



EL LÉXICO URBANO DE LA CIUDAD DE GRANADA: PROPUESTA DE INFORMATIZACIÓN DEL CORPUS DEL PROYECTO DE LA NORMA CULTA Y DISEÑO DE UNA NUEVA ENCUESTA LÉXICA

THE URBAN LEXICON OF THE CITY OF GRANADA: PROPOSAL FOR THE DIGITALIZATION OF THE EDUCATED NORM CORPUS AND THE DESIGN OF A NEW LEXICAL SURVEY

Gonzalo Águila Escobar
Universidad de Granada
gonzaloaguila@ugr.es

RESUMEN

El artículo presenta los fundamentos, metodología y objetivos del proyecto Léxico urbano de Granada, desarrollado a partir del corpus del Estudio coordinado de la norma lingüística culta del español hablado en las principales ciudades de Iberoamérica y de la Península Ibérica. La investigación se articula en dos líneas complementarias: la informatización del léxico del Proyecto de la Norma Culta de Granada, con el fin de preservar, digitalizar y hacer accesible el repertorio original de Francisco Salvador, y la construcción de un nuevo léxico urbano, que permita observar la evolución diacrónica y la variación actual del español culto granadino. Esta doble dimensión combina la recuperación patrimonial con la innovación metodológica, mediante la integración de bases de datos relacionales y herramientas digitales aplicadas a la geolingüística léxica. El trabajo se enmarca en la red de proyectos internacionales VARILEX y Disponibilidad Léxica (LD), y ofrece un modelo de actualización y reaprovechamiento de los corpus históricos del español culto en el contexto urbano contemporáneo.

Palabras clave: norma culta, léxico urbano, informatización, Granada, variación léxica, corpus digital, proyecto Lope Blanch, VARILEX, disponibilidad léxica.

ABSTRACT

This paper outlines the theoretical and methodological framework of the Urban Lexicon of Granada project, developed from the corpus of the Coordinated Study of the Educated Linguistic Norm of Spanish Spoken in the Main Cities of Ibero-America and the Iberian Peninsula. The research is structured around two complementary goals: the digitalization and informatization of the educated lexicon of Granada, aimed at preserving and providing open access to Francisco Salvador's original corpus, and the creation of a new urban lexicon that reflects current lexical dynamics and diachronic variation. This dual approach combines linguistic heritage recovery with methodological innovation, through the use of relational databases and digital tools applied to lexical geolinguistics. The project is linked to the international networks VARILEX and Lexical Availability (LD) and proposes a model for updating and reusing historical corpora of educated Spanish in contemporary urban contexts.

Keywords: educated norm, urban lexicon, digitalization, Granada, lexical variation, digital corpus, Lope Blanch Project, VARILEX, lexical availability.



1. INTRODUCCIÓN

El léxico constituye el nivel de la lengua más dinámico y sensible a los cambios sociales, culturales y tecnológicos. Sin embargo, el español carece todavía de un corpus léxico urbano actual, informatizado y comparable a escala panhispánica que permita observar su evolución, medir sus procesos de convergencia y divergencia y analizar la realidad del vocabulario en el siglo XXI, tal y como alienta Moreno Fernández (2020, p. 129-130). Los estudios existentes, tales como atlas lingüísticos, léxicos de la *norma culta* (Lope Blanch, 1986) o proyectos sociolingüísticos, ofrecen materiales valiosísimos, pero dispersos, elaborados con metodologías heterogéneas y, en la mayoría de los casos, no digitalizados en bases de datos.

El problema que motiva este proyecto es, por tanto, la ausencia de un repertorio léxico moderno, urbano y coordinado que dé continuidad al Proyecto de la Norma Culta¹ y que permita conocer el léxico del español actual con criterios científicos, técnicos y comparables. A pesar de su magnitud, los catorce léxicos publicados entre 1978 y 2010 (México, Madrid, San Juan, Santiago, Granada, Buenos Aires, Caracas, Lima, Sevilla, entre otros) permanecen sin un tratamiento digital en bases de datos que facilite su análisis multivariante. A ello se suman los avances desiguales en los atlas lingüísticos, aún en proceso de informatización, y las limitaciones de los corpus orales, como PRESEEA (*Presentación*, en línea), que no fueron concebidos para el estudio sistemático del vocabulario.

El presente trabajo propone crear un nuevo acopio léxico urbano que sirva como corpus de referencia del español contemporáneo y que combine la herencia metodológica del Proyecto de la Norma Culta con las posibilidades de las tecnologías actuales y, para ello, propone comenzar con el léxico de la ciudad de Granada. El proyecto se plantea en dos fases complementarias. En la primera, se llevará a cabo la informatización y normalización de los léxicos existentes,

¹ Nos referiremos así al proyecto denominado originariamente como Proyecto de Estudio Coordinado de la Norma Lingüística Culta de las Principales Ciudades de Hispanoamérica y rebautizado en la actualidad como Proyecto de la Norma Culta Hispánica “Juan M. Lope Blanch”.

generando una base de datos común que integre todos los materiales del Proyecto de la Norma Culta y permita su explotación digital. En la segunda, se desarrollará un nuevo estudio del léxico urbano del siglo XXI, con un cuestionario revisado, reducido y adaptado a la realidad actual, y con una metodología híbrida de recogida de datos (presencial y digital) basada en encuestas informatizadas mediante plataformas como Limesurvey, Voiceform o Redcap.

El nuevo léxico urbano adoptará un diseño muestral común por edad, sexo e instrucción social, siguiendo las pautas del Proyecto preselea y los datos se almacenarán en una base de datos relacional con codificación uniforme, apta para análisis estadísticos y cartográficos (análisis de clúster, correlación, distancia léxica, etc.), compatible con otras plataformas digitales de cartografiado léxico como CORPAT o Varilex-R.

La finalidad última del proyecto es construir un corpus léxico urbano panhispánico que refleje la realidad lingüística del español del siglo XXI, integrando los distintos niveles de análisis (local, regional y nacional). Este corpus permitirá no solo describir el léxico actual, sino también observar sus transformaciones respecto al repertorio culto del siglo XX, establecer patrones de convergencia hacia un estándar común o de diversificación regional, y ofrecer datos aplicables a la enseñanza del español, la lexicografía diferencial, la traducción o la planificación lingüística.

En suma, esta propuesta parte de la convicción de que la informatización del pasado y la creación de un nuevo corpus del presente son dos acciones inseparables. Solo desde una perspectiva conjunta retrospectiva y prospectiva será posible reconstruir el mapa léxico del español y avanzar hacia un léxico urbano del siglo XXI representativo, comparativo y abierto a la comunidad científica internacional.

1.1. ANTECEDENTES

El estudio del léxico hispánico ha experimentado desde mediados del siglo XX un desarrollo sostenido e impulsado por una serie de proyectos panhispánicos de carácter coordinado. El Proyecto de la Norma Culta, propuesto por

RILEX. VOLUMEN MONOGRÁFICO

El Léxico urbano de la ciudad de Granada: propuesta de informatización del corpus del proyecto de la norma culta y diseño de una nueva encuesta léxica

Juan M. Lope Blanch en el marco del Programa interamericano de lingüística y enseñanza de idiomas (PILEI) en 1964 en Bloomington, constituye el primer intento sistemático de descripción comparada del habla culta urbana en el mundo hispánico. A partir de los cuestionarios unificados publicados por el CSIC entre 1971 y 1973 (fonética y fonología, morfosintaxis i y ii y léxico), se desarrollaron corpus homogéneos de las principales ciudades hispánicas, cuyos resultados se publicaron entre 1978 y 2005. El componente léxico, con 4452 ítems distribuidos en 21 campos temáticos, permitió construir repertorios comparables del vocabulario culto de catorce ciudades de habla hispana (véase Tabla 1). Los trabajos derivados, como los de Bentivoglio (1978), Lope Blanch (1977), Carranza Vásquez (1988), Contreras (1983, 1988), Rabanales (1983), Huyke (1977), López Morales (1977, 1991), Quilis (1983), Valencia (1984), Luna Traill (2007), Marrero (1999), Santana (2014) y Fernández de Molina (2017), entre otros, consolidaron la dimensión léxica del proyecto como base para la comparación diatópica y diastrática del español culto.

Año	Título del léxico	Autor (es)
1978	<i>Léxico del habla culta de México</i>	Juan M. Lope Blanch (dir.)
1981	<i>Encuestas léxicas del habla culta de Madrid</i>	Juan Carlos de Torres
1986	<i>Léxico del habla culta de San Juan de Puerto Rico</i>	Humberto López Morales y Héctor R. Cardona Santana (coords.)
1987	<i>Léxico del habla culta de Santiago</i>	Ambrosio Rabanales y Lidia Contreras
1991	<i>Léxico del habla culta de Granada</i>	Francisco Salvador
1996	<i>Léxico del habla culta de La Paz</i>	José Mendoza
1997	<i>Léxico del habla culta de Santafé de Bogotá</i>	Hilda Otálora de Fernández (dir.)
1998	<i>Léxico del habla culta de Las Palmas de Gran Canaria</i>	José Antonio Samper (dir.)
1998	<i>Léxico del habla culta de Caracas</i>	Mercedes Sedano y Z. Pérez
1998	<i>Léxico del habla culta de Buenos Aires</i>	Academia Argentina de Letras
2000	<i>Léxico del habla culta de Córdoba</i>	Alicia Malanca de Rodríguez, María T. Toniolo y María E. Zurita
2000	<i>Léxico del habla culta de Lima</i>	Rocío Caravedo (dir.)
2005	<i>Léxico del habla culta de Sevilla</i>	Pedro Carbonero (dir.) y Asunción Ortiz (dir.)
2010	<i>Léxico del habla culta de La Habana</i>	Ana María González Mafud, Marisela Pérez Rodríguez

TABLA 1. *Relación de léxicos publicados en el Proyecto de la Norma Culta.*

RILEX. VOLUMEN MONOGRÁFICO

El Léxico urbano de la ciudad de Granada: propuesta de informatización del corpus del proyecto de la norma culta y diseño de una nueva encuesta léxica

Durante las décadas siguientes, esta línea se amplió con los estudios de disponibilidad léxica, coordinados por Humberto López Morales², que permitieron caracterizar el léxico potencial de los hablantes a partir de asociaciones semánticas y análisis estadísticos. Este programa, representado por autores como Alba (1995), Hernández y Samper (2001) o Ávila y Villena (2010), introdujo nuevas variables cognitivas y socioculturales que enriquecieron la comprensión del léxico activo y de su variación intergeneracional.

En continuidad con el espíritu coordinado en el Proyecto de la Norma Culta, el Proyecto varilex (Variación léxica del español), coordinado desde 1993 por Hiroto Ueda (Universidad de Tokio) (2009, 2011), representa la fase contemporánea y digitalizada de la investigación léxica panhispánica. Su objetivo, como se recoge en su página web (VARILEX, *Variación Léxica del Español En el Mundo*, s. f.), es reunir, actualizar y difundir datos del léxico moderno urbano mediante encuestas informatizadas y cartografía interactiva, con la creación de un banco de datos léxicos internacional:

Este proyecto nace de nuestro anhelo de conocer la situación actual del léxico español del mundo [...]. Y como todavía no existe un estudio exhaustivo e informatizado del español actual de todas las regiones hispanohablantes, proponemos, como primera tentativa, formar una red internacional de investigación sobre el español.

Las sucesivas encuestas, realizadas en más de sesenta ciudades hispanohablantes y con más de mil conceptos analizados, han permitido estudiar la variación diatópica en tiempo real y observar tendencias de convergencia y divergencia léxica en el mundo hispánico. En su segunda fase, a partir de 2011, varilex incorporó un sistema de encuestas web dinámicas (Ueda, 2011), que ofrecen participación directa a los hablantes y resultados visualizados en mapas bidimensionales, convirtiéndose así en una plataforma abierta y colaborativa para la descripción actualizada del léxico español. No obstante, los datos, como

² Actualmente siguen desarrollándose estudios en esta línea, como la disponibilidad léxica de Granada en el siglo xxi en el marco del Proyecto de Investigación «Estudio de los condicionantes sociales del español actual en Granada: nuevas identidades, nuevos retos, nuevas soluciones». Código: PID2023-148371NB-C41.

indica Moreno Fernández (2020, p. 135), sufrieron en 2016 una reconfiguración parcial y revisión lo que llevó a “reunir por países los datos procedentes de distintas ciudades de cada uno de ellos, de modo que la ordenación final prescindiera de la localización por núcleos de población, en favor de una ordenación por territorios nacionales”. Además, “los expertos procedieron a modificar aquellos elementos lingüísticos que no podían considerarse como generales o mayoritarios dentro de cada país, lo que implicaba la retirada de arcaísmos, dialectalismos o voces jergales, especialmente”. Estas dos decisiones favorecen, claro está, la determinación de aquellas unidades léxicas que son comunes de cada país y que pertenecen al nivel estándar, pero ocultan el hecho de que, como el propio Moreno Fernández (2020, p. 135) reconoce también:

La consideración de las naciones o países como unidades de referencia para etiquetar el léxico no deja de ser un artificio que pasa por encima de la propia realidad geolingüística, que revela a menudo usos regionales o locales que, cuando no son tratados como generales, son ignorados por minoritarios, difuminándose así el paisaje dialectal.

Al mismo tiempo, los corpus sociolingüísticos contemporáneos, como el Proyecto para el estudio del español de España y América (presea), han desarrollado metodologías de muestreo y codificación de enorme precisión, aunque centradas sobre todo en fenómenos fonéticos, morfosintácticos o discursivos. El componente léxico, en cambio, aparece de manera tangencial en las entrevistas semidirigidas, lo que impide construir inventarios exhaustivos o estudiar la variación léxica con la misma solidez cuantitativa que otros niveles de la lengua (Sánchez García, 2018, p.99). Los corpus orales son adecuados para en análisis fonético, morfosintáctico o pragmático, pero no para el léxico: “Estos corpus orales no están enfocados a la cuestión léxica más que a su aparición de forma espontánea dentro del discurso y, por supuesto, aún menos a la naturaleza dialectal, ya que la temática planteada en la conversación semidirigida no suele ser propicia para ello” (Águila, 2021, p. 306). A este respecto, expone Moreno Fernández (2020, p. 141) que “los corpus no suelen ofrecer la posibilidad de construir inventarios o paradigmas

completos, lo que obliga a complementarlos mediante búsquedas en otras fuentes”.

A nivel regional, proyectos como converlex (Ávila Muñoz, 2017) o vitalex (Águila, 2011) han retomado la tradición geolingüística de los atlas regionales españoles, como el *Atlas lingüístico y etnográfico de Andalucía (alea)* (1961-1973), ofreciendo datos actualizados sobre la vitalidad y atrición del léxico andaluz. Estas investigaciones locales, unidas a los macroproyectos panhispánicos, contribuyen a delinear un panorama coherente de los procesos de nivelación, convergencia y conservación léxica en el español contemporáneo.

En este contexto de renovación metodológica y de convergencia entre los grandes proyectos panhispánicos de investigación léxica, surge el proyecto de estudio del léxico urbano de la ciudad de Granada (ECOS-GR), concebido como una iniciativa local integrada en el marco general del nuevo léxico urbano hispánico. Su objetivo principal es construir un corpus léxico representativo del español culto granadino del siglo XXI, que permita analizar la evolución del vocabulario respecto al registrado en el *Léxico del habla culta de Granada* (Salvador, 1991) y establecer comparaciones sistemáticas con otras ciudades del mundo hispánico.

La metodología adoptada combina los principios del Proyecto de la Norma Culta, en especial, la organización temática del cuestionario y el carácter onomasiológico de las encuestas, con la estructura sociolingüística del proyecto preseea, basada en la estratificación por edad, sexo y nivel de instrucción. Esta hibridación metodológica permite preservar la continuidad con el modelo clásico del Proyecto de la Norma Culta y, al mismo tiempo, incorporar la representatividad y precisión sociológica del preseea. El resultado es un diseño mixto que aspira a integrar la perspectiva léxica y la sociolingüística dentro de un mismo marco operativo y comparable.

En conjunto, el proyecto se configura como un laboratorio de innovación metodológica para la investigación del léxico urbano hispánico y aspira a integrarse en una futura red comparativa panhispánica que permita documentar,

desde la perspectiva local de Granada, los procesos globales de cambio, nivelación y convergencia del español actual.

2. EL PROYECTO PARA EL ESTUDIO DEL LÉXICO URBANO DE LA CIUDAD DE GRANADA

2.1. FASE 1: INFORMATIZACIÓN DEL CORPUS DEL PROYECTO DE LA NORMA CULTA

La primera fase del proyecto tiene como objetivo la informatización integral de los catorce léxicos urbanos publicados dentro del Proyecto de la Norma Culta Hispánica “Juan M. Lope Blanch” entre 1978 y 2010 que corresponden a 14 ciudades del mundo hispánico, 4 de España (Madrid, Las Palmas de Gran Canaria, Granada y Sevilla) y 10 de América (México D.F, San Juan de Puerto Rico, Santiago de Chile, La Paz, Santafé de Bogotá, Caracas, Buenos Aires, Córdoba, Lima y La Habana).

Esta acción constituye el punto de partida para la creación de una base de datos unificada que conserve, normalice y haga accesibles los materiales léxicos originales del siglo xx. La informatización no se plantea como una simple tarea de digitalización, sino como un proceso científico de estructuración y estandarización de datos orientado a su futura explotación lingüística y estadística; ya lo requería Caravedo (2000, pp. 13-14) en la introducción del *Léxico del habla culta de Lima*:

(...) trascendiendo los límites del propio proyecto, en el marco de una actividad cada vez más creciente y fructífera de investigación desarrollada en los últimos decenios en materia de lenguaje. Me refiero a la formación de corpóra³ como objetivos en sí mismos, en la forma de bancos de datos exentos en lo posible de procesamiento analítico, los cuales se ofrecen a la comunidad especializada para una infinidad de propósitos de análisis y de conocimiento.

El corpus del Proyecto de la Norma Culta presenta una estructura común derivada del cuestionario de 1971, compuesto por 4445 ítems distribuidos en veintiún campos semánticos. Sin embargo, la diversidad de equipos

³ Precisamente, los autores del léxico de Puerto Rico (López & Cardona, 1986, p. 11) se refieren a su obra como un “corpus léxico de 7.304 vocablos (unos pocos repetidos)”.

nacionales y de criterios editoriales dio lugar a una considerable heterogeneidad metodológica, que dificulta la integración automática de los datos:

1. El número de informantes varía según la ciudad: 12 en Puerto Rico, La Paz, Las Palmas, Caracas, Córdoba y Sevilla; 13 en Santiago de Chile; 16 en Madrid, y 25 en México, Granada y Bogotá.
2. El reparto por generaciones y sexo suele ser equitativo, aunque presenta algunas particularidades. Así, Madrid distingue cuatro generaciones al incluir una franja muy joven por debajo de los 18 años. En Santiago de Chile, la distribución por sexo es equilibrada salvo en la segunda generación, donde hay dos mujeres y cuatro hombres. En Puerto Rico, además de los doce informantes plenos, se trabajó con cuarenta y ocho sujetos adicionales, divididos en cuatro submuestras rigurosamente paralelas.
3. Criterios de lematización y presentación de los datos. La forma de lematizar y exponer la información léxica también presenta notables diferencias entre ciudades. Algunos equipos son especialmente detallados en el uso de marcas gramaticales o estilísticas (como ocurre en Caracas), mientras que otros, como Puerto Rico o Granada, apenas ofrecen indicaciones al respecto. En cuanto al tratamiento de la sinonimia, también se observan pequeñas variaciones: Las Palmas, por ejemplo, señala la preferencia de determinadas voces, mientras que Granada no refleja esa jerarquización.
4. Diferencias metodológicas en la presentación de los datos. La divergencia metodológica más significativa se encuentra en el caso de Puerto Rico, cuyo léxico presenta únicamente los totales globales sin detallar los informantes que los producen ni sus características sociales preestratificadas. Esta ausencia de correlación entre las formas léxicas y los datos sociolingüísticos de los hablantes limita las posibilidades de análisis comparativo respecto a otros corpus del proyecto.

Para solventar estas disparidades, el proyecto propone un protocolo de codificación y armonización que permita transformar los materiales impresos en datos estructurados y comparables. Este proceso se desarrollará en tres etapas:

1. Conversión digital mediante técnicas de reconocimiento óptico de caracteres (OCR) y revisión manual posterior.
2. Normalización léxica, con la unificación de criterios de lematización, variantes y marcas gramaticales según los estándares del *Corpus del Español del Siglo XXI* (REAL ACADEMIA ESPAÑOLA, s. f.).
3. Estructuración relacional, organizando los datos en una base de datos SQL o XML compatible con los estándares de intercambio lingüístico (TEI-LEX0, ISOCat).

La informatización permitirá vincular cada forma léxica con sus metadatos asociados (informante, ciudad, campo semántico, frecuencia y observaciones) y generar consultas avanzadas que faciliten la comparación entre ciudades y sincronías. Además, esta base servirá como puente tecnológico con otros corpus léxicos en desarrollo, como el Corpus de los atlas lingüísticos (corpat, s. f.) o Varilex, posibilitando análisis cruzados y cartografías interactivas. Un ejemplo del aprovechamiento de estos materiales sería la posibilidad de generar mapas léxicos contrastivos como el ejemplo del ítem 507, *guisantes* (véase el Gráfico 1).

Desde el punto de vista metodológico, esta fase constituye la base empírica sobre la que se apoyará el nuevo proyecto de encuesta léxica. La informatización de los materiales históricos no solo garantiza la preservación del patrimonio lingüístico, sino que ofrece un marco de referencia indispensable para evaluar el cambio léxico en tiempo real y aparente. En este sentido, la recuperación digital del *Léxico del habla culta de Granada* (Salvador, 1991) desempeña un papel esencial, al servir como corpus de contraste frente a los datos que se recogerán en la nueva investigación de campo.

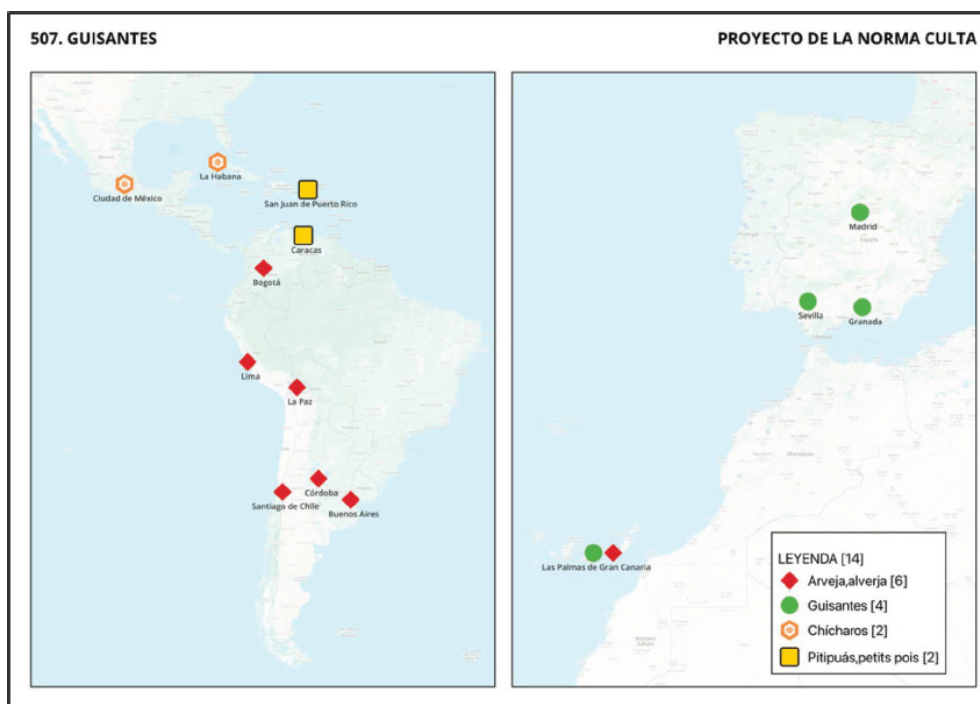


GRÁFICO 1. Mapa del ítem 507, denominación de guisantes.

2.2. FASE 2: NUEVO PROYECTO DE ENCUESTAS LÉXICA (PROYECTO DE LA NORMA CULTA Y PRESEEA)

La segunda fase del proyecto representa el núcleo metodológico y científico de la propuesta. Su finalidad es actualizar el estudio del léxico culto urbano mediante la elaboración y aplicación de una nueva encuesta que combine la tradición lexicográfica del Proyecto de la Norma Culta con la estructura socio-lingüística del preseea. Se trata de un diseño renovado que permite observar el cambio léxico en tiempo real, describir la distribución social y diatópica de las variantes y ofrecer una visión integral del español culto del siglo XXI.

2.2.1. Justificación de la nueva encuesta

Desde la publicación del primer léxico del habla culta de la ciudad de México (Lope Blanch, 1978) han transcurrido más de cuatro décadas, un periodo lo suficientemente extenso como para que el cambio léxico sea plenamente observable. En este tiempo, la lengua española ha experimentado transformaciones profundas en todos los niveles, pero el léxico, dado su carácter

permeable y su relación directa con la innovación cultural y tecnológica, ha sido el ámbito más dinámico. El léxico urbano de finales del siglo XX refleja una sociedad analógica, mientras que el de las primeras décadas del XXI se desarrolla en un contexto digital, global y multicultural.

Por ello, la actualización de los estudios léxicos urbanos no puede limitarse a la digitalización de los materiales existentes, sino que requiere un nuevo trabajo de campo. La nueva encuesta propuesta retoma la estructura semántica del cuestionario del Proyecto de la Norma Culta, pero la adapta a las condiciones de investigación contemporáneas: duración limitada, equilibrio temático y aplicabilidad tecnológica. De este modo, el proyecto no solo preserva la continuidad con la tradición del PILEI, sino que la renueva metodológicamente al integrarla en un marco sociolingüístico y digital.

2.2.2. Diseño general y principios metodológicos

En primer lugar, reivindicamos la encuesta mediante cuestionario léxico como instrumento esencial para la recogida del léxico. Como señalaba Moreno Fernández (2020, p. 141), “los corpus no suelen ofrecer la posibilidad de construir inventarios o paradigmas completos”, de ahí que el cuestionario previamente establecido y heredado de la dialectología tradicional sea, todavía, una metodología válida y, yo diría, esencial para llevar a cabo cualquier estudio léxico completo. Caravedo (2000, p. 18) abunda en ello:

(...) no podría esperarse recoger un inventario léxico ordenado según los campos onomasiológicos a partir de una entrevista espontánea encuadrada en temas limitados, como se hace para abordar la fonética, pues tales temas no propiciarían la aparición exhaustiva y ordenada de las palabras que se busca investigar.

El nuevo cuestionario léxico se concibe como un instrumento modular, digital y sociolingüísticamente controlado, orientado a obtener datos comparables entre ciudades y sincronías. La encuesta se organiza en torno a cinco principios fundamentales:

1. Continuidad estructural: se mantiene la organización onomasiológica del cuestionario de 1971, aunque reducida y actualizada, garantizando la posibilidad de contrastar los resultados con los léxicos históricos.

RILEX. VOLUMEN MONOGRÁFICO

El Léxico urbano de la ciudad de Granada: propuesta de informatización del corpus del proyecto de la norma culta y diseño de una nueva encuesta léxica

2. Actualización temática: se sustituyen o complementan campos obsoletos (por ejemplo, “el correo” o “los medios de comunicación tradicionales”) por otros de relevancia actual (“tecnología y redes”, “vida digital”, “movilidad urbana”).
3. Optimización temporal: el número total de ítems se reduce de los 4445 originales a un número en torno a los 2000, lo que permite completar la encuesta en sesiones más reducidas, equilibrando exhaustividad y viabilidad.
4. Integración sociolingüística: se adoptan las variables de muestreo del preseea (edad, sexo y nivel de instrucción), con una muestra de 54 informantes por ciudad y se estudia la viabilidad y rendimiento de incluir la estratificación por barrios o distritos de manera sistemática.
5. Codificación digital automática: cada respuesta se vincula a una entrada léxica mediante un identificador único, lo que permite su almacenamiento directo en la base de datos general del proyecto.

2.2.3. *Elaboración del cuestionario*

El proceso de elaboración del nuevo cuestionario parte del análisis de los 21 campos semánticos originales del Proyecto de la Norma Culta (Rabanales, 2004) y de su rendimiento léxico en los distintos léxicos urbanos. A partir de esa revisión se seleccionan los campos de mayor productividad léxica y sociocultural (alimentación, vivienda, vida social, tecnología, transportes, enseñanza, cuerpo humano), reduciendo la redundancia interna y adaptando los contenidos a los nuevos contextos de uso.

El objetivo no es reproducir el inventario original, sino garantizar representatividad conceptual y variación suficiente. Así, por poner un ejemplo, el campo de la alimentación incluirá productos tradicionales como guisantes o pimiento y podrá añadir nuevas cuestiones como persimón o *sushi*, mientras que en el campo de la tecnología y comunicación se abordará términos propios de la era digital como *whatsapp*, *chat*, *videollamada*, *selfie* o *emoticono*. Esta revisión se desarrolla en colaboración con especialistas de las diferentes

ciudades participantes, quienes validan la pertinencia de cada ítem y proponen equivalencias culturales locales.

2.2.4. Modalidad híbrida de recogida de datos

Una de las innovaciones principales del proyecto es el modelo híbrido de recolección digital, que combina la entrevista presencial con la encuesta virtual asistida. Este sistema busca conservar el control metodológico y el rigor científico propio de la investigación sociodialectal tradicional (encuesta *sur place*) y, al mismo tiempo, aprovechar las ventajas tecnológicas de la recogida automatizada.

En el caso de Granada, ciudad piloto del proyecto, se prevé realizar las sesiones en un entorno controlado, con grupos de informantes que completarán la encuesta en ordenadores o tabletas conectadas a una plataforma digital. La herramienta, diseñada sobre software como Limesurvey, REDCap o Voiceform, permitirá registrar simultáneamente las respuestas escritas y las grabaciones orales, garantizando la fidelidad fonética y la posibilidad de análisis acústico posterior. Cada sesión contará con la asistencia de personal investigador encargado de resolver dudas terminológicas o de interpretación.

De manera complementaria, se contempla la aplicación remota de una parte del cuestionario, distribuida entre hablantes de perfiles sociolingüísticos equivalentes. Este formato asíncrono permitirá ampliar la cobertura y contrastar la fiabilidad de los datos obtenidos presencialmente (*Patrimonio Léxico da Gallaecia*, s. f.). Todos los resultados se almacenarán automáticamente en la base de datos general, con codificación normalizada e identificación de las variables sociolingüísticas pertinentes.

2.2.5. Codificación y análisis de los datos

El diseño digital del cuestionario permite una codificación inmediata de las respuestas, que se almacenan en una base de datos relacional con estructura interoperable. Cada registro léxico incluye los siguientes campos: lema,

categoría gramatical, respuesta textual, tipo de variante (sinónima, formal, regional), variables sociales del informante, frecuencia y observaciones contextuales. Esta estructura facilita el análisis cuantitativo y la comparación automática entre ciudades y generaciones.

Una vez completada la recogida, los datos podrán someterse a diversos tipos de análisis:

1. Análisis estadístico descriptivo (frecuencias, distribuciones, correlaciones).
2. Análisis multivariante (componentes principales, clúster, métodos de concentración léxica).
3. Análisis geográfico mediante visualización cartográfica, integrando la información de las distintas ciudades en mapas dinámicos interactivos.

El procesamiento de la información se realizará con software libre (R, Python, QGIS), garantizando transparencia y reproducibilidad científica. Este enfoque digital permitirá explorar con rapidez cuestiones clásicas de la variación léxica: identificación de áreas léxicas, detección de procesos de cambio, convergencia hacia normas locales o panhispánicas y diferenciación sociolingüística de usos, entre otras.

3. APLICACIÓN PILOTO EN GRANADA Y RESULTADOS ESPERADOS

Granada desempeñará el papel de ciudad piloto para validar la metodología y el cuestionario. La investigación de campo, prevista como proyecto inicial, permitirá comprobar el rendimiento de los ítems seleccionados, ajustar la duración y evaluar la fiabilidad de la plataforma digital. La comparación con los datos del *Léxico del habla culta de Granada* (Salvador, 1991) ofrecerá un contraste diacrónico de más de cuarenta años, que permitirá observar de forma directa los procesos de sustitución léxica, innovación y pérdida.

Este estudio piloto servirá también como base para la expansión del proyecto al resto de ciudades del Proyecto de la Norma Culta, con la posibilidad de establecer convenios con los equipos preseleccionados correspondientes. De esta manera, se aspira a configurar una red panhispánica de observación léxica

urbana, que combine rigor metodológico, actualización tecnológica y apertura colaborativa.

Entre los resultados más relevantes de esta fase destacan:

1. La creación de un nuevo corpus léxico urbano del español culto contemporáneo, interoperable y de acceso controlado.
2. La posibilidad de comparar con criterios homogéneos dos sincronías como son la del siglo XX y la del XXI.
3. El desarrollo de un modelo metodológico replicable que integre investigación sociolingüística, tecnología digital y lexicografía.
4. La aportación de materiales empíricos útiles para la lexicografía académica, la enseñanza del ELE y la industria lingüística.

En conjunto, esta fase permitirá pasar de la conservación pasiva del legado léxico a su renovación activa mediante nuevos instrumentos de observación y análisis. La informatización del corpus histórico, iniciada en la fase I, adquiere ahora pleno sentido como base de comparación y como sustento metodológico para la exploración del cambio léxico urbano en el siglo XXI.

4. PROYECCIÓN Y APLICACIONES

El desarrollo del nuevo corpus léxico urbano y la informatización del legado del Proyecto de la Norma Culta abren múltiples líneas de investigación y de transferencia del conocimiento. En el plano académico, la combinación de ambas fases permitirá construir una infraestructura panhispánica de observación léxica que integre la dimensión histórica, sociolingüística y tecnológica del estudio del léxico del español.

Desde una perspectiva teórica, el proyecto ofrece un marco metodológico replicable para otras lenguas y variedades, al tiempo que redefine los límites entre dialectología, sociolingüística y lexicografía. La posibilidad de cruzar variables sociales con datos onomasiológicos amplía significativamente las herramientas de análisis disponibles para el estudio del cambio lingüístico.

RILEX. VOLUMEN MONOGRÁFICO

El Léxico urbano de la ciudad de Granada: propuesta de informatización del corpus del proyecto de la norma culta y diseño de una nueva encuesta léxica

En el plano aplicado, los resultados del proyecto tendrán repercusiones directas en tres ámbitos principales:

1. Lexicografía y normalización: los datos del corpus servirán como base para la elaboración de diccionarios diferenciales y para la marcación diatópica en obras normativas, contribuyendo a una descripción más precisa del español culto contemporáneo.
2. Enseñanza del español como lengua extranjera (ELE): el corpus permitirá seleccionar léxico de alta frecuencia y de uso general en el ámbito culto, favoreciendo una enseñanza basada en datos empíricos y en la realidad sociolingüística panhispánica (Águila, 2025).
3. Tecnología lingüística y humanidades digitales: la estructura digital del corpus posibilitará su integración en aplicaciones de procesamiento del lenguaje natural (PLN), minería de datos y visualización interactiva, convirtiéndolo en un recurso estratégico para la innovación tecnológica en el ámbito hispánico.

Además, la creación de una plataforma web con acceso académico facilitará la cooperación entre equipos de investigación de distintos países y el desarrollo de proyectos complementarios. La interoperabilidad del corpus con bases existentes como las de Varilex, CORPAT o VITALEX, permitirá construir un sistema integrado de recursos léxicos del español, contribuyendo así a una visión unificada y dinámica del léxico urbano en el mundo hispanohablante y contestar a muchas de las preguntas que se plantean en la actualidad en torno al léxico y que se centran en conocer el área de uso de cada una de las unidades léxicas: “Un análisis conjunto de la variación en todo el mundo hispánico podría arrojar nuevos datos e informaciones para comprender mejor la dinámica interna de este importante complejo dialectal” (Moreno Fernández, 2020).

5. CONCLUSIONES

El proyecto presentado en este trabajo plantea un modelo de investigación que combina recuperación patrimonial e innovación metodológica.

La informatización del corpus del Proyecto de la Norma Culta no se concibe como una meta aislada, sino como la condición necesaria para el desarrollo de una nueva etapa en el estudio del léxico urbano. A partir de esta base, la nueva encuesta léxica diseñada bajo el marco del preseea permitirá analizar con rigor y comparabilidad el léxico culto del siglo XXI, observando los procesos de cambio y estandarización que configuran el español actual.

La elección de Granada como ciudad piloto responde tanto a su relevancia histórica en el Proyecto de la Norma Culta como a su posición estratégica dentro del espacio andaluz. Su léxico culto, ya documentado por Salvador (1991), ofrece una base ideal para contrastar sincronías y validar metodologías. La informatización de ese material y la aplicación de la nueva encuesta convertirán a Granada en el punto de partida de una red internacional de observación léxica.

En definitiva, el proyecto busca situar nuevamente a la investigación léxica en el centro de la lingüística hispánica, dotándola de herramientas actualizadas y de una base empírica sólida. La articulación entre tradición y modernidad, entre el legado del PILEI y la innovación digital del preseea, constituye la clave para comprender y describir el léxico urbano del español en el siglo XXI. En la actualidad se avanza simultáneamente en la informatización del corpus de Granada y en la elaboración del nuevo cuestionario, con el propósito de iniciar la encuesta y contrastar los resultados obtenidos con los de 1991, al tiempo que se determina el léxico vigente en la Granada actual.

FINANCIACIÓN

Este trabajo y la idea de este proyecto se incluyen dentro de un proyecto de investigación denominado “Estudio de los condicionantes sociales del español actual en Granada: nuevas identidades, nuevos retos, nuevas soluciones” con código PID2023-148371NB-C41 y con el acrónimo ECOS-GR.

RILEX. VOLUMEN MONOGRÁFICO

El Léxico urbano de la ciudad de Granada: propuesta de informatización del corpus del proyecto de la norma culta y diseño de una nueva encuesta léxica

REFERENCIAS BIBLIOGRÁFICAS

- Academia Argentina de Letras. (1998). *Léxico del habla culta de Buenos Aires*.
- Alba, O. (1995). *El léxico disponible de la República Dominicana*. Pontificia Universidad Católica Madre y Maestra.
- ALEA = Alvar López, M., Llorente Maldonado, A. & Salvador Caja, G. (1961-1973). *Atlas lingüístico y etnográfico de Andalucía*. Universidad de Granada/CSIC.
- Águila Escobar, G. (2011). Vitalidad léxica y etnográfica en La Alpujarra (1950-2010). Análisis de la vitalidad del léxico de La Alpujarra en comparación con el ALEA. En A. Cestero, I. Molina & F. Paredes (eds.), *La lengua, lugar de encuentro. Actas del XVI Congreso Internacional de la ALFAL* (pp. 1847-1856). Editorial Universidad de Alcalá.
- Águila Escobar, G. (2021). Aproximación al estudio del léxico dialectal en el corpus PRE-SEEA-Granada. En M. M. Soto Melgar & A. Zholobova (eds.), *El español de Granada*. Estudio sociolingüístico (pp. 301-329). Peter Lang.
- Águila Escobar, G. (2025). Los léxicos de la norma culta y la enseñanza de la variedad léxica en ELE: el caso de las denominaciones de las legumbres en el mundo hispánico. *Porta Linguarum*. (En prensa).
- Ávila Muñoz, A. M. (2017). Entre limonetas y malaguitas. estudio del léxico divergente de proximidad. *Hispania*, 100(4), 538-553. <https://doi.org/10.1353/hpn.2017.0097>
- Ávila Muñoz, A. M. & Villena Ponsoda, J. A. (2010). *Variación social del léxico disponible en la ciudad de Málaga*. Sarriá.
- Bentivoglio, P. (1978). Observaciones sobre el léxico del cuerpo humano en el habla culta de la ciudad de Caracas. En *Lingüística y educación. Actas del IV Congreso Internacional de la ALFAL* (pp. 229-235). Universidad Nacional Mayor de San Marcos.
- Caravedo, R. (2000). *Léxico del habla culta de Lima*. Pontificia Universidad Católica del Perú. <https://doi.org/10.18800/9789972423413>
- Carbonero Cano, P. (dir.) & Ortiz Torres, A. (ed.). (2005). *Léxico del habla culta de Sevilla*. Publicaciones de la Universidad de Sevilla.
- Carranza Vásquez, T. (1988). Galicismos y anglicismos en el léxico del vestuario en el español de México. En *Actas del VI Congreso Internacional de la ALFAL* (pp. 575-592). Universidad Nacional Autónoma de México.
- Contreras, L. (1983). Extranjerismos e indigenismos en el léxico chileno relativo a la alimentación. En *Philologica hispaniensia in honorem Manuel Alvar* (Vol. 1, pp. 159-174). Gredos.
- Contreras, L. (1988). Los anglicismos en el léxico del habla culta de Santiago de Chile. En *Actas del VI Congreso Internacional de la ALFAL* (pp. 593-654). Universidad Nacional Autónoma de México.
- CORPAT – Corpus de los Atlas Lingüísticos (s. f.). <https://corpat.es/>
- Fernández de Molina, E. (2017). Estudio contrastivo de la norma culta de tres ciudades peninsulares. Análisis del campo semántico de la vivienda. *Onomázein*, 37, 90-111. <https://doi.org/10.7764/onomazein.37.09>

RILEX. VOLUMEN MONOGRÁFICO

El Léxico urbano de la ciudad de Granada: propuesta de informatización del corpus del proyecto de la norma culta y diseño de una nueva encuesta léxica

- González Mafud, A. M.^a, Pérez Rodríguez, M., Perdomo Carmona, M. & Gutiérrez Fuentes, G. (2010). *Léxico del habla culta de La Habana*. Editorial UH. Universidad de La Habana.
- Hernández Cabrera, C. E. & Samper Hernández, M. (2001). Léxico disponible y norma culta. En A. Valencia (ed.), *Actas del Congreso Internacional "El español culto en el mundo hispánico"* (Vol. II, pp. 79-89). Universidad Bolivariana y ALFAL.
- Huyke Freiría, I. (1977). Anglicismos en el vocabulario culto de San Juan: cuatro campos léxicos. En J. M. Lope Blanch (ed.), *Estudios sobre el español hablado en las principales ciudades de América* (pp. 63-84). UNAM.
- Lope Blanch, J. M. (1977). *Estudios sobre el español hablado en las principales ciudades de América*. UNAM.
- Lope Blanch, J. M. (1978). *Léxico del habla culta de México*. UNAM.
- Lope Blanch, J. M. (1986). *El estudio del español hablado culto. Historia de un proyecto*. UNAM.
- López Morales, H. (1977). Tres calas léxicas en el español de La Habana (indigenismos, afro-negrismos, anglicismos). En J. M. Lope Blanch (ed.), *Estudios sobre el español hablado en las principales ciudades de América* (p. 49-63). UNAM.
- López Morales, H. (1991). Muestra de léxico panantillano. El cuerpo humano. En E. Luna Traill (coord.), *Scripta philologica in honorem Juan M. Lope Blanch* (Vol. 2, pp. 593-625). UNAM.
- López Morales, H. & Cardona Santana, H. R. (1986). *Léxico del habla culta de San Juan de Puerto Rico*. Academia Puertorriqueña de la Lengua Española.
- Luna Traill, E. (2007). Coincidencias y discrepancias léxicas en cuatro ciudades de habla hispana. *Revista de Filología de la Universidad de La Laguna*, 25, 353-360.
- Malanca de Rodríguez, A., Toniolo, M.^a T. & Zurita, M.^a E. (2000). *Léxico del habla culta de Córdoba*. Universidad Nacional de Córdoba.
- Marrero Pulido, V. (1999). *La estructura del léxico en la norma lingüística culta de Las Palmas de Gran Canaria*. Universidad de Las Palmas de Gran Canaria.
- Mendoza, J. (1996). *Léxico del habla culta de La Paz*. Talleres Stampa.
- Moreno Fernández, F. (2020). Los ismos nacionales de la lengua española. *Boletín de la Real Academia Española*, 100(321), 115-146.
- Otálora de Fernández, H. (dir.) (1997). *Léxico del habla culta de Santafé de Bogotá*. Instituto Caro y Cuervo.
- Patrimonio Léxico da Gallaecia | Estudo da variación lingüística*. (s. f.). Patrimonio Léxico de Gallaecia. Recuperado 15 de octubre de 2025, de <https://ilg.usc.gal/plg/>
- PRESEEA | *Presentación*. (s. f.). <https://preseea.uah.es/>
- Quilis, A. (1983). Galicismos en la lengua hablada de Madrid. En *Philologica hispaniensia in honorem Manuel Alvar* (Vol. 1, pp. 537-548). Gredos.

- Rabanales, A. (1983). Términos de base indígena y extranjera en el léxico relativo al cuerpo humano del habla culta de Santiago de Chile. En *Philologica hispaniensi: in honorem Manuel Alvar* (Vol. 1, pp. 549-564).
- Rabanales, A. (2004). La encuesta como técnica dialectológica. *Onomázein*, 9, 75-93. <https://doi.org/10.7764/onomazein.9.04>
- Rabanales, A. & Contreras, L. (1987). *Léxico del habla culta de Santiago*. México. Universidad autónoma de México.
- Real Academia Española. Banco de datos (CORPES XXI) (s. f.). *Corpus del Español del Siglo XXI (CORPES)* <https://www.rae.es/>
- Salvador Salvador, F. (1991). *Léxico del habla culta de Granada*. Granada. Universidad de Granada.
- Samper, J. A. (Dir.) (1998). *Léxico del habla culta de Las Palmas de Gran Canaria*. Cabildo Insular de Gran Canaria y Universidad de Las Palmas de Gran Canaria.
- Sánchez García, F. J. (2018). El índice de riqueza léxica en el nivel de estudios bajo del corpus PRESEEA-Granada. *Itinerarios*, 28, 95-107.
- Santana Marrero, J. (2014). Diferencias léxicas en el campo de la alimentación entre la norma culta de Sevilla y de Las Palmas de Gran Canaria. *Anuario de Letras. Lingüística y Filología*, 2(2), 59-99. <https://doi.org/10.19130/iifl.adel.2.2.2014.78>
- Sedano, M. & Pérez, Z. (1998). *Léxico del habla culta de Caracas*. Universidad Central de Venezuela. Consejo de Desarrollo Científico y Humanístico.
- Torres, J. C. de (1981). *Encuestas léxicas del habla culta de Madrid*. Consejo Superior de Investigaciones Científicas.
- Ueda, H. (2009). Resultados y proyectos en las investigaciones sobre variación léxica del español. *Dialectología*, 2, 51-80.
- Ueda, H. (2011). Nuevo método para la recogida de datos de variación léxica. Encuestas en web en el proyecto VARILEX. En A. Cestero, I. Molina & F. Paredes (eds.), *La lengua, lugar de encuentro. Actas del XVI Congreso Internacional de la ALFAL* (pp. 3733-3740). Editorial Universidad de Alcalá.
- Valencia, A. (1984). Las comidas de un día ordinario. Análisis de un campo léxico. *Anuario de Letras. Lingüística y Filología*, 22, 235-250.
- VARILEX. *Variación léxica del español en el mundo*. (s. f.). <https://h-ueda.sakura.ne.jp/varilex/index.html>