

E S P

A Ñ

**Preseca-Medellín:
¿Cómo se habla el
español en la
capital
antioqueña?**

O ~ I

La forma como los habitantes de Medellín hablan el español puede estudiarse gratis y desde cualquier parte del mundo gracias al Corpus Preseea-Medellín, un compendio de entrevistas y estudios sobre el tema, y a las plataformas informáticas que facilitan y apoyan la consulta.

Por:

Jorge Mauricio Molina Mejía

Doctor en Informática y Ciencias del Lenguaje, coordinador del Grupo de Estudios Sociolingüísticos y del Semillero Corpus Ex Machina.

María Claudia González Rátiva

Doctora en Lingüística, coordinadora del Laboratorio de Fonética y Filología, profesora e investigadora del Grupo de Estudios Sociolingüísticos.

*Facultad de Comunicaciones,
Universidad de Antioquia.*

Caminar por las calles de la ciudad de Medellín puede ser una experiencia lingüística muy interesante. En la ciudad, la mayoría de las personas hablan el característico dialecto “paisa”, pero con muchas variantes según el barrio, la edad de los hablantes, el nivel educativo y otros factores. Así, es curioso notar cuantas veces se escuchan expresiones como “pues” o “avemaría”, y también construcciones como “hubieron tres personas”.

Estas singularidades pueden ser estudiadas hoy de una manera muy

práctica lo que permite mayor rigor a las investigaciones en sociolingüística, área que se encarga del estudio del lenguaje humano en diferentes contextos

sociales. En esa área, desde el Grupo de Estudios Sociolingüísticos –GES–, adscrito a la facultad de Comunicaciones de la Universidad de Antioquia, hemos realizado durante años investigaciones y herramientas. Y hoy, la información de varios de esos estudios está disponible gracias a la tecnología.

Se trata del proyecto Corpus Lingüístico de Medellín, que cuenta con una plataforma informática especializada que permitirá que investigadores y especialistas en la lengua española tengan acceso desde cualquier parte del mundo.

Con ella, los investigadores podrán realizar estudios de diferentes aspectos del español hablado en Medellín, tales como la fonética (estudio de los sonidos del habla), la morfología (la estructura de las palabras y de sus elementos constitutivos), la sintaxis

(modo en que se combinan las palabras para formar oraciones o enunciados), la fraseología (estudio de las frases, los refranes, proverbios, etc.) y la pragmática (estudio del

lenguaje en relación con los hablantes de un determinado entorno).

En Medellín, la mayoría de las personas hablan el característico dialecto “paisa”, pero con muchas variantes según el barrio, la edad de los hablantes, el nivel educativo y otros factores.

¿En qué consiste el Corpus Preseea-Medellín?

El Proyecto para el Estudio Sociolingüístico del Español de España y América, PRESEEA, busca facilitar el estudio del lenguaje oral en diferentes ciudades de Hispanoamérica, tanto en el ámbito social como en el lingüístico. Es liderado por el profesor Francisco Moreno Fernández de la Universidad Alcalá de Henares en España.

Has hasta el momento, 43 equipos de trabajo se han sumado al desarrollo de investigaciones en España y en América. Siete de estos equipos se encuentran ubicados en ciudades de Colombia: Bogotá, Tunja, Cartagena de Indias, Barranquilla, Pereira, Valledupar y Medellín.

La fase del proyecto dedicada al español de Medellín se inició en el año 2006; desde entonces Preseea-Medellín ha venido mejorándose cada vez con distintos proyectos financiados por el Comité para el Desarrollo de la Investigación de la Universidad de Antioquia —CODI—.

Para construir el Corpus Preseea-Medellín central, integrantes del GES entrevistaron a 119 personas que llevaran mínimo 10 años viviendo en la ciudad de Medellín, seleccionados para representar cuatro grandes variables sociales: sexo, edad, clase social y nivel educativo. Con ellos hablaron sobre temas cotidianos, grabaron las entrevistas y las categorizaron para mirar en ellas fenómenos del habla.



Fotografía: Sergio Urquijo Morales, revista Experimenta

La enorme cantidad de información lingüística, organizada en el Corpus Preseea-Medellín, se consolidó en una plataforma web que se puede visitar en el sitio <http://comunicaciones.udea.edu.co/corpuslinguistico/>. La plataforma permite utilizar dos herramientas: la primera para acceder a los documentos en audio y a las transcripciones de las entrevistas; la segunda herramienta, de uso morfológico y sintáctico, para hacer consultas en el corpus y obtener palabras o grupos de palabras.

¿Qué se puede estudiar a partir de la herramienta general?

Cualquier persona interesada en analizar los audios para realizar estudios tanto fonéticos como fonológicos (es decir, que permiten analizar los diferentes sonidos del lenguaje humano) puede descargar los audios y utilizar luego otras herramientas que les permitan estudiar los diferentes sonidos (como *Praat*, un programa informático que permite recortar, segmentar y analizar los diferentes sonidos del habla humana). Así pueden estudiar como los medellinenses pronuncian la “s”, o determinada vocal, y si lo hacen de forma diferente a otras regiones del país y de Hispanoamérica.

Los investigadores también pueden descargar las transcripciones de cada una de las entrevistas y cotejarlas con los respectivos audios, o realizar cualquier tipo de trabajo a partir de estas transcripciones. Por ejemplo, pueden analizar las funciones semánticas y pragmáticas del “pues” o de la expresión “Eh, Ave María” aparecen en el corpus; u observar las veces que alguien emplea el verbo “haber” en plural, como por ejemplo en: “hubieron muchas personas” o “habían muchos huecos en las calles”, etc. A este tipo de trabajo se le llama análisis de concordancias o de ocurrencias de una palabra o de un grupo de palabras.

Este tipo de análisis nos permite verificar la forma en la que determinada parte de la población habla según su nivel socioeconómico o educativo, el sexo o la edad. Para ello, al igual que en la etapa anterior, se pueden emplear otras herramientas que nos ayudan a contar estas palabras (por ejemplo, *Antconc*, que es un programa informático para hacer conteo de palabras en un gran corpus de textos).

¿Qué nos permite estudiar la herramienta de análisis morfológico y sintáctico?

Una segunda plataforma (<http://grupotnt.udea.edu.co/Preseca/index.php>) permite realizar análisis mucho más finos en el corpus del habla de los habitantes de Medellín, desde dos aspectos lingüísticos: el morfológico y el sintáctico.

Cuando nos referimos a la morfología, estamos hablando de la forma de las palabras y de su estructura. Es así como la plataforma permite hacer búsquedas de palabras su lema —forma en el que la palabra aparece en el diccionario—, o su categoría gramatical (si se trata de un verbo, un sustantivo, un adjetivo, un adverbio, una preposición, etc.). Si queremos buscar todas las posibles conjugaciones del verbo “haber”, lo buscaremos por su lema, y así obtendremos: “han”, “habían”, “había”, “hubimos”, “hubieron”, “hubo”, etc.

El segundo aspecto que permite la plataforma es la búsqueda de formas sintácticas, que es cómo las palabras se organizan en enunciados y oraciones. Podemos buscar un lema determinado, como lo es “estar”; luego podemos buscar la forma “lloviendo” y esto puede estar seguido de la categoría adverbio. El sistema recuperará de Corpus segmentos como “está lloviendo allá”, “estuvo lloviendo acá”, “está lloviendo mucho”, etc. Esto nos permite, además, hacer búsquedas de tipo fraseológico (proverbios, frases, etc.), con lo que se pueden analizar expresiones como: “Eh, Ave María pues”, “Quiubo hombre”, etc.; o refranes como: “No hay mal que por bien no venga” o “En boca cerrada no entra mosca”.

Para realizar todas estas búsquedas, el semillero de investigación en lingüística computacional y de corpus *Corpus Ex Machina*, adscrito al GES, se ha encargado de etiquetar y de anotar de manera morfosintáctica todos los documentos que contienen las 119 transcripciones de las entrevistas. Gracias a ello, disponemos de todas las posibles formas, todos los lemas de las palabras y de sus respectivas categorías gramaticales. Este trabajo posibilita que contemos con una herramienta de gran valor para los lingüistas y los sociolingüistas, para realizar estos análisis lingüísticos de una manera bastante fácil, ya que la plataforma es bastante intuitiva para su utilización.

Como vemos, nos encontramos frente a un grupo de herramientas que facilita a cualquier investigador del mundo estudiar la forma en la que los habitantes de la ciudad de la Eterna Primavera se expresan en la lengua de Cervantes. Es muy importante que estos recursos se exploten y trabajen en las diferentes instituciones, tanto de educación superior como en el bachillerato y en la escuela.

Además, este tipo de herramientas puede convertirse en un insumo importante para la enseñanza y el aprendizaje de la lengua española, como lengua materna o como de lengua extranjera. Para las personas que estudian el Español Lengua Extranjera (ELE), es muy interesante poder acceder a audios y a transcripciones de una variedad como la de la ciudad de Medellín. ✖

Este artículo es resultado parcial de la investigación Caracterización lingüística y sociolingüística del español hablado en Medellín, financiada por el Comité para el Desarrollo de la Investigación —CODI— de la Universidad de Antioquia.

Este es un grupo de herramientas que facilita a cualquier investigador del mundo estudiar la forma en la que los habitantes de la ciudad de la Eterna Primavera se expresan en la lengua de Cervantes.