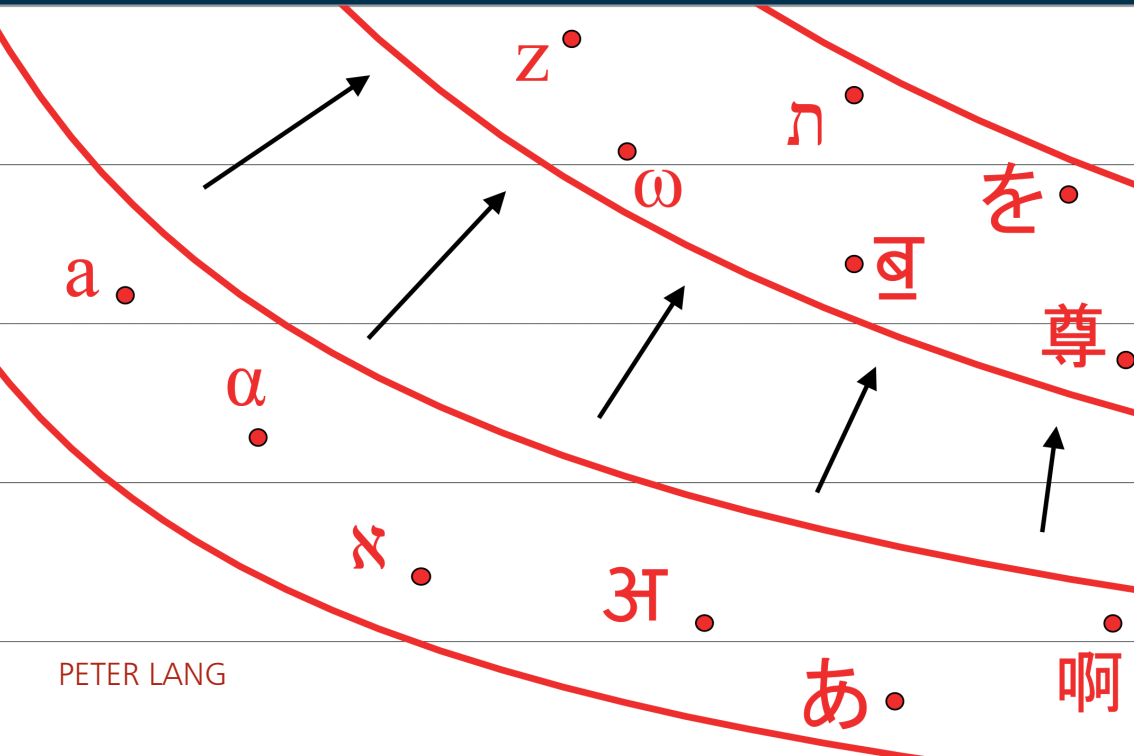


La complejidad de los idiomas

Germán Coloma



Este libro recorre distintos campos de la lingüística (fonología, morfología, sintaxis, semántica) en busca de las principales características que hacen que los idiomas sean más simples o más complejos. Luego procura medir dichas características, a través de indicadores tales como el número de sonidos, la extensión promedio de los enunciados o la frecuencia con la que se repiten las palabras. Tales medidas son a su vez comparadas entre sí, a fin de detectar posibles “efectos de compensación” que indiquen, por ejemplo, si un idioma que tiene palabras más largas tiende por otro lado a usar oraciones más cortas. En todo este recorrido, el autor introduce una serie de conceptos tomados de la lingüística cuantitativa, es decir, del estudio del lenguaje a través de métodos estadísticos. Para ello hace un uso extensivo del Atlas Mundial de Estructuras Lingüísticas (WALS), que es probablemente la base de datos más completa sobre características de los idiomas a nivel internacional. También relaciona sus resultados con teorías elaboradas por diferentes corrientes del pensamiento lingüístico, que hacen referencia a factores de tipo biológico, funcional, histórico y geográfico.

Germán Coloma es graduado de la Universidad Nacional de La Plata y de la Universidad de California, Los Ángeles. Actualmente se desempeña como profesor titular en la Universidad del CEMA (Buenos Aires, Argentina), en la cual tiene a su cargo un proyecto de investigación sobre el uso de métodos estadísticos para el análisis de problemas lingüísticos. El doctor Coloma ha sido también investigador visitante en la Universidad de California, Santa Bárbara. Si bien su principal área de investigación tiene que ver con el empleo de métodos estadísticos en economía, desde el año 2010 ha escrito varios trabajos sobre lingüística cuantitativa. Algunos de ellos han aparecido en revistas internacionales tales como *American Journal of Linguistics* (EEUU), *Language Sciences* (Holanda), *SKY Journal of Linguistics* (Finlandia), *Journal of Quantitative Linguistics* (Reino Unido), *Revista de Investigación Lingüística* (España), *Revista de Lingüística Teórica y Aplicada* (Chile) y *Glottometrics* (Alemania).



La complejidad de los idiomas

La complejidad de los idiomas

Germán Coloma



PETER LANG

Oxford • Bern • Berlin • Bruxelles • Frankfurt am Main • New York • Wien

Bibliographic information published by Die Deutsche Nationalbibliothek.
Die Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available on the Internet at <http://dnb.d-nb.de>.

A catalogue record for this book is available from the British Library.

Library of Congress Control Number: 2016959087

Cover image: © Germán Coloma.

ISBN 978-1-78707-272-5 (print) • ISBN 978-1-78707-278-7 (ePDF)
ISBN 978-1-78707-279-4 (ePub) • ISBN 978-1-78707-280-0 (mobi)

© Peter Lang AG 2017

Published by Peter Lang Ltd, International Academic Publishers,
52 St Giles, Oxford, OX1 3LU, United Kingdom
oxford@peterlang.com, www.peterlang.com

Germán Coloma has asserted his right under the Copyright, Designs and Patents Act, 1988, to be identified as Author of this Work.

All rights reserved.

All parts of this publication are protected by copyright.
Any utilisation outside the strict limits of the copyright law, without the permission of the publisher, is forbidden and liable to prosecution. This applies in particular to reproductions, translations, microfilming, and storage and processing in electronic retrieval systems.

This publication has been peer reviewed.

Printed in Germany

Para Mariana,
Javier y Carolina

Índice de contenidos

Prólogo	ix
CAPÍTULO 1	
Conceptos de complejidad	1
1.1. Aspectos generales	1
1.2. Componentes del lenguaje	9
1.3. Patrones universales y tipología lingüística	15
1.4. Efectos de compensación	18
CAPÍTULO 2	
Medición de la complejidad	27
2.1. Medidas teóricas o tipológicas	28
2.2. Medidas empíricas	31
2.3. Factores filogenéticos, geográficos y poblacionales	40
2.4. La ley de Menzerath	50
CAPÍTULO 3	
Complejidad fonológica	57
3.1. Inventario de fonemas	58
3.2. Acento y tono	74
3.3. Estructura silábica	80
CAPÍTULO 4	
Complejidad morfológica	85
4.1. Síntesis y fusión	85
4.2. Composición, derivación y flexión	96
4.3. Categorías nominales	102
4.4. Categorías verbales	107

CAPÍTULO 5

Complejidad sintáctica	113
5.1. Orden de las palabras	114
5.2. Alineamiento morfosintáctico	125
5.3. Enunciados simples y complejos	134

CAPÍTULO 6

Complejidad verbal y léxica	143
6.1. Sistemas verbales	144
6.2. Variables de complejidad léxica	154

CAPÍTULO 7

Relación entre medidas de complejidad	169
7.1. Correlación simple y parcial	170
7.2. Correlación entre variables numéricas	176
7.3. Correlación y regresión	180
7.4. Sistemas de ecuaciones y lingüística sinérgica	184
7.5. Medidas empíricas parciales y complejidad global	193

CAPÍTULO 8

Conclusiones y comentarios	205
8.1. Conclusiones basadas en el texto	206
8.2. Otros comentarios	213
8.3. Consideraciones finales	226

Referencias Bibliográficas	233
----------------------------	-----

APÉNDICE 1

Muestra de 100 idiomas del WALS	241
---------------------------------	-----

APÉNDICE 2

Muestra de 50 idiomas de “El viento norte y el sol”	245
---	-----

Índice alfabético	247
-------------------	-----

Prólogo

Este libro tiene esencialmente dos objetivos. Por un lado, busca presentar una serie de aspectos que tienen que ver con los distintos rasgos por los cuales los idiomas humanos pueden ser más simples o más complejos. Por otro lado, intenta introducir al lector en el mundo de la llamada “lingüística cuantitativa”, o sea, en el estudio de temas relacionados con el lenguaje a través de un enfoque basado en métodos estadísticos.

El texto está estructurado en ocho capítulos. El primero contiene una descripción general del tema de la complejidad de los idiomas y de los distintos componentes del lenguaje a los cuales el mismo puede aplicarse. El segundo introduce al lector en las principales formas de medición estadística de dicha complejidad, utilizando tanto medidas teóricas como empíricas. Luego vienen una serie de capítulos (3, 4, 5 y 6) en los cuales se analizan por separado distintas medidas de la complejidad de los idiomas, según el aspecto de la lengua al cual tales medidas se refieran. En el capítulo 7, en cambio, se estudian de manera sistemática las relaciones cuantitativas que pueden establecerse entre las medidas de complejidad, y con dichas relaciones se intenta desentrañar una serie de fenómenos que vinculan unas medidas con otras. En el capítulo 8, por último, se enuncian una serie de conclusiones y comentarios finales sobre la complejidad de las lenguas que se hablan en el mundo.

Si bien yo espero que este libro pueda resultar útil para especialistas en lingüística que quieran profundizar sobre la aplicación de métodos estadísticos a su campo de estudio, y que también le sirva a personas dedicadas a la estadística que quieran ver cómo su disciplina puede usarse para el análisis de fenómenos del lenguaje, mi aspiración es que además sea de interés para gente que no es necesariamente experta ni en estadística ni en lingüística. Por eso es que el modo en el cual ha sido escrito no presupone que el lector tenga una formación específica en ninguna de las dos áreas.

Lo que sí se toma como un supuesto es que quien lee este libro conoce los temas de lengua y de matemática que se enseñan habitualmente en las

escuelas de nivel medio (y que se acuerda relativamente bien de ellos). Por ese hecho, no aparecen aquí definiciones de conceptos tales como “adverbio”, “objeto directo”, “promedio” o “raíz cuadrada”. Sí me tomo cierto espacio, en cambio, para definir qué cosa es un fonema, o una familia lingüística, o una regresión estadística, o un coeficiente de correlación parcial. Eso se debe a que tales conceptos están ligados directamente con conocimientos específicos de lingüística o de estadística, y que por lo tanto van más allá de la formación habitual que puede tener una persona que dejó de estudiar lengua o matemática una vez que egresó del colegio secundario.

Otra cosa que a veces presuponen las obras de lingüística es que el lector es políglota. No es este el caso. Como este libro está escrito en castellano, lo único que yo supongo al respecto es que el lector domina bien dicho idioma, y trato por lo tanto de exprimirlo al máximo a fin de extraer la mayor cantidad posible de ejemplos de nuestra propia lengua. Cuando eso no me alcanza (porque hay muchos aspectos en los cuales otros idiomas son más simples o más complejos que el español), trato de ejemplificar usando otras lenguas a las cuales el lector promedio puede haber estado expuesto (por ejemplo, inglés, portugués o francés). Si en ellos no encuentro nada que me sirva, paso a buscar en idiomas más alejados de nuestro ámbito pero que son hablados por muchos millones de personas en el mundo (por ejemplo, chino, japonés, ruso, hindi o árabe), o bien en lenguas menos populares pero más cercanas a nosotros por razones de carácter geográfico (por ejemplo, vasco, quechua, guaraní o mapuche).

Como el libro trata el tema de la complejidad de los idiomas, no hay prácticamente ninguna página del mismo en la cual no aparezca varias veces la palabra “idioma”. Para no repetirla tanto, de vez en cuando la reemplazo por la palabra “lengua”, que opera así como un sinónimo estricto de “idioma”. Lo mismo pasa con los dos nombres que suelen usarse para nuestro idioma: español y castellano. Ambos son utilizados alternativamente para referirse al mismo concepto, y también aparecen combinados con el concepto anterior. De ese modo, en esta obra, las expresiones “idioma español”, “idioma castellano”, “lengua española” y “lengua castellana” significan exactamente lo mismo.

No quiero terminar este prólogo sin dedicarle algunos párrafos a agradecer a una serie de personas que hicieron posible que escribiera este libro. En primer lugar, debo mencionar a la Universidad del CEMA, que me

permitió dedicarle buena parte de mi tiempo laboral a esta obra, como parte del proyecto de investigación sobre el uso de métodos econométricos para el análisis de problemas lingüísticos que tengo a mi cargo en dicha casa de estudios desde el año 2014. También quiero agradecerle a la Editorial Peter Lang, que confió en la propuesta que le presenté y aceptó publicar esta obra.

Otro grupo de personas que me brindó mucha ayuda es el que puede denominarse de manera genérica “la comunidad lingüística”. Eso es algo que yo valoro mucho, porque no es común que gente que trabaja en una disciplina se ponga muy contenta cuando viene alguien de otro lado (por ejemplo, del área de la estadística económica) y empieza a escribir de temas que “les pertenecen a ellos”. En este caso en particular, sin embargo, todo lo que recibí en estos años de los especialistas en lingüística con los que me contacté fueron elogios y comentarios constructivos. Y en ese punto quiero agradecer especialmente a tres grandes figuras de la lingüística cuantitativa internacional, como son Gabriel Altmann, Reinhard Köhler y Geoffrey Sampson, y a los profesores del departamento de lingüística de la Universidad de California en Santa Bárbara (en particular a Matthew Gordon, Stefan Gries y Fermín Moscoso).

Asimismo, entre los expertos argentinos en lingüística que me ayudaron a mejorar mis trabajos sobre el tema, no puedo dejar de nombrar a Laura Colantoni, Damián Blasi, Guiomar Ciapuscio y Angelita Martínez. Las últimas dos personas en dicha lista, además, tuvieron la amabilidad de invitarme a exponer parte de los resultados de mi trabajo en la Universidad de Buenos Aires y en la Universidad Nacional de La Plata (respectivamente), y allí obtuve varias ideas y sugerencias que luego sirvieron para mejorar mis escritos.

Por último, no puedo dejar de mencionar el apoyo y la colaboración de mi esposa, Mariana Conte Grand. No solo ella me ayudó con su cariño, e impulsándome para llevar a cabo varias de las tareas que luego me conducirían a la redacción de este libro (por ejemplo, la relacionada con mi visita a Santa Bárbara), sino que también me dio ideas ligadas con varios métodos estadísticos que terminé utilizando en distintos capítulos.

Buenos Aires, octubre de 2016.

Germán Coloma

Conceptos de complejidad

1.1. Aspectos generales

1.1.1. *Complejidad absoluta y relativa*

Cuando una persona intenta intuitivamente calificar a cierto idioma como “simple” o “complejo”, suele partir del marco de referencia que le da la lengua de la cual es hablante nativo. A los que hablamos español, por ejemplo, nos resulta natural pensar que el idioma portugués es relativamente simple, y que el chino mandarín, en cambio, es más complejo. Más aún, cuando nos topamos con el idioma gallego, solemos pensar que el mismo es más simple todavía que el portugués.

De hecho, juicios como los contenidos en el párrafo anterior derivan directamente de que el portugués y el gallego son idiomas parecidos al castellano, en tanto que el mandarín es totalmente distinto. Si el portugués nos parece simple y el mandarín complejo es porque aquel tiene muchas características comunes con nuestra lengua (que, por lo tanto, no tenemos necesidad de aprender), en tanto que el mandarín tiene muchas características diferentes. Pero la pregunta respecto de qué idioma es más complejo tendrá seguramente una respuesta distinta si, en vez de hacérsela a un hablante de español, se la hacemos a un hablante de chino shanghainés. A dicha persona el mandarín le parecerá sin duda mucho más simple que el portugués, entre otras cosas porque se escribe con los mismos símbolos de su propio idioma y porque tiene una gramática virtualmente idéntica.

Los ejemplos expuestos son evidencias directas de que el concepto de complejidad idiomática puede pensarse (y, de hecho, comúnmente se piensa) como algo relativo a determinado hablante o grupo de hablantes. Pero también es posible concebir a la complejidad de un idioma en términos

absolutos, es decir, evaluar si un idioma es más complejo que otro no porque sea más o menos parecido al idioma que uno habla sino por la presencia o ausencia de ciertas “complejidades objetivas”. Comparemos, por ejemplo, la conjugación del verbo “amar” en portugués y en inglés. En portugués, este verbo se escribe igual que en castellano, y se pronuncia de una manera muy parecida. El inglés, en cambio, utiliza una palabra totalmente distinta (*love*). Sin embargo, para conjugar el verbo “amar”, en inglés solamente necesitamos tres formas (*love – loves – loved*), y una vez que las aprendemos y las combinamos con algunas reglas generales que sirven para conjugar cualquier verbo, podemos amar en presente, pasado y futuro a cualquier persona singular o plural que queramos.

En portugués, en cambio, el verbo “amar” adopta 49 formas distintas (“*amo*”, “*ama*”, “*amei*”, “*amavam*”, “*amado*”, etc.). Algunas de ellas se usan solo para una determinada función (por ejemplo, “*amo*” solo sirve para la primera persona del singular del presente del indicativo, igual que en castellano), y otras para más de una (por ejemplo, “*ama*” se usa para la tercera persona singular del presente del indicativo y para la segunda persona del singular del imperativo, también igual que en castellano).

La enorme diferencia que hay entre las pocas variantes verbales del inglés y las muchísimas que presenta el portugués (y nótese que lo que estamos describiendo es un verbo “regular”) hacen que aun un hablante de español o de gallego, que encuentra que casi todas las formas portuguesas son muy similares a las de su propio idioma, concuerde que, en términos absolutos, la conjugación del verbo “amar” en portugués es más compleja que la conjugación del verbo equivalente en inglés. Partiendo de esa conclusión, solo hay un paso para llegar a otra que nos dice que la conjugación del verbo “amar” no solo es más simple en inglés que en portugués, sino que también es más simple en inglés que en castellano. Y eso es así porque en este tema estamos comparando cierto nivel de “complejidad absoluta” y no de “complejidad relativa”.

Una definición posible de complejidad absoluta, que se utiliza para evaluar distintos aspectos de los idiomas, aparece en un artículo del lingüista finlandés Matti Miestamo (2008), y hace referencia al “número de partes de un sistema”. Otra definición alternativa es la propuesta por el estadounidense John McWhorter (2001), quien considera que un idioma es más

complejo que otro si posee “más distinciones y/o reglas explícitas”. Ambas definiciones pretenden evaluar la complejidad de manera independiente de la persona que esté llevando a cabo la evaluación: la conjugación del verbo “amar” es más compleja en portugués que en inglés porque forma un sistema que consta de 49 “partes” (contra las tres que tiene el sistema equivalente en inglés), y también es más compleja porque surge como consecuencia de la aplicación de numerosas distinciones entre formas que sirven para los distintos tiempos y personas verbales (en oposición a las dos o tres distinciones que hay que hacer en inglés para saber cuándo usar “love”, “loves” y “loved”).

Pero en lingüística existe también un concepto de complejidad relativa que intenta librarse de la trampa que se genera cuando uno evalúa un idioma muy parecido al propio y lo compara con otro que es muy diferente. Dicho concepto considera que un fenómeno idiomático es más complejo si es “más difícil de procesar o de aprender”, y su definición está asociada con la obra del holandés Wouter Kusters (2003). Para él, la mayor o menor complejidad relativa de un idioma debe analizarse desde la perspectiva de un “extranjero generalizado” (*generalized outsider*), o sea, de alguien que no habla la lengua en cuestión ni está familiarizado con la cultura de las personas que hablan dicha lengua como idioma nativo, y que solo la necesita para comunicarse con esas personas. Dado eso, Kusters considera que la complejidad relativa de un idioma es el esfuerzo que dicho extranjero generalizado debe hacer para aprender el idioma en cuestión.

1.1.2. Complejidad local y global

Los conceptos de complejidad absoluta y relativa pueden aplicarse a la evaluación de los idiomas en su conjunto o bien a la evaluación de algún componente de dichos idiomas. Cuanto más acotado sea dicho componente, más precisa será la evaluación, pero también es más probable que la misma tenga un nivel de generalidad menor. Si queremos, por ejemplo, comparar la complejidad de la conjugación del verbo “cantar” en español y en inglés, un punto que podemos observar es que todas las formas de

dicho verbo (“canto”, “cantas”, “cantaba”, “cantarían”, etc.) comienzan en nuestro idioma con la expresión “cant-” (que es lo que se conoce como la “raíz” de todas esas palabras). En inglés, en cambio, el verbo equivalente presenta solo cuatro formas (*sing – sings – sang – sung*), pero las últimas dos cambian la raíz respecto de las dos primeras. Esto nos indica que, en cierto sentido, la conjugación del verbo “cantar” es más compleja en inglés que en castellano (que usa la misma raíz para todas las formas verbales). Pero esto es un fenómeno relacionado con cierta “complejidad local”, aplicable únicamente al verbo que estamos analizando.

Cuanto más general es la comparación que realicemos, sin embargo, la evaluación de la complejidad se vuelve menos local y más “global”. La comparación entre las conjugaciones de “amar” en portugués y en inglés que hicimos en el apartado anterior, por ejemplo, es más general que la que acabamos de hacer para el verbo “cantar”, ya que “amar” es un verbo regular en los dos idiomas contrastados, en tanto que “cantar” es regular en castellano (y en portugués) e irregular en inglés.

Más general aún es la comparación entre las distintas formas del tiempo pasado que puede adoptar cualquier verbo en un idioma. En inglés, por ejemplo, se utiliza una única forma para varias situaciones en las que el castellano nos pide que distingamos entre pretérito perfecto del indicativo, pretérito imperfecto del indicativo y pretérito imperfecto del subjuntivo. Compárense, por ejemplo, las versiones en español y en inglés de los tres enunciados que aparecen más abajo:

- | | | | |
|-----|--|---|---|
| (1) | <i>Ellos <u>estuvieron</u> en España</i> | – | <i>They <u>were</u> in Spain</i> |
| (2) | <i>Ellos <u>estaban</u> en España</i> | – | <i>They <u>were</u> in Spain</i> |
| (3) | <i>Ojalá <u>estuviesen</u> en España</i> | – | <i>I wish they <u>were</u> in Spain</i> |

Los ejemplos expuestos nos sirven para ilustrar que, en términos mucho más globales, el sistema verbal del español es más complejo que el del inglés, y en eso concuerdan tanto una visión basada en el concepto de complejidad absoluta (porque el sistema verbal español tiene más “partes” y más “reglas”, y hace más distinciones que el inglés) como una basada en el concepto de complejidad relativa (porque requiere más esfuerzo de aprendizaje por parte de un “extranjero generalizado”).

El mismo tipo de comparaciones es susceptible de hacerse respecto de otros aspectos del lenguaje, como puede ser la pronunciación de las palabras. Supongamos que analizamos primero la complejidad de la palabra “sombbrero” en español y en inglés (idioma en el cual la palabra equivalente es “*hat*”). Tanto desde el punto de vista absoluto como relativo, esta palabra parece ser más simple en inglés que en español, ya que consta de solo tres sonidos (contra ocho que tiene en español) y de una sola sílaba (contra tres de la versión española). Las tres sílabas de la palabra “som-bre-ro”, además, son bastante diferentes entre sí (la primera tiene una vocal en medio de dos consonantes, la segunda tiene dos consonantes y luego una vocal, y la tercera tiene una consonante y luego una vocal). Si bien algunos de los sonidos que se usan en español para pronunciar “sombbrero” se repiten (la “o” y la “r” aparecen dos veces cada una), de cualquier modo el número de sonidos distintos que se necesitan para pronunciar dicha palabra es el doble de los que se necesitan para pronunciar “*hat*”. Todo eso hace también que desde el punto de vista del esfuerzo de procesamiento y de aprendizaje que se requiere para dominar la pronunciación de esta palabra, casi cualquier extranjero encuentre más sencillo decir “*hat*” en vez de decir “sombbrero”.

Este análisis absolutamente local de la complejidad de la pronunciación, sin embargo, se invierte totalmente si nos movemos a un entorno más general. En inglés, por ejemplo, existen 10 u 11 sonidos vocálicos que sirven para distinguir entre el significado de las palabras, mientras que en castellano solo tenemos cinco vocales ([a], [e], [i], [o] y [u]). Para el oído de un inglés, las palabras “*hat*” (“sombbrero”), “*heat*” (“calor”), “*hit*” (“golpear”), “*hut*” (“cabaña”), “*hurt*” (“dañar”), “*heart*” (“corazón”) y “*hot*” (“caliente”) solo difieren entre sí por la vocal que aparece entre [h] y [t], y existen además otras vocales adicionales (por ejemplo, las que aparecen en “*head*”, “*hoard*”, “*hood*” y “*whoèd*”) que también sirven para distinguir palabras entre sí.

En inglés, asimismo, es relativamente común encontrar vocablos tales como “*sprints*” (“corridas”). Dicha palabra, que solo tiene una sílaba, consta de una vocal precedida por tres consonantes y seguida de otras tres (con lo cual la sílaba en cuestión está compuesta por siete sonidos). Nada similar existe en español, donde el número máximo de sonidos que puede tener una sílaba es cinco (por ejemplo, la primera sílaba de la palabra “transporte”), y

aún en esos casos lo más común es que la pronunciemos omitiendo alguno de dichos sonidos (por ejemplo, que digamos “trasporte”, sin pronunciar la “n”).

Las particularidades mencionadas en los dos párrafos anteriores (y unas cuantas más) hacen que, en general, cualquier “extranjero generalizado” considere que la pronunciación del inglés es más compleja que la del español, y eso puede objetivarse a través del recuento del número de vocales de ambos idiomas y de la caracterización de sus posibles estructuras silábicas. Cuanto más grande es el nivel en el cual uno quiere definir la complejidad de un idioma, sin embargo, más difícil se vuelve hacerlo. Esto ha llevado a que no sea habitual utilizar una medida única para la complejidad global de las distintas lenguas, y que en la mayoría de los casos se trabaje con medidas de complejidad local.

1.1.3. *Marcación, estructura jerárquica y eficiencia*

En su libro sobre complejidad gramatical, el lingüista estadounidense Peter Culicover (2013) considera que, a efectos de su uso en los análisis interlingüísticos (es decir, para la comparación entre varias lenguas) e intralingüísticos (es decir, para la comparación entre estructuras de una misma lengua) lo más útil es aplicar conceptos que se computen en términos relativamente locales pero que puedan definirse de manera general. Uno de dichos conceptos es el de “marcación” (*markedness*), que proviene de la lingüística teórica y cuyo uso se asocia con la obra de Noam Chomsky (1965). Según esta concepción, una forma es más compleja si está “marcada” y menos compleja si no lo está, y la propia lógica del idioma guía hacia el uso de la forma más simple (menos marcada). Por ejemplo, si partimos del sustantivo “cómputo”, vemos que el mismo está menos marcado en su forma singular que plural (“cómputos”), porque el plural aparece marcado con el sufijo “-s”. Cuando el idioma español deriva un verbo en base a dicho sustantivo, parte de la forma menos marcada (singular) y no de la más marcada (plural), y construye la palabra “computar” (en vez de “computosar”). Lo mismo hace al volver a formar un sustantivo (“computadora”), y también al crear un nuevo verbo (“computarizar”). Cada una de estas formas, sin embargo, está más marcada que la anterior y es, por lo tanto, más compleja.

Otro concepto mencionado por Culicover como una forma de medir la complejidad local de ciertos fenómenos lingüísticos tiene que ver con la denominada “estructura jerárquica” de los mismos. Dicho concepto se ejemplifica normalmente con la oposición entre oraciones simples y complejas, en la cual las primeras aparecen aisladas y las segundas aparecen integradas por varios enunciados jerárquicamente ordenados. Cuanto más niveles jerárquicos haya dentro de una oración, más compleja será la misma, como puede comprobarse en los siguientes ejemplos que van incluyendo cada vez más enunciados “incrustados” dentro de la oración principal:

- (4) *El perro come la comida.*
- (5) *El perro come la comida [que le dio el hombre].*
- (6) *El perro come la comida [que le dio el hombre [que tenía otro perro]].*
- (7) *El perro come la comida [que le dio el hombre [que tenía otro perro [que ladraba mucho]]].*

Nótese sin embargo que la mayor simplicidad de (4) respecto de las siguientes oraciones sucesivamente más complejas se obtiene a costa de un menor contenido informativo. Si, por ejemplo, quisiéramos decir lo mismo que expresa (7) utilizando exclusivamente oraciones simples, deberíamos escribir algo así:

- (8) *El perro come la comida. Un hombre se la dio. Ese hombre tenía otro perro. Ese otro perro ladraba mucho.*

Si ahora comparamos (7) con (8), vemos que, en cierto sentido, este último texto es más complejo que el primero, ya que consta de cuatro oraciones distintas (en vez de una sola “gran oración”), que en total tienen 20 palabras (en vez de 17), 38 sílabas (en vez de 31) y 75 sonidos (en vez de 60). Desde el punto de vista de la “eficiencia informativa” de ambos textos, por lo tanto, el primero parece ser mejor que el segundo.

Es precisamente dicha idea de eficiencia la que se encuentra implícita en otro concepto que sirve para evaluar la complejidad relativa de las estructuras lingüísticas, y que tiene que ver con el procesamiento de los componentes de tales estructuras. Dicho concepto, propuesto por el inglés

John Hawkins (2004), considera que una estructura es más compleja si, dado el contenido que quiere expresar, es menos eficiente. Dicha eficiencia, por su parte, tiene que ver con la cantidad de memoria que debe utilizarse para retener el contenido que se está procesando, la cual a su vez depende de la cantidad de componentes que se están utilizando y de las propiedades que se le asignan a dichos componentes. Comparemos, por ejemplo, estas dos oraciones:

- (9) *Juan rompió el televisor [que me pidió prestado].*
 (10) *Juan rompió el [que me pidió prestado] televisor.**

Tal como puede observarse, (9) es una oración que en castellano consideramos correcta, en tanto que (10) es una oración que consideramos incorrecta (y por ese hecho la estamos señalando con un asterisco). Sin embargo, tanto (9) como (10) están construidas usando exactamente las mismas palabras (es decir, los mismos componentes) y, en principio, buscan expresar el mismo significado. Pero mientras (9) es más “eficiente” para expresar dicho significado, (10) lo es menos, básicamente porque obliga a recordar por más tiempo que Juan rompió algo, antes de saber qué es exactamente lo que rompió. Es probablemente por eso que el idioma español nos fuerza a usar (9) en vez de (10), dado que esa última forma de expresar lo mismo es menos eficiente y, por ende, más compleja.

En el caso de la comparación de (7) con (8), en cambio, la mayor eficiencia (y, por ende, la menor complejidad de procesamiento) de la primera alternativa depende del hecho de que usa menos componentes que la segunda. Nótese además que, aunque (7) es una sola oración “gramaticalmente compleja”, respeta la lógica implícita en el razonamiento de Hawkins de ir procesando los componentes sin necesidad de retenerlos por demasiado tiempo en nuestra memoria. La misma oración sería mucho menos eficiente (y más compleja) si la escribiéramos del siguiente modo:

- (11) *El perro come la [que le dio el hombre [que tenía otro perro [que ladraba mucho]]] comida.**

por el mismo motivo por el que (10) es menos eficiente que (9).

1.2. Componentes del lenguaje

El lenguaje humano está compuesto esencialmente por sonidos, que se agrupan en palabras, que se agrupan a su vez en enunciados.¹ Cada uno de dichos elementos representa una unidad que se combina con otras unidades equivalentes, con el objeto de formar construcciones más complejas. Los sonidos, por ejemplo, son las unidades mínimas de pronunciación de un idioma, y se combinan entre sí de manera secuencial para formar palabras. Las palabras son las unidades mínimas con significado independiente, que también se combinan de manera secuencial para formar enunciados (que son, a su vez, las unidades mínimas de comunicación en las cuales se dividen los textos).

Entre los sonidos, las palabras y los enunciados existen ciertas unidades intermedias que tienen importancia para analizar distintos fenómenos lingüísticos. A mitad de camino entre el sonido y la palabra se encuentra la sílaba, que es un conjunto de sonidos que se pronuncian de manera relativamente simultánea y que están organizados en torno a un determinado núcleo (sonido básico), que es generalmente una vocal. También a mitad de camino entre el sonido y la palabra se encuentra el morfema, que es un conjunto de sonidos cuya unidad no tiene que ver con su pronunciación sino con su significado. La palabra “inevitable”, por ejemplo, está formada por diez sonidos (que en este caso corresponden cada uno a una letra), los

1 Esta caracterización, obviamente, se aplica a las lenguas habladas. En el caso de las lenguas de señas, no existe una unidad relacionada con el sonido, y la mayoría de los signos que se utilizan en ellas tienen un significado en sí mismo (y son, en ese sentido, más asimilables a palabras que a unidades de sonido). En las lenguas de señas, sin embargo, existen formas de descomponer a las señas en distintos elementos o “articuladores” (gestos manuales, posición de la mano, dirección del movimiento, etc.) que podrían cumplir las funciones que tienen los sonidos como elementos que forman las palabras en las lenguas habladas. Para un estudio más detallado de estos temas, puede consultarse el manual de lenguas de señas publicado en el año 2012 por la editorial De Gruyter, cuyos compiladores son Roland Pfau, Markus Steinbach y Bencie Woll.

cuales se agrupan de manera diferente en cinco sílabas y en tres morfemas. Mientras que la agrupación de los sonidos en sílabas nos indica el modo en el cual la palabra se pronuncia (i-ne-vi-ta-ble), la agrupación en morfemas nos indica cuáles son los componentes que hacen al significado final de dicha palabra (in-evit-able).

En cuanto a las unidades intermedias entre las palabras y los enunciados, vale la pena mencionar al sintagma. Este es un conjunto de palabras organizado jerárquicamente, que contiene una palabra que constituye su núcleo y, eventualmente, otras que operan como complementos de dicho núcleo o como especificadores del sintagma como un todo. Así como las sílabas y los morfemas pueden tener un solo sonido, y las palabras pueden tener una sola sílaba o un solo morfema, los sintagmas también pueden tener una sola palabra, y los enunciados tener un solo sintagma. Por ejemplo, el enunciado “Juan come” está formado por dos sintagmas: “Juan” (sintagma nominal) y “come” (sintagma verbal), y cada uno de ellos consta de una única palabra. En cambio, el enunciado “¡Qué mala suerte!” está formado por un único sintagma nominal, pero el mismo tiene tres palabras. Una de ellas es el núcleo (“suerte”); otra es un complemento de dicho núcleo (“mala”); y la tercera (“qué”) actúa como especificador de todo el sintagma, indicando que se trata de una exclamación.

1.2.1. Fonología y fonética

La descomposición del lenguaje humano en sonidos, palabras y enunciados (y su posterior agrupamiento y división en sílabas, morfemas y sintagmas) permite analizar a la lengua desde distintas perspectivas. Una de dichas perspectivas tiene que ver con la estructura del sistema de sonidos que se usa para construir las palabras, y recibe el nombre de fonología. Dicho campo de análisis estudia los sonidos como unidades que tienen valor distintivo para formar palabras, y los agrupa en “paquetes” más o menos similares que se pueden usar con dicha función. Cada uno de dichos paquetes recibe el nombre de “fonema”, y dentro de él puede haber un único sonido o varios sonidos más o menos parecidos cuya distinción no es relevante para el idioma que se está analizando.

Tomemos, por ejemplo, la palabra española “endulzado”. En dicha palabra, la letra “d” aparece dos veces, pero el sonido que se usa para pronunciarla es distinto en cada situación. En su primera aparición, la pronunciación estándar consiste en apoyar la lengua un poco por encima de los dientes y luego separarla, haciendo vibrar al mismo tiempo las cuerdas vocales. Dicho sonido se representa fonéticamente a través del símbolo [d]. Para la “d” que aparece por segunda vez (es decir, la que está entre medio de las vocales “a” y “o”), la pronunciación más común consiste en hacer vibrar las cuerdas vocales pero sin llegar a apoyar la lengua por encima de los dientes, dejando un pequeño espacio para que pase el aire. Ese segundo sonido se representa fonéticamente a través del símbolo [ð].

Si bien para los que hablamos español la diferencia entre los dos sonidos descritos en el párrafo anterior nos parece algo completamente esotérico, hay idiomas en los cuales sirve para distinguir entre palabras con significados diferentes. Cuando en inglés alguien pronuncia la palabra “*day*” (“día”), utiliza un sonido parecido al de la primera “d” de “endulzado”. En cambio, si la misma persona pronuncia la palabra “*they*” (“ellos”), usa un sonido bastante semejante al de la segunda “d” de “endulzado”.

La causa por la cual nosotros “no nos damos cuenta” de que pronunciamos distinto la letra “d” en diferentes posiciones es que para nosotros [d] y [ð] son dos “alófonos” del mismo fonema (es decir, dos sonidos que tenemos mentalmente agrupados dentro del mismo paquete). En el idioma inglés, en cambio, /d/ y /ð/ son dos fonemas distintos. El hecho de que puede haber sonidos distintos que están en un mismo idioma y sin embargo no tienen valor distintivo vuelve aconsejable utilizar una nomenclatura diferente para representar a los sonidos en sí (que usualmente se escriben entre corchetes) y para representar a los fonemas (que usualmente se escriben entre barras). El símbolo que se usa para representar a determinado fonema, sin embargo, se elige siempre entre los que están disponibles en el alfabeto fonético, y en general es el del sonido (alófono) más característico del fonema en cuestión.

La distinción entre fonema y sonido es en cierto modo la clave de la diferencia que existe entre la fonología y la fonética. Mientras la primera tiene que ver con la estructura de los sonidos de un idioma en función de formar palabras que luego tengan algún significado, la fonética es una

disciplina que estudia la pronunciación en sí, y el conjunto de sonidos y combinaciones de ellos que se usan en el lenguaje. La fonética es, por lo tanto, una de las herramientas que usa la fonología para definir el sistema de sonidos de una lengua. Desde una perspectiva inversa, la fonología es una especie de “fonética funcional”, que solo se preocupa por analizar la pronunciación en tanto cumpla un papel en el sistema conjunto de la lengua.

1.2.2. Morfología y sintaxis

Nótese que, en el lenguaje humano, ni los sonidos ni los fonemas tienen en sí ningún significado, más que el de servir para combinarse y formar palabras. El estudio de la palabra desde el punto de vista de sus constituyentes mínimos con significado separable es, en cambio, el campo de la morfología. Dicha manera de analizar el lenguaje se concentra en identificar los distintos morfemas que pueden constituir una palabra, así como las relaciones entre los mismos. También sirve para distinguir entre morfemas léxicos (es decir, los que tienen un significado conceptual) y morfemas gramaticales (que solo cumplen una función auxiliar).

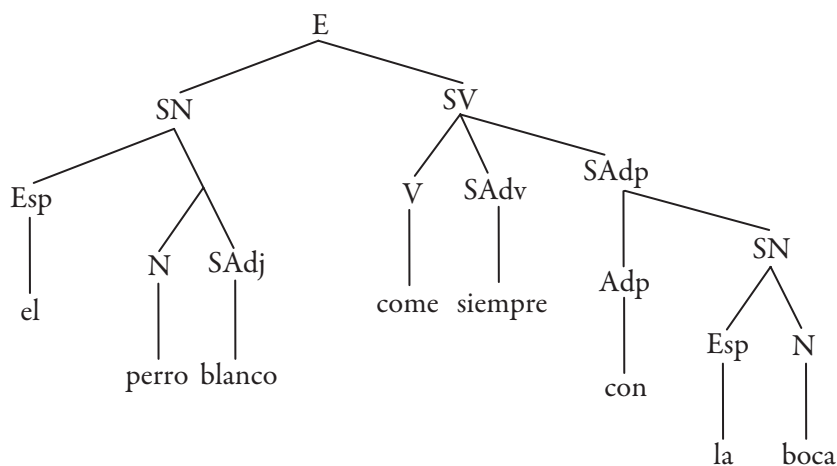
Muchas veces, los morfemas gramaticales adoptan formas que solo les permiten aparecer como componentes de una palabra junto a un morfema léxico. Dichos morfemas gramaticales se unen a los morfemas léxicos en un proceso que se denomina “afijación”, y que consiste en formar palabras con una raíz (morfema léxico) y con uno o varios afijos (morfemas gramaticales). En el caso de la palabra “in-evit-able”, por ejemplo, “-evit-” es la raíz que hace referencia al verbo “evitar”, en tanto que “in-” (negación) y “-able” (capacidad) son afijos que modifican el concepto principal denotado por la raíz.

La morfología estudia también casos en los cuales los morfemas se unen en una relación de composición, en la cual dos o más morfemas léxicos forman una nueva palabra (por ejemplo, “mal-trato”). Pueden darse también casos combinados, en los que hay al mismo tiempo composición y afijación (por ejemplo, “mal-trat-ador”).

Si ahora pasamos a un nivel de análisis más agregado, y nos concentramos en el modo en el cual las palabras se combinan para formar enunciados,

entramos en el campo de la sintaxis. Dicha rama del análisis lingüístico parte de la idea de que las palabras se combinan en sintagmas o “frases”, y estos a su vez forman enunciados más o menos complejos. Si bien los enunciados pueden estar formados por un único sintagma nominal (por ejemplo, “Buenos días”) o verbal (por ejemplo, “Llueve”), lo normal es que los mismos consistan en una combinación entre un sintagma nominal (SN) y un sintagma verbal (SV).² Dentro de dichos sintagmas, a su vez, pueden aparecer otros que forman parte de ellos, y que pueden ser sintagmas nominales, verbales, adjetivos, adverbiales y adposicionales. Todos ellos, sin embargo, se combinan dentro de una estructura que tiene una forma jerárquica, la cual se puede representar gráficamente a través de un “diagrama de árbol”.

Tomemos, por ejemplo, el enunciado “El perro blanco come siempre con la boca”, que puede representarse a través del siguiente diagrama:



Tal como puede observarse, este enunciado (E) se compone de un sintagma nominal (“el perro blanco”) que se combina con un sintagma verbal (“come siempre con la boca”). Dentro del primero de ellos, sin embargo, aparece un adjetivo (“blanco”) que forma un sintagma adjetivo (SAdj).

2 A este tipo de enunciados con sintagma nominal y verbal se los suele denominar “cláusula” u “oración”. Sobre esta terminología, véase Garrido (2009), capítulo 5.