Fecha de presentación del Informe: Día 6 Mes 12 Año 2024

1. Datos generales del Proyecto

Código del proyecto: 4379			
Título del proyecto: Estudio sociolingüístico del habla de Cali. Construcción del corpus y análisis preliminares.			
Facultad o Instituto Académico: Facultad de Humanidades			
Departamento o Escuela: Departamento de Lingüística y Filología			
Grupo (s) de investigación:			
Entidades:			
Palabras claves: Sociolingüística, corpus, variación sociolingüística			
Investigadores [1]	Nombre	Tiempo asignado	Tiempo dedicado
Investigador Principal	Róbinson Grajales Alzate	220	220
Coinvestigadores	Margareth Marmolejo Caicedo	0	220
Otros participantes			

2. Resumen ejecutivo:

Santiago de Cali constituye un dialecto urbano del español colombiano que no ha sido estudiado ampliamente ni de manera sistemática en términos sociolingüísticos. De ahí la necesidad de construir un corpus sociolingüístico para la descripción y el análisis de fenómenos de variación lingüística en relación con factores lingüísticos y sociales. En el proyecto del Estudio Sociolingüístico del español de Cali se describieron y analizaron

[1] Todas las personas relacionadas en el informe y que participen en el proyecto deben haber suscito el acta de propiedad intelectual de acuerdo con los formatos establecidos.

fenómenos lingüísticos y sociolingüísticos presentes en el habla de Cali a partir de la creación y sistematización de un corpus sociolingüístico. Este proyecto se desarrolló en el marco de la sociolingüística variacionista y la lingüística de corpus. Para ello se estableció la muestra representativa de hablantes, se diseñó una entrevista sociolingüística semiestructurada que se aplicó y se recogió en grabaciones de audio. Lo anterior se hizo teniendo en cuenta la metodología de estudios internacionales como PRESEEA, lo que permite garantizar la compatibilidad con otros estudios sociolingüísticos variacioncitas y así poder hacer estudios comparativos en el futuro. El principal resultado fue la creación del corpus PRESEEA-Cali que consta de 72 entrevistas sociolingüísticas de igual número de informantes preestratificados según las variables de edad, género y nivel educativo. Este corpus se encuentra disponible en el sitio web del proyecto PRESEEA: https://preseea.uah.es/corpus-preseea. A partir de este corpus se hicieron unos primeros análisis para caracterizar el habla de Cali. Los fenómenos estudiados fueron las formas de tratamiento pronominal, el marcador discursivo *pues* y la expresión de sujetos pronominales. Con respecto a las formas de tratamiento pronominal se encontró que en Cali prevalece el uso de *usted* frente a *tú* y *vos*, lo cual indica una tendencia a la cortesía de distanciamiento; no obstante, la cortesía de solidaridad es impulsada por las generaciones jóvenes, los niveles educativos altos y los estratos socioeconómicos medios; la variable sexo es significativa en la elección de forma de tratamiento según el sexo del interlocutor y en las relaciones no familiares. En relación con el marcador discursivo *pues* los resultados revelan que pues es un marcador polifuncional que está presente en todos los tipos de discurso, posee dos funciones pragmáticas principales, un total de diez subfunciones, y de estas, las más frecuentes son la de retrasar el discurso, la comentadora y la continuativa. Con respecto a la expresión de sujetos pronominales se encontró una tasa pronominal de 28,6%, la más alta de comunidades continentales andinas, un condicionamiento de la persona y el número gramaticales y un fuerte efecto de las variables de edad y género en las que los hablantes más jóvenes y las mujeres promueven los sujetos expresos. Estos resultados no solo contribuyen a la descripción del habla de Cali sino además al conocimiento de la sociolingüística hispánica.

Santiago de Cali is an urban dialect of Colombian Spanish that has not been widely or systematically studied in sociolinguistic terms. Hence the need to build a sociolinguistic corpus for the description and analysis of linguistic variation phenomena in relation to linguistic and social factors. In the Sociolinguistic Study of Cali Spanish project, linguistic and sociolinguistic phenomena present in the speech of Cali were described and analyzed based on the creation and systematization of a sociolinguistic corpus. This project was developed within the framework of variationist sociolinguistics and corpus linguistics. To this end, a representative sample of speakers was established, a semi-structured sociolinguistic interview was designed, applied and recorded in audio recordings. This was done taking into account the methodology of international studies such as PRESEEA,

which allows for ensuring compatibility with other variationist sociolinguistic studies and thus being able to carry out comparative studies in the future. The main result was the creation of the PRESEEA-Cali corpus, which consists of 72 sociolinguistic interviews with the same number of informants pre-stratified according to the variables of age, gender and educational level. This corpus is available on the PRESEEA project website: https://preseea.uah.es/corpus-preseea. From this corpus, initial analyses were carried out to characterise the speech of Cali. The phenomena studied were the forms of pronominal address, the discourse marker pues and the expression of pronominal subjects. Regarding the forms of pronominal address, it was found that in Cali the use of usted prevails over tú and vos, which indicates a tendency towards politeness of distancing; however, politeness of solidarity is promoted by younger generations, high educational levels and middle socioeconomic strata; the variable sex is significant in the choice of form of address according to the sex of the interlocutor and in non-family relationships. Regarding the discourse marker, the results reveal that it is a polyfunctional marker that is present in all types of discourse, it has two main pragmatic functions, a total of ten subfunctions, and of these, the most frequent are the delaying of the discourse, the commentator and the continuative. Regarding the expression of pronominal subjects, a pronominal rate of 28.6% was found, the highest of Andean continental communities, a conditioning of the grammatical person and number and a strong effect of the age and gender variables in which younger speakers and women promote the expressed subjects. These results not only contribute to the description of the speech of Cali but also to the knowledge of Hispanic sociolinguistics.

3. Síntesis del proyecto:

Tema

Sociolingüística variacionista y lingüística descriptiva.

Objetivo general

Construir un corpus sociolingüístico para la descripción y el análisis de fenómenos de variación lingüística en relación con factores lingüísticos y sociales.

Objetivos específicos

1. Definir la ciudad de Cali como comunidad de habla.
2. Establecer la muestra representativa de hablantes de Cali.
3. Diseñar una entrevista semiestructurada que permita elicitar diferentes tipos de discurso.
4. Aplicar y grabar la entrevista semiestructurada para la muestra representativa con el fin de construir un corpus oral sociolingüístico de Cali.

5. Transcribir las muestras de habla oral recolectadas a través de las entrevistas.

6. Sistematizar el corpus de entrevistas con un etiquetado mínimo y alojarlo en un servidor que permita consultarlo y procesarlo.

7. Analizar por lo menos un fenómeno de variación lingüística para iniciar, de manera preliminar, la caracterización del habla de Cali.

Metodología

Esta investigación fue de carácter descriptivo y se enmarcó en la sociolingüística variacionista, haciendo uso de las metodologías construidas dentro de este campo de estudio. Se partió del supuesto de que Cali es un centro urbano que constituye una comunidad de habla estratificada socialmente y con variantes lingüísticas cuyo uso está distribuido según los grupos sociales que la componen.

Para explicitar esto en términos sociolingüísticos y recopilar materiales que permitieran iniciar una caracterización amplia del habla de Cali, se siguieron los siguientes pasos:

- Se recopiló información histórica y demográfica de Cali, y se amplió la búsqueda de investigaciones sobre esta variedad de habla, para definirla como un centro urbano que constituía una comunidad de habla. Esto se hizo por medio de una investigación documental a partir de fuentes académicas y gubernamentales.
- A partir de esta información, junto con los referentes teóricos y metodológicos, se estableció el tipo de muestreo para la construcción del corpus, teniendo en cuenta modelos cuantitativos y cualitativos para determinar la representatividad de la muestra.
- Una vez se definió lo anterior, se diseñaron los instrumentos para la recolección de los datos lingüísticos, tomando como referencia las principales técnicas usadas para recoger información en otros corpus e investigaciones. En primer lugar, se diseñó una entrevista semiestructurada con criterios sociolingüísticos que permitiera recoger la actuación lingüística de los informantes, que contuviera diferentes ejes temáticos para elicitar aspectos lingüísticos (fonético-fonológicos, morfosintácticos, semántico-pragmáticos) y diferentes tipos de discurso (descriptivo, explicativo, argumentativo, narrativo). En caso de identificar fenómenos de variación que no pudieran ser recogidos a través de la entrevista, se diseñaron cuestionarios que permitieran observarlos, como en el caso de las formas de tratamiento pronominales.
- Después de esto se procedió a la recolección de las entrevistas por cada uno de los informantes determinados para la muestra. Este paso incluyó una prueba piloto del instrumento para optimizar su aplicación, así como una capacitación a los entrevistadores para que aprovecharan de la mejor manera el procedimiento según los propósitos y aseguraran una buena calidad del material. Las entrevistas se recogieron con equipos de grabación de alta calidad para garantizar su

legibilidad, almacenamiento y la posibilidad de hacer estudios fonéticos con el corpus.

- Las entrevistas obtenidas fueron transcritas y revisadas para luego proceder a la organización del corpus con los audios de las entrevistas y su correspondiente transcripción. Estas transcripciones fueron revisadas por un investigador diferente al entrevistador para garantizar su idoneidad.
- Este material fue codificado y anotado (inicialmente con unas etiquetas mínimas) para sistematizarlo según las variables pertinentes para hacer búsquedas en el sitio web donde se alojó el corpus. Después de hacer las pruebas necesarias para garantizar el buen funcionamiento del corpus e identificar errores en su sistematización, se puso a disposición de otros investigadores interesados en estudiar el habla de Cali y se puede consultar en el siguiente enlace: https://preseea.uah.es/corpus-preseea
- Con el corpus sistematizado, se identificaron tres fenómenos lingüísticos que, según estudios dialectológicos y sociolingüísticos anteriores, tenían una alta significación sociolingüística en términos variacionistas: las formas de tratamiento pronominales, los marcadores del discurso y la expresión de los sujetos pronominales. Para cada uno de ellos se estableció la distribución sociolingüística y las condiciones de variabilidad de sus variantes. En primer lugar, se hicieron una descripciones y análisis cualitativos de los fenómenos estudiados para establecer sus características formales y funcionales. Luego, se hicieron los análisis cuantitativos, a través de estadística descriptiva e inferencial, para establecer la significatividad, la correlación y la probabilidad de las variables independientes con respecto a la variable dependiente. Para este propósito se usaron pruebas no paramétricas utilizando el programa *R* en diferentes versiones.

Resultados obtenidos

Como se puede apreciar a partir de los objetivos del proyecto, el resultado principal es la construcción del corpus PRESEEA-Cali, y el segundo resultado importante son los análisis iniciales para la descripción sociolingüística del habla de Cali. Con respecto al corpus, se construyeron documentos de trabajo y se diseñaron materiales para la recolección y la sistematización de los datos, proceso que se describe a continuación.

El corpus sociolingüístico de Cali es un corpus oral estratificado socialmente, constituido por 72 archivos sonoros correspondiente a entrevistas sociolingüísticas semi-dirigidas. El objetivo de construir este corpus fue contar con materiales auténticos representativos del habla caleña que permitieran diferentes análisis lingüísticos y sociolingüísticos para caracterizar dicha variedad de habla. Este corpus cuenta además con las transliteraciones de las entrevistas en archivos en formato .txt y .docx. Estos archivos poseen un marcado .xml, que da información sobre las características del archivo, como

tiempo de grabación, lugar de grabación, responsable de la grabación, información sobre el entrevistado, entre otros. También poseen el etiquetado inicial propuesto por PRESEEA.

Para la elaboración del corpus, en primer lugar, se definió a Cali como comunidad de habla. Si bien la definición de este concepto es compleja y no hay consenso en el mundo sociolingüístico, se decidió tener en cuenta dos definiciones. Cali puede considerarse como comunidad de habla, ya que cumple con las características recogidas por Johnson (2013):

1. Una población que comparte y se comunica a través de la misma lengua,
2. que se identifica como comunidad y
3. que a su vez se subdivide en otras comunidades dependiendo de factores sociales y contextuales.

Labov (1972: 120-1) por su parte define comunidad de habla como “participation in a set of shared norms... [which] may be observed in overt types of evaluative behavior, and [in] the uniformity of abstract patterns of variation”. Gracias a estudios como los de Montes (2000), Vallés (2003) y Ramírez y Almira (2016), se puede afirmar que Cali comparte un conjunto de normas lingüísticas, un comportamiento evaluativo diferente de hablantes de otras regiones y una “uniformidad de modelos abstractos de variación que son invariables respecto a particulares niveles de uso” (Silva-Corvalán, 2001). Atendiendo las directrices de PRESEEA, Cali es un núcleo urbano hispanohablante, con una población tradicionalmente asentada en el lugar y que presenta cierta heterogeneidad sociológica.

Luego de resolver este asunto conceptual, se procedió a hacer la caracterización social de la población de la ciudad con base en el Censo de 2018. Para 2018, Cali contaba con un total de 2’275.380 habitantes, de lo que 1’079.530 eran hombres y 1’195.850 eran mujeres. Del total de habitantes, 1’475.791 tenían más de 20 años. Esta información es necesaria para poder establecer la población representativa para el corpus. Así pues, según PRESEEA (2003), para ciudades con alrededor de dos millones de habitantes, una muestra de 72 hablantes se considera representativa dada la proporción de hablantes de 1/25.000.

Con el fin de facilitar la comparación estadística desde el punto de vista sociológico, lingüístico y estilístico, decidimos acogernos a la recolección de los datos por cuotas de afijación uniforme. Este método consiste en dividir el universo que conforma la muestra en subpoblaciones o cuotas que representen uniformemente las variables sociales establecidas en el estudio. Como es de esperarse, dada la naturaleza sociolingüística del estudio, las variables sociales a tenerse en cuenta fueron la edad, el sexo y el nivel educativo.

La muestra se dividió en 36 hombres y 36 mujeres. La edad se estratificó en tres niveles, hablantes entre 20 y 34 años, 35 y 54 años y mayores de 55 años. También se

reconocieron tres niveles de educación, bajo (entre 0 y 9 años de instrucción), medio (entre 10 y 14 años de instrucción) y alto (más de 15 años de instrucción). Para PRESEEA, el grado de instrucción empieza con personas analfabetas o con máximo 5 años de escolarización. Para el momento de redacción del proyecto, este tipo de hablantes era tal vez aún abundante. No obstante, de acuerdo con datos del DANE, la tasa de alfabetización en Colombia para 2018 era del 94 % en adultos, por lo que encontrar personas, especialmente en una capital como Cali, con pocos estudios o analfabetas es cada vez menos probable.

La Tabla muestra la distribución uniforme de los informantes.

	Hombres			**Mujeres**		
	Bajo	Medio	Alto	Bajo	Medio	Alto
20-34	4	4	4	4	4	4
35-54	4	4	4	4	4	4
<55	4	4	4	4	4	4

De esta manera, queda definida la comunidad de habla, la muestra representativa que se tuvo como base para los análisis, el instrumento de recolección de datos y la base de datos con las 72 entrevistas alojada en el sitio web https://preseea.uah.es/corpus-preseea, con lo que se cumplen 6 de los 7 objetivos específicos.

El último objetivo específico se cumplió con los análisis de las formas de tratamiento pronominales (FTP), algunos marcadores del discurso y la expresión de sujetos pronominales.

Los principales hallazgos sobre las FTP son los siguientes: Cali es una comunidad de habla en la que prima el uso de usted sobre las formas tú y vos. Esto indica que hay una predominancia de la cortesía de distanciamiento. Las mujeres son quienes usan más la forma de tratamiento de distancia usted y los hombres prefieren las formas de solidaridad tú y vos. Aunque la comunidad en general tiende a la cortesía de distanciamiento, algunos grupos impulsan el uso de las formas de tratamiento de solidaridad. Así, el uso de los pronombres tú y vos es promovido por los hombres, las dos generaciones más jóvenes, los niveles educativos más altos y los estratos socioeconómicos medios.

En el estudio del marcador discursivo *pues* se encontró, entre otros aspectos, lo siguiente: Aunque cumpliendo funciones diferentes, pues no parece estar restringido en cuanto a posición ni tipo de discurso. En el corpus analizado se encontraron dos grandes funciones pragmáticas y una función gramatical. Las funciones pragmáticas presentaron a su vez

subfunciones, siete metadiscursivas y tres de intensificación. La subfunción más frecuente en PRESEEA-Cali fue la de retraso del discurso, seguida de la comentadora, en la que pues introduce un nuevo comentario para iniciar el turno o dentro de este en secuencias interactivas. La tercera subfunción más frecuente fue la continuativa, en la que pues sirve como estrategia para continuar con el discurso luego de una pausa o vacilación y para unir unidades internas de la oración. También, pues se encontró como reformulador de explicación, corrección y recapitulación. Además, funciona como intensificador o refuerzo de actos de habla y de otros marcadores, con los que comparte una función pragmática, o con enfocadores de la alteridad.

Por último, el análisis de la expresión de sujetos pronominales arrojó, entre otros aspectos, lo siguiente: Los resultados correspondientes al condicionamiento social revelan los efectos de algo similar a la gradación etaria. Encontramos que las tasas pronominales aumentan proporcionalmente con la edad, lo cual constituye una tendencia adquisitiva atestiguada en español. Los hallazgos correspondientes al género del hablante muestran diferencias entre las conductas sociolingüísticas de mujeres y hombres caleños que pueden marcar divisiones dentro de la zona macrodialectal andina colombiana y parecen contribuir a la apertura de senderos investigativos.

Principales conclusiones y/o recomendaciones

Como se puede ver, se cumplió el objetivo de construir un corpus sociolingüístico para la descripción y el análisis de fenómenos de variación lingüística en relación con factores lingüísticos y sociales. De esta manera, se suple una necesidad identificada en los estudios lingüísticos de Cali: contar con materiales validados e idóneos para el estudio sociolingüístico de esta comunidad de habla, así como avanzar en la descripción lingüística y sociolingüística de esta comunidad. La metodología empleada ha permitido hacer discusiones contrastivas con los resultados obtenidos en otras variedades del español, lo cual es una contribución importante a la lingüística descriptiva y a la sociolingüística hispánicas. Es importante promover más estudios a partir del corpus PRESEEA-Cali para continuar avanzando en estos propósitos.

4. Impactos actual o potencial:

En el ámbito académico el corpus está siendo usado en la docencia como material de las clases de sociolingüística, a partir de él se trabaja la construcción de corpus sociolingüísticos, la recolección y sistematización de datos y el análisis de fenómenos sociolingüísticos. De igual manera, el corpus se ha puesto a disposición para el desarrollo de trabajos de grado de pregrado en el área de sociolingüística, acompañado de la aplicación de guías metodológicas internacionales desarrolladas por PRESEEA para el análisis de fenómenos sociolingüísticos, lo que lo convierte en una herramienta formativa bastante productiva.

En el ámbito investigativo, como se verá más adelante en los productos, se ha tenido impacto en el ámbito hispánico al aportar un corpus de estudio de Cali, se han llevado los resultados a congresos nacionales e internacionales y se han publicado artículos en revistas indexadas.
Asimismo, el proyecto ha tenido un impacto cultural porque ha despertado el interés por conocer el habla caleña y el equipo investigador ha participado en eventos culturales y de divulgación sobre la forma en que se habla en Cali.
El corpus sigue siendo una fuente importante para el desarrollo de otros estudios descriptivos y sociolingüísticos de la variedad de habla caleña que se pueden desarrollar en el futuro. De igual manera, puede ser usado para otros propósitos como la enseñanza del español como lengua extranjera, así como para análisis de contenido de las entrevistas en el marco de los estudios culturales y del análisis del discurso.
Se espera que este corpus pueda complementarse con otros más espontáneos y en situaciones diversas para hacer estudios más amplios sobre el habla de Cali.

5. Productos:

Tabla No. 1**. Cantidad y tipo de productos pactados en el *Acta de Trabajo y Compromiso* y productos finalmente presentados**

TIPO DE PRODUCTOS	**No. de PRODUCTOS PACTADOS**				**No. de PRODUCTOS PRESENTADOS**			
Productos de nuevos conocimientos								
Artículo en revista ISI-SCOPUS:	**Q1**	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**
Artículo completo publicado en revistas indexadas	**A1**	**A2**	**B**	**C**	**A1**	**A2**	**B**	**C**
			1				3	
Libros de autor que publiquen resultados de investigación								
Capítulos en libros que publican resultados de investigación								

TIPO DE PRODUCTOS	No. de PRODUCTOS PACTADOS		No. de PRODUCTOS PRESENTADOS	
Productos o procesos tecnológicos patentados o registrados				
• Prototipos y patentes				
• Software				
Productos o procesos tecnológicos usualmente no patentables o protegidos por secreto industrial	1		1	
Normas basadas en resultados de investigación				
Formación de recursos humanos	No. de estudiantes vinculados	No. de tesis	No. De estudiantes Vinculados	No. De tesis
Estudiantes de pregrado				
Semillero de Investigación				
Estudiantes de maestría				
Estudiantes de doctorado				
Joven investigador				
Productos de divulgación				
Publicaciones en revistas no indexadas				
Ponencias presentadas en eventos (congresos, seminarios, coloquios, foros)	No. de ponencias nacionales	No. de ponencias internacionales 1	No. de ponencias nacionales	No. de ponencias internacionales 3
Propuesta de investigación				
Propuestas presentadas en convocatorias externas para búsqueda de financiación.				

Tabla No. 2. Detalle de productos

Tipo de producto:	Artículo
Nombre General:	Cuadernos de Lingüística Hispánica, (37), e12694
Nombre Particular:	Formas de tratamiento pronominales en el español de Cali, Colombia
Ciudad y fechas:	Tunja, 14 de junio de 2021
Participantes:	Róbinson Grajales Alzate Margareth Marmolejo Caicedo
Sitio de información:	https://revistas.uptc.edu.co/index.php/linguistica_hispanica/article/view/12694/10760
Formas organizativas:	Universidad del Valle

Tipo de producto:	Artículo
Nombre General:	Cuadernos de Lingüística Hispánica, (40), e14250
Nombre Particular:	Funciones pragmáticas del marcador discursivo pues en el habla de Cali, Colombia.
Ciudad y fechas:	Tunja, 12 de julio de 2022
Participantes:	Margareth Marmolejo Caicedo
Sitio de información:	https://revistas.uptc.edu.co/index.php/linguistica_hispanica/article/view/14250/12096
Formas organizativas:	Universidad del Valle

Tipo de producto:	Artículo

Nombre General:	Lingüística, Vol. 40, No. 1, 2024
Nombre Particular:	Estudio sociolingüístico de la expresión pronominal en Cali
Ciudad y fechas:	Montevideo, 16 de abril de 2024
Participantes:	Rafael Orozco Margareth Marmolejo Caicedo Róbinson Grajales Alzate
Sitio de información:	https://www.mundoalfal.org/ojs/index.php/Revista/article/view/161/142
Formas organizativas:	Louisiana State University Universidad del Valle

Tipo de producto:	Sitio web
Nombre General:	Corpus Proyecto de Estudio Sociolingüístico del Español de España y América (PRESEEA)
Nombre Particular:	Corpus PRESEEA-Cali
Ciudad y fechas:	Alcalá de Henares, 2 de enero de 2021
Participantes:	Róbinson Grajales Alzate Margareth Marmolejo Caicedo
Sitio de información:	https://preseea.uah.es/equipos/cali
Formas organizativas:	Universidad del Valle

Tipo de producto:	Ponencia internacional
Nombre General:	Congreso Internacional de Lingüística General
Nombre Particular:	Análisis sociolingüístico de los marcadores de reformulación en el español de Colombia

Ciudad y fechas:	Sevilla, 25 de junio de 2021
Participantes:	Margareth Marmolejo Caicedo Róbinson Grajales Alzate
Sitio de información:	
Formas organizativas:	Universidad del Valle

Tipo de producto:	Ponencia internacional
Nombre General:	XIXI Congreso Internacional ALFAL
Nombre Particular:	El marcador discursivo pues en el español de Cali, Colombia
Ciudad y fechas:	Virtual. 8 al 13 de agosto de 2021.
Participantes:	Margareth Marmolejo Caicedo
Sitio de información:	
Formas organizativas:	Universidad del Valle

Tipo de producto:	Ponencia internacional
Nombre General:	Simposio Internacional PRESEEA
Nombre Particular:	La expresión de sujetos pronominales en el Corpus PRESEEA-Cali
Ciudad y fechas:	Alcalá de Henares, 13 de junio de 2023
Participantes:	Rafael Orozco Margareth Marmolejo Caicedo Róbinson Grajales

Sitio de información:	
Formas organizativas:	Louisiana State University Universidad del Valle

Robinson Grajales A.

Firma del investigador principal

VoBo. Vicedecano de Investigaciones