

**ELABORACIÓN DE UNA BASE DE DATOS EN XML PARA UN
DICCIONARIO BRIBRI–ESPAÑOL ESPAÑOL–BRIBRI EN LA WEB¹**

ELABORAÇÃO DE UMA BASE DE DADOS EM XML PARA UM DICIONÁRIO
BRIBRI–ESPANHOL ESPANHOL–BRIBRI NA WEB

ELABORATION OF A DATABASE IN XML FOR A BRIBRI–SPANISH SPANISH–
BRIBRI DICTIONARY ON THE WEB

Haakon Krohn²

Universidad de Costa Rica

Resumen: Este artículo describe la elaboración de una base de datos en XML para un diccionario bribri–español español–bribri consultable por medio de una página HTML dinámica en la web. El escrito se enfoca en las distintas soluciones tomadas con respecto a la estructura de los árboles XML, los elementos de información incluidos y la incorporación de enlaces y remisiones entre diferentes entradas. La base de datos consiste en dos documentos XML, uno para cada dirección del diccionario. En ambas estructuras, las entradas constituyen el primer nivel debajo del elemento radical. A su vez, cada entrada puede contener cualquier cantidad de acepciones, las cuales abarcan la información perteneciente a la microestructura. Las remisiones, utilizadas para evitar la redundancia de datos, se registran a nivel de las acepciones y se muestran como enlaces suaves o duros en la página web, dependiendo de ciertos factores. La estructura XML presentada aquí puede adaptarse fácilmente para otros diccionarios digitales y la visualización de los datos es completamente personalizable.

Palabras clave: lexicografía; XML; lengua bribri.

Resumo: Este artigo descreve a elaboração de uma base de dados em XML para um dicionário bribri–espanhol espanhol–bribri consultável por meio de uma página HTML dinâmica na web. O texto se concentra nas distintas soluções tomadas a respeito da estrutura das árvores XML, os elementos de informação incluídos e a incorporação de *links* e remissões entre entradas diferentes. A base de dados consiste em dois documentos XML, um para cada direção do dicionário. Em ambas as estruturas, as entradas constituem o primeiro nível abaixo do elemento radical. Por sua vez, cada entrada pode conter qualquer número de significados, que abrangem as informações pertencentes à microestrutura. As remissões, utilizadas para evitar redundância de dados, são registradas no nível das aceções e são mostradas como *links* suaves ou duros, dependendo de certos fatores. A estrutura XML apresentada aqui pode ser adaptada facilmente para outros dicionários digitais e a visualização dos dados é completamente personalizável.

Palavras-chave: lexicografia; XML; língua bribri.

Abstract: This article describes the elaboration of a database in XML for a Bribri–Spanish Spanish–Bribri dictionary that can be consulted through a dynamic HTML page on the web. The paper focuses on the different solutions taken regarding the structure of the XML trees, the

¹ Este artículo forma parte del proyecto de investigación 745-B8-131 “Diccionario digital bilingüe bribri (fase 1)”, inscrito en el Instituto de Investigaciones Lingüísticas de la Universidad de Costa Rica, con presupuesto de la Vicerrectoría de Investigación de dicha institución.

² Universidad de Costa Rica. E-mail: hkrohn@gmail.com.

information elements included and the incorporation of links and redirects between distinct entries. The database consists of two XML documents, one for each direction of the dictionary. In both structures, the entries constitute the first level below the root element. Each entry can in turn contain any number of acceptations, which include the information pertaining to the microstructure. The redirects, used in order to avoid data redundancy, are registered at the level of the acceptations and are displayed as soft or hard links on the web page, depending on certain factors. The XML structure presented here can easily be adapted for other digital dictionaries and the visualization of the data is completely customizable.

Keywords: lexicography; XML; Bribri language.

Introducción

La web constituye un canal óptimo para la publicación de diccionarios. No solo minimiza el espacio físico ocupado por la obra y facilita el acceso a la información, sino que también permite una presentación mucho más dinámica e interactiva de los datos que en un diccionario tradicional impreso. Con el fin de simplificar la elaboración y la publicación digital de los diccionarios, se han creado varios softwares y plataformas gratuitos que se encuentran disponibles en la web. En esta área destaca el Summer Institute of Linguistics, que ha distribuido varias herramientas lexicográficas digitales de uso libre, tales como Lexique Pro (SIL INTERNATIONAL, 2012), Language Forge (SIL INTERNATIONAL, 2020a) y Webonary (SIL INTERNATIONAL, 2020b).

Sin embargo, en todos estos casos, la libertad del lexicógrafo está delimitada por la estructura y las funcionalidades del software o de la plataforma. Aunque este marco generalmente permite cierta flexibilidad en cuanto a la estructura del diccionario y la presentación de la información, es muy difícil que se ajuste perfectamente a las necesidades de un proyecto en particular. Este inconveniente es con el que nos hemos enfrentado en el proyecto de un diccionario bilingüe digital bribri-español español-bribri. Dicho reto, en última instancia, nos instigó a crear una base de datos y nuestra propia plataforma desde cero.

Específicamente, elaboramos una base de datos consistente en dos documentos XML (Extensible Markup Language) para almacenar información lexicográfica que se visualiza de manera dinámica en un documento HTML (HyperText Markup Language) por medio de JavaScript, una técnica conocida bajo las siglas AJAX (Asynchronous JavaScript and XML). En el presente artículo, se describe la estructura de esta base de datos XML y se explican las distintas decisiones tomadas durante el proceso. Dado que existen muy pocas publicaciones acerca de la elaboración de bases de datos para diccionarios en línea, este trabajo puede servir como referencia para proyectos

lexicográficos similares. Por la misma razón, para los lectores no familiarizados con XML, se incluyen una introducción a este lenguaje de marcado y una breve explicación de cómo se extraen los datos mediante AJAX.

1. La lengua bribri

El bribri es una lengua de la familia chibchense. De acuerdo con la clasificación de Constenla (2008), pertenece al grupo ístmico, dentro del cual forma un subgrupo junto con el cabécar. La mayoría de los hablantes del bribri habitan en el sur de Costa Rica, pero la lengua también es hablada en territorio panameño. En el Censo Nacional de 2011, más de 18 000 personas en Costa Rica se autoidentificaron como bribris, de las cuales el 45,1 % afirmó hablar bribri (FUENTES, 2014, p. 332-333). A su vez, los Censos de Población y Vivienda de Panamá del 2010 muestran que aproximadamente 1000 bribris habitan en dicho país, pero no se incluyen datos lingüísticos (INSTITUTO NACIONAL DE ESTADÍSTICA Y CENSO DE PANAMÁ, 2015, p. 40). Relevante para este proyecto lexicográfico es el hecho de que el bribri presenta una notable variación dialectal. Se suelen reconocer tres áreas dialectales principales: el de Amubre y el de Coroma en la vertiente del mar Caribe, y el de Salitre en la vertiente del océano Pacífico (JARA, 2018, p. 11-13).

2. El proyecto de investigación

El proyecto se inscribió bajo el nombre “Diccionario digital bilingüe bribri (fase 1)” en el Instituto de Investigaciones Lingüísticas de la Universidad de Costa Rica en el 2018, con el propósito de elaborar un diccionario bribri–español español–bribri de uso libre en la web. Forma parte de un esfuerzo en aras de la conservación del patrimonio lingüístico de Costa Rica y pretende facilitar el acceso a la información lexicográfica del bribri para los hablantes, los investigadores, los estudiantes y cualquier otra persona interesada en esta lengua.

Durante la primera fase del proyecto, además de elaborarse la base de datos y el sitio web para visualizar la información del diccionario, se registraron todos los lexemas hallados en un corpus, junto con sus categorías gramaticales, variaciones ortográficas y definiciones. Dicho corpus comprende un conjunto de diccionarios, vocabularios y

textos en bribri, así como artículos de investigación sobre este idioma. En la segunda fase, programada para iniciar en el 2020, se trabajará con hablantes nativos para corroborar y corregir los datos registrados, así como para agregar información faltante.

3. Estado de la cuestión

Antes de dar inicio al proyecto, existía un diccionario impreso bribri–español español–bribri, desarrollado por Margery (1982, 1996). Este contiene 2619 entradas en la sección bribri–español, además de una sección español–bribri con entradas no enumeradas. También se habían publicado vocabularios más limitados en otras obras, de las que se pueden destacar Constenla y Margery (1978, 1979), Constenla, Elizondo y Pereira (1998), Jara y García (2003, 2009, 2013), Jara (2018) y varios diccionarios pictográficos temáticos desarrollados por el proyecto “Diversidad lingüística de Costa Rica” (DIPALICORI, 2020), coordinado por Carlos Sánchez Avendaño. Los materiales de Dipalicori, así como los libros de Jara y García (2003, 2013) y Jara (2018), están disponibles en la web en formato PDF. No obstante, todas estas obras, con la excepción de algunas de las producidas por Dipalicori, tienen el formato de libros impresos tradicionales y en ninguna se extrae la información de una base de datos, por lo que la presentación es estática.

4. Terminología

Debido a que los términos lexicográficos no siempre se comprenden de la misma manera, es conveniente definir brevemente la terminología utilizada en este artículo. En primer lugar, se hace uso de la noción *lexema* para aludir a una unidad de significado con abstracción de sus componentes gramaticales (cf. PORTO, 2002, p. 138). Esto quiere decir que el lexema abarca el conjunto de todas las formas flexivas de dicha unidad de significado. Por tratarse de un elemento abstracto, se representa en el diccionario por medio de uno o varios lemas, los cuales son formas concretas del lexema. La elección de lemas es conocida como lematización. Para las lenguas con una tradición lexicográfica significativa, suele existir una convención con respecto a la lematización, pero este no es el caso para el bribri, como se explica en la sección 5.3.1.

En lo que concierne a los términos *macroestructura* y *microestructura*, se siguen las definiciones de Porto (2002). La macroestructura de un diccionario “está constituida

por todas sus entradas dispuestas de acuerdo con un determinado criterio ordenador” (2002, p. 38). Un mismo diccionario puede presentar varias macroestructuras, sobre todo cuando se trata de diccionarios digitales. En el diccionario bribri–español español–bribri, existe un índice alfabético que puede considerarse una macroestructura estática, mientras que una función de búsqueda proporciona una macroestructura dinámica. Por otro lado, la microestructura es el “conjunto de informaciones —también dispuestas de acuerdo con un determinado patrón o patrones— que se ofrecen dentro del artículo lexicográfico” (PORTO, 2002, p. 135).

5. Diccionario en XML

5.1. Fundamentos técnicos

XML es un lenguaje³ de marcado derivado de SGML (Standard General Markup Language; ISO-8879), del cual constituye un subconjunto (FAWCETT; QUIN; AYERS, 2012, p. 7). Fue creado originalmente para la estructuración, el almacenamiento y el intercambio de información, y en los últimos años se ha convertido en la herramienta más importante para la transmisión de datos entre aplicaciones (CAREY; VODNIK, 2015, p. 4). La primera versión recomendada por el World Wide Web Consortium, XML 1.0, fue publicada en febrero del 1998 (WORLD WIDE WEB CONSORTIUM, 1998). Hoy en día, XML presenta un uso extendido tanto en la web como en otros ambientes, tales como la contaduría, la banca, los recursos humanos y la medicina; incluso, es un componente importante en las aplicaciones para sistemas Android e iOS (CAREY; VODNIK, 2015, p. 4-6).

XML no es, desde luego, la única alternativa para el almacenamiento de datos lexicográficos. Otra opción es JSON (JavaScript Object Notation), en la que la información se estructura de manera similar. Decidimos utilizar XML para este diccionario principalmente porque lo consideramos un lenguaje ligeramente más legible para un ser humano, lo cual facilita la edición manual. Por otro lado, si se optara por trabajar con estructuras más complejas, se recomendaría hacer uso de una base de datos relacional con SQL (Structured Query Language).

³ O, en realidad, un estándar que describe la sintaxis utilizada para crear lenguajes (HUNTER et al., 2007, p. 7).

Los documentos XML pueden ser elaborados en cualquier editor de texto simple y se guardan con la extensión “.xml”. Los caracteriza el uso de etiquetas (ing.: *tags*), cuyos nombres se escriben entre paréntesis angulares. Los bloques de construcción básicos son los “elementos”, que consisten en una etiqueta de apertura, la información contenida y la correspondiente etiqueta de cierre (la cual lleva una barra inclinada al principio). Una definición en el diccionario puede representarse como en el ejemplo 1:⁴

(1) <def>hermano de mujer</def>

En lugar de datos, los elementos pueden contener otros elementos. De esta manera, los documentos XML presentan una estructura arbórea. Un elemento en un nivel inferior a otro se conoce como *hijo*, mientras que uno en un nivel superior es llamado *padre*. El árbol XML siempre cuenta con un elemento radical que encierra todos los demás y funciona como el nodo superior de la jerarquía de datos. Por otro lado, los elementos que contienen datos no pueden incluir otros elementos y se conocen como *hojas*, con referencia a la metáfora arbórea.

Uno de los propósitos de XML, al igual que otros lenguajes de marcado, es la separación entre la información y la presentación (SCHMIDT, 2010, p. 23). En nuestro caso, todos los datos lexicográficos se registran en un documento en el que no existe ninguna indicación acerca de la presentación visual. Por lo tanto, pueden ser visualizados de muchas maneras diferentes sin que sea necesario modificar el documento XML. Asimismo, la información en sí puede ser transferida de un dispositivo a otro sin inconvenientes. Gracias al alto grado de interoperabilidad del XML, los datos pueden ser accedidos desde diferentes tipos de sistemas (SHARMA et al., 2012). De este modo, el mismo archivo que contiene la información lexicográfica puede fungir como base para visualizar el diccionario en, por ejemplo, documentos PDF y aplicaciones para dispositivos móviles, además de las páginas web.

5.2. Sistema de escritura del bribri

En el diccionario se hace uso de la ortografía más comúnmente empleada por los bribris para escribir su lengua. Se trata de una versión del alfabeto latino/romano que

⁴ Este ejemplo corresponde a una de las definiciones de la palabra bribri *akè*.

fue desarrollada por Jack Wilson en los años setenta y ochenta del siglo XX (WILSON, 1982), y ligeramente modificada por Adolfo Constenla Umaña en los años noventa (CONSTENLA; ELIZONDO; PEREIRA, 1998). En este sistema, varias consonantes son representadas por dígrafos (*ch, sh, tk, ts*) y una por un apóstrofo ('). Asimismo, una letra lleva virgulilla (*ñ*), dos llevan diéresis (*ë, ö*) y se emplean acentos agudo, grave y circunflejo sobre las vocales para marcar tonos. Las letras básicas con cualquiera de estos diacríticos son representadas por un solo carácter en Unicode; por ejemplo, *á* es U+00E1 y *ö* es U+00F6. En cambio, para otro diacrítico, el subrayado, que marca nasalidad vocálica en la ortografía bribri empleada, se utiliza el macrón combinable U+0331 en el diccionario. Similarmente, para las vocales que simultáneamente llevan marca de tono y otro diacrítico, se hace uso de los acentos combinables U+0300, U+0301 y U+0302 para el tono.

5.3. Estructura de la base de datos

La base de datos de este diccionario consiste en dos documentos XML diferentes: uno para la sección bribri–español y otro para la dirección recíproca. En ambos documentos, el elemento radical es <diccionario>. El siguiente nivel de la estructura arbórea es <e>, correspondiente a cada una de las entradas. El lema que representa a cada entrada se especifica por medio de un atributo, de ahí que cada entrada se abra de la siguiente manera, con el lema particular en lugar de los puntos suspensivos:

(2) <e lema="...">

Para que la información pueda accederse correctamente, cada entrada tiene que tener un identificador único. En este diccionario, se utiliza el lema como identificador, por lo que todos los lemas deben ser distintos. En consecuencia, en este sistema simplificado, los homógrafos se registran como distintas acepciones de una misma entrada.⁵

⁵ Esta estrategia también se emplea en algunos diccionarios impresos, tal como el *Clave. Diccionario de uso del español actual* (MALDONADO; GARCÍA; ALMARZA, 2012).

Cada entrada puede tener un número ilimitado de acepciones, las cuales se registran dentro del elemento <ac>. De esta manera, una entrada con dos acepciones tiene la estructura mostrada a continuación:

```
(3) <e lema="...">
      <ac>
      ⋮
      </ac>
      <ac>
      ⋮
      </ac>
    </e>
```

En los siguientes subapartados, se describe la microestructura específica para cada una de las dos secciones del diccionario.

5.3.1. Sección bribri–español

A pesar de que el enfoque de este artículo sea la estructura XML del diccionario, cabe hacer hincapié en el tema de la lematización en la sección bribri–español. La morfología flexiva del bribri es relativamente simple, con la excepción de la verbal. Los lemas escogidos para representar a los sustantivos y los adjetivos son, generalmente, las formas singulares, pero también se incluyen algunas formas plurales irregulares como lemas.

En lo que atañe a los verbos, para facilitar la búsqueda en el diccionario, decidimos representarlos mediante dos lemas distintos. De esta manera, el buscador arroja resultados para cualquiera de las dos formas, y ambas aparecen en el índice alfabético. Uno de los lemas corresponde al infinitivo, que es como se representan los verbos en el diccionario de Margery (1982, 1996). Por otro lado, el segundo lema verbal corresponde al perfecto remoto. Este se incluye debido a que es la única forma a partir de la que se pueden predecir las demás formas regulares (Constenla, Elizondo y Pereira, 1998, p. 15), de ahí que, morfológicamente, se pueda considerar más básica que el infinitivo.

No obstante, las formas morfológicas utilizadas como lemas (de cualquier categoría gramatical), que aquí llamaremos *formas lematizables*, presentan diferencias ortográficas en el corpus; esto se debe tanto a la variación fonológica dialectal y estilística como al hecho de que la lengua no cuenta con una ortografía estandarizada.⁶ Dado que deseamos minimizar el carácter prescriptivo del diccionario, se incluyeron todas las variantes observadas de cada forma lematizable. El registro de varios lemas para un mismo lexema se resolvió por medio de remisiones, como se explicará más adelante. Gracias al formato digital, la gran cantidad de remisiones no representa ningún inconveniente para los usuarios del diccionario, sino una ventaja, ya que facilita la búsqueda de la información.

Entre un diccionario y otro suele haber muchas disimilitudes en cuanto a la microestructura. En la obra tratada aquí, todas las acepciones completas (es decir, las que no son remisiones) contienen tres elementos de información obligatorios:

- <cat> – categoría gramatical
- <def> – definición lexicográfica
- <fue> – fuentes (citas bibliográficas de las obras en las que se encontró el lexema y/o la definición; más adelante, incluirá también las fuentes orales)

Un ejemplo de una entrada con una sola acepción se presenta a continuación:

```
(4) <e lema="kúiki">
      <ac>
          <cat>adv.</cat>
          <def>dentro de siete días</def>
          <fue>Jara y García (2013)</fue>
      </ac>
</e>
```

⁶ Las variaciones se observan, por ejemplo, en vocales inacentuadas (*dakúr* ~ *dukúr* ~ *dkúr* ‘murciélagos’), en la aproximante labiovelar (*duwás* ~ *dawás* ~ *duás*), en las vocales nasales *a* y *o* (*ká* ~ *kó* ‘lugar, espacio, país, terreno, mundo, tiempo, día’), en sílabas átonas originalmente terminadas en fricativa velar (*ajkò* ~ *jkò* ~ *kkò* ~ *kò* ‘boca, borde’), en los tonos (p.ej. *dók* ~ *dòk* ‘llegar’), etcétera.

Además, se utiliza un elemento optativo, <sin>, para registrar sinónimos. En la página web, cada sinónimo funciona como un enlace al artículo respectivo, por lo que su forma ortográfica tiene que ser idéntica a la de un lema registrado en la base de datos. Cuando una misma acepción presenta varios sinónimos, decidimos separarlos con un punto y coma. De esta manera, resultó sencillo programar una función para visualizarlos como enlaces separados en la página web. Por lo tanto, tres sinónimos de una misma acepción —en este caso: òkpu, sàlpu y kaèpu— se registran como ilustrado en el ejemplo 5:⁷

(5) <sin>òkpu;sàlpu;kaèpu</sin>

Si en un futuro se decidiera agregar más elementos de información, tales como ejemplos, imágenes o audios, simplemente se crearían nuevas etiquetas. En el caso de imágenes y audios, lo que se registraría en la base XML serían los nombres de los archivos, mientras que los archivos en sí se guardarían en una carpeta particular dentro del sitio web.

Ahora, como ya se ha señalado, para evitar el almacenamiento por duplicado de la información, distinguimos entre dos tipos de acepciones: las completas y las remisiones.⁸ Las acepciones completas incluyen la microestructura descrita anteriormente, mientras que las remisiones solo contienen un enlace a otra entrada. A su vez, creamos dos clases de remisiones: (1) las que vinculan lemas que son variantes de una misma forma lematizable (ligeramente distintos entre sí debido a la variación en la pronunciación o la ortografía) y (2) las que enlazan una forma morfológica con otra del mismo lexema.

El primer tipo de remisiones se especifica mediante el atributo `rem`, que se añade a la etiqueta <ac>. En el ejemplo 6, la única acepción de la entrada representada por el lema *tbè* consiste en una remisión a *tabè*, cuya entrada incluye toda la información lexicográfica sobre este lexema.⁹ En tales casos, la página web se programó para visualizar inmediatamente la información de la entrada correspondiente. Las

⁷ Estos sinónimos corresponden a la única acepción del sustantivo bribri *ákpu*, que significa ‘águila arpía (*Harpia harpyja*)’.

⁸ Las remisiones son tradicionalmente conocidas como “entradas de referencia”, término evitaremos aquí, puesto que, en nuestro diccionario, se registran a nivel de las acepciones, no de las entradas.

⁹ Como no se agrega ninguna otra información acerca de la acepción de *tbè*, no se utiliza una etiqueta de cierre en este elemento, sino que <ac> se cierra de una vez por medio de una barra inclinada al final de la etiqueta.

remisiones de esta naturaleza se conocen como *enlaces duros*. Por motivos de consistencia, las acepciones completas se registraron bajo el lema que consideramos la forma etimológicamente más conservadora en cada caso, mientras que las demás variantes son remisiones; sin embargo, esta diferenciación no necesariamente tiene que visibilizarse en la página web.

```
(6) <e lema="tbè">
      <ac rem="tabè" />
    </e>
```

La inclusión del atributo `rem` a nivel de acepción, y no a nivel de entrada, permite que una misma entrada contenga tanto acepciones completas como remisiones. Tal caso se ilustra en el ejemplo 7, el cual ha sido abreviado para efectos de este artículo, pues en realidad incluye varias acepciones completas más. En estas ocasiones, en lugar de visualizarse inmediatamente la información de la entrada a la que hace referencia la remisión, se muestra en la página web un enlace que abre el artículo correspondiente al ser presionado, lo cual se conoce como un *enlace suave*. En lo que concierne al ejemplo en cuestión, *kò*, además de presentar los significados consignados en las acepciones completas, es una variante dialectal de *ajkò* ('boca, borde'), por lo que su última acepción se visualiza como un enlace suave a la entrada representada por dicho lema.

```
(7) <e lema="kò">
      <ac>
        <cat>sust.</cat>
        <def>hoja</def>
        <fue>Margery (1996)</fue>
      </ac>
      <ac>
        <cat>sust.</cat>
        <def>pelo, vello</def>
        <fue>Constenla et al. (1998)</fue>
      </ac>
      <ac rem="ajkò" />
```

</e>

Asimismo, se programó una función en la plataforma web que, al abrirse un artículo, confecciona una lista de las entradas que remiten al lema en cuestión. De esta manera, en el artículo también se pueden mostrar todas las variantes formales registradas de ese lema. Esta información, por lo tanto, no tiene que registrarse directamente en la microestructura de la entrada, de ahí que se evite el almacenamiento por duplicado.

El segundo tipo de remisiones es el que enlaza distintas formas flexivas de un mismo lexema. Por ejemplo, para los verbos, la información completa se registra en las entradas correspondientes a las formas infinitivas, por lo que las formas de perfecto remoto son remisiones. Estas acepciones incluyen dos atributos: `remf`, que consigna el lema que representa la entrada donde la información completa se halla guardada (la forma infinitiva en el caso de los verbos), y `f`, que especifica de qué forma flexiva se trata. De esta manera, en un futuro se podrán incluir como lemas todas las formas conjugadas de los lexemas sin que se tenga que modificar la estructura de la base de datos. Al abrirse la entrada del ejemplo 8, se muestra en la página web el texto “Perfecto remoto activo de yulök”, en el que la última palabra es un enlace que lleva al artículo completo.¹⁰

```
(8) <e lema="yulö' ">
      <ac remf="yulök" f="perfecto remoto activo" />
    </e>
```

Cabe aclarar que los dos tipos de remisiones también pueden aparecer en una misma entrada, como sucede en varios casos en el diccionario; un ejemplo se presenta a continuación:

```
(9) <e lema="ña' ">
      <ac rem="iña' " />
      <ac remf="ñúk" f="perfecto remoto activo" />
    </e>
```

¹⁰ El significado es ‘buscar, servir, repartir, ofrecer’.

5.3.2. Sección español–bribri

La estructura de la sección español–bribri es similar a la de bribri–español, pero presenta algunas diferencias que se explicarán en este subapartado. Estas se deben principalmente al hecho de que la metalengua del diccionario es el español. En primera instancia, la estructura XML de entradas y acepciones es la misma. Ahora bien, en la sección español–bribri, los elementos obligatorios en cada acepción son:

- <bri> – el o los lemas correspondientes en bribri
- <cat-bri> – la categoría gramatical en bribri
- <fue> – fuentes

Los lemas bribris se muestran en la plataforma web como enlaces a las entradas de la sección bribri–español, de manera que se entrelazan los dos documentos XML, por lo que se separan con punto y coma en la base de datos. Por tanto, las entradas y las acepciones tienen el formato ejemplificado a continuación:

```
(10) <e lema="enviar">
      <ac>
          <bri>apàtkök;katàtök;ktàtök;pàtkök</bri>
          <cat-bri>v.tr.</cat-bri>
          <fue>Margery (1996), Jara (2013)</fue>
      </ac>
</e>
```

De manera facultativa, se emplea el elemento <espec> para agregar cualquier especificación a las acepciones, por ejemplo, para aclarar o delimitar la referencia del término español o para señalar que los términos bribris corresponden a una locución que incluye el lema español. Los ejemplos 11, 12 y 13 ilustran algunos de sus usos. En el primero, la preposición española *con* es polisémica, por lo que se proporcionan significados más específicos para las posiciones bribris correspondientes. En el ejemplo 12, la segunda acepción corresponde a una locución que contiene el lema *feliz*

(*ponerse feliz*), ya que el bribri tiene un verbo para este concepto. En cuanto al ejemplo 13 (el cual ha sido abreviado con respecto a la entrada original), el lema *tucán* es un término genérico, de ahí que la entrada incluya todos los nombres bribris registrados para distintas especies de esta familia de aves.

```
(11) <e lema="con">
      <ac>
          <espec>(comitativo)</espec>
          <bri>ta</bri>
          <cat-bri>posp.</cat-bri>
          <fue>Jara (2018), Constenla et al. (1998),
          Margery (1996)</fue>
      </ac>
      <ac>
          <espec>(instrumental)</espec>
          <bri>wa</bri>
          <cat-bri>posp.</cat-bri>
          <fue> Jara (2018), Constenla et al. (1998),
          Margery (1996)</fue>
      </ac>
      <ac>
          <espec>[alguien o algo que guía]</espec>
          <bri>wétsë;wös</bri>
          <cat-bri>posp.</cat-bri>
          <fue>Jara (2018), Jara (1993), Margery
          (1996)</fue>
      </ac>
  </e>
```

```
(12) <e lema="feliz">
      <ac>
          <bri>bua'</bri>
          <cat-bri>adj.</cat-bri>
          <fue>Jara (1993)</fue>
```

</ac>
 <ac>
 <espec>ponerse feliz</espec>
 <bri>tsènuk</bri>
 <cat-bri>v.med.</cat-bri>
 <fue>Jara y García (2013), Sánchez
 (2009)</fue>
 </ac>
 </e>

- (13) <e lema="tucán">
 <ac>
 <espec>(genérico) (ave) (familia
 Ramphastidae)</espec>
 <bri>urrèk</bri>
 <cat-bri>sust.</cat-bri>
 <fue>Jara y García (2013), Flores (2009)</fue>
 </ac>
 <ac>
 <espec>tucán pico arcoíris (ave) (Remphastos
 sulphuratus)</espec>
 <bri>urrèk</bri>
 <cat-bri>sust.</cat-bri>
 <fue>Flores (2009)</fue>
 </ac>
 <ac>
 <espec>tucán orejiamarillo (ave) (Selenidera
 spectabilis)</espec>
 <bri>kálibtsik</bri>
 <cat-bri>sust.</cat-bri>
 <fue>Margery (1996)</fue>
 </ac>
 <ac>

```

<espec>quioro, tucán de pico castaño (ave)
(Ramphastos swainsonii)</espec>
<bri>tsiö̀</bri>
<cat-bri>sust.</cat-bri>
<fue>Sánchez (s.f.), Jara y García (2003), Flores
(2009)</fue>
</ac>
</e>

```

Un segundo elemento optativo, <cat-esp>, se usa para aclarar la categoría gramatical del lema español cuando este presenta homografía o polisemia. Decidimos separar esta información del elemento <espec> para poder representarla en un formato diferente en la página web. De la misma manera, otros tipos de información registrados en <espec> podrían haberse registrado en elementos separados. El ejemplo 14 muestra el uso del elemento <cat-esp> en la entrada del lema *cerca*, el cual representa dos homógrafos: un sustantivo y un adverbio.

```

(14) <e lema="cerca">
  <ac>
    <cat-esp>sust.</cat-esp>
    <bri>aká;akó;kó</bri>
    <cat-bri>sust.</cat-bri>
    <fue>Margery (1996), Jara y García (2013),
    Jara (1987)</fue>
  </ac>
  <ac>
    <cat-esp>adv.</cat-esp>
    <bri>kùkù̀e;kúkú́r;tsí́nét</bri>
    <cat-bri>adv.</cat-bri>
    <fue>Margery (1996), Constenla et al.
    (1998), Jara y García (2013)</fue>
  </ac>
</e>

```

Por último, en la sección español–bribri solo se hace uso de un tipo de remisión. Esta se emplea para entrelazar lemas que son sinónimos en español, para evitar el almacenamiento por duplicado. De este modo, solo en una de las entradas se registra la información completa, pero todas las remisiones funcionan como enlaces duros, por lo que la entrada completa se muestra inmediatamente, sin importar cuál sea abierta por el usuario. Debido a este funcionamiento, el atributo de la remisión se registra a nivel de la entrada. A continuación se ofrece un ejemplo, el cual es una remisión de *rostro* a *cara*:

(15) `<e lema="rostro" rem="cara" />`

6. Visualización del diccionario en HTML mediante AJAX

Como ya se ha señalado, la información guardada en los documentos XML puede ser extraída y visualizada de muchas maneras. Para este diccionario en particular, se optó por el uso de AJAX, con el fin de crear una página dinámica e interactiva. De esta forma, se pueden agregar, modificar y eliminar partes del contenido de una página HTML sin volver a cargar la página completa, lo cual es muy conveniente para un diccionario. Una vez cargada la base de datos, la información se visualiza inmediatamente cuando el usuario abre un artículo, inclusive si pierde el acceso a internet durante la sesión.

AJAX utiliza una combinación de XMLHttpRequest, JavaScript y HTML DOM (Document Object Model) para solicitar datos del servidor web y visualizárselos al cliente (W3SSCHOOLS, 2018). Los datos transportados pueden ser de formato XML, como en nuestro caso, pero actualmente es más común emplear AJAX con otros formatos (FAWCETT; QUIN; AYERS, 2012, p. 616).

XMLHttpRequest es una interfaz que fue introducida como un control ActiveX en Internet Explorer en 1999 (ASLESON; SCHUTTA, 2006, p. 13), pero ahora está presente en todos los navegadores recientes. JavaScript, por su parte, es un lenguaje de *script* empleado principalmente en páginas web. Todos los navegadores modernos contienen un motor para interpretar y ejecutar JavaScript inmediatamente, y hoy en día es uno de los lenguajes de programación más utilizados (JONES, 2017, p. 3). Por último, HTML DOM es un modelo que representa, de manera jerárquica, la estructura lógica de los objetos de un documento (ROBIE, 1998).

En lo que concierne al uso de AJAX en la plataforma web, una función JavaScript como la del ejemplo 16 debe ser ejecutada al cargarse el documento HTML:

```
(16) function cargar() {  
        xmlhttp = new XMLHttpRequest();  
        xmlhttp.open("GET", "diccionario.xml", false);  
        xmlhttp.send();  
        diccionario = xmlhttp.responseXML;  
    }
```

Esta función abre uno de los documentos XML (llamado “diccionario.xml”) en el fondo, del que cualquier dato puede ser extraído por medio de la variable `diccionario` y visualizado en la página HTML. Sin embargo, queda fuera del alcance del presente escrito proporcionar más detalles al respecto.

En la interfaz web del diccionario bribri–español español–bribri, el usuario puede acceder a un artículo de dos maneras: mediante un índice alfabético de todos los lemas registrados en el diccionario o por medio de un buscador que arroja una lista de lemas semejantes al término de búsqueda ingresado por el usuario. La implementación y el funcionamiento del buscador se describen con detalle en Krohn (2019).

Conclusiones

Este artículo ha sido un reporte de caso de la elaboración de una base de datos XML para un diccionario en particular. La estructura presentada es relativamente sencilla y fácil de modificar, por lo que puede adaptarse para otros proyectos lexicográficos. El uso de una base de datos en lugar de un documento estático tradicional facilita sustancialmente el mantenimiento del diccionario y permite gran flexibilidad e interactividad en la visualización de los datos. Además, los archivos tienden a ser más ligeros que los de formatos estáticos, por lo que el diccionario se carga con mayor rapidez.

Referencias

- ASLESON, R.; SCHUTTA, N. *Foundations of AJAX*. Nueva York: Springer, 2006.
- CAREY, P.; VODNIK, S. *New perspectives on XML*. 3ª ed. Boston: Cengage Learning, 2015.
- CONSTENLA UMAÑA, A. Estado actual de la subclasificación de las lenguas chibchenses y de la reconstrucción fonológica y gramatical del protochibchense. *Estudios de Lingüística Chibcha*. Vol. 27, 2008, p. 117-135. 2008.
- CONSTENLA UMAÑA, A.; MARGERY PEÑA, E. *Bribri I*. San José: Universidad de Costa Rica, 1978.
- CONSTENLA UMAÑA, A.; MARGERY PEÑA, E. *Bribri II*. San José: Universidad de Costa Rica, 1979.
- CONSTENLA UMAÑA, A.; ELIZONDO FIGUEROA, F.; PEREIRA MORA, F. *Curso básico de bribri*. San José: Editorial de la Universidad de Costa Rica, 1998.
- DIVERSIDAD Y PATRIMONIO LINGÜÍSTICO DE COSTA RICA [DIPALICORI]. Bribri. [S.l.]: Universidad de Costa Rica, 2020. Disponible en: <https://www.dipalicori.ucr.ac.cr/lengua-cultura/bribri/>. Acceso: 12 de abril de 2020.
- FAWCETT, J.; QUIN, L.; AYERS, D. *Beginning XML*. 5ª ed. Indianapolis: Wiley, 2012.
- FLORES SOLÓRZANO, S. Los mamíferos en la clasificación etnobiológica de la comunidad de Amubre. *Estudios de Lingüística Chibcha*. Vol. 28, 2009, p. 7-46.
- FUENTES RODRÍGUEZ, E. Características demográficas y socioeconómicas de las poblaciones indígenas de Costa Rica (Censo 2011). En: INSTITUTO NACIONAL DE ESTADÍSTICA Y CENSOS (Ed.). *Costa Rica a la luz del Censo 2011*. San José: INEC, 2014, p. 309-343. Disponible en: https://www.inec.cr/sites/default/files/documentos/inec_institucional/publicaciones/anp_oblaccenso2011-01.pdf_2.pdf. Acceso: 12 abril 2020.
- HUNTER, D.; RAFTER, J.; FAWCETT, J.; VAN DER VLIST, E.; AYERS, D.; DUCKETT, J.; WATT, A.; MCKINNON, L. *Beginning XML*. 4ª ed. Indianapolis: Wiley, 2007.
- INSTITUTO NACIONAL DE ESTADÍSTICA Y CENSO DE PANAMÁ. *Diagnóstico de la Población Indígena en Panamá*. [S.l.], 2015. Disponible en: https://www.contraloria.gob.pa/inec/archivos/P6571INDIGENA_FINAL_FINAL.pdf. Acceso: 12 abril 2020.
- JARA MURILLO, C. V. El campo léxico de la vivienda en el bribri de Coroma. *Estudios de Lingüística Chibcha*. Vol. 6, 1987, p. 109-187.
- JARA MURILLO, C. V. *I ttè. Historias bribris*. San José: Editorial de la Universidad de Costa Rica, 1993.

JARA MURILLO, C. V. Morfología verbal de la lengua bribri. *Estudios de Lingüística Chibcha*, Vol. 32, 2013, p. 95-15.

JARA MURILLO, C. V. *Gramática de la lengua bribri*. San José: E-Digital, 2018.

JARA MURILLO, C.; GARCÍA SEGURA, A. *Diccionario de mitología bribri*. San José: Editorial de la Universidad de Costa Rica, 2003.

JARA MURILLO, C. V.; GARCÍA SEGURA, A. *Se' ttó bribri ie. Hablemos en bribri*. San José: E-Digital, 2013.

JONES, D. *JavaScript: Novice to Ninja*. 2ª ed. Melbourne: SitePoint, 2017.

KROHN, H. S. Programación de buscadores en JavaScript para diccionarios digitales. *Cuadernos de Lingüística Hispánica*. Vol. 34, 2019, p. 109-130.

MALDONADO, C.; GARCÍA MÁRQUEZ, G.; ALMARZA ACEDO, N. (Eds.). *Clave. Diccionario de uso del español actual*. Madrid: Editorial SM, 2012.

MARGERY PEÑA, E. *Diccionario fraseológico bribri-español español-bribri*. San José: Editorial de la Universidad de Costa Rica, 1982.

MARGERY PEÑA, E. *Diccionario fraseológico bribri-español español-bribri*. 2ª ed. San José: Editorial de la Universidad de Costa Rica, 1996.

PORTO DAPENA, J. Á. *Manual de técnica lexicográfica*. Madrid: ARCO/LIBROS, 2002.

ROBIE, J. What is the Document Object Model? En: World Wide Web Consortium, *Level 1 Document Object Model Specification*. [S.l.], 1998. Disponible en: <https://www.w3.org/TR/WD-DOM/cover.html>. Acceso: 12 abril 2020.

SÁNCHEZ AVENDAÑO, C. La voz media en bribri y la hipótesis de la elaboración relativa de los eventos. *Estudios de Lingüística Chibcha*. Vol. 28, 2009, p. 47-73.

SÁNCHEZ AVENDAÑO, C. (Coord.). *Íyiwak. Diccionario pictográfico de los animales en bribri*. San José: Vicerrectoría de Acción Social de la Universidad de Costa Rica, [S.f.]. Disponible en: <https://www.dipalicori.ucr.ac.cr/wp-content/uploads/Diccionario-Animales-Bribri.pdf>. Acceso: 12 abril 2020.

SCHMIDT, T. Another extension of the stylesheet metaphor. En: WITT, A.; METZING, D. (Eds.). *Linguistic modeling of information and markup languages*. Heidelberg/Londres/Nueva York: Springer 2010, p. 23-44.

SHARMA, S.; GOYAL, S. B.; SHANDLIYA, R.; SAMADHIYA, D. Towards XML interoperability. En: WYLD, D.; ZIZKA, J.; NAGAMALAI, D. (Eds.). *Advances in Computer Science, Engineering & Applications (ICCSEA 2012)*. Vol. 1, 2012, p. 1035-1043.

SIL INTERNATIONAL. *Lexique Pro*. Versión 3.6. [S.l.]: SIL International, 2012. Disponible en: <http://www.lexiquepro.com/download.htm>. Acceso: 12 abril 2020.

SIL INTERNATIONAL. *Language Forge*. [S.l.]: SIL International, 2020a. Disponible en: <https://languageforge.org>. Acceso: 12 abril 2020.

SIL INTERNATIONAL. *Webonary*. [S.l.]: SIL International, 2020b. Disponible en: <https://www.webonary.org>. Acceso: 12 abril 2020.

WILSON, Jack. El alfabeto bribri. En: MARGERY PEÑA, E. *Diccionario fraseológico bribri-español español-bribri*. San José: Editorial de la Universidad de Costa Rica, 1982, p. xi-xx.

WORLD WIDE WEB CONSORTIUM. *Extensible Markup Language (XML) 1.0. W3C Recommendation 10-February-1998*. [S.l.], 1998. Disponible en: <https://www.w3.org/TR/1998/REC-xml-19980210>. Acceso: 12 abril 2020.

W3SCHOOLS. *AJAX Introduction*. [S.l.], 2018. Disponible en: https://www.w3schools.com/js/js_ajax_intro.asp. Acceso: 12 abril 2020.