

CRISTIAN DÍAZ RODRÍGUEZ

EL DICCIONARIO DE VOCES AMERICANAS DE AYALA (1777) EN EL TLEAM: SACRIFICAR EN ARAS DE HOMOGENEIZAR

Universidad de La Laguna

Resumen

El *Tesoro lexicográfico del español en América (TLEAM)* aspira a reunir, en una misma plataforma, todas las obras lexicográficas que den cuenta del español hablado en América, lo que exige la homogeneización, tanto a nivel micro como macroestructural, de las obras repertoriadas. A partir del *Diccionario de voces americanas de Ayala (1777)*, describiremos cómo se ha operado para conseguir un equilibrio entre el respeto a la fidelidad de los originales y la armonización exigida por la editorialización.

palabras clave: lexicografía, *TLEAM*, lematización, editorialización, armonización

Abstract

The Diccionario de voces americanas by Ayala (1777) in the TLEAM: to sacrifice in the name of homogenization

*The Tesoro lexicográfico del español en América (TLEAM) aims to bring together, on a single platform, all lexicographic works that account for the Spanish spoken in America, which requires the homogenization, both at the micro and macrostructural levels, of the listed works. Starting from Ayala's *Diccionario de voces americanas (1777)*, we will describe how efforts have been made to achieve a balance between respecting the fidelity of the originals and the harmonization required by editorialization.*

keywords: lexicography, *TLEAM*, lemmatization, editorialization, harmonisation

0. Introducción

En 2018, los lexicógrafos hispanistas Dolores Corbella, Alejandro Fajardo y Jutta Langenbacher-Liebgott editaron un libro que aspiraba a reunir la descripción de los proyectos más punteros en el ámbito de las Humanidades digitales pilotados por «labrantes de palabras»: *Historia del léxico español y Humanidades digitales*. Además de señalar las ventajas que presentaba la conjunción Informática-Filología, insistían ya en el prólogo de la obra en la necesidad de estar atentos a los desafíos que este maridaje iría revelando con el tiempo:

Estos avances, que han traído consigo un indiscutible progreso, no están exentos de algunos problemas que se derivan de su concepción y de su funcionamiento. La dispersión de los recursos, la falta de interconexión, la carencia de estándares, la rápida obsolescencia, etc., dificultan el acceso a la información que contienen. Su diseño, por otra parte, no siempre resulta adecuado para determinadas investigaciones. Todo esto hace que sea necesaria una crítica especializada, desde el punto de vista del usuario filólogo, que contribuya a señalar sus puntos débiles para poder mejorar las próximas versiones y los nuevos productos (2018: 5-6)

Ciertamente, uno de los mayores retos a los que nos enfrentamos a la hora de crear herramientas o recursos digitales es el de la homogeneización exigida por la *editorialización*, entendida en su sentido más estricto¹. En efecto, las enormes posibilidades que nos ofrece el manejo informatizado de cantidades ingentes de datos conllevan, por desgracia, una pérdida tributaria a la especificidad de cada fuente original, un pequeño precio que hay que pagar en aras de garantizar el tratamiento eficaz de la información manejada, pero que tiene una impronta mayor desde una perspectiva patrimonial.

Évidemment, cela a aussi un impact sur les contenus eux-mêmes : le concept d'*éditorialisation* souligne comment la technologie façonne les contenus. Selon cette définition, on serait tenté d'assimiler l'*éditorialisation* à la curation des contenus (*digital curation*) – qui désigne le processus d'organisation des contenus dans un environnement numérique déterminé. Mais il y a une distinction fondamentale entre les deux processus : le concept d'*éditorialisation* implique une dimension culturelle qui n'est pas présente dans l'idée de curation. Cette dernière renvoie plutôt aux pratiques visant à rassembler, organiser et afficher des contenus dans un environnement donné – ce qui met d'abord l'accent sur l'ensemble des compétences nécessaires pour réaliser une

¹ [L']*éditorialisation* est une instance de mise en forme et de structuration d'un contenu dans un environnement numérique. On pourrait dire, en ce sens, que l'*éditorialisation* est ce que devient l'édition sous l'influence des technologies numériques. (Vitali Rosati 2020 § 31)

bonne curation. En revanche, l'éditorialisation fait référence à la façon qu'ont les outils, les pratiques émergentes et les structures déterminées par les outils d'engendrer une relation différente avec les contenus eux-mêmes. On pourrait dire que la curation désigne l'action d'un individu spécifique ou d'un groupe d'individus défini, tandis que l'éditorialisation met l'accent sur la façon dont cette action est structurée par les caractéristiques de l'environnement numérique. Il faut enfin souligner que ces caractéristiques ne sont pas seulement techniques, mais aussi culturelles. (Vitali Rosati 2020 § 32)

En este artículo nos centraremos en el análisis de cómo se ha tenido que proceder, en el caso de la elaboración del *Tesoro lexicográfico del español en América*, al que nos referiremos por sus siglas *TTEAM*², para conciliar la homogeneización de los materiales reunidos con la salvaguarda de la especificidad metalexicográfica de cada una de las fuentes almacenadas. En efecto, el ejercicio de digitalización homogeneizadora, indispensable para garantizar la eficiencia de las eventuales consultas, exige una reflexión profunda en torno a diversos aspectos relacionados tanto con la variedad identificada en el nivel macroestructural como con aquella propia de la microestructura de las obras lexicográficas: órdenes macroestructurales no necesariamente alfabéticos, ortografías vacilantes, lematización geminada, mezcla de categorías gramaticales en la lematización, falta de sistematización de las lematizaciones anidadas... Describir cómo se ha procedido en el *TTEAM* para dar respuesta a cada uno de estos desafíos es el objetivo principal de este artículo. En concreto, nos valdremos de la forma en la que ha sido adaptado el *Diccionario de voces americanas* de Ayala (1777) para su introducción en la susodicha plataforma tesorística.

Tras hacer una pequeña descripción del proyecto *TTEAM*, así como de la obra que hemos seleccionado para ilustrar el proceso de homogeneización, nos centraremos en la presentación y discusión de las diferentes decisiones metodológicas que se han tenido que implementar para poder garantizar su inclusión en la plataforma de hospedaje y aumentar la eficiencia de su consulta a través de los procedimientos de búsqueda previstos.

2 Esta publicación forma parte de los proyectos de I+D+i PID2020-117659GB-100 y PID2023-149847NB-I00, financiados por el MCIN/AEI/10.13039/501100011033.

1. El *TLEAM*: una estructura para salvaguardar el léxico del español de América

El *Tesoro léxico del español en América-TLEAM* (cfr. Corbella Díaz; Díaz Rodríguez 2023; Corbella Díaz, Fajardo Aguirre, Díaz Rodríguez 2024) es un proyecto que, siguiendo la estela de la tradición tesorística hispánica, inaugurada por Gili Gaya y su *Tesoro Lexicográfico (1492-1726)*, aspira a convertirse en un repositorio lexicográfico del español en América y, por consiguiente, en un depositario de la memoria patrimonial léxica de las regiones hispanófonas del Nuevo Mundo³. El proyecto se articula en tres fases de volcado de materiales lexicográficos a la plataforma: la primera, ya concluida, incluía las obras publicadas hasta 1884, fecha en la que la Real Academia Española sistematizó, por primera vez, la marcación diatópica, introduciendo, además, el concepto de *americanismo*; la segunda, actualmente en marcha, incluye las obras datadas entre 1884 y 1925, año marcado por la publicación de la decimoquinta edición del *Diccionario de la Lengua Española* (conocido tradicionalmente como *DRAE* hasta su edición de 2014); la tercera, por su parte, tomará como fecha de cierre la creación de la Asociación de Academias de la Lengua Española (ASALE) en 1951⁴.

En esta infraestructura se almacenan, en formato digital, los testimonios diferenciales que han visto la luz en la América hispana, con independencia de que puedan catalogarse exactamente como obras lexicográficas *stricto sensu*: desde diccionarios generales o vocabularios a glosas a listados con disposición lexicográfica. En efecto, el *TLEAM* aspira a ser una herramienta de la que se puedan valer tanto el usuario profano como el especializado, lo que exige un alto grado de versatilidad en el diseño de sus prestaciones. Por ello, además de contar con un sistema de búsqueda por lema al uso, o guiado por índice, se ha adoptado un rico sistema de marcación textual, basada en estándares de codificación TEI basado en metalingüaje XML, implementado a través de la plataforma TEITOK (Arrabal Rodríguez 2022), que aumente, por un lado, las posibilidades de interrogación del corpus (etimología, información gramatical, lemas, forma normalizada, variantes,

³ En este sentido, bien es cierto que se cuenta ya con algunos proyectos —finalizados o *in fieri*— que tienen por objetivo el crear «diccionarios de diccionarios» que den buena cuenta del español de ciertas zonas geoléctales, con una perspectiva diacrónica, p. ej.: *Tesoro Lexicográfico del español de Puerto Rico (Tesoro. pr)* o el *TLEX-Tesoro Lexicográfico de Cuba* (Camacho Barreiro 2010) pero ninguno aspira a cubrir toda el área americana de habla hispana.

⁴ Aunque sería deseable completar la horquilla temporal cubierta con obras posteriores a 1951, lo cierto es que, esto implicaría el tener que depender enormemente de los permisos de propiedad intelectual o explotación de cada una de las obras recogidas, lo que nos hace desistir ante tal ímpetu desafío.

ejemplos, acepciones, información fraseológica, marcas de variación, categoría ontológica, zona geolingüística...); y, por otro, la personalización del corpus de búsqueda en función del objetivo perseguido. En otras palabras, al posibilitar la parametrización de los subcorpus de trabajo, así como las posibilidades de búsqueda, para adaptarlas a los objetivos individuales de cada usuario, el *TTEAM* funcionaría como un auténtico *hub* multicorpus y multipropósito. Es justamente esta faceta de corpus consultable y explotable —concebido en términos de Rastier (2004)⁵—, la que exige una homogeneización o, si se prefiere, una armonización, a varios niveles, del material almacenado, de manera que se permita una búsqueda y una explotación eficaz, al mismo tiempo que se asegure una visualización clara y ergonómica de la información volcada, que salve las especificidades tributarias a la diversidad introducida por la disparidad de documentos reunidos, cuya materialización puede afectar a distintos niveles: diferencias diamésicas, multiplicidad de criterios para organizar la macroestructura, falta de sistematicidad en la lematización, descarte de los topónimos y antropónimos, variación ortográfica en diacronía, discrepancias geolingüísticas, etc.

Ahora bien, a pesar de que cualquier ejercicio de homogeneización exige un sacrificio en detrimento de la fidelidad absoluta de la reproducción de los originales, no es menos cierto que:

En cualquiera de los casos, esta renuncia procedimental no es óbice para que el investigador interesado en estos aspectos pueda acceder a la disposición primigenia, puesto que el *TTEAM* ofrece la posibilidad de acceder a la ficha documental [...] de cada una de las obras que acoge, en la que [...] se permite el acceso tanto a la información paratextual [...]⁶ como a la copia facsímil o escaneada del documento original, cuando esta obra esté digitalizada. (Corbella Díaz, Fajardo Aguirre, Díaz Rodríguez 2024: 281)

1.1 *El Diccionario de voces americanas de Manuel José de Ayala (1777)*

Antes de proceder a escrutar algunas de las intervenciones realizadas durante el proceso de editorialización del *TTEAM*, haremos una pequeña descripción de la

5 “Un corpus est un regroupement structuré de textes intégraux, documentés, éventuellement enrichis par des étiquettes et rassemblés : (i) de manière théorique réflexive en tenant compte des discours et des genres, et (ii) de manière pratique en vue d'une gamme d'applications”.

6 La versatilidad a la que se aspira es tal que también se pone a disposición del usuario la consulta del material liminar paralexicográfico, como pueden ser las introducciones, las notas al lector, los prólogos o los prefacios.

obra de la que nos valdremos para exemplificarlas: el *Diccionario de voces americanas* de Manuel José de Ayala (1777). Este es el título que se le asignó a un manuscrito, compuesto por dos tomos —II-884 y II-855, de 455 folios numerados y 141 folios sin numerar respectivamente— conservado en la Biblioteca del Palacio Real de Madrid dentro del catálogo de *Manuscritos de América* con el título de *Diccionario de América* y que fue editado por Miguel Ángel Quesada Pacheco para Arco/Libros en 1995. Aunque se desconoce con certeza el autor de esta obra, se suele atribuir a Manuel José de Ayala⁷, ya que, además de haber estado estrechamente ligado a la Corte⁸—lo que habría posibilitado que su obra se conservara en la Biblioteca del Palacio Real—, existen similitudes gráficas y estilísticas, amén de una interreferencialidad notable entre esta obra y su *Cedulario Índico* y su *Miscelánea*. No en vano, este diccionario figura en la relación de méritos y servicios de Ayala.

A pesar de que no llegó a publicarse en el momento de su redacción, a saber, el siglo XVIII, se puede presuponer que se elaboró con esta finalidad pues, como nos recuerda en el prólogo de la edición Quesada Pacheco, pese a que hay folios completamente en blanco: “[e]l libro fue concebido de antemano, ya que casi todos los folios contienen letras de ordenamiento, donde se irían a poner los artículos de acuerdo con el criterio metodológico establecido por el que concibió la obra” (1995: x), lo que implica que los folios se iban llenando progresivamente a medida que se fuese recopilando el material. De hecho, “a lo largo del manuscrito hay diferentes fechas que consignan el día en que se introdujo la palabra, lo cual fortalece la idea de su composición gradual y paulatina” (Quesada Pacheco 1995: xi). Justamente, gracias a la consignación de estas fechas, ha sido posible establecer la fecha, o, mejor dicho, la horquilla temporal, de elaboración de este proyecto lexicográfico: entre 1751 (s.v. *bejuquillo*) y 1777 (s.v. *holpatán*).

7 Otras hipótesis consideran a Alcedo como potencial autor de esta obra fundamental para la lexicografía hispanoamericana, pues se arguyen coincidencias parciales entre esta y otra cuya autoría sí que está comprobada, a saber: *Diccionario geográfico histórico de las Indias Occidentales* 1786-1789. Sin embargo, las discrepancias estructurales y estilísticas parecen descartar esta autoría.

8 Manuel José de Ayala fue ilustre recopilador de las leyes de las Indias, así como archivero de la Secretaría de Estado del Despacho Universal de Indias. Para más inri, en el artículo dedicado a la voz *anaquine* se hace referencia explícita a la profesión del supuesto artífice de este diccionario: “Esta noticia comprobé en los papeles del Archivo de mi cargo, en carta de aquel govern[ad]or d[o]n Jossep L Antonio Gil de 26 de sep[tiem]bre de 1763, escrita al ex[celentísi]mo S[eñ]or Bo. F. d[o]n Julián de Arraiga remitiendo un cajoncito con 14 beas de ella que tuve en mi mano”. (Tomo I, f. 49 v. *apud* Quesada Pacheco 1995: 13).

2. Retos de la homogeneización exigida por la *editorialización*

Sin que podamos abordar de manera exhaustiva todas las adaptaciones que, para la elaboración del *TLEAM*, hemos tenido que operar en los diferentes textos-meta de carácter lexicográfico con la intención de poder garantizar un equilibrio entre fidelidad / autenticidad del documento original, por una parte, y la eficacia de búsqueda / la organicidad de visionado de la información, por otra; presentaremos, a continuación, a título de ejemplo representativo, una serie de adaptaciones que han tenido que operarse en el texto transcritto a partir de la obra original de Manuel José de Ayala, *Diccionario de voces americanas*, para poder garantizar su integración efectiva en la plataforma de custodia y consulta *TLEAM*.

2.1 Sobre el orden determinado por la macroestructura

Una de las principales características del *Diccionario de voces americanas de Ayala* es la forma en la que se disponen los lemas en la macroestructura, pues no siguen el orden alfabético al uso, sino según una aplicación parcial del mismo: se ordenan las entradas según la disposición gráfica de las dos primeras letras que componen cada vocablo —aa, ab, ac, ad...—, sin importar si conforman o no una única sílaba; pero, a partir de ahí, no se sigue un orden marcado por el abecedario⁹, lo que produce que nos encontramos con secuencias como: AL {*alcribís, aloes, alhueuetes, algarrobas, alabastro*}; JA {*jagüelles, jaboncillos, jalsontles, jayo, Jaén, Janta, Jardín del rey, Java Mayor y Menor, Juana / Cuba / Fernandina, Jayna, Joujuas, Janos, Jalapa, Jabalí, Jayba*}. Consciente de esta dificultad, en su edición, Quesada Pacheco, completa la reproducción del diccionario con un índice alfabético de palabras:

Para facilitar la búsqueda de un término, se ha hecho un índice alfabético de todas las entradas del manuscrito, al final del libro. En él se han incluido también los sinó-

9 Al optar por esta particular disposición del lemario, Ayala dotaba de flexibilidad a su obra, pues proponía un sistema de repertoriado a caballo entre el diccionario y un archivo vertical o base de datos *avant la lettre*, idóneo para una alimentación paulatina, no necesariamente secuencial. Su contrapartida, como hemos señalado, la dificultad de su consulta dada la imprevisibilidad de la secuenciación de lemas. En cualquiera de los casos, si tenemos en cuenta otros aspectos que también apuntan a la poca sistematización de la obra —p. ej.: diversidad de caligrafías, presencia de rasgos ortotipográficos vacilantes, variedad de tintas, uso caprichoso de cursiva y letrea de imprenta, disparidad en el esmero caligráfico...—, todo parece apuntar a que esta no sería la disposición final prevista por el autor, sino tan solo una forma de facilitar su trabajo de recopilación.

nimos y otras palabras de interés (nombres propios, americanismos, indigenismos) que figuran dentro de un artículo, pero que no aparecen como entrada, y las entradas que presentan variaciones gráficas en la actualidad. En vista de que se ha respetado la escritura y el orden de entradas del manuscrito, se recomienda consultar el índice antes de pasar al diccionario [...]. Con esto se facilita la búsqueda sin correr el riesgo de pasar por alto algún artículo. (1995: xliv)

Desde el punto de vista de la editorialización, esta peculiaridad no presenta ningún problema a la hora de integrar la información en el *TLEAM*, pues el almacenamiento de los diferentes lemas puede llevarse a cabo siguiendo la disposición propuesta por Ayala, lo que permite garantizar cierta fidelidad al consultar el lemario de manera íntegra, como se muestra en la Imagen 1.

The screenshot shows a search results page for the term "guardas". The results are listed in a hierarchical structure:

- edit guardas**
 - 1. Piedras que están a los lados de la veta y en el cielo de labor. Nueva España.
- edit guijo**
 - 1. Perno de hierro sobre que estriva el mortero.
- edit guatanas**
 - 1. Son las ligas para atarse las medias en la sierra del Perú.
- edit guapapas**
 - 1. Llaman los naturales de Potosí y los de otras partes de aquel Reino de el Perú a los hornos para el beneficio de los metales. Son semejantes a los hornos castellanos, y solo se diferencia en que por todas partes están aquellos llenos de agujeros por donde entra el ayre quando el viento sopla y entonces solo pueden fundir.
- edit guavina**
 - 1. Es un pescado muy regalado de la América.
- edit guajiros**
 - 1. Son unos yndios comunitarios a la ciudad del río de la Hacha, que al menor antojo, le hacen hostilidades por que no guardan lealtad ni consecuencia.

Imagen 1: ejemplo de secuenciación de lemas que no siguen el orden alfabético.

Ahora bien, en principio, el común de los usuarios nunca accederá al contenido de las palabras de manera secuencial (tampoco alfabética) sino a través de un motor de búsqueda que permita mostrar en pantalla la información solicitada. La única salvedad la podríamos encontrar a la hora de realizar una búsqueda a través de una eventual rueda de deslizamiento (ver imagen 2) en la que aparecieran todas las palabras contenidas en nuestra plataforma; posibilidad, a nuestro juicio, poco operante, dado que este lemario general forzosamente tendría que seguir un orden alfabético estricto, que iría en contra de la disposición seguida por Ayala.

The screenshot shows the TLEAM search interface. At the top, there's a logo with 'TLEAM' and language links 'ES | EN'. Below it is a navigation menu with links like 'Inicio', 'Equipos', 'Fuentes', 'Publicaciones', 'Ayuda', 'Contacto', 'Modo de cita', 'Congresos', 'user: CD', 'Admin', 'Help', 'XML Files', 'Búsqueda en XDXF', 'Example SQL', 'Editor de diccionarios', and 'Búsqueda'. The main area has a title 'Buscar en los diccionarios' and a search bar with 'guayna' entered. A 'Búsqueda' button and a 'mostrar opciones' link are next to it. The results are presented in three cards:

- 1777 Ayala AM - edit**: Definition of 'guayna' as '1. Es un pescado muy regalado de la América.'
- 1876 García RD - edit**: Definition of 'guavina' as '1. Pez de río.'
- 1894 Gutiérrez González CO-ANT - edit**: Definition of 'guavina' as '1. Canción provincial.'

To the right, there's an 'Índice' sidebar with a list of words: guauro, guausí, guava, guavaloca, guavica, guavilla, **guavina** (highlighted in green), guaxarapos, guaxiro, guaxoti, and guay.

Imagen 2: interfaz de búsqueda a través de un motor de búsqueda y de un lemario deslizante global ordenado alfabéticamente.

2.2 Sobre la ortografía vacilante

En general, uno de los mayores retos a los que nos enfrentamos durante un proceso de editorialización que abarque una horquilla temporal tan amplia como la propuesta por el *TLEAM* es el de intentar repertoriar, y brindar un trato similar, a todas las variantes gráficas de una palabra recogidas en diacronía, en muchos casos atribuibles a la existencia de una norma poco marcada o, incluso, totalmente ausente. En el caso del *Diccionario de voces americanas*, las principales razones de la abundante variación gráfica se deben a la falta de definición clara de los valores y usos de ciertos grafemas, así como a una vacilante correlación entre fonemas y grafemas:

- Confusión entre “i” e “y”, pues no se reserva el primero de estos grafemas para los valores vocálicos y el segundo, para los consonánticos. Así, por ejemplo, si observamos simplemente el caso de las palabras que comienzan gráficamente por “i” o “y”, podemos constatar la cohabitación de lemas que efectivamente respetan la repartición funcional preconizada en el siglo XVIII —valor vocálico de “i”: *ixistineses*, *iztaes*, *icotea*, *inca* o *inga*, *incienso*, *indio timaba*; y valor consonántico de “y”: *yanaconas*, *yareta*, *yamas*, *yato*, *yacaré*, *yamacuá*, *yalamacuá*, *yapayaes*, *tabaqué*, *Yabucoa*, *yaquino*, *yaqui*, *yare*, *yayrepa*, *yantayanta*, *yavar chunca*, *yeguana*, *yerba*— con numerosos ejemplos que contravienen esta distinción y que, por tanto, reflejan un valor consonántico de “i” —p. ej.: *iantem*— y un valor vocálico de “y”:

Ybagui, ybagué, Ybueras (o *Higüeras*), *ycho, ychuris, ycacos, yca, ycho, yllo, ylbol, ymagua, ymperial, yncorporadero, yndehé, ynguarán, ynfieles, yngenos, indios*. En algunas ocasiones, hay que añadir un elemento suplementario que también interviene en el paradigma, ora la “j” —*icotea* vs. *jicotea*—, ora la “ll”: *yamas, iantén*.

- Uso vacilante de la “h”, tanto inicial como interno-etimológica: *hamana, icotea, Carthago, uipeles, higuana*.
- Confusión de la “b” y la “v”: *bálsamo perubiano, báquiras, Córdova, Varco* (Santiago del Estero).
- Inestabilidad en el sistema de sibilantes: *senzontle, passote, Epazote*.
- Falta de sistematicidad en el uso de la “e” protética: *Spíritu Santo*.
- Hesitación para representar gráficamente el fonema /j/: *Xeniquén / He-nequén, jurón / hurón, Xaca, Xagua, Xaguay, Xalisco, Xamayca, xengibre, Xihuitl (Jigüite)*.

Esta situación se resuelve, en la mayoría de casos, consignando una de las variantes gráficas —generalmente la aceptada sincrónicamente por los diccionarios de referencia, en nuestro caso particular el *DLE* o el *DAmer*— como forma normalizada (<k reg>), por ejemplo *pava*, y enlazar todas las variantes registradas a esta, en este caso (cfr. tercera línea de código en la Imagen 3) *pabas*, de manera que, al lanzar una búsqueda, ese lema sea recuperable tanto si se introduce esa variante en el motor de búsqueda como si se introduce la forma normalizada.

```
</deftext></def></ar><ar id="ar-999"><k reg="payagu&#xE1;n">payaguanes</k><def id="ar-999.1" n="1"><
</deftext></def></ar><ar id="ar-1005"><k reg="mora">paniquiris</k> o <k>moras</k><def id="ar-1005.1"
</deftext></def></ar><ar id="ar-1007"><k reg="pava">pabas silvestres </k><def id="ar-1007.1" n="1"><
</deftext></def></ar><ar id="ar-1010"><k reg="palometa">paguara</k> o <k>palometá</k> o <k>payara</k
</deftext></def></ar><ar id="ar-1011"><k reg="payaguarima">payaguarima</k><def id="ar-1011.1" n="1">
</deftext></def></ar><ar id="ar-1012"><k reg="palo">palo amarillo</k><def id="ar-1012.1" n="1"><gr v
```

Imagen 3: código XML crudo para vincular las variantes de un lema con una forma normalizada común.

2.3 Sobre la homogeneización de la lematización

El asociar cada una de las variantes de los lemas a una forma normalizada también es la solución que se ha implementado para dar respuesta a la falta de sistematici-

cidad de la lematización. En efecto, contrariamente a lo que se preconiza actualmente en la práctica lexicográfica, según la cual el lema debe aparecer, si es un verbo, en su forma infinitiva; o en su forma masculina singular, eventualmente acompañada de los morfemas de género y, rara vez, de número, cuando se trata de adjetivos o sustantivos, se comprueba que, en el proceso de editorialización del *TTEAM*, se ha apostado por mantener la variante original del lema que sirve de entrada para cada artículo. Así, por ejemplo, como se observa en la Imagen 3, se ha conservado como lema *payaguanes*, en plural, pero se ha asociado a la forma normalizada *payaguán*, en singular, de manera que la interfaz de volcado de resultados mostrará los mismos resultados con independencia de que el usuario introduzca *payaguán* o *payaguanes* en el motor de búsqueda. Esto no se cumple en el sentido inverso, es decir, mientras que una búsqueda a partir de la forma normalizada *payaguán*¹⁰ proporcionará al usuario todas las entradas encabezadas tanto por *payaguán* como por *payaguanes* (cfr. Imagen 4), una búsqueda a través de la forma *payaguanes* solo volcará la información relacionada con las entradas cuyo lema también aparezca en plural, obviando, pues, las que solo aparezcan en singular.

The screenshot shows the TTEAM search interface. On the left, there's a sidebar with links for Inicio, Equipos, Fuentes, Publicaciones, Ayuda, Contacto, Modo de cita, Congresos, user: CD, Admin, Help, XML Files, and Búsqueda en XDXF. The main area has a search bar with 'payaguán', a 'Búsqueda' button, and a 'mostrar opciones' link. Below the search bar is a dropdown menu showing '1777 Ayala AM - edit'. Underneath is a section for 'payaguanes' with the definition '1. Nación de yndios con los antecedentes habitantes las mismas provincias.' To the right is an 'Índice' sidebar with a list of words: paya, payaco, payada, payada de contrapunto, payador, payadura, payaguá, and payaguanes. The words 'payaguá' and 'payaguanes' are highlighted in green.

Imagen 4: volcado de la información del lema *payaguanes* al que se ha accedido a través de la forma normalizada *payaguán*.

10 Sin embargo, como se puede apreciar en la Imagen 4, la variante *payaguán*, al no haber sido reconocida como lema de entrada en ninguna de las obras registradas hasta ahora, no aparece en el lemario deslizante global, por lo que, el acceso a la información a través de este procedimiento —clicar en el lemario deslizante— estaría bloqueado para las formas normalizadas que no cuentan con una variante de un lema isomórfico registrado.

2.4 Sobre la lematización geminada

La asociación de las distintas variantes con una forma normalizada, que, en los casos anteriores, ha resultado ser tan satisfactoria, en el caso de la lematización geminada —p. ej.: *acmuli / amole; almorajifazgo / diezmos; antena / entena*—, en cambio, presenta limitaciones. Este escollo es especialmente significativo dada la alta frecuencia de entradas con lematización geminada a las que recurre Ayala, adoptando, para ello, distintas disposiciones:

- Enumeración yuxtapuesta: *Juana, Cuba, Fernandina.*
- Enumeración copulativa: *Java Mayor y Menor.*
- Enumeración disyuntiva total: *Arbinas, o tembladeras; Bejuquillo, hebecaguana, o raicilla; Pacae o pacay.*
- Enumeración disyuntiva parcial: *bálsamo negro, o [bálsamo] peruviano.*
- Reenvío hiperonímico o paronímico: *Páquira / báquiras: vide Jabalí*
- Amalgama de distintas categorías gramaticales: *demora o demorados*

Si observamos, nuevamente, la Imagen 3, en concreto la segunda y la cuarta líneas de código, se puede comprobar cómo las entradas vinculadas respectivamente con las formas normalizadas *mora* y *palometa*, corresponden a las entradas con lematización geminada: *panquiris / moras y paguara / palometa / payara*, cada una marcada como co-lemas a través de la etiqueta “k” entre las que, por cuestiones de legibilidad, se ha introducido la conjunción disyuntiva “o”. La limitación a la que acabamos de aludir consistiría en la imposibilidad de asociar cada uno de los co-lemas a una forma normalizada casi-isomorfa, teniendo que seleccionar una única *k reg.*

The screenshot shows the TTEAM search interface. At the top left is the TTEAM logo with 'ES | EN' options. The main search bar contains the word 'mora'. Below it are three search results:

- 1697 Páramo y Cepeda PA - edit**
mora.
1. Vestidura que usan los Yndios de echura de sobrepelliz, sin mangas.
- 1777 Ayala AM - edit**
paniquiris o moras
1. Este es un ramo de data de la proveeduría de Panamá (que oy corre al cargo de oficiales *reales* por la extinción del empleo de proveedor que hizo el Rey) y significa el gasto que causan a la Real Hacienda los yndios de paz, assí dáríenes de las vertientes del norte y sur, como los guaymies de las Bocas del Almirante y Toro a barlovento del Exculo de Veraguas, a quienes por venir a avisar la pacificación y novedades que ocurren en sus poblaciones se les assiste por quatro días con dos *rreales* diarios y quatro al graduado de capitán, y al tiempo de su regreso con dos varas de bayeta, dos de coleta, un manojo de tabaco, un cuchillo o machete, y un poco para gastos del viage, y todo se entiende por paniquiris y moras.
- 1789 Velasco EC - edit**
moras
1. Il libro 2º Deino venetiale in 3216.º De las frutas comestibles de diversas plantas y árboles in 621 Moras. Δ

Imagen 5: interfaz de salida de búsqueda en el caso de buscar a través de una palabra normalizada ligada a un lema geminado.

La consecuencia de esta limitación es especialmente significativa en el momento de volcar los datos resultantes de la búsqueda. Así, por ejemplo, la información contenida en el artículo encabezado por el lema geminado *paniquiris o moras*, se mostrará cuando el usuario introduzca en el motor de búsqueda los valores “paniquiris” y “moras”—coincidentes con las variantes originales de los lemas, o el valor “mora”—carente de formas flexivas, que ha sido identificada como forma normalizada. Sin embargo, si el usuario interroga el motor de búsqueda a partir del valor “paniquiri”, forma, en principio, esperable según las preconizaciones de la lematización contemporáneas, no obtendríamos ningún resultado, puesto que no hay nada que lo vincule con nuestro artículo.

2.5 Sobre la duplicidad de lemas

La naturaleza *in fieri* del proyecto lexicográfico de Ayala bastaría para justificar la inclusión de repeticiones, que, presumiblemente, habrían desaparecido una vez se hubiera procedido a corregir sistemáticamente la obra con la finalidad de dotarla de su forma definitiva. En algunos casos, no hay duda de que se trata de duplicados no deseados por el autor, ya que se reproducen de manera exacta, o práctica-

mente sin alteraciones, las definiciones vinculadas a ciertos lemas. En otros casos, en cambio, la repetición es totalmente voluntaria, pues es la forma adoptada por el autor para enfrentarse tanto a los fenómenos de homonimia como a los de polisemía. Este es el caso, por ejemplo, de la voz *víbora*, que se encuentra definida tanto en el folio 100v como en el folio 101r, con sendas acepciones: “isletas junto a Xamayca, donde se han perdido por su peligro muchos navíos” y “en todos los países fríos y aun en los cálidos y templados de Yndias las hay en mucha abundancia, y de tan mortífera picadura como las de España; pero al mismo tiempo abundan aquellos parajes de antídotos contra su ponzoña.”; o de la voz *patí*, cuyas acepciones aparecen definidas hasta en tres páginas diferentes, cada una constituyendo un nuevo duplicado lemático: “es un pez de las veinte y dos especies de que se haze relación en la antecedente voz.” (f. 25v, s. v.); “río caudaloso, que pasa por la ciudad de Santa Fee en el Nuevo Reyno de Granada” (f. 30r, s. v.) y “es un lienzo de algodón que fabrican los yndios de la provincia de Yucatán; cuyas piezas tienen dos varas de largo y tres quartas de ancho, y su valor regular son doce reales. De estas dan dos anualmente todos los indios a sus curas por vía de limosna que han introducido estos.” (f. 32r, s. v.).

Para intentar paliar esta situación, en el *TLEAM* se ha optado por mantener la multiplicidad de entradas con independencia de que su separación responda o no a un caso de homonimia, lo que, atendiendo a la lexicografía moderna, sí que implicaría la duplicidad de las entradas. Si bien esta opción, en el caso del *Diccionario de voces americanas* de Ayala, parece responder a la necesidad de respetar fielmente la obra original, esta decisión no deja de tener consecuencias bajo una percepción global del proyecto, ya que, como resultado, en el *TLEAM*, podemos encontrar duplicidad de entradas que reflejan etimologías diferentes, duplicidad de entradas que nada tienen que ver con la etimología, como en Ayala, sino que dan respuesta a un tratamiento organizado por acepciones y, por último, entradas que se articulan en diferentes acepciones. Desde el punto de vista de la consulta de la información, esta situación no ofrece mayor dificultad, pues la solución pasa por considerar la acepción y no el artículo lexicográfico en sí, como unidad de muestra. Desde el punto de vista de la búsqueda simple, al no exigirse ninguna restricción, si se introduce una palabra, se mostrarán todas las acepciones ligadas a un lema coincidente con el valor de entrada, pero, desde el punto de vista de la búsqueda compleja, sí que se permite parametrar las restricciones, con lo que, al estructurar la búsqueda a nivel de las acepciones y no de los lemas, se garantiza una mayor precisión y exactitud en la selección de resultados que se han de mostrar en la interfaz de volcado de resultados. Como era de esperar, esta decisión

metodológica también conlleva algún efecto no deseable: el falseo del lemario, especialmente desde el punto de vista cuantitativo, aunque también, desde el punto de vista cualitativo, si el usuario decidiera optar por realizar una búsqueda a través del lemario deslizante, opción, en principio, como hemos apuntado, poco operante.

2.6 Sobre la anidación de elementos dependientes de un lema ordenante

En relación con el punto anterior, encontramos la imposibilidad de anidar, en el caso de la obra de Ayala, elementos dependientes, especialmente compuestos y locuciones, bajo un lema ordenante. Como sabemos, la lexicografía contemporánea, en particular aquella que se edita en papel, dispone de ciertas reglas estables para incluir los elementos poliléxicos dentro de artículos encabezados por lemas que funcionan como palabras ordenantes. Así, por ejemplo, tanto el compuesto *bola criolla* como la locución *dar bola* [a alguien] se incluirían dentro del artículo encabezado por *bola*, el nombre que funciona como núcleo o palabra ordenante. Esta estructuración, desde una perspectiva digital, es irrelevante, pues la información será recuperable siempre que se introduzca alguno de los elementos que integran la polilexía, con independencia del lugar en el que se almacene, pero como la filosofía de la editorialización que subyace al *TLEAM* aspira a encontrar un equilibrio entre la obra original, la tradición lexicográfica y la eficiencia en la explotación de la herramienta, se ha optado por consignar, como en la obra original, una entrada independiente a los términos poliléxicos, pero, al mismo tiempo, se han vinculado, mediante una falsa forma normalizada, al elemento que según la tradición lexicográfica tendría que funcionar como palabra ordenante. Esto se puede comprobar si se observa, una vez más, la Imagen 3, en la que se puede apreciar cómo los compuestos *pabas silvestre* o *palo amarillo* se vinculan respectivamente a *pava* y *palo*. Como se señalaba en el apartado anterior, esto produce una multiplicación en la casuística repertoriada, ya que, en las entrañas del *TLEAM*, conviven elementos poliléxicos que encabezan entradas independientes y otros que se encuentran anidados dentro de los artículos lexicográficos encabezados por los lemas que funcionan como palabras ordenantes, en función de la opción elegida para la elaboración de cada una de las obras originales aquí editorializadas. Sin embargo, nuevamente, esta decisión metodológica parece estar justificada, pues el único efecto pernicioso que produce es de tipo cuantitativo, ya que, como señalábamos en el apartado anterior, se falsean los números referidos al lemario

global, falsedad que bien vale la enorme eficiencia y exactitud en la búsqueda de información que se consigue a cambio.

3. Conclusiones

La editorialización de cualquier obra implica priorizar ciertos elementos frente a otros, pero, cuando el proyecto al que se aspira es tan ambicioso como el *TLEAM*, en el que coincide la diversidad de materiales reunidos con la dificultad inherente al abarcamiento de una horquilla temporal tan amplia, puesta en correlación con una tradición lexicográfica en constante evolución, la priorización puede implicar directamente el renunciar a plasmar ciertos aspectos, generalmente relacionados con el respeto de la autenticidad y la fidelidad de la obra original, con el propósito de garantizar una gran eficiencia en el almacenaje, la búsqueda y la forma en la que se muestra la información al usuario. A través del escrutinio de la obra *Diccionario de voces americanas*, atribuida a Ayala, hemos podido comentar los mayores obstáculos a los que nos hemos enfrentado durante la implementación de la estructura en la que se sostiene el *TLEAM*, así como las soluciones que se han propuesto. Evidentemente, la enorme casuística que, casi de manera inherente, caracteriza este proyecto, hace que sea imposible anticipar todas las situaciones posibles, por lo que, en más de una ocasión nos vemos obligados a renunciar a los objetivos y motivaciones que sustentan este proyecto, pero, al mismo tiempo, creemos que una buena planificación, basada en el estudio detallado de los materiales y la anticipación ante eventuales adversidades, de las que en este artículo solo hemos presentado una pequeña muestra, dejando de lado otras cuestiones igualmente medulares en la concepción de cualquier estructura de este tipo (inclusión de los elementos onomásticos, etimologías, desarrollo de abreviaturas, comentarios internos de autores, tratamiento de ejemplos, etc.) son la base que garantizan el éxito de un proyecto de editorialización de tal envergadura.

Bibliografía citada

Arrabal Rodríguez, Pilar (2022), «TEITOK, a visual solution for XML/TEI encoding: editing, annotating and hosting linguistic corpora», *RIDE – A review journal for digital editions and resources* 15, doi: 10.18716/ride.a.15.5.

- Ayala, José Manuel de (1777), *Diccionario de voces americanas*. Edición de Miguel Ángel Quesada Pacheco (1995). Madrid, Arco Libros.
- Camacho Barreiro, Aurora (2010), «Tesoro lexicográfico de Cuba: un recorrido a través de la historia de las palabras (siglos XIX-XXI)», *Revista digital LibrínSula* 1-3 https://librinSula.bnjm.cu/secciones/300/nombrar/300_nombrar_2.html [15/12/2024].
- Corbella Díaz, Dolores; Díaz Rodríguez, Cristian (2023), “El Tesoro lexicográfico del Español en América: entre repositorio lexicográfico y memoria patrimonial”, *Perspectives de recherche en linguistique et philologie romanes*, eds. Dolores Corbella; Josefina Dorta; Rafael Padrón, Strasbourg, ELiPhi: 689-702.
- Corbella Díaz, Dolores; Fajardo Aguirre, Alejandro; Díaz Rodríguez, Cristian (2024), “TTEAM: Tesoro lexicográfico del español en América o cómo atesorar el patrimonio léxico en la era digital”, *Lexicografía del español: panhispanismo e internacionalización*, eds. Alejandro Fajardo Aguirre; Dolores Torres Medina; Cristian Díaz Rodríguez, Berlin, Peter Lang: 269-286.
- Corbella Díaz, Dolores; Fajardo Aguirre, Alejandro; Langenbacher-Liebgott, Jutta, eds. (2018), *Historia del léxico español y Humanidades digitales*, Berlin, Peter Lang, Studia Romanica et Lingüística.
- DAmer = Asociación de Academias de la Lengua Española, (2010). *Diccionario de americanismos*. <https://www.asale.org/damer/>
- DLE = Real Academia Española, [versión 23.8 en línea]. <https://dle.rae.es>.
- Rastier, François (2004), «Enjeux épistémologiques de la linguistique de corpus», *Texto!*, sección Dits et inédits. www.revue-texto.net/Inedits/Rastier/Rastier_Enjeux.html [01/12/2024].
- Tesoro.pr = Academia Puertorriqueña de La Lengua Española, (2016), *Tesoro lexicográfico del español de Puerto Rico*. <https://tesoro.pr> [24/12/2024]. Es la versión electrónica del texto de Vaquero, María / Morales, Amparo, 2005, *Tesoro lexicográfico del español de Puerto Rico*, San Juan: Academia Puertorriqueña de la Lengua Española.
- TTEAM = Corbella, Dolores, coord. (2021-), *Tesoro lexicográfico del español en América* [en línea]. Disponible en www.ull.es/tteam. [23/01/2025].
- Vitali-Rosati, Marcelo (2020), «Pour une théorie de l'éditorialisation», *Humanités numériques* 1. DOI : <https://doi.org/10.4000/revuehn.371>

Cristian Díaz Rodríguez es doctor en Estudios Filológicos y Ciencias del Lenguaje. Actualmente es profesor ayudante doctor en el Departamento de Filología Clásica, Francesa, Árabe y Románica de la Universidad de La Laguna. Sus investigaciones se orientan, principalmente, hacia los ámbitos de la fraseología y la paremiología francesa y española, así como de la lexicografía bilingüe. Ha sido coeditor de las obras *Lexicografía del español: intercomunicación y diálogos* y *Lexicografía del español: panhispanismo e internacionalización*.

cdiarodr@ull.edu.es