

ISSN 2500-4840

ISSN EN LÍNEA 2745-1100

# PROLEGÓMENOS

VOL.

8



## Revista Prolegómenos

VOLUMEN 8 · SEM 2020

ISSN 2500-4840 · ISSN EN LÍNEA 2745-1100

UNIVERSIDAD NACIONAL DE COLOMBIA  
FACULTAD DE CIENCIAS HUMANAS  
DEPARTAMENTO DE LINGÜÍSTICA  
SEDE BOGOTÁ



### Rectora

Dolly Montoya Castaño

### Vicerrector

Jaime Franky Rodríguez

### Director Bienestar Sede Bogotá

Oscar Arturo Oliveros

### Jefe de División de Acompañamiento Integral

Zulma Edith Camargo Cantor

### Coordinador Programa Gestión de Proyectos

William Gutiérrez

### Decana Facultad de Ciencias Humanas

Luz Amparo Fajardo

### Directora Bienestar Facultad Ciencias Humanas

Esperanza Cifuentes Ardila

### Corrección de Estilo

Diana Luque Villegas (PGP)

### Diseño y Diagramación

Fernando Rodríguez (PGP)

proyectoug\_bog@unal.edu.co

/gestiondeproyectosUN

@PGPunal

issuu.com/gestiondeproyectos



### Dirección

Rafael Antonio Gutiérrez Martínez

Jennifer Paola Vega Barragán

Joan Sebastián Gutiérrez Rodríguez

### Comité Editorial

César Darío Martínez Cáceres

Katherine Valentina Méndez Barón

Sebastián Díaz Escobar

Marko Geuseppe Piñeros Crespo

Dayana Melisa García Beltrán

Erika Mesa Díaz

### Autores

Leidy Sophia Sandoval Camargo

Bryan Steven Loaiza-Camacho

David Felipe Guerrero Beltrán

Manuela Rondón Triana

Rafael Antonio Gutiérrez Martínez

**Prolegómenos** - Es una revista estudiantil de Lingüística de jóvenes pertenecientes al grupo de estudios -Ema. El objetivo principal de la revista es ser una herramienta de divulgación de trabajos de carácter empírico y teórico, enmarcados en el ámbito lingüístico disciplinar e interdisciplinar, los cuales hayan sido realizados por estudiantes y/o profesionales de Lingüística u otras áreas afines al estudio del lenguaje. Con esto, la revista pretende promover la escritura, el debate y la publicación de temáticas diversas enfocadas en el análisis del lenguaje.

### Contacto

✉ revprol\_fchbog@unal.edu.co

📌 /Resprolegomenos

🌐 resprolegomenos.blogspot.com.co

📍 @Resprolegomenos

### Universidad Nacional de Colombia

Sede Bogotá

Cra 45 No 26 - 85

Edificio Uriel Gutiérrez

www.unal.edu.co

Los textos presentados en la siguiente publicación expresan la opinión de sus respectivos autores y la Universidad Nacional de Colombia no se compromete directamente con la opinión que ellos puedan suscitar



# CONTENIDO



<b>EDITORIAL</b>	pág. 7
<b>MORFEMAS DE CLASIFICACIÓN EN KAMËNTSÁ: APROXIMACIÓN MORFOSINTÁCTICA Y SEMÁNTICA</b> Leidy Sophia Sandoval Camargo	pág. 9
<b>RETOS Y LOGROS EN LA CONSTRUCCIÓN DE CORPUS ELECTRÓNICOS: EL CASO DEL CREOLE DE SAN ANDRÉS</b> Bryan Steven Loaiza-Camacho	pág. 23
<b>LA GRAMÁTICA DE LA ENFERMEDAD: APROXIMACIÓN A LA EXPRESIÓN LINGÜÍSTICA DE LA ENFERMEDAD EN EL KARIJONA (CARIBE, REGIÓN AMAZÓNICA)</b> David Felipe Guerrero Beltrán	pág. 37
<b>REFLEXIONES EN TORNO A LA NOCIÓN DE CONOCIMIENTOS INNATOS DE LA GRAMÁTICA UNIVERSAL</b> Manuela Rondón Triana y Rafael Antonio Gutiérrez Martínez	pág. 51
<b>NORMAS DE PUBLICACIÓN</b>	pág. 61



## EDITORIAL

“Creo que el rol de los intelectuales no es seguir la corriente, sino perseguir la libertad, preguntarse por ella, y transmitir los resultados de su pesquisa. Y no tener miedo”.

*Tzvetan Todorov*

El viaje de *Prolegómenos* ha llegado a su octavo puerto y con él sus aprendizajes y las historias que han quedado grabadas en cada volumen. Tanto los autores como los editores, lectores o cualquier persona que haya contemplado los primeros pasos de esta revista han comprendido cómo la voluntad, la entrega y un amor descomunal al conocimiento se han convertido en los pilares de esta publicación. Dichos pilares nos han permitido, en tan poco tiempo, lograr el reconocimiento de una buena parte de la comunidad académica, así como un apoyo incondicional a las diferentes iniciativas que surgen desde nuestro nido.

Uno de los eventos más importantes en nuestra corta historia como grupo y revista estudiantil, se encuentra consignado en estas páginas en las que quedan por escrito las memorias del “Foro Ling20: ¿Cuál es el estado de la actual de la investigación lingüística en Colombia?”. Este evento, realizado con el fin de conmemorar los 20 años de la fundación de la carrera de Lingüística en la Universidad Nacional, tuvo como propósito invitar a la reflexión respecto del quehacer del lingüista en las múltiples esferas donde se desenvuelve en la actualidad, junto con su proyección y responsabilidad en el desarrollo científico y social del país. Para ello, fue grato contar con la participación de reconocidos investigadores y académicos como Jon Landaburu, Oliver Müller, Olga Ardila, María Emilia Montes, Silvia Baquero, Alfredo Rivera y Rubén Flórez, quienes aportaron una valiosa cuota de experiencia. No obstante, esta editorial quiere exaltar, especialmente, la participación de cada uno de los estudiantes de la carrera, miembros del comité editorial de *Prolegómenos*, expositores y asistentes al evento, ya que sin su intervención y apoyo este evento no hubiese sido posible.

De esta manera, el grupo estudiantil EMA, representado en su proyecto *Prolegómenos*, continúa extendiendo a toda la comunidad estudiantil su idea de trabajar colectivamente en la realización de actividades creadoras de puentes entre los diferentes actores presentes en la academia. A partir de la premisa de que nuestro presente nos exige caminar unidos, aún más cuando nos hemos convencido de la necesidad de ahogar el miedo y aventurarnos hacia lo desconocido. Asimismo, *Prolegómenos* se con-

vierte en un escenario en donde se busca estimular la curiosidad del investigador, valor indispensable en los sujetos entregados a la ciencia. Lo anterior a partir de los testimonios de quienes ya han emprendido el vuelo en la vastedad del universo lingüístico y gracias a las remembranzas sus observaciones y sus experiencias, que han sido asombrosos procesos de descubrimiento y autodescubrimiento.

Finalmente, queremos expresar nuestros más profundos agradecimientos a todas las personas que han rodeado a la revista con el cariño y la dedicación con los que se sostienen los sueños. *Prolegómenos*, hoy por hoy, es parte del patrimonio de los estudiantes; por ello, quienes hacemos parte de este proyecto y quienes en adelante comenzaremos a dirigir los nuevos rumbos de la revista, queremos decirles que siempre estaremos orgullosos de pertenecer a esta familia.

Comité editorial

## MORFEMAS DE CLASIFICACIÓN EN KAMËNTSÁ: APROXIMACIÓN MORFOSINTÁCTICA Y SEMÁNTICA<sup>1</sup>

Leidy Sophia Sandoval Camargo<sup>2</sup>

### Resumen

El presente artículo esboza las principales características morfosintácticas y semánticas de los morfemas de clasificación nominal en la lengua kamëntsá. El análisis propuesto se llevó a cabo a partir de los datos recogidos en campo a través de una herramienta de elicitación diseñada para este fin. Así, se encontraron nueve marcas de clasificación que tienen características morfosintácticas como la concordancia, la sufijación a distintas clases de palabras al interior de la cláusula y que tienen funciones tanto derivativas como gramaticales. Asimismo, estas marcas expresan parámetros semánticos de forma, consistencia, materia y configuración. Este sistema no se adapta completamente a la tipología tradicional, aunque presenta rasgos comunes con los sistemas de la Amazonía.

**Palabras clave:** clasificación nominal, morfosintaxis, semántica, kamëntsá.

<sup>2</sup> Lingüista y estudiante del pregrado en Antropología de la Universidad Nacional de Colombia. Correo: lssandovalc@unal.edu.co

## Abstract

This article outlines the main morphosyntactic and semantic characteristics of nominal classification morphemes in the kaměntsá language. The proposed analysis was carried out from the data collected in the work field through an elicitation tool designed for this purpose. As a result, nine classification morphemes were found that have morphosyntactic characteristics such as concordance, suffixing to different classes of words within the clause and they have both derivative and grammatical functions; likewise, these marks express semantic parameters of form, consistency, material and configuration. This system is not completely adapted to the traditional typology, although it has common features with the systems of the Amazon.

**Key words:** nominal classification, morphosyntax, semantics, kaměntsá language.

## 1. Introducción

Este artículo tiene como principal objetivo caracterizar los morfemas de clasificación nominal del kaměntsá (a.k.a. camsá, kamsá) a través de la descripción de su comportamiento morfosintáctico y semántico. Esta es una lengua aislada hablada en el Valle de Sibundoy, el cual se encuentra ubicado en el Alto Putumayo, al suroccidente de Colombia. De acuerdo con las cifras oficiales, la lengua cuenta con alrededor de 4.000 hablantes (Juajibioy Chindoy, 2008, p. 22), aunque, de acuerdo con los diagnósticos sociolingüísticos realizados, tan solo un 10 % de la población, aproximadamente, tiene conocimiento pleno de la lengua (Mora, 2004).

A partir de los trabajos sobre la gramática de esta lengua, se conoce que esta cuenta con una morfología compleja en la cual las raíces nominales y verbales pueden afijar un número amplio de morfemas. Igualmente, tiene límites morféminos bastante definidos y permite incorporar raíces nominales a los verbos. En cuanto al orden de constituyentes, el orden dominante es SV para intransitivas y VO para transitivas. No se establece un orden dominante para S, V y O, teniendo en cuenta que es una lengua Pro-drop, por lo que presenta concordancia anafórica a través de índices de persona en el verbo y casi nunca se encuentran estos

tres elementos en constituyentes separados. El sistema de clasificación nominal de la lengua kaměntsá ya había sido mencionado por algunos autores, aunque someramente, a través de la presentación de propuestas del inventario de morfemas.

## 2. Metodología

La recolección de datos se hizo a partir del trabajo de campo realizado en los municipios de Sibundoy y San Francisco (Alto Putumayo) entre junio y julio del 2017. Durante la preparación de este trabajo se diseñó una herramienta de elicitación que propiciara que los hablantes produjeran oraciones o frases en las que los morfemas de clasificación estuvieran presentes, sin que se tuviera que hacer una traducción desde el español. Para ello, fueron de gran utilidad los aportes metodológicos de la investigación de Seifart (2009) sobre la clasificación en miraña y las estrategias metodológicas presentadas por Ospina Bozzi y Orjuela (2019).

Esta herramienta de elicitación corrobora algunos de los postulados hechos por autores que han investigado la lengua kaměntsá previamente, al igual que indaga sobre aspectos que podrían tener relevancia para el sistema, pero de los cuales no se había hablado. La primera parte de la herramienta busca obtener oraciones simples con frases nominales de determinación, en donde se diga el nombre de un objeto y se ofrezcan caracterizaciones sencillas de este, con el fin de saber cuál es la estructura básica de la frase nominal. En esta parte también se contempla obtener oraciones de cuantificación simples presentando para su elicitación varios objetos del mismo tipo.

La segunda parte busca obtener información sobre los criterios semánticos que son relevantes en el sistema de clasificación. En esta parte también se espera obtener cláusulas determinativas, posesivas y cuantificadoras para indagar sobre el comportamiento de las marcas de clasificación en este tipo de construcciones. Para ello, esta sección se estructura a partir de los parámetros de clasificación, de acuerdo con los análisis previos en la literatura existente; sin embargo, también se tienen en cuenta otros parámetros con el fin de corroborar su relevancia para el sistema. En consecuencia, se ofrecen situaciones con variables fijas (controladas) y una cambiante, lo cual permite indagar sobre los parámetros de clasificación del sistema y las características morfosintácticas de este. Algunos de las variables tenidas en cuenta son el tamaño, la consistencia, la forma, el material, la textura, entre otros.

Una vez en campo, se contó con la participación de 9 colaboradores, 6 mujeres y 3 hombres, entre 33 y 80 años, quienes fueron contactados a partir de la técnica de muestreo de bola de nieve. Con cada uno de los colaboradores se realizó una sesión de grabación de alrededor de media hora y otra sesión dedicada a traducir los datos recogidos. En total, se obtuvieron 4 horas y 29 minutos de grabación para el análisis.

La transcripción de las grabaciones se hizo a partir de la fonología propuesta por Juajibioy Chindoy (2008). En el proceso de sistematización se optó por separar por muestras de acuerdo con cada hablante. Este proceso arrojó un total de 192 muestras que después se dividieron de acuerdo con los tipos de oraciones encontradas, a saber: determinativas, comparativas, cuantificadoras o posesivas. Esto permitió tener criterios de comparación más efectivos para encontrar regularidades en las muestras obtenidas. No obstante, para hacer generalizaciones sobre la morfología de la lengua fue necesario estudiar las estructuras al interior de la frase nominal (FN).

De este modo, se realizó un análisis de dos tipos: uno morfológico, en el que se analizaron los paradigmas del nombre y las variaciones de alomorfos, y otro sintáctico, donde se tuvieron en cuenta las estructuras sintácticas al interior de la FN. Para el análisis morfológico, en primer lugar, se separaron las cláusulas que tenían marcas de clasificación frente a las que no, con el fin de estudiar los factores que podrían incidir en su uso. Posteriormente, se categorizaron las cláusulas de acuerdo con cada morfema de clasificación y con el tipo de palabra en donde se encontraban sufixados (adjetivos, nombres, verbos, etc.), para corroborar que cada morfema propuesto como parte del sistema de clasificación tuviera las características generales del sistema. En cuanto al análisis sintáctico, una vez categorizadas las oraciones por el tipo de frase (posesiva, cuantificadora, etc.), se establecieron regularidades de acuerdo con el orden entre núcleo y dependientes, con si había concordancia o no y con la obligatoriedad de las marcas de clasificación.

### 3. Los sistemas de clasificación nominal

Los sistemas de clasificación nominal son el mecanismo por medio del cual se categorizan los conceptos que representan las entidades con las que interactuamos, como pueden ser los objetos, las personas y otros seres vivos, de acuerdo con la manera como las percibimos. Por esta razón, los parámetros de clasificación utilizados en las lenguas para categorizar los nombres varían en función del ambiente, la tecnología utilizada, los

alimentos que consumen, y demás disposiciones culturales y sociales. A su vez, puede haber subcategorizaciones que respondan a parámetros de forma, consistencia, configuración, función, tipo de posesión, entre otros.

Además, Allan (1977) realizó un estudio comparativo entre aproximadamente 50 lenguas y llegó a la conclusión de que los morfemas clasificadores se agrupan a partir de los rasgos semánticos de materia (animacidad, objetos abstractos o materia), forma (una sola dimensión saliente, bidimensional o tridimensional), consistencia (flexible, duro o rígido, no-discreto), talla, locación (locación inherente, locación contingente), disposición (objetos en específico sin configuración inherente, posición u objetos sin distribución inherente) o cantidad.

En cuanto a la expresión lingüística, existen distintas formas en las que se expresan los sistemas de clasificación nominal: estas van desde lo más léxico hasta lo más gramatical (Grinevald, 2000), es decir, sistemas que se expresan por medio de palabras como los términos de clase o de medida, o sistemas que se expresan a partir de la morfología de la lengua. Los sistemas que utilizan morfemas para expresar tales categorizaciones, tradicionalmente, se han dividido en dos tipos: clases nominales y clasificadores nominales. Respecto a los primeros, usualmente, se caracterizan por tener un inventario pequeño de morfemas, ser obligatorios, tener morfemas ligados, presentar concordancia en las distintas palabras de la cláusula y cada nombre siempre es categorizado por la misma marca; por el contrario, los clasificadores suelen tener un número elevado de morfemas, no ser obligatorios, aparecer una única vez en el nombre, ser morfemas libres y un nombre puede ser categorizado por distintas marcas (en función de la voluntad del hablante) (Dixon, 1986).

### 4. Presentación general del sistema

El sistema de clasificación nominal del kaměntsá divide los nombres en aquellos que comúnmente afijan morfemas de clasificación y aquellos que no. Los nombres que afijan morfemas de clasificación suelen ser inanimados, aunque también se encuentran nombres de algunas plantas y animales, pero el significado prototípico de los morfemas cambia. Algunos referentes son caracterizados a través de distintos parámetros semánticos, por lo que es común que se presente variación en el uso de los morfemas de clasificación de acuerdo con el registro, el propósito o la edad de los hablantes. Un panorama tentativo de los morfemas que conforman el sistema de clasificación y su semántica se observa en la Tabla 1.

Tabla 1. Morfemas de clasificación del kamĕnsá.

Número	Marca de clasificación	Alomorfos	Semántica
1	-bé		Forma redondeada
2	-x	-xe, -xa, -xua	Forma alargada y flexible
3	-fx	-fxe, -fxa, -fxua	Forma alargada y rígida
4	-tʃ		Forma plana
5	-tʂ		Materia carnuda
6	-ʃ	-ʃa, -ʃe	Materia seca
7	-ʂ		Materia desmoronable
8	-i		Materia líquida
9	-ǰá		Parte de otro(s) elemento(s)

Fuente: Elaboración propia

Este inventario de morfemas se obtuvo de acuerdo con el análisis realizado, aunque difiere de los inventarios propuestos previamente por Juajibioy Chindoy (1973), Jamioy Muchavisoy (1989), Juajibioy Mutumbajoy (1995) y McDowell (1994), tanto en los morfemas pertenecientes al sistema como en la semántica de algunos de estos. Para proponer este inventario se corroboró que todos los morfemas cumplieran con los criterios tanto morfosintácticos como semánticos que caracterizan al sistema, los cuales se exponen a continuación.

### 4.1 Comportamiento morfosintáctico

En esta lengua, los morfemas de clasificación se sufijan a las palabras, es decir, se trata de morfemas ligados y están altamente gramaticalizados en la lengua. Pese a esto, la presencia de las marcas de clasificación depende de la decisión del hablante, por lo que la presencia de la marca de clasificación no es obligatoria como se evidencia en (1). En 1.a. el nombre *ndətʂ* “piedra” está sufijado solo por el morfema de diminutivo, aunque en 1.b. este mismo nombre sí presenta la marca de clasificación.

(1)	a.	Uta	<b>ndətʂ-tem</b>	fʃfendə-tema-t
		Dos	<b>pedra-dim</b>	azul-dim-dl

“dos piedritas azules”

b.	Ak	Kontsebomna	Uta	<b>ndətʂ-bé,</b>
	2.sg	Tener	dos	<b>pedra-clI</b>

“Tú tienes dos piedras”

Además, los nombres pueden sufijar distintos morfemas de clasificación (pertenecer a más de una categoría) para expresar otras de sus características o nombrar otros referentes. Esto se evidencia en (2) donde en 2.a. el mollejón (canica) está caracterizado como un objeto redondeado, por lo que sufija la marca *-bé*, la cual es su marca de clasificación prototípica. Sin embargo, en 2.b. al estar en un contexto donde lo importante no es la redondez del objeto, sino que este es brillante, el mismo nombre es sufijado con la marca *-ʃ* “materia seca”.

(2)	a.	[Pos]	[Adj]	[[N-CL]	Adj-CL]fɪ1] fɪ2] fɪ3]
		atʂ-be	Base	mojexon-bé	buangan-bé
		1.sg-pos	Pequeño	mollejón-clI	rojo-clI
		“mi pequeño mollejón rojo”			
	b.	Num	N-CL	Adj-CL	
		Uta	mojexon-o-ʃ	wafnindʒan-o-ʃ	
		dos	mollejón-v.ep-cl6	brillantes-v.ep-cl6	
		“dos mollejones brillantes”			

Los morfemas de clasificación pueden ser sufijados a nombres, adjetivos, numerales, demostrativos, algunos determinantes y verbos. Esto puede evidenciarse en 3.a. donde el morfema de clasificación se sufija al numeral y al nombre; también está presente en 3.b. en el demostrativo y en 3.c. donde está sufijada al verbo. Cabe señalar el hecho de que estos morfemas tengan también una función anafórica, pues se hace referencia a la entidad sin necesidad de que aparezca el nombre en la cláusula (3.b. y 3.c.)

(3)	a.	Num-CL	Adv	N-CL
		uta-be	casi	bol-bé,
		dos-CL1	Casi	bola-CL1
		“dos casi bolas” (186)		

- b. Dem-CL  
mo-fx  
DEM-CL3  
“este (el lápiz)” (96)
- c. Adv. V-CL  
mor t-o-n-x-o-ta-bé.  
Después PERF-3.SG.AGEN-ACER-DIST-VAL-sentarse-CL1  
“Después se sentó [en la pelota]” (26.3)

La concordancia de **marca de clasificación** solo se presenta al interior de la frase nominal entre el adjetivo y el núcleo que, comúnmente, es un nombre, pero también puede ser un posesivo o un numeral. La concordancia es obligatoria solo cuando el adjetivo sucede al elemento que determina (4.b), pero no cuando lo precede (4.a).

- (4) a. Adj N-CL  
Base nijə-fxa-tema,  
pequeño madera-CL3-DIM  
“pequeña maderita (alargada)” (148).
- b. N-CL Adj-CL  
mojexon-be-ka buangan-bé,  
mollejón-CL1-EQUIP rojo-CL1  
“como un mollejón rojo” (103.1)

Ambos casos pueden estar presentes en una FN compleja y se continúa manteniendo la estructura básica, como se muestra en el ejemplo 5.

- (5) [Dem [Adj [[N-CL] Adj-CL]]]  
muenj base pelota-bé şaxans-bé,  
“aquella pequeña pelota blanca”

## 4.2 Semántica de las marcas de clasificación

El inventario de morfemas presentado como parte del sistema de clasificación nominal puede caracterizarse en términos de cuatro parámetros

semánticos, a saber, dimensión, consistencia, materia y configuración, tal como se muestra a continuación.

### 4.2.1 Dimensión

**a. Sin dimensión saliente: forma redondeada: -bé CL1.** Este morfema se usa para clasificar objetos de forma redondeada en los que no sobresale ninguna dimensión; es decir, que no se pueden catalogar como largos o planos, como se aprecia en 6. Además de esto, también puede ser utilizado para referir a entidades que no necesariamente son redondas, pero en las que se quiere resaltar su carácter de redondez, como cuando se quiere expresar que algo/alguien es grueso o gordo.

- (6) ʃəmn-bé wabwan-bé,  
huevo-CL1 cocido-CL1  
“Huevo cocido”

### b. Una dimensión saliente y consistencia

**i. Forma alargada y flexible: -x CL2.** El morfema -x se usa para clasificar objetos de forma alargada y flexible, como se observa en (7), donde este morfema se sufixa al nombre “cadena” (7.a.) y también al de “planta de fríjol” (7.b.)

- (7) a. caden-ə-xua  
cadena-V.EP-CL2.  
“cadena”
- b. tsəm-bia-xua  
fríjol-det-cl2  
“fríjol [planta]”

**ii. Forma alargada y rígida: -fx CL3.** En el ejemplo (8) se presenta el morfema -fx que tiene una semántica que también da cuenta de formas alargadas (con saliencia de una sola dimensión), pero de consistencia rígida.

- (8) ntɕaba-fx-nga      mas      bana-fxa  
dedos-CL3-PL      Más      largo-CL3  
“dedos más largos” (56)

Este mismo morfema sufijado a nombres animados aporta la noción de delgadez o flacura, como se muestra en (9).

- (9) bojeto-fx-a-nga  
pollo-CL3-V.EP-PL  
“Los polluelos (flacos)” (Juajibioy Chindoy, p. 75)

**c. Dos dimensiones salientes: forma plana: -tʃ CL4.** El morfema -tʃ se sufija a nombres cuyo referente es “plano” o tiene prominencia en dos dimensiones, como lo muestra el ejemplo (10).

- (10) tʃombia-tʃ  
chumbe-cl4  
“chumbe”

#### 4.2.2 Materia

**a. Carnuda: -tɕ CL5.** Este morfema se sufija a nombres de entidades compuestas de carne, tal como se observa en (11). Es importante señalar que el principal criterio es que se trate de elementos que se conforman de carne y no es aplicable a todas las partes del cuerpo, como *tamo-ɕa-xa* “cuello” o *xuaɕa-bé* “nariz” que no son carnudas.

- (11) a. mn-tɕ-en  
carne-CL5-en  
“carne”  
b. buakua-tɕ  
brazo-CL5  
“brazo”

**b. Seca: -ʃ CL6.** La marca -ʃ clasifica elementos de materia seca. Esto se evidencia en el ejemplo (12), en donde esta marca se sufija a la raíz *niŋ-* “madera”, y el significado de la palabra es madera seca, es decir, “leña”.

- (12) Kem      uta-soi      niŋ-a-ʃ  
DEM      dos-cosa      madera-V.EP-CL6  
“Estas dos cosas [son] leña”

Este morfema también se asigna a nombres que hacen referencia a nuevos objetos, los cuales generalmente son préstamos del español, como en *televisor-fa* “televisor” y *abión-fa* “avión”.

**c. Desmoronable: -ɕ CL7.** El morfema -ɕ expresa la noción de que el objeto al que hace referencia el nombre está compuesto de una materia que puede desmoronarse o desintegrarse, independientemente de la consistencia de la entidad. En (13) se presentan ejemplos de ello.

- (13) a. naxoma-ɕ  
carbón-CL7  
“carbón”  
b. bidra-ɕ  
vidrio-CL7  
“vidrio”

**d. Líquida: -i CL8.** Este morfema se utiliza para designar elementos de materia líquida o algunos fluidos (como el viento). Ejemplos de nombres a los que se sufija este morfema se presentan en (14).

- (14) a. bexa-i  
agua-CL8  
“agua”  
b. letʃe-i  
leche-CL8  
“leche”

### 4.2.3 Configuración: parte de un todo -fá CL9

Con el morfema -fá sufijado a un nombre se hace referencia a elementos que se configuran como parte de un elemento mayor, como en el caso de (14.a.), donde el “arete” se concibe como parte de la oreja, o bien, a objetos que están compuestos de múltiples elementos, como en (14.b.): una corona tradicional kaměntsá está compuesta por plumas o tejidos.

- |       |    |  |
|-------|----|--|
| (14)  | a. | wa-mats-axonjan-fá<br>cosa-oreja-colgar-CL9<br>“arete” |
| <hr/> |    |  |
|       | b. | b. korona-fá<br>corona-CL9<br>“corona”                 |

A modo de síntesis, en el sistema de clasificación del kaměntsá se presentan cuatro parámetros: dimensión (ninguna dimensión saliente, una dimensión saliente, dos dimensiones salientes), consistencia (flexible o rígido), materia (carnuda, líquida, seca, desmoronable) y configuración (parte de un todo).

## 5. Consideraciones tipológicas

Los sistemas que utilizan morfemas para expresar categorizaciones, tradicionalmente se han dividido en dos tipos: clases nominales y clasificadores nominales. El kaměntsá, por su parte, no puede catalogarse como uno de estos dos tipos de sistemas, puesto que posee características de ambos. Por un lado, se asemeja a los sistemas de clases porque presenta un pequeño número de morfemas ligados, tiene concordancia al interior de la FN, las marcas se sufijan a palabras distintas al nombre y es relevante el criterio de animacidad; por otro lado, la asignación de morfemas se da de acuerdo con la voluntad del hablante y con factores externos asociados, por lo que existe variación en el uso de las marcas, no es obligatoria su presencia y son relevantes los parámetros de forma, consistencia y materia.

De acuerdo con lo anterior, la lengua kaměntsá no se adecua completamente a la tipología propuesta. Como respuesta a los nuevos datos

provenientes de lenguas que no pueden encasillarse plenamente en la tipología tradicional, se ha caracterizado este tipo de clasificación nominal como sistemas mixtos: sistemas que integran varios subsistemas o sistemas de clase que se encuentran en una etapa temprana de gramaticalización. No obstante, para el caso de las lenguas amazónicas, algunos teóricos como Seifart y Payne (2007) argumentan que este tipo de clasificación es generalizada en la zona y presenta completa coherencia con el resto del funcionamiento de la lengua, por lo que no debería describirse en los términos de otros modos de clasificación, sino como sistemas prototípicos de la región amazónica.

Si bien la lengua kaměntsá no es plenamente amazónica, sí comparte varias características tipológicas de esta región, lo que se aprecia en su sistema de clasificación. No obstante, algunas de sus particularidades podrían haber estado influenciadas por el área andina, con la cual hay un contacto importante. De forma similar, este no parece ser un sistema mixto o uno por el cual se integre a varios sistemas, ya que las mismas marcas se utilizan en todos los tipos de construcciones de oración y tienen las mismas funciones, tanto gramaticales como derivativas.

## 6. Conclusiones

La lengua kaměntsá posee una riqueza lingüística compleja que no ha sido estudiada profundamente en todos sus niveles. Como se mostró a lo largo de este artículo, el sistema de clasificación nominal de esta lengua tiene características que contribuirían a ampliar la tipología y a dar claves importantes sobre la difusión areal de ciertos rasgos en esta región. De igual forma, divulgar este conocimiento contribuye a vislumbrar las múltiples posibilidades que tienen las lenguas para expresar la percepción del mundo de sus hablantes y, de esta manera, valorizar y apreciar la diversidad. Tales hallazgos permiten enriquecer la teoría lingüística, al igual que los estudios etnolingüísticos y de tipo interdisciplinar que indaguen sobre las estrechas relaciones entre la lengua, el territorio, la cultura y la forma en la que pensamos y categorizamos las entidades del mundo.

En el caso del kaměntsá, es necesario seguir profundizando en el tema de la categorización nominal, como es el caso de los contextos de uso de los morfemas, tanto a nivel sintáctico como pragmático, su relación con los tipos de oración, la afijación de morfemas de clasificación a verbos, la influencia areal y la actual dinámica del sistema.

## Referencias

- Allan, K. (1977). Classifiers. *Language*, 53(2), 285-311.
- Dixon, R. M. (1986). Noun classes and noun classification in typological perspective. En C. G. Craig (ed.), *Noun classes and categorization*, (pp. 105-112). Amsterdam: John Benjamins.
- Grinevald, C. (2000). A morphosyntactic typology of classifiers. En: G. Senft (Ed.), *Systems of nominal classification 4* (pp. 50-92). Cambridge: Cambridge University Press.
- Jamioy Muchavisoy, N. (1989). *Estructura verbal Kamëntsá*. Tesis de maestría. Bogotá: CCELA - Universidad de los Andes.
- Juajibioy Chindoy, A. (2008). *Lenguaje ceremonial y narraciones tradicionales de la cultura kamëntsá*. Bogotá: Fondo de Cultura Económica.
- Juajibioy Chindoy, A., y Wheeler, A. (1973). Bosquejo etnolingüístico del grupo Kamsá de Sibundoy, Putumayo, Colombia. Bogotá: Instituto Lingüístico de Verano.
- Juajibioy Mutumbajoy, M. C. (1995). *Actancia verbal en la lengua Kamentsa*. Tesis de maestría. Bogotá, Colombia: CCELA, UNIANDES.
- McDowell, J. H. (1994). *“So Wise Were Our Elders”: Mythic Narratives of the Kamsá*. Kentucky: University Press of Kentucky.
- Mora, M. A. (2004). Una mirada a la vitalidad de la lengua indígena Kamëntsá a través de la descripción sociolingüística. *Forma y Función*, (17), 34-56.
- Ospina-Bozzi, A. M. y Orjuela, L. (2019). Herramienta para la recolección de información sobre clases léxicas. *Forma y Función*, 32(2), 223-254.
- Seifart, F. (2009). Towards a typology of unitization: Miraña noun classes compared to numeral classifiers and singulatives. *Max Planck Institute for Evolutionary Anthropology*, Leipzig.
- Seifart, F. y Payne, D. L. (2007). Nominal classification in the North West Amazon: Issues in areal diffusion and typological characterization. *International Journal of American Linguistics*, 73(4), 381-387.

## RETOS Y LOGROS EN LA CONSTRUCCIÓN DE CORPUS ELECTRÓNICOS: EL CASO DEL CREOLE DE SAN ANDRÉS<sup>1</sup>

Bryan Steven Loaiza-Camacho<sup>2</sup>

### Resumen

Este artículo presenta una experiencia piloto realizada durante el semestre 2018-I para la recopilación de un corpus pequeño de la lengua creole de San Andrés, la cual estuvo enmarcada en los campos de la lingüística de corpus, la documentación lingüística y la criollística. Expone los principales retos y logros del piloto. Pese a que no se cumplió con todas las metas planteadas, el corpus resultante permitió examinar ciertas cuestiones lingüísticas de interés en la comunidad de habla raizal. Propone ser un insumo para el avance de la lingüística de corpus en Colombia, especialmente con respecto a las lenguas indígenas y criollas.

**Palabras clave:** creole de San Andrés, lingüística de corpus, documentación lingüística.

<sup>1</sup> El presente artículo está basado en el trabajo de grado *Experiencia piloto en el levantamiento de un corpus especial de lengua oral para la descripción lingüística del creole de San Andrés*, presentado por el autor para optar al título de lingüista y dirigido por la profesora Raquel Sanmiguel Ardila, a quien se extiende el más sentido agradecimiento.

<sup>2</sup> Lingüista de la Universidad Nacional de Colombia. Correo: bsloaizac@unal.edu.co

## Abstract

This paper presents a pilot experience that took place during the 2018-I semester to build up a small corpus for Islander Creole English, framed within the fields of corpus linguistics, language documentation, and creolistics. We present the main challenges and successes within the pilot. Even though not all goals were met, the resulting corpus allowed certain elements of interest, present in the Raizal speech community, to be examined. We hope this experience can serve as input for the progress of corpus linguistics in Colombia, especially in regards to indigenous and creole languages.

**Keywords:** Islander Creole English, corpus linguistics, language documentation.

## 1. Introducción

En el presente artículo, se exponen los resultados de una experiencia piloto llevada a cabo en la Sede Caribe de la Universidad Nacional de Colombia durante el semestre 2018-I. Este piloto tenía como finalidad la formación de un grupo local de colaboradores para la recolección de un corpus de la lengua creole de San Andrés, el posterior análisis de los datos recolectados y la evaluación de la viabilidad de un proyecto más amplio en esta sede.

El interés por este proyecto surgió de la línea Educación, Lengua y Cultura del grupo Estado y Sociedad, el cual ha venido formulando acciones investigativas que contribuyan al entendimiento de la conformación social, cultural y lingüística del archipiélago de San Andrés, Providencia y Santa Catalina en el contexto del Gran Caribe. Este grupo identificó que, en cuanto a los estudios acerca de la lengua creole de San Andrés, estos no han sido ni sistemáticos ni constantes; los últimos estudios de enfoque descriptivo datan de hace más de quince años; no se menciona de manera explícita la forma de recolectar los datos en varios de ellos y los datos sobre los que se basaron estos estudios son escasos, de difícil acceso o están desfasados y requieren actualización (Sanmiguel, Schoch y Pelufo, 2014).

Para este piloto, se escogió la metodología de la lingüística de corpus por la capacidad de procesamiento, sistematización y análisis que ofrece.

Ello condujo a una reflexión acerca del estado de la lingüística de corpus en el país. En Colombia, este es aún un campo incipiente, especialmente con respecto a las lenguas indígenas y criollas. Los corpus electrónicos existentes suelen concentrarse en el español y específicamente en determinados dialectos o registros; e.g., el Corpus Sociolingüístico de Medellín (Grupo de Estudios Sociolingüísticos, 2008) y el proyecto de Corpus Lingüísticos del Instituto Caro y Cuervo (2018). Cuando se llevó a cabo el piloto, no se tenía conocimiento de otros proyectos similares en el país y se afirmó en su momento, que no existía ningún corpus electrónico de una lengua nativa de Colombia (J. Bernal Chávez, comunicación personal, 11 de mayo de 2018). Posteriormente, se conoció el trabajo sobre el nasa yuwe y el nam trik adelantado por medio de herramientas de corpus electrónico, expuesto en Rojas (2017). Por tanto, ya hay avances en Colombia desde la lingüística de corpus para los cuales la experiencia piloto expuesta aquí puede configurar un aporte.

En este sentido, el presente artículo tiene el objetivo de brindar una reflexión acerca de los principales retos y logros en la construcción de los corpus electrónicos por medio de un caso particular.

## 2. Marco conceptual

El enfoque de la experiencia piloto estuvo enmarcado tanto en la descripción, como en la documentación de una lengua, por lo cual pertenece conceptualmente a dos campos principales: la documentación lingüística y la criollística. La lingüística de corpus, la cual determinó la metodología del piloto, guarda amplia relación con la documentación lingüística y, en varios aspectos, estas dos comparten aspectos fundamentales. Por otro lado, la criollística aparece como un campo que presenta diferentes retos para ambas (Loaiza-Camacho, 2018).

### 2.1 Lingüística de corpus

Es de notar que la lingüística de corpus no se puede considerar en sí una teoría ni una rama de la lingüística, sino una metodología y un enfoque que informa otros campos y cuya aplicación puede darse desde diferentes teorías del lenguaje (Gries, 2009). Sin embargo, se requiere de una conceptualización acerca de lo que constituye un corpus que vaya más allá de estas teorías.

El término corpus denota diferentes conceptos. La definición cotidiana reza que es un “conjunto lo más extenso y ordenado posible de datos o textos científicos, literarios, etc., que pueden servir de base a una investigación” (RAE, 2014), definición común entre las ciencias humanas y naturales, y de la cual surgen otros términos como *corpus de conocimiento* para referirse al compendio conceptual de una disciplina (Gimeno, 1998, citado en Álvarez-Álvarez, 2015, p. 174).

El sentido de este término en la lingüística —el cual se adopta aquí— se refiere específicamente a los corpus lingüísticos, los cuales son “una colección de piezas de una lengua que se seleccionan y se ordenan de acuerdo con criterios lingüísticos explícitos para ser utilizadas como una muestra de dicha lengua” (Sinclair, 1996, p.4, traducción propia). Estos, a su vez, se pueden clasificar como corpus electrónicos al tener un soporte informático. En lo subsiguiente, esta será la clase de corpus a la cual se hará referencia en este artículo. Este tipo de corpus se define de la siguiente manera:

[Es] un conjunto de textos de lenguaje natural e irrestricto, almacenados en un formato electrónico homogéneo, y seleccionados y ordenados de acuerdo con criterios explícitos para ser utilizadas como modelo de un estado o nivel de lengua determinado, en estudios o aplicaciones relacionados en mayor o menor medida con el análisis lingüístico (Santulla, 2005, citado en Villayande Llamazares, 2010, p. 325).

Los corpus se clasifican de diversas maneras. Aquí, se mencionan tres: (a) por la modalidad o el medio de producción de los textos, (b) por el tamaño y la especificidad de los textos y (c) por el número de lenguas. La clasificación de un corpus depende en gran medida de la metodología empleada, por lo cual esta cuestión se tratará en la sección 3.

## 2.2 Documentación lingüística

En la medida en que los corpus se ocupan de datos lingüísticos, ellos también hacen parte de la labor de documentación lingüística. El objetivo del piloto era, además de recopilar y analizar un corpus pequeño, sentar las bases para una documentación duradera y versátil (Himmelman, 2006) de la lengua creole.

La documentación lingüística se basa en un enfoque empírico que se propone conocer el uso de la lengua por medio de “grandes cantidades

de texto” (Boas, 1917, citado en Woodbury, 2011, p. 163, traducción propia). Para esto, se interesa especialmente por los datos primarios, i.e., el comportamiento lingüístico observable de los hablantes (Himmelman, 2006) y, para la sistematización de estos datos, considera indispensable el registro de metadatos, los cuales han de constituir una documentación en sí acerca de los datos primarios.

Lo idóneo en la documentación lingüística es recolectar la mayor cantidad posible de datos lingüísticos; sin embargo, esta meta es frecuentemente inalcanzable, por razones tanto materiales (de tiempo y recursos) como éticas. Por ello, la recolección de datos lingüísticos debe ser sistemática y el corpus documental de una lengua (i.e. todo el material existente sobre una lengua, incluyendo diccionarios y gramáticas) debe basarse en un conjunto de ideas conocido como la teorización del corpus (Woodbury, 2011). Esta teorización depende en gran medida del marco metodológico de un proyecto, por lo cual la teorización del corpus en la experiencia piloto se definirá con mayor precisión en la sección 3

## 2.3 Criollística

Las lenguas criollas presentan una serie de retos frente a la documentación lingüística y la lingüística de corpus.

En primer lugar, las lenguas criollas suelen presentar una gran variabilidad. Estas lenguas son el resultado de una serie de reestructuraciones de una lengua de superestrato con una lengua (o lenguas) de substrato y se dan en el contexto de un contacto forzoso entre diferentes grupos sociales. En ocasiones, este contacto con la lengua de superestrato se mantiene y se forma un continuo poscriollo, en el cual, entre la variedad “más criolla” (basilecto) y la lengua estándar (acrolecto), se ubican los diferentes ideolectos (mesolectos) de los hablantes. Esto plantea una problemática en la selección y la transcripción del corpus.

En cuanto a la selección, este continuo problematiza la identificación de aquello que es creole y aquello que no lo es. En San Andrés, no hay un contacto tan estrecho con el inglés como a principios del siglo XX e incluso se afirma que no sería apropiado caracterizar la situación lingüística de la isla ya como un continuo (Bartens, 2002); sin embargo, se mantienen huellas de dicho continuo en la actualidad. Para superar esto, se aceptó la concepción de la documentación lingüística de que los corpus deben ser lo más diversos posible y de que, incluso, se deben basar en

una comunidad de habla y no en un único código (Himmelmann, 1996, citado en Woodbury, 2011).

La segunda problemática que surge se relaciona con la transcripción de la lengua. No se ha estandarizado el creole y, aunque existe una propuesta ortográfica de ella, no se ha difundido masivamente (Moya, 2006). Incluso, si se decidiera sobre un sistema de escritura para la recolección de un corpus, la variabilidad lingüística afecta la decisión de utilizar, bien sea formas estándares, bien sea formas criollas. Esto, en términos de lingüística de corpus, también afecta la relación *type-token*, ya que, en ocasiones, podría resultar vago decidir si dos *tokens* similares corresponden a dos *types* diferentes o al mismo *type* (Sebba y Dray, 2007). Esta cuestión, sin embargo, se puede superar al recopilar el corpus, ya que el análisis de este permitiría obtener una perspectiva más comprensiva acerca de la lengua y, por lo tanto, una mejor identificación de las formas que son estándares y las que son criollas.

### 3. Marco metodológico

En la teorización del corpus, se hizo un acercamiento a los agentes interesados que eran miembros de la comunidad raizal, para conocer su punto de vista acerca de esta iniciativa de investigación. Hubo opiniones muy diversas acerca del piloto, tanto positivas, como negativas. Este acercamiento, sin embargo, llamó la atención de algunas investigadoras nativas de la isla cuya lengua materna es el creole y que estaban interesadas en aportar sus conocimientos y colaborar con el proyecto. Con estas participantes, se estableció que el propósito del piloto sería explorar las herramientas de corpus para posteriormente expandir el alcance del proyecto.

Para esto, se diseñó el corpus como un “corpus especial de lengua oral”. Por el tamaño y la especificidad del corpus, no se podría decir que sería general ni especializado, ya que no sería lo suficientemente representativo de una variedad lingüística. Sería un corpus especial, ya que, por la naturaleza de las sesiones que integrarían el corpus, habría una intervención explícita de las colaboradoras en la producción de los hablantes (Sinclair, 1996). Con relación al número de lenguas, se esperaba que el corpus fuera monolingüe, ya que, independientemente de si algunas sesiones contenían un mínimo de alternancia con el español o el inglés, el corpus seguiría siendo monolingüe. Por otro lado, si había presencia de alguna sesión con contenido excesivamente acrolectal, no se consideró que esto sería suficiente para constituir un corpus comparable

de dos variedades lingüísticas diferentes (Sinclair, 1996). La decisión de recopilar un corpus de lengua oral obedeció al hecho de que hay pocos textos escritos en lengua creole, los cuales no son de temática diversa (Ramírez-Cruz, 2017), y conseguir las licencias para utilizar estos textos no habría sido viable por cuestiones de tiempo.

Este corpus contendría sesiones que versaran sobre una temática en específico: la crianza en *di yaad* (“el patio”), un espacio social importante en San Andrés. Por la naturaleza de la temática, los entrevistados serían personas entre los 45 y los 65 años, sin restricción sobre el sexo, cuyo trasfondo lingüístico incluyera el creole como lengua nativa o segunda lengua, además de que hubieran vivido en San Andrés de forma continua gran parte de su niñez y de su adolescencia, y que hubieran estado residiendo en la isla de forma continua durante los últimos dos años (Loaiza-Camacho, 2018).

Para la recolección de los datos lingüísticos, se había sugerido conseguir equipos especializados de grabación; sin embargo, por cuestiones administrativas, no se obtuvieron los equipos a tiempo y cada participante llevaría a cabo las sesiones con un teléfono móvil propio.

Cada sesión vendría acompañada de un formato de consentimiento de uso y un formato de metadatos por medio del cual se llevaría un registro de aspectos importantes acerca de las sesiones y de los participantes. Para nombrar cada sesión, se acordó un método que codificaría el barrio en el que tomó lugar, la edad del entrevistado, el sexo del entrevistado y la fecha de la sesión.

Para la transcripción de los datos lingüísticos, se decidió adoptar la ortografía para el creole desarrollada por la Islander Spelling Committee con el apoyo de la Corporación Universidad Cristiana (con algunas adaptaciones). Este sistema ortográfico fue escogido por su carácter semigrafemático y por la facilidad de usarlo en la transcripción por computadora. Esta transcripción se haría por medio del programa ELAN, del Instituto Max Planck.

Posteriormente, se realizaría un análisis cuantitativo del corpus por medio del análisis de frecuencias de *types* y *tokens*, concordancias y colocaciones. Para este análisis, se convertirían las sesiones transcritas en formato de texto plano .txt y luego se utilizaría el programa AntConc. Adicionalmente, para algunos cálculos, se recurriría al lenguaje de programación R.

## 4. Resultados

En total, se llevaron a cabo seis sesiones. Estas sesiones y los nombres asignados a ellas se encuentran ilustradas en la Tabla 1.

Tabla 1. Sesiones.

Sesión	Nombre
1	mhf57_270418
2	phm54_280418
3	Lolia_Yard
4	Pabli_Yard
5	macf61_280418
6	maslf52_210418

Fuente: Elaboración propia, con base en el piloto realizado.

Inicialmente, se había propuesto que cada sesión durara treinta minutos aproximadamente. Con esto en mente, el corpus esperado habría alcanzado tres horas. Sin embargo, la duración de cada sesión de trabajo fue inferior a lo esperado y el corpus observado fue de un poco más de una hora, como se ilustra en la Tabla 2.

Tabla 2. Duración de sesiones.

Sesión	Nombre	Duración
1	mhf57_270418	8:08
2	phm54_280418	6:22
3	Lolia_Yard	6:39
4	Pabli_Yard	4:45
5	macf61_280418	21:27
6	maslf52_210418	20:26
<b>Duración total</b>	<b>1:07:47</b>	

Fuente: Elaboración propia, con base en el piloto realizado.

Los principales retos que surgieron durante las sesiones fueron (a) la temática, (b) la experiencia en trabajo de campo y (c) los equipos de grabación.

En cuanto a la temática, las participantes del piloto expresaron que la temática resultaba algo restrictiva, lo cual provocaba que los entrevistados no pudieran expresar demasiada información.

En cuanto al trabajo en campo, algunas de las participantes expresaron sentir inseguridad frente a las sesiones, ya que, en lugar de un diálogo fluido, la sesión se tornaba en un cuestionario de preguntas cerradas en algunas ocasiones. Esto se reflejó también en cierta dificultad para apropiarse de las convenciones para nombrar las sesiones.

En cuanto a los equipos de grabación, al tener que utilizar sus teléfonos móviles, esto resultaba algo intrusivo para los entrevistados, quienes se limitaban a dar respuestas cortas y concisas a las preguntas planteadas al ver el dispositivo. Esto se reflejó también en que, según las participantes, en ocasiones los entrevistados comenzaban a mantener una conversación más fluida y prolongada una vez la entrevista hubiera acabado y el equipo de grabación hubiera sido desactivado.

Para cada sesión se hizo una transcripción ortográfica. En la Tabla 3, se aprecia el número de *tokens* de cada transcripción.

Tabla 3. Número de tokens.

Sesión	Nombre	Nro. de Tokens
1	mhf57_270418	863
2	phm54_280418	915
3	Lolia_Yard	569
4	Pabli_Yard	321
5	macf61_280418	144
6	maslf52_210418	2193
<b>Total de tokens</b>	<b>5005</b>	

Fuente: Elaboración propia, con base en el piloto realizado.

Originalmente, se había esperado que, si el corpus era de una duración aproximada de tres horas, entonces tendría un número aproximado de 18.000 *tokens*, asumiendo una cantidad de cien tokens por minuto. Una nueva aproximación se realizó con base en la duración aproximada de una hora y ocho minutos, lo cual daba un número de tokens esperados de 6.800. El corpus es más pequeño que eso. Sin embargo, se debe tener en cuenta que la transcripción realizada para la sesión **macf61\_280418** se hizo sobre una porción de aproximadamente cuatro minutos. Por ello, el total de minutos transcritos fue 52; con esto, el número esperado de *tokens* era 5.200. El tamaño observado del corpus sigue siendo más pequeño que el tamaño esperado; sin embargo, es posible determinar estadísticamente si ambos tamaños son lo

suficientemente similares para que su diferencia no sea significativa, como se ilustra más abajo.

La normalización de la desviación entre el número de *tokens* esperado y el número de *tokens* observado se calcula de la siguiente manera, donde

$$DP_{norm} = \frac{DP}{1-s} = \frac{195}{1-5005} \approx -0.02$$

$DP_{norm}$  es la normalización, DP es el valor absoluto de la diferencia entre el número observado y el número esperado de tokens y s es el número de *tokens* del corpus más pequeño (i.e. el observado) (Loaiza-Camacho, 2018): Puesto que el valor resultante se acerca a 0, es razonable decir que el número de *tokens* observado es similar al número de tokens esperado y la diferencia entre ambos no es significativa (Lijffijt y Gries, 2012).

En la transcripción de las sesiones, el mayor reto fue la apropiación de la ortografía utilizada. Por ser el creole una lengua mayoritariamente oral y por ser San Andrés una sociedad que se podría considerar multigráfica (Lüpke, 2011) debido al contacto de dos lenguas mayoritarias (el inglés y el español), se suele recurrir al uso de convenciones ortográficas de uno o de otro para escribirlo. Considérense los siguientes ejemplos:

- a. <c> en lugar de <k>
  - i. ...*Ahn when da crab taim now...* (maslf52\_210418)
  - ii. ...*dem gat wan pat deh mek krab suup...* (phm54\_280418)
- b. <x> para el grupo consonántico <ks>
  - i. ...*wii yuuztu plie jaks wi yuuztu plie stik...* (mhf57\_270418)
  - ii. ...*wii non gat non jax wii tek a yong laim...* (Lolia\_Yaad)
- c. palabras ortográficas, e.g., <once> en lugar de <wans>
  - i. ...*dem waz a lat difarent dan de once dem tideh bikaaz...* (Pabli\_Yaad)

En el análisis de las frecuencias, se identificaron los veinte tokens más frecuentes en el corpus, como se ilustran en la Tabla 4.

Puede observarse que se cumple el principio general de que los *tokens* más frecuentes en un corpus son las palabras funcionales (Manning y Schütze, 1999): ‘ahn’, ‘dem’, ‘yuuztu’; incluso el *token* ‘go’, el cual, aunque puede funcionar como una palabra léxica, es una palabra funcional, prominente en el creole. No obstante, hay una palabra léxica plena que se destaca: el verbo *plie*, en el puesto 14. Cuando se calcula la normalización de la desviación de la proporción esperada del *token* con la proporción observada, de la forma expuesta anteriormente, el valor resultante (0.282) es cercano a 0, por lo cual se puede decir que los valores de ambas proporciones son similares. Esto indica que, en todas las sesiones, este *token* se encuentra de forma balanceada y no es el caso de que se encuentre presente de forma prevalente en una sesión específica. Además, hay una

Tabla 4. Lista de frecuencias

Puesto	Frecuencia	Token	Puesto	Frecuencia	Token
1	230	Wi	11	72	wii
2	191	Di	12	71	Go
3	139	ahn	13	68	ai
4	134	dem	14	61	plie
5	124	An	15	55	Tu
6	95	A	16	55	Yu
7	94	yuuztu	17	54	deh
8	81	Fi	18	44	taim
9	81	So	19	44	Yo
10	78	now	20	40	wehn

Fuente: Elaboración propia.

fuerte correlación entre el número de *tokens* en cada una de las sesiones y la frecuencia de *plie*; i.e., entre más *tokens* hay en cada sesión, hay mayor frecuencia del verbo *plie* (Kendall Tau-B test,  $p < 0.05$ ).

Esta alta frecuencia de *plie* y su dispersión son indicativos de un fenómeno quizás discursivo. Es posible que los miembros de la comunidad raizal asocien de forma recurrente la crianza en *di yaad* con el juego, más que con otros ámbitos. Esta hipótesis se vería respaldada por la comparativamente baja frecuencia de *tokens* relacionados con otros ámbitos, e.g., *mama* “madre”, que ocurre con una frecuencia de 15 y ocupa el puesto 63, y *frenz* “amigos” que ocurre una única vez y se encuentra en el puesto 607. Las colocaciones de *plie* también indicarían esta asociación y, a su vez, son útiles para estudiar el dominio semántico del juego: *marbl*, *stik*, *liberty*, *lata*, *jaks*, *biesbaal*, etc. Esta generalización, sin embargo, se aplica a un corpus limitado en tamaño y, por lo tanto, no pretende ser definitivo. Existe la posibilidad de que la alta frecuencia de *plie* se deba a las preguntas realizadas por las participantes. En cualquier caso, un corpus más grande en torno a la temática de la crianza en *di yaad* podría confirmar o desmentir esta hipótesis.

Otras posibilidades de estudios que surgen del corpus son las siguientes:

1. la variación lingüística
  - i. ...*ihn gat fi muoto di televizhan si tingz in deh...* (phm54\_280418)
  - ii. ...*osea televizhan woz neva di distrokshan...* (maslf52\_210418)
2. la alternancia de código y los préstamos lingüísticos
  - i. ...*now dem hav a lat a aparatos dem now tu kiip...* (mhf57\_270418)
  - ii. ...*dem yuuztu kom from farad barrio an sidong...* (phm54\_280418)
3. el cambio lingüístico y la innovación léxica
  - i. ...*dis mont now spesifik dis April...* (maslf52\_210418)
  - ii. ...*shii yuuztu yuuz ihn strategy pan mi...* (maslf52\_210418)

## 5. Conclusiones

El corpus recolectado, aunque no cumplió con todas las metas propuestas, dio cuenta de ciertas particularidades al interior de la comunidad de habla raizal. Además, permitió formular una serie de hipótesis a partir de las cuales se pueden adelantar mayores estudios de la lengua a partir de la recopilación de corpus electrónicos (Loaiza-Camacho, 2018).

Los mayores retos enfrentados durante el piloto fueron (a) la inseguridad frente al trabajo en campo, (b) la falta de equipos de grabación especializados para las sesiones, (c) la implementación de la ortografía del creole y (d) la falta de apropiación de las convenciones para la sistematización de los datos lingüísticos.

En general, se demostró que el uso de herramientas informáticas permite realizar determinados tipos de análisis de forma eficaz y sistemática. Estas aplicaciones pueden apoyar futuros proyectos de documentación y descripción lingüística.

## 6. Recomendaciones

Es importante incentivar el acercamiento de este proyecto con la comunidad de habla raizal y así abrir el diálogo para definir las metas, los propósitos y la ruta que se debe seguir. En la medida en que los miembros de la comunidad son agentes interesados en su propia lengua, es necesario buscar un vínculo más estrecho con ellos (Loaiza-Camacho, 2018).

Un proyecto de esta naturaleza se beneficiaría del aporte de profesionales de diferentes áreas del saber que estén interesados en la documentación de las prácticas de la comunidad de habla raizal. Para llevar a cabo una nueva versión del proyecto, es necesario reformular algunas de las fases, tratar de obtener los recursos necesarios y llevar a cabo una planificación adecuada con el tiempo del que se dispone, así como fortalecer aquellos aspectos exitosos en el piloto y mejorar aquellos que presentaron dificultades.

Se espera que esta experiencia informe a otros proyectos de alcance similar y aporte al avance de la lingüística de corpus para el estudio de lenguas criollas e indígenas en Colombia.

## Referencias

- Álvarez-Álvarez, C. (2015). Teoría frente a práctica educativa: algunos problemas y propuestas de solución. *Perfiles Educativos*, 37(148), 172-190. <https://doi.org/10.1016/j.pe.2015.11.014>
- Bartens, A. (2002). Another Short Note on Creoles in Contact with Non-Lexifier Prestige Language. *Journal of Pidgin and Creole Languages*, 17 (2), 273–278.
- Gries, S. T. (2009). What Is Corpus Linguistics? *Language and Linguistics Compass*, (3), 1–17. <https://doi.org/10.1111/j.1749-818x.2009.00149.x>
- Grupo de Estudios Sociolingüísticos. (2008). *Corpus Sociolingüístico de Medellín*. Recuperado de: <http://comunicaciones.udea.edu.co/corpuslinguistico/>
- Himmelmann, N. P. (2006). Language documentation: What is it and what is it good for? En: Gippert, J., Himmelmann, N. y Mosel, U. (Eds.), *Essentials of Language Documentation* (pp. 1–30). Berlin.
- Instituto Caro y Cuervo. (2018). Corpus Lingüísticos del Instituto Caro y Cuervo (CLICC). Recuperado el 1 de octubre de 2018 de: <http://clicc.caroycuervo.gov.co/>
- Lijffijt, J. y Gries, S. (2012). Correction to Stefan Th. Gries' "Dispersions and adjusted frequencies in corpora", *International Journal of Corpus Linguistics*, 13:4 (2008). *International Journal of Corpus Linguistics*, 17 (1), 147–149.
- Loaiza-Camacho, B. S. (2018). *Experiencia piloto en el levantamiento de un corpus especial de la lengua oral para la descripción lingüística del creole de San Andrés* (tesis de pregrado). Universidad Nacional de Colombia, Bogotá, Colombia.
- Lüpke, F. (2011). Orthography development. En: Austin, P. & Salla-bank, J. (Eds.), *The Cambridge Handbook of Endangered Languages* (pp. 312–336). Cambridge.
- Manning, C. y Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. Cambridge & London: Massachusetts Institute of Technology.
- Moya, S. (2006). Fi Wii News: A Creole Writing Experience. *Cuadernos Del Caribe*, 8, 89–96. <https://doi.org/10.15446/cc>
- Ramírez-Cruz, H. (2017). *Ethnolinguistic Vitality in a Creole Ecology: San Andrés and Providencia*. University of Pittsburgh.
- Real Academia Española. (2014). *Diccionario de la lengua española* (23a ed.). Madrid.

- Rojas, T.** (Comp.) (2017). *Corpus lingüísticos. Estudio y aplicación en revitalización de lenguas indígenas*. Popayán: Universidad del Cauca.
- Sanmiguel, R., Schoch, M. y Pelufo, A.** (2014). *Proyecto de Investigación sobre el creole (lengua criolla de base inglesa) del Archipiélago de San Andrés y Providencia*. San Andrés.
- Sebba, M. y Dray, S.** (2007). Developing and Using a Corpus of Written Creole. En: J. Beal, K. Corrigan, y H. Moisl (Eds.), *Creating and Digitalizing Language Corpora* (Vol 1: Synchronic Databases) (pp. 181–204). New York: Palgrave Macmillan. <https://doi.org/10.1057/9780230223936>
- Sinclair, J.** (1996). *Preliminary recommendations on Corpus Typology* (EAGLES Document EAG-TCWG-CTYP/P).
- Villayande Llamazares, M.** (2010). *Aproximación a la lingüística computacional*. Universidad de León.
- Woodbury, A.** (2011). Language Documentation. En: P. Austin & J. Sallabank (Eds.), *The Cambridge Handbook of Endangered Languages* (pp. 159–186). Cambridge: Cambridge University Press.

## LA GRAMÁTICA DE LA ENFERMEDAD: APROXIMACIÓN A LA EXPRESIÓN LINGÜÍSTICA DE LA ENFERMEDAD EN EL KARIJONA (CARIBE, REGIÓN AMAZÓNICA)<sup>1</sup>

David Felipe Guerrero Beltrán<sup>2</sup>

### Resumen

Este artículo pretende socializar una propuesta teórica para el análisis de la gramática de la enfermedad, que involucra las estrategias lingüísticas utilizadas para expresar los estados de enfermedad. Se presentan las consideraciones iniciales para la consolidación de un programa de investigación de la gramática de la enfermedad basadas en el trabajo fundacional de Aikhenvald (2018). Posteriormente, se desarrolla un análisis inicial para el caso de la lengua Karijona<sup>3</sup> (familia Caribe). Finalmente, se concluye que este es un dominio en el cual existen estrechas relaciones entre lengua, cultura e historia, que requiere una indagación más profunda y exhaustiva.

**Palabras clave:** gramática de la enfermedad, karijona, lenguas caribe, tipología lingüística, lingüística antropológica.

<sup>1</sup> Agradezco a Alexandra Aikhenvald, Kasia Wojtylak y a Camilo Robayo por sus comentarios, conversaciones y (especialmente) por su paciencia, las cuales fueron fundamentales no solo para concebir este artículo, sino para fortalecer mi proceso de formación como lingüista. Agradezco también a la comunidad Karijona de Puerto Nare, cuyos miembros, especialmente los abuelos sabedores, han estado siempre abiertos a enseñarme su lengua y a brindarme un espacio de mutuo respeto y cariño. Asimismo, agradezco al equipo de organización del Foro LING 20, quienes brindaron admirablemente el espacio para que este trabajo pudiera ser expuesto y compartido. Esta investigación fue financiada por la Beca de Posgrado de la Universidad Nacional de Colombia (UNAL), por el proyecto "Descripción y Revitalización de la Lengua Karijona" dirigido por Camilo Robayo COD: 41670 – HERMES, de la UNAL; y al Research Grant de la Australian Linguistic Society (ALS).

<sup>2</sup> Estudiante de Maestría en Lingüística de la Universidad Nacional de Colombia. Correo: [dfguerrerob@unal.edu.co](mailto:dfguerrerob@unal.edu.co)

<sup>3</sup> A lo largo de este artículo, Karijona va a referir al grupo étnico, mientras que Karijona va a referir a su lengua tradicional.

## Abstract

This paper aims to share a theoretical proposal for the analysis of the grammar of disease, which involves the linguistic strategies employed on the expression of the state of being sick. Following the foundational paper of Aikhenvald (2018), I introduce the initial considerations for the developing of a research program focused on the grammar of disease. Subsequently, I present a first analysis of the grammar of disease in the Karijona language (Carib family). Finally, I conclude that the analysis of the grammar of disease highlights strong relations between language, culture and history that require further inquiry.

**Key words:** grammar of disease, karijona language, carib languages, linguistic typology, anthropological linguistics.

## 1. Introducción<sup>4</sup>

La enfermedad, fisiológica y socialmente construida, hace parte de la experiencia de cada ser y comunidad humana. Por lo tanto, es de esperar que cada persona cuente con diferentes estrategias lingüísticas para expresar y comunicar los estados y procesos que involucran tener una enfermedad. La relación entre medicina y cultura ya ha sido explorada previamente en la literatura especializada en el campo de la medicina (Eisenberg y Kleinman, 1981; Kleinman, Eisenberg y Good, 1978; Lupton, 2012); asimismo, varios trabajos etnográficos se han realizado desde la antropología sobre la relación entre la medicina tradicional y la cosmogonía de pueblos indígenas dentro del territorio colombiano (ver: Estrada, 2014; Hildebrand, 1983; Portela Guarín y Portela García, 2018). No obstante, el papel del lenguaje como puente conciliador entre fenómenos fisiológicos y constructos socioculturales aún no ha sido explorado con profundidad dentro de la lingüística.

<sup>4</sup> Abreviaciones: 1,2,3: primera, segunda y tercera persona; 1+2: primera persona inclusiva; ADVZ: Advercializador; ANAF: anafórico; ANIM: animado; AUG: número aumentado; COP: cópula; DIST: distante; DUR: durativo; ENF: enfático; FRUST: frustrativo; HAB: habitual; INAN: inanimado; INST: Instrumental; IPFV: imperfectivo; MIN: número mínimo; NEG: negación; NFUT: tiempo no-futuro; NMZ: nominalizador; OBL: argumento oblicuo; PFV: perfectivo; PREV: posesión previa; POSS: posesión actual; RED: reduplicación; REP: evidencia reportada; Sa: Sujeto intransitivo alineado con sujeto transitivo; So: Sujeto intransitivo alineado con objeto; Sp: préstamo del español; TEMP: posposición temporal; VLZ: verbalizador.

Recientemente, Aikhenvald (2018) presentó, en el marco del *Round-table discussion on 'The grammar of disease'* del Language and Culture Research Centre (James Cook University), una propuesta inicial para el estudio de la expresión lingüística de la enfermedad (en adelante: gramática de la enfermedad) desde una perspectiva tipológica. Dicha propuesta contiene un conjunto de consideraciones teóricas para la definición de un nuevo tópico de investigación para la lingüística descriptiva y tipológica. En esta propuesta, la autora brinda las categorías y consideraciones básicas para analizar translingüísticamente la gramática de la enfermedad. No obstante, la búsqueda de un tópico de investigación translingüístico implica algunos retos y consideraciones por superar, tales como:

- a. El concepto de enfermedad involucra aspectos ideológicos y culturales; por lo tanto, no es homogéneo en todas las sociedades.
- b. Estudiar la gramática de la enfermedad translingüísticamente implica aproximarse a la concepción de la enfermedad transculturalmente. No hacerlo podría conllevar a la imposición de categorías exógenas en el proceso de documentación y análisis lingüístico, e invisibilizaría las concepciones y epistemologías propias de las respectivas comunidades de habla, especialmente en el caso de comunidades minoritarias.
- c. El involucramiento de aspectos ideológicos y culturales en la concepción de enfermedad implica que los fenómenos lingüísticos asociados a la gramática de la enfermedad no puedan estudiarse exclusivamente en términos lingüísticos, debido a que requieren también de consideraciones antropológicas y sociológicas.

Esto lleva a que sea relevante introducir teóricamente la discusión acerca de cómo definir el concepto de enfermedad de manera tal que sea aplicable transculturalmente. Por lo mismo, la ponencia pretende que este tema, presentado como un tópico de investigación inexplorado, se extienda hacia la consolidación de un programa de investigación.

Este artículo se estructura así: en el numeral 2, se presenta la propuesta tipológica de Aikhenvald (2018) y se incluyen algunas consideraciones teóricas que permiten acotar el concepto de enfermedad. En el numeral 3, se presenta el estudio de caso de la lengua Karijona, en el cual se observa la aplicación de las categorías presentadas. Finalmente, en el numeral 4, se muestran las conclusiones del análisis realizado.

## 2. La gramática de la enfermedad

### 2.1 Sobre el concepto de enfermedad y su expresión lingüística

De acuerdo con Kleinman et al. (1978, p. 251) es posible diferenciar entre:

**Enfermedad fisiológica** (*disease*): esta consiste en la anormalidad en el funcionamiento y la estructura de los sistemas y órganos corporales.

**Enfermedad social** (*illness*): la cual comprende las experiencias de cambios negativos en los estados de ser, los cuales cumplen una función social; es decir, considera la experiencia humana de tener una enfermedad y sus implicaciones sociales.

La división entre enfermedad fisiológica y social de Kleinman *et al.* (1978) brinda un primer paso hacia la acotación del concepto de enfermedad, puesto que hace explícita la división entre los procesos fisiológicos humanos y los constructos sociales que giran en torno a ellos. Sin embargo, el concepto de “enfermedad social” sigue siendo demasiado heterogéneo como para aplicarse transculturalmente.

Por tal razón, con base en el concepto de anormalidad (*anormalité*) de Foucault (2005), se consideran los siguientes criterios para acotar el concepto de enfermedad social:

- Hay una estructura social de conocimiento que distingue entre normal (**persona sana**) y anormal (**persona enferma**);
- la anormalidad (**enfermedad**) se crea como categoría asociada a individuos socialmente peligrosos (**enfermos**), lo cual implica que dichos individuos tienen que ser corregidos (**curados**);
- la búsqueda de corrección implica que se desarrollen **prácticas sociales** en las cuales la anormalidad (enfermedad) sea reconocida (tenga **síntomas**), corregida (tenga **tratamiento**), estudiada (tenga **causas** u **orígenes**); y tenga consecuencias o connotaciones negativas asociadas (como la **muerte**, el **dolor**, la **desgracia**, etc.).

Estas consideraciones presuponen que los comportamientos sociales son categorizados en términos estructurales. Sin embargo, estas no presuponen categorías universales del concepto de enfermedad (como podría ser su caracterización en términos fisiológicos). En contraste, estas se enfocan en las relaciones sociales que llevan a que una persona sea considerada como enferma, según los criterios propios de una determinada comunidad. Se propone entonces que el análisis de la gramática de la enfermedad esté fundamentado en el concepto de enfermedad social desde la anormalidad, más que en el de enfermedad fisiológica.

### 2.2 Consideraciones de análisis

Basado en los planteamientos de Aikhenvald (2018; comunicación personal), se proponen los siguientes criterios para analizar la gramática de la enfermedad:

- a. Conocimiento cultural involucrado: en qué medida una determinada enfermedad puede tener síntomas, tratamientos, orígenes o connotaciones asociados.
- b. Inventario, taxonomía y clasificación de las enfermedades.
- c. Clases de palabras utilizadas: este criterio considera cuáles clases de palabras son utilizadas para codificar las enfermedades, tal como lo son nombres (enfermedad, gripa, etc.), verbos (enfermar, contagiar), adjetivos (agripado, enfermo), formas compuestas (dolor de cabeza), entre otras.
- d. Estructuras de cláusula utilizadas: en qué medida se utilizan cláusulas intransitivas, transitivas, ditransitivas, copulativas o no verbales.
- e. Funciones sintácticas de la enfermedad (Ed): Si la enfermedad es un predicado, sujeto intransitivo o transitivo, objeto, complemento de cópula o argumento periférico.
- f. Funciones sintácticas del enfermo (Eo): Si quien porta la enfermedad es codificado gramaticalmente como un sujeto intransitivo o transitivo, objeto, complemento de cópula o argumento periférico.

## 2.3 Esquemas semántico-gramaticales

Con base en la propuesta de Aikhenvald (2018) y los criterios presentados en la sección 2.2, se proponen los siguientes esquemas semántico-gramaticales para ser considerados dentro del análisis descriptivo<sup>5</sup>:

**a. Esquema de predicación.** Este esquema está basado en clase de palabra y función sintáctica de la enfermedad (Ed) y considera los casos en los cuales esta es codificada por medio de un predicado:

- i.* Adjetival: Ed es un predicado adjetival dentro de una cláusula copulativa o no verbal.
- ii.* Verbal I: Ed es predicado verbal dentro de una cláusula intransitiva.
- iii.* Verbal II: Ed es predicado verbal dentro de una cláusula transitiva.

**b. Esquema de ubicación.** Este esquema considera la extensión metafórica de construcciones locativas en la que el enfermo (Eo) es un argumento periférico o complemento de cópula introducido por una adposición espacial o marca de caso locativa, el cual se divide en dos subtipos, dependiendo el estado de movimiento:

- i.* Dinámico: Ea es agente y sujeto de un verbo de movimiento y Eo es la meta del movimiento.
- ii.* Estático: Ed está ubicado en Eo introducido por una cópula o un predicado de postura o posición.

**c. Esquema relacional.** Ed es un argumento periférico o un complemento de cópula introducido por una adposición o una marca de caso. Este esquema también puede considerar extensiones metafóricas de construcciones metafóricas (tal como en c-i); sin embargo, en el caso de las construcciones del esquema relacional, a diferencia de las del esquema de ubicación, la enfermedad y no el enfermo es el argumento periférico.

*i.* Contenedora: Ed es el contenedor de Eo.

*ii.* Comitativo: Ed es concomitante con Eo.

*iii.* Instrumental: Ed es instrumento de afectación de Eo.

*iv.* Malefactivo: Ed afecta negativamente a Eo.

**d. Esquema de posesión.** Este esquema involucra un predicado de posesión:

- i.* Eo agentivo: el enfermo posee a la enfermedad.
- ii.* Ed agentivo: la enfermedad posee al enfermo.

**e. Esquema de adquisición** (involucra esquemas complejos como tomar, dar, coger):

- i.* Eo agentivo: Eo es agente y Ed es paciente de la adquisición.
- ii.* Ed agentivo: Ed es agente y Eo es paciente de la adquisición.

**f. Esquemas no canónicos** (involucran construcciones con otras características gramaticales):

- i.* Tópico: Eo es tópico en posición inicial y Ed es sujeto de cláusula siguiente.
- ii.* Verbo estático-parte del cuerpo: Eo es tópico en posición inicial y Ed está codificado en verbo estático y en parte del cuerpo del Eo afectada.

En la siguiente sección se presenta una aplicación de estas consideraciones de análisis en el caso particular de la lengua karijona.

<sup>5</sup> Estos esquemas son presentados con mayor detalle y profundidad en Aikhenvald (2018); la presente propuesta considera algunas modificaciones en su clasificación, tal como la organización en subgrupos de acuerdo con las características estructurales y funcionales de los esquemas.

### 3. La gramática de la enfermedad en el karijona

#### 3.1 La comunidad Carijona y su lengua

Los karijona son una comunidad indígena de entre 120 y 200 personas. Están ubicados en diferentes lugares de la región amazónica colombiana. Hay dos resguardos en Miraflores (Guaviare) y en La Pedrera (Amazonas), los cuales son los principales asentamientos de la comunidad. Adicionalmente, algunas familias están ubicadas en las ciudades de Villavicencio (Meta), Leticia (Amazonas), San José del Guaviare (Guaviare) y Bogotá. Se cree que hay karijonas viviendo en aislamiento voluntario en el territorio del Parque Natural Chiribiquete.

El karijona [iso 639-3 cbd] es una lengua de la familia lingüística Caribe, la única de esta familia hablada en la región del noroccidente amazónico. Actualmente, la lengua cuenta con cinco personas (una familia) que usan la lengua en el contexto familiar. Hay diez personas que conocen la lengua, pueden sostener conversaciones y contar historias tradicionales, y la usan ocasionalmente de manera espontánea. Adicionalmente, hay alrededor de 16 personas que tienen un conocimiento pasivo de la lengua. Todos los hablantes son bilingües con el español y mayores de 45 años. No obstante, un diagnóstico sociolingüístico se está llevando a cabo actualmente por el estudiante Sebastián Porras (en proceso), con el cual se espera tener información más precisa del estado de vitalidad de la lengua y las actitudes lingüísticas de las personas.

Los datos de primera mano fueron recogidos en trabajos de campo en el resguardo karijona de Puerto Nare (Miraflores, Guaviare) y Leticia, entre 2014 y 2017. Los datos de segunda mano fueron recogidos por Camilo Robayo en los resguardos de Puerto Nare y La Pedrera entre 1982 y 1986.

El corpus consiste en alrededor de treinta horas de grabaciones transcritas de primera y segunda mano, las cuales incluyen historias tradicionales, historias de vida, narraciones de eventos históricos, conversaciones y entrevistas. En los datos de primera mano participaron los sabedores Ana Benjumea, Lilia Gómez, Ernesto Carijona y José Romero. Gran parte de las grabaciones fueron transcritas y traducidas al español por el profesor Víctor Narváez.

#### 3.2 Análisis exploratorio

Una primera aproximación al karijona muestra que los orígenes de las enfermedades juegan un rol importante en cómo estas son expresadas gramaticalmente.

Hay una diferencia entre las enfermedades que tienen origen no-indígena (relacionadas con los blancos) las cuales están directamente asociadas con la muerte, y aquellas de origen indígena que pueden curarse por medio de un rezo, del uso de yerbas medicinales o con reposo. Las enfermedades de origen indígena en el contexto amazónico también están asociadas a la violación de reglas sociales de comportamiento (tabúes), tal como menciona Wojtylak (2018) para el caso de la comunidad Murui.

En el corpus analizado, se encontró una referencia recurrente a las enfermedades no-indígenas, debido a las epidemias que la comunidad karijona tuvo que afrontar a comienzos del siglo XX. Se observa un esquema de predicación verbal (a-iii) con el verbo *ahēh*, que suele asociarse a dos significados: ‘morir’ y ‘estar enfermo’:

- |       |  |                         |                           |
|-------|--|-------------------------|---------------------------|
| (1)   | [mure-me   | dʒ-atawə] <sub>pp</sub> | dʒ-ahēh-i <sub>pred</sub> |
|       | niño-ADVZ  | 1-TEMP                  | 1.So-morir-PFV            |
|       | “cuando era niño (cuando mi niñez) yo me enfermé”    |                         |                           |
| <hr/> |  |                         |                           |
| (2)   | ararər-ake   | n-ahē-dʒa-epred         |                           |
|       | ahora-NEG  | 3.Sa-morir-NFUT-IPFV    |                           |
|       | “él no se va a morir ahora (él va a morir no ahora)” |                         |                           |
| <hr/> |  |                         |                           |

Este verbo también se combina con la partícula frustrativa =rehe para aquellos casos en los que se sobrevive a una enfermedad grave (de la cual se espera la muerte):

- |       |   |                                |                             |
|-------|---|--------------------------------|-----------------------------|
| (3)   | irətɪwə   | dʒ-ahēh-i=rehe <sub>pred</sub> | dʒ-awona=ha <sub>pred</sub> |
|       | Entonces  | 1.So-morir-PFV=FRUST           | 1.So-curar=ENF              |
|       | “Entonces yo me enfermé (casi muero) y me volví a curar otra vez” |                                |                             |
| <hr/> |   |                                |                             |

En otras ocasiones, se observa este verbo *ahēh* nominalizado junto con la posposición espacial *həkə*, de acuerdo con el esquema relacional (c):

- (4) [mama kehiri]<sub>pp,vcs</sub> [ahēh-i<sub>pred</sub>-tī<sub>np</sub> hōkō=gōrō=ha]<sub>pp,vcc</sub>  
 mamá.Sp Finado morir-PFV-NMZ SOBRE=TODAVÍA= ENF  
 “finadita mi mamá todavía estaba enferma (estaba sobre la enfermedad)”

Cuando las enfermedades de origen no-indígena son mencionadas (como nombres) en las narraciones sobre epidemias, estas aparecen con la posición instrumental =ke, de acuerdo con el esquema relacional (c-iii):

- (5) isanom-ake=gōrō=ha=ti nai<sub>cop</sub>  
 poco-NEG=TODAVÍA=ENF=REP COP  
 [karijona kahi-tjā]<sub>pp,cs</sub> giri  
 Carijona tumba-PREV.AUG Compañero  
 “Hay muchas tumbas de los carijona, compañero”
- (6) isanom-ake nai<sub>cop</sub> [mōrō bijīga=ke]<sub>pp,vcc</sub>  
 poco-NEG COP 3.INAN sarampión=INST  
 [mōrō ahēh-i<sub>pred</sub>=tjā-tjā]<sub>pp,cs</sub>,  
 3.INAN morir-PFV=FRM.AUG-REDUP  
 “Hay varios que murieron por (con) ese sarampión,”

- (7) toferina=ke<sub>pp,vcc</sub> [ahēh-i<sub>pred</sub>=tjā-tjā]<sub>pp,cs</sub>  
 toferina.Sp=INST morir-PFV=FRM.AUG-REDUP  
 “...los que murieron por (con) toferina,”

- (8) [mōrō ‘paludismo’ ka-toho=ke]<sub>pp,vcc</sub> [ahēh-i<sub>pred</sub>=tjā-tjā]<sub>pp,cs</sub>  
 3.INAN paludismo.Sp decir-NMZ.OBL=INST morir-PFV=FRM.AUG-REDUP  
 “...los que murieron por ese que llaman ‘paludismo’...”

- (9) [t-ehi-ri-ko ehōr-i<sub>pred</sub>]-hā<sub>np</sub>=gōrō=ha nai<sub>cop</sub>,  
 3-remedio-MIN.POSS-AUG.R encontrar-PFV-NEG=TODAVÍA= ENF COP  
 “No había (no se había encontrado) todavía el remedio de ellos (los blancos),”

- (10) [irō mesa]<sub>obl</sub> t-edudō-ake<sub>pp,vcc</sub> ki-na-kōtakō<sub>cop</sub> [mōrō ehī]<sub>cc</sub>.  
 3. ANAF Tiempo 3-CONOCER-NEG 1+2.SO-COP-HAB 3.INAN remedio  
 “en esa época no se conocía ese remedio.”

Por otro lado, se observa cómo la expresión de enfermedad por fuera del contexto de las epidemias utiliza diferentes estrategias. En el ejemplo (11), se tiene que la fiebre se expresa por medio de un proceso de verbalización que toma como base el adjetivo atu ‘estar caliente’ (esquema de predicación a ii):

- (11) ‘ōwi d3-atu-na-e<sub>pred</sub> do!  
 1.MIN 1.So-caliente-VLZ-IPFV Interjección  
 ga-nō<sub>pred</sub> mōkō Kuwai  
 3.decir-DUR 3.ANIM.DIST Kuwai  
 “«¡Estoy con fiebre, hombre!» dijo (está diciendo) ese Kuwai.”

### 3.3 Limitaciones y retos por superar

El trabajo se realizó con bastantes limitaciones en relación con los conocimientos culturales asociados a las enfermedades y sin un inventario previo de ellas. Esto llevó a que no fuera posible indagar detalladamente los aspectos lingüísticos y culturales presentados en la sección 2. Hace falta investigar con mayor profundidad los siguientes aspectos lingüísticos:

- El inventario y taxonomía de enfermedades endógenas y exógenas.
- Los esquemas semántico-gramaticales asociados a enfermedades endógenas.
- Los esquemas semántico-gramaticales utilizados en la variedad del español de Puerto Nare.
- La relación con otros dominios semánticos (i.e. espacialidad, emociones, estados anímicos o necesidades fisiológicas), tal como menciona Aikhenvald (2018).

También hace falta indagar en los siguientes aspectos culturales:

- Los aspectos de sintomatología y tratamiento de las enfermedades.
- Las connotaciones asociadas a los distintos tipos de enfermedades.
- La relación con otros campos de conocimiento (conocimiento del territorio, de la etnobotánica y de la medicina tradicional; roles sociales; reglas de comportamiento; entre otros).

También hace falta considerar en la tipología si pueden definirse tipos o factores básicos de enfermedades que influyan en la expresión lingüística o clasificación endógena de las enfermedades y sean aplicables transculturalmente, por ejemplo:

- a. Es probable que las enfermedades que involucran partes específicas del cuerpo requieran explicitar la parte del cuerpo adolorida en la construcción. Llama la atención que los ejemplos presentados por Aikhenvald (2018) para los esquemas denominados aquí como no canónicos<sup>6</sup> refieren en su mayoría a enfermedades que involucran partes del cuerpo específicas (ver Sección 2.3).
- b. La expectativa de los hablantes (Aikhenvald, 2018) con respecto a las consecuencias de la enfermedad. Es posible que haya una diferencia entre las enfermedades que causan la muerte y las que causan dolor o malestar. En los ejemplos (5)-(8) de la Sección 3.2 se puede observar el uso de la posposición instrumental =*ke* en los casos en los cuales las enfermedades fueron causa de muerte.
- c. Las enfermedades que pueden ser reconocidas visualmente (tales como herida, infecciones o brotes), en contraposición con aquellas que no son visibles (disfunción de órganos, maldiciones, virus, entre otras).
- d. Hace falta indagar también si existen asimetrías en el caso de enfermedades no fisiológicas (enfermedades mentales) o en el

caso de enfermedades mayores (la malaria, por ejemplo), y los síntomas asociados dichas enfermedades (tal como la fiebre).

Estas consideraciones, más que limitaciones, son oportunidades para enfocar futuros proyectos e iniciativas de investigación que permitan consolidar esta propuesta, aún inicial y endeble, en un programa de investigación robusto.

#### 4. Conclusiones

Este trabajo se acercó a las bases teóricas propuestas por Aikhenvald (2018) para el análisis de la gramática de la enfermedad, un tema hasta ahora muy poco estudiado en la lingüística. Esta primera aproximación a la gramática de la enfermedad en el karijona evidencia los vínculos existentes entre la expresión lingüística de la enfermedad, la historia y la cultura de la comunidad karijona. Se observó que las categorías y consideraciones presentadas en la sección 2 fueron aplicadas parcialmente en el análisis exploratorio de la gramática de la enfermedad en el karijona, pese a que aún se requiere un estudio más profundo sobre el tema.

La reflexión en torno al concepto de enfermedad y el análisis exploratorio de la gramática de la enfermedad muestran que esta está estrechamente ligada a las epistemologías de cada comunidad de habla, es decir que, existe una relación intrínseca entre la expresión lingüística de las enfermedades y el constructo de conocimientos locales con los que una comunidad las concibe y categoriza.

En el contexto del reconocimiento y visibilización de las comunidades minoritarias, el estudio de la gramática de la enfermedad (incluyendo la concepción de las enfermedades y sus clasificaciones, sus tratamientos y significados asociados, así como sus motivaciones sociales e históricas) permitirá apoyar las iniciativas cuyo interés sea el de preservar sus epistemologías propias y su patrimonio lingüístico.

Se invita a profesores, investigadores, activistas y lingüistas en formación, indígenas y no indígenas, a ver en este campo de estudio un referente para direccionar programas de investigación que involucren a los miembros de las comunidades como agentes de conocimiento y que visibilice la investigación lingüística nacional y regional en el panorama internacional.

<sup>6</sup> La autora no hace esta agrupación entre esquemas no canónicos, sino que presenta el esquema de tópico (*topic scheme*) y el esquema de verbo estático-parte del cuerpo (*stative verb-body part schema*) como esquemas independientes.

## Referencias

- Aikhenvald, A. Y.** (2018). *The grammar of disease: an introduction* (Global Workshop, 6 June). Cairns: LCRC James Cook University.
- Eisenberg, L., y Kleinman, A.** (1981). *The Relevance of Social Science for Medicine. Studies in Comparative Cross-Cultural Research 1*. <https://doi.org/10.1007/978-94-009-8379-3>
- Estrada, H.** (2014). *Medicina y magia en el sur del Tolima*. Bogotá: Instituto Caro y Cuervo.
- Foucault, M.** (2005). *Abnormal. Lectures at the College de France 1974-1975*. London: Verso.
- Hildebrand, M.** (1983). Cosmovisión y el concepto de enfermedad entre los Ufaina. In F. C. Colombianas (Ed.), *Medicina, shamanismo y botánica*. Bogotá: Editorial Presencia.
- Kleinman, A., Eisenberg, L., y Good, B.** (1978). Culture, illness, and care. Clinical Lessons from Anthropologic and Cross-Cultural Research. *Annals of Internal Medicine*, 88(2), 251–258.
- Lupton, D.** (2012). *Medicine as Culture*. London: Sage Publications Ltd.
- Portela Guarín, H., y Portela García, S. C.** (2018). *El cuerpo, el arco y la seña*. Popayán: Universidad del Cauca.
- Wojtylak, K.** (2018). *How did this grow in you? - About the expression of disease in Murui (Witotoan)* (Global Workshop, 25 September). LCRC James Cook University.

# REFLEXIONES EN TORNO A LA NOCIÓN DE CONOCIMIENTOS INNATOS DE LA GRAMÁTICA UNIVERSAL<sup>1</sup>

Manuela Rondón Triana<sup>2</sup>

Rafael Antonio Gutiérrez Martínez<sup>3</sup>

## Resumen

En este artículo, se mostrará que, en el Programa Minimalista (PM), Chomsky renuncia a la noción de conocimiento gramatical, supuesta en Principios y Parámetros (P&P). Este tema parece no haber llamado la atención de los estudiosos de la lingüística chomskyana y esto es interesante porque, en el modelo de P&P, la existencia de conocimientos lingüísticos tiene un rol explicativo fundamental. Inicialmente, se expondrá el modelo de P&P, en el cual se defiende la existencia de un conocimiento innato; posteriormente, se abordará el PM, para mostrar la introducción de un nuevo elemento en la teoría. Finalmente, se sostendrá la hipótesis de que la idea central de Chomsky no era defender una noción de conocimiento innato, sino encontrar la mejor explicación a varios fenómenos lingüísticos.

**Palabras clave:** Conocimientos innatos, *gramática* universal, programa minimalista, teoría de principios y parámetros.

<sup>1</sup> Consideramos necesario hacer dos aclaraciones. Primero, este texto es un avance de nuestra investigación, en el marco del proyecto Fundamentos de la Normatividad, que está adscrito al grupo de investigación Filosofía y Cognición del Departamento de Filosofía de la Universidad Nacional de Colombia. Agradecemos los significativos aportes que los otros integrantes del proyecto han hecho a él. Segundo, esta versión del texto es posterior a la que presentamos en el Foro Ling20 (a cuyos participantes también agradecemos por sus preguntas y comentarios) y, en esa medida, contiene algunas modificaciones que se corresponden con desarrollos posteriores en nuestra investigación.

<sup>2</sup> Filósofa, Estudiante de la Licenciatura en Español y Filología Clásica de la Universidad Nacional de Colombia, sede Bogotá. Correo: mroundont@unal.edu.co

<sup>3</sup> Estudiante de Lingüística de la Universidad Nacional de Colombia, sede Bogotá. Correo: raagutierrezma@unal.edu.co

## Abstract

In this article it will show that in the Minimalist Program (MP) Chomsky renounces the notion of grammatical knowledge that he assumed in the Principles and Parameters model (P&P). This topic does not seem to have attracted the attention of scholars that study Chomskyan linguistics. This is interesting because in the P&P model the existence of linguistic knowledge has a fundamental explanatory role in the theory. First, the P&P model, in which the existence of an innate knowledge is defended, will be presented. Next, the MP will be address to show the introduction of a new element in the theory. Finally, the hypothesis according to which the central point for Chomsky was not to defend a notion of innate knowledge, but to find the best explanation for several linguistic phenomena will be advance.

**Key Words:** Innate knowledge, Universal Grammar, Minimalist Program, Principles and Parameters.

## 1. Introducción

En sus primeras formulaciones de la gramática generativa, Noam Chomsky postulaba que toda teoría legítima del lenguaje debía satisfacer dos condiciones: la adecuación descriptiva y la adecuación explicativa. La primera corresponde a la pregunta ¿cuál es la naturaleza del sistema de las lenguas naturales?, y se responde por medio de la formulación de gramáticas que lo describen, es decir, que muestran las propiedades y rasgos de cada una de las lenguas (e. g., del francés, del español o del ika) (cf. Chomsky 2012, p. 4; 1988, p. 67). La segunda corresponde a la pregunta sobre cómo la lengua se adquiere a lo largo del desarrollo del niño, dada la pobreza y asistematicidad del *input* lingüístico, esto es, al llamado “problema de la pobreza del estímulo” (también conocido como el problema de Platón), y se responde esclareciendo el estado inicial de la facultad de lenguaje. Aquí nos interesa el segundo criterio de satisfacción.

De acuerdo con Chomsky, “[...] en todo dominio de crecimiento, incluida la adquisición de una lengua particular, hay una enorme brecha entre los datos disponibles y el estado que se obtiene” (2012, p. 8). Los estímulos ambientales que recibe un niño cuando está adquiriendo su lengua son escasos, poco informativos y se ofrecen de una manera desor-

ganizada. Por esto, no podrían dar cuenta, por sí solos, de la riqueza del conocimiento que el niño termina teniendo en un corto periodo de tiempo (cf. Lorenzo, 2007, p. 211). A partir de esta premisa, el argumento postula que se debe, entonces, suponer la existencia de un estado inicial que contiene ciertos conocimientos innatos, los cuales permiten explicar la brecha. Dicho en pocas palabras, el argumento de la pobreza del estímulo establece que el conocimiento que necesita ser adquirido para desarrollar la capacidad cognitiva del lenguaje excede ampliamente la información disponible en el entorno, de manera que es necesario suponer que el organismo aporta información innata (cf. Skidelsky, p. 144). Este es el problema en el que se central el modelo de P&P (problemas y parámetros).

## 2. La noción de conocimiento en principios y parámetros

El modelo de P&P, desarrollado por Chomsky en la década de los ochenta, postula que existe una gramática universal, que constituye el estado inicial de la facultad del lenguaje. Esta puede ser entendida como un componente innato de la mente humana, compuesto por un sistema de principios y condiciones, a partir del cual se produce una lengua particular en interacción con la experiencia (cf. Chomsky 1986, p. 3; 1975 (citado en Ramos)). Así pues, desde este modelo, Chomsky defiende la existencia de “principios generales, que son propiedades de la facultad lingüística, a los que se suman ligeras opciones de variación, que son los llamados «parámetros»” (2003, p. 30). En primer lugar, los principios son innatos, determinados biológicamente y, como consecuencia, son comunes a todos los miembros de la especie humana. Además, se manifiestan vigentes en todas las lenguas y en todas las construcciones (2003, p. 30). En segundo lugar, los parámetros son variaciones particulares de cada lengua, determinados de manera sistemática a partir de la experiencia. En este caso, la GU establece los posibles valores que ellos podrían asumir. Veamos un par de ejemplos.

### 2.1 Principio de dependencia de la estructura

De acuerdo con el principio de dependencia de la estructura, “para determinar el sistema gramatical apropiado con las predicciones correctas es

necesario considerar las propiedades jerárquicas de las expresiones lingüísticas, en lugar de la posición de las palabras [...]” (Fortuny, *s/f.*, p. 4). Consideremos uno de los ejemplos clásicos que propone Chomsky a propósito de la formación de oraciones interrogativas. Dada una oración declarativa como (1), para formar la interrogativa (2), es necesario desplazar el primer verbo a la posición inicial. Con base en esto, un niño que está en proceso de adquisición de la lengua podría formular la regla *word-by-word linear rule*, según la cual se debe encontrar el primer verbo de la oración afirmativa y moverlo a la posición inicial para construir la pregunta.

(1) El hombre está comiendo patatas

(2) ¿Está el hombre comiendo patatas?

Si el niño siguiera la regla mencionada en toda construcción de interrogaciones, formaría la oración interrogativa (4) a partir de (3). Esto, sin embargo, nunca ocurre; los niños jamás cometen este error (Chomsky, 2012, p. 10). Por tanto, debe haber un principio (innato), según el cual la interpretación de las oraciones es estructural y no lineal. Se debe, entonces, formar (5) y no (4).

(3) El hombre que está comiendo patatas en el jardín es mi padre

(4) \*¿Está el hombre que comiendo patatas en el jardín es mi padre?

(5) ¿Es el hombre que está comiendo patatas en el jardín mi padre?

El niño advierte que, a pesar de que el primer verbo de (3) sea “está”, no es este el que debe ponerse en primera posición para formar la interrogativa respectiva, dado que hace parte de un constituyente de la clase nominal que cumple la función de sujeto (“el hombre que está comiendo patatas”). El verbo principal es “es” y, por lo tanto, el niño repara en que este es el que debe desplazar. Esto, de acuerdo con una regla que instancia el principio de dependencia de la estructura, a saber, “para formar una oración interrogativa, muévase a una posición inicial el primer auxiliar que aparezca en la oración que suceda al sintagma nominal que actúe como sujeto” (Usón *et al.*, 2012, p. 70).

## 2.2 El parámetro del sujeto nulo

El parámetro del sujeto nulo regula la variación entre las lenguas en cuyas oraciones no es posible omitir el sujeto (como el inglés y el francés) y las lenguas en las que sí es posible hacerlo (como el español y el italiano) (Chomsky, 1988, p. 67; Usón *et al.*, 2012, p. 71), siempre y cuando, por la información pragmática, sea claro a quién se hace referencia. Las opciones de este parámetro, cabe resaltar, se fijan a partir de la experiencia, pero la GU establece que solamente hay dos valores posibles predeterminados. Así, por ejemplo, mientras que en español son posibles las construcciones (6) y (7); en inglés (8) es posible, pero (9) no.

(6) Anastasia quiere comer

(7) Quiere comer

(8) Anastasia wants to eat

(9) \*Wants to eat

Ahora bien, principios y parámetros como estos constituyen el estado inicial, con el cual, en el marco de este modelo, se soluciona el problema de la pobreza del estímulo<sup>4</sup>. En este punto, cabe preguntarse qué tipos de elementos son estos principios y parámetros que hacen parte de nuestra herencia biológica, esto es, de qué clase son aquellas cosas innatas gracias a las cuales podemos adquirir una lengua. De acuerdo con este modelo, son conocimientos. A propósito, dice Chomsky:

[...] el niño se acerca a la tarea de adquirir una lengua con un rico marco conceptual que ya está puesto en lugar y, también, con un rico sistema de suposiciones sobre la estructura del sonido y la estructura de preferencias más complejas. Estas constituyen las partes de nuestro conocimiento que vienen “de la mano original de la naturaleza”, en palabras de Hume (1988, p. 34).

<sup>4</sup> Cabe mencionar que, gracias a la postulación de los principios y los parámetros, también se planteó una respuesta al problema de las similitudes y las diferencias entre todas las lenguas. Los principios, que, como dijimos, están presentes de manera invariable en todas las lenguas, dan cuenta de sus semejanzas. Los parámetros, por su parte, permiten un número finito de permutaciones que dan cuenta de las diferencias (Chomsky, 2016, p. 15).

Así pues, el estado inicial ya está compuesto por una serie de suposiciones y por una parte del conocimiento lingüístico necesario para que emerja nuestra lengua. En términos de Fodor, Chomsky propone la existencia de “[...] verdades (sobre la estructura de posibles primeras lenguas) que los seres humanos comprenden de manera innata y [...] contenidos proposicionales innatamente especificados” (Fodor, p. 7). Así, “es la integración de este conocimiento innato con un corpus de «datos lingüísticos primarios» (e. g. con las observaciones del niño de las preferencias de los miembros adultos de su comunidad de habla) lo que explica la eventual asimilación de las capacidades lingüísticas maduras” (Fodor, p. 4).

Cabe señalar que este conocimiento ha sido objeto de crítica de varios filósofos. Esta crítica ha llevado a Chomsky a refinar su idea de conocimiento y a plantear un neologismo: *cognocer* (que hace referencia a un *saber qué* tácito y no justificado). Con todo, no es nuestro propósito en este texto entrar en los pormenores de la discusión. Lo importante por notar aquí es que la respuesta al problema de Platón, pregunta fundacional de la gramática generativa, está ligada a la suposición de cierto tipo de conocimiento.

### 3. El programa minimalista

En los últimos veinte años, Chomsky ha propuesto un nuevo enfoque al programa de investigación empírica de la facultad del lenguaje, el PM, que busca responder una pregunta que va “más allá de la adecuación explicativa” (Chomsky 2003, 2012). Esta es ¿por qué la facultad del lenguaje tiene precisamente las propiedades que tiene? Desde este momento, la lingüística chomskyana ya no solamente se propone dos criterios de satisfacción (adecuación descriptiva y adecuación explicativa), sino tres.

El PM plantea que los principios que explican por qué el lenguaje tiene las propiedades que tiene son los que pertenecen al tercer factor (Eguren, 2014, p. 40). En él, se incluyen condiciones no específicas de la facultad del lenguaje, consideradas leyes de la biología o de la naturaleza, que están presentes en varios procesos de los organismos vivos. Estas son de dos tipos: principios de análisis de datos (que se utilizan tanto para analizar datos lingüísticos como, por ejemplo, datos visuales), por un lado, y principios de arquitectura estructural y restricciones sobre el desarrollo, por otro lado (Chomsky 2005 6; Eguren, 2014, pp. 40-41). Los segundos, que, según Chomsky, son los realmente relevantes a la hora de determinar la naturaleza de las lenguas, incluyen condiciones impuestas por los siste-

mas limítrofes (a saber, el sensoriomotor y el intencional-conceptual)<sup>5</sup> y diferentes principios de economía o eficiencia computacional (por ejemplo, la prohibición de que haya elementos superfluos en las representaciones o pasos inmotivados en las derivaciones) (Chomsky, 2005, p. 6; Lorenzo, 2007, p. 220).

Ahora bien, el punto central para los propósitos de nuestro trabajo tiene que ver con cómo la inclusión de este nuevo elemento trae consigo cambios significativos en la respuesta que se había dado al problema de la pobreza del estímulo. Como mencionamos en la sección anterior, este problema, en el modelo de P&P, se solucionaba apelando a un estado inicial compuesto exclusivamente por la GU (primer factor de la adquisición de la lengua), que, junto con la experiencia (segundo factor de la adquisición de la lengua), determinaba el conocimiento lingüístico al que se llegaba. En contraste, el PM plantea que el estado inicial con el que soluciona el problema también incluye elementos del tercer factor, que en ningún caso son conocimientos, los cuales exigen un empobrecimiento de la GU. Como consecuencia, esta no es sino un componente más de dicho estado inicial<sup>6</sup>.

¿A qué se debe dicho empobrecimiento? En la medida en la que se pretende dar cuenta de las características del lenguaje acudiendo a principios del tercer factor, el PM termina por extraer de la GU todas aquellas propiedades que se pueden explicar por medio de estos (Eguren, 2014, pp. 42, 43). En otras palabras, principios que antes se consideraban parte de la GU, o sea, conocimientos, se comprenden ahora como resultados de restricciones del tercer factor, una vez se asume el proyecto de dar cuenta de las propiedades del lenguaje a partir de leyes naturales.

Esto pone en evidencia que el PM explica los mismos fenómenos que P&P explicaba acudiendo a conocimientos por medio de los elementos del tercer factor. Un ejemplo de esto es lo que ocurre con el parámetro de posición de núcleo, que, en la teoría de P&P, hacía parte de la GU. Este pa-

<sup>5</sup> Para entender mejor el rol de los sistemas limítrofes hay que considerar que, de acuerdo con Chomsky, el uso y conocimiento del lenguaje depende de tres sistemas: el sistema cognitivo lingüístico (que almacena la información de la lengua que uno conoce) y dos sistemas de actuación que acceden a esa información: uno tiene que ver con la exteriorización (el sensoriomotor) y otro con el significado que se les asigna a las expresiones lingüísticas (el conceptual-intencional) (2003, pp. 18-19). Según él, el primero claramente hace parte de la facultad del lenguaje; los otros dos, se supone, también, aunque no son exclusivos de ella.

<sup>6</sup> Anteriormente, habíamos mencionado que, con base en el hecho de que los niños adquieren su lengua con relativa rapidez y sin estímulos suficientes, se planteaba la existencia de una GU rica y compleja, es decir, sobredimensionada. Ahora vemos que esta premisa no era suficiente para sostener esta GU, puesto que solamente nos obliga a postular que hay un estado inicial que provee elementos ausentes de la experiencia, pero este estado inicial no debe ser igual a la GU.

rámetro divide las lenguas entre aquellas en las que los núcleos se sitúan delante de sus complementos, como el español —e. g. (10) y (11)—, y las lenguas en las que se sitúan detrás, como el vasco —e. g. (12) y (13)— (Eguren, 2014, p. 44).

(10) Lazó la piedra (verbo + frase nominal)

(11) Sin cabeza (preposición + frase nominal)

(12) Harria bota zuen (frase nominal + verbo)  
piedra.la lanzar aux.

(13) Bururik gabe (frase nominal + preposición)  
cabeza.part sin

En este modelo las dos opciones posibles del parámetro están determinadas por la GU. En el PM, en cambio, que existan estas dos posibilidades no está preestablecido por ella, sino por una propiedad física del tercer factor:

[...] la naturaleza de nuestro aparato articulatorio hace que el núcleo de un sintagma y su complemento, como cualquier otra unidad lingüística, se deban pronunciar uno tras otro, dando lugar a tan solo dos opciones (Holmberg, 2010). El constatado hecho de que los núcleos tienden a preceder o seguir a sus complementos de manera uniforme en una lengua determinada sería, a su vez, consecuencia de la aplicación, en el proceso de adquisición de las lenguas maternas, de mecanismos generales de aprendizaje no exclusivos del lenguaje, unos mecanismos pertenecientes de nuevo, por tanto, al tercer factor, como la “generalización del input” (Roberts y Holmberg, 2010) o el “sesgo del superconjunto” (Boeckx, 2011), que favorecen la consolidación de sistemas armónicos u homogéneos (Eguren, 2014, p. 44).

Así, propiedades como la posición del núcleo, que antes se explicaban a partir de principios y parámetros (o sea, conocimientos), ahora están determinadas por elementos del tercer factor; en particular, por una condición impuesta por el sistema de exteriorización. Teniendo en cuenta lo anterior, tiene sentido la creencia de Chomsky de acuerdo con la cual los elementos del tercer factor (postulados para responder la tercera pregunta) hacen parte del problema de la adquisición de la lengua materna: de

cierta manera, parece que sobreviven los mismos principios, aunque ya no afirmemos que hacen parte de la GU o que sean conocimientos, sino que dependen de leyes naturales no exclusivas del lenguaje.

#### 4. Conclusión

Ahora bien, ¿qué nos puede decir sobre la teoría de Chomsky la renuncia a la idea de conocimientos innatos? Para Chomsky no era esencial que se tratara de conocimientos, sino de algo que fuera un buen *explanans*. Más allá de que la renuncia a la postulación de conocimientos innatos contribuya a la integración de la teoría lingüística y la biología (en tanto hace que sea más sencillo dar cuenta de la aparición del lenguaje en la evolución del ser humano) o de que implique una simplificación de la teoría (que, según Chomsky, corresponde a la manera estándar de hacer ciencia), es posible afirmar que ella nos sirve para notar (o, más bien, recordar) un hecho importante: Chomsky, considerado como la figura central que revivió la tesis innatista en el siglo XX, no tenía un objetivo epistemológico. Pretendía, más bien, dar cuenta de la facultad del lenguaje, lo cual implica explicar cómo adquirimos la lengua a pesar de no contar con estímulo suficiente. El objetivo de Chomsky, así, no era la postulación de ideas filosóficas, sino dar cuenta de fenómenos lingüísticos de alta complejidad.

#### Referencias

- Chomsky, N. (1988). *Language and Problems of Knowledge. The Managua Lectures*. Massachusetts: The MIT Press.
- Chomsky, N. (2003). *La arquitectura del lenguaje*. Barcelona: Editorial Kairós.
- Chomsky, N. (2005). “Three Factors in Language Design”. *Linguistic Inquiry*. 36(1), 1-22.
- Chomsky, N. (2012). “Poverty of Stimulus: Unfinished Business”. *Studies in Chinese Linguistics*. 33(1), 3-16.
- Chomsky, N. (2016). *¿Por qué solo nosotros? Evolución y lenguaje*. Barcelona: Editorial Kairós.
- Eguren, L. (2014). “La Gramática Universal en el Programa Minimalista”. *Revista de Lingüística Teórica y Aplicada*. 52(1), 35-58.

- Fodor, J.** (1983). *The Modularity of the Mind*. Cambridge: The MIT Press.
- Fortuny, J.** “Structure Dependence and Linear Order: Clarifications and Foundation”. 1-24. En: [ling.auf.net/lingbuzz/002227/v3.pdf](http://ling.auf.net/lingbuzz/002227/v3.pdf)
- Lorenzo González, G.** (2007). “El Programa Minimalista y el Problema de Platón. Reflexiones sobre la aportación del minimalismo a los desafíos del desarrollo lingüístico”. *ELUA* 21(1), 209-228.
- Ramos, J.** (1992). Reflexiones sobre el innatismo. *Ideas y Valores*, (89), 15-48.
- Skidelsku, L.** (2016). The poverty of the Stimulus Argument Once Again. *Análisis filosófico*, 143-170.
- Usón, R. M., Cervel, S. P., Rodríguez, F. J. C., y de Mendoza Ibáñez, F. J. R.** (2012). *Teoría lingüística: Métodos, herramientas y paradigmas*. Madrid: Editorial Universitaria Ramón Areces.

## NORMAS DE PUBLICACIÓN

El grupo estudiantil *Ema* invita a toda la comunidad académica a participar en la convocatoria de la revista **PROLEGÓMENOS – Revista Estudiantil de Lingüística**. Los interesados pueden enviar sus escritos al correo electrónico [revprol\\_fchbog@unal.edu.co](mailto:revprol_fchbog@unal.edu.co), teniendo en cuenta los siguientes criterios:

### Criterios de contenido

- Abordar temáticas enmarcadas en la lingüística y sus vertientes interdisciplinarias.
- Pueden ser escritos desde diferentes disciplinas que tengan un enfoque lingüístico.
- Estar dentro de los siguientes tipos de escritos: ensayos, reseñas, artículos científicos, artículos de revisión, textos de discusión, crónicas de trabajo de campo.
- Estar redactados de manera coherente, utilizar un léxico propio del registro escrito y respetar las normas ortográficas establecidas por la Real Academia Española.
- Incluir los datos del autor (nombre, carrera, institución y correo electrónico).
- Los artículos podrán ser presentados en inglés o español.

## Crterios gráficos

- La extensión de acuerdo con el texto presentado es: artículos científicos y de revisión 5000 palabras, ensayos 3000 palabras, crónica 2000 palabras, reseñas 2000 palabras y discusión 1500 palabras (dentro del número de palabras se excluyen las referencias).
- Archivo en formato Word.
- Texto escrito a una sola columna.
- Tipo de fuente Arial 12 e interlineado de 1,5.
- Márgenes de 3cm a cada lado.
- Las tablas no pueden ser más grandes que media página y juntas no pueden sobrepasar las tres páginas.
- Seguir los criterios de citación y referencia bibliográfica establecidos por la APA.
- Si el artículo postulado es el resultado de alguna investigación o proyecto, indicar los datos necesarios de reconocimiento.
- Si se presentan agradecimientos, se deben presentar por medio de una nota al pie de página.
- Si el texto incluye imágenes, estas deben enviarse como adjuntos en el correo y deberán tener una calidad de mínimo 300dpi.
- Si el texto incluye tablas es preferible que estén elaboradas en el documento como tal y no insertadas como imagen. Si son imágenes, deberán cumplir con los criterios establecidos en el anterior ítem.

## Crterios de seleccón

- Pertinencia de la información.
- Relevancia.
- Veracidad/verosimilitud.
- Coherencia y cohesión.

**Nota 1:** Todos los textos deben presentarse con resumen, *abstract* (máximo 120 palabras), palabras clave, *key words* (máximo cinco) y referencias.

**Nota 2:** La revista *Prolegómenos* no excluye ninguna línea de investigación.

**Nota 3:** Los textos preseleccionados se enviarán a una evaluación de pares académicos. Se rechazarán aquellos que no cumplan con los requisitos o que realicen plagio.

**Nota 4:** Hacer plagio acarreará las sanciones especificadas por la Universidad Nacional de Colombia para tal caso.



La revista **PROLEGÓMENOS** se diagramó en la ciudad de Bogotá, Colombia, en el mes de octubre de 2020, 82 años después de la transmisión radial de "La guerra de los mundos" por Orson Welles.

Las fuentes tipográficas usadas fueron

DIN Pro  
Times LT Std  
Times New Roman



