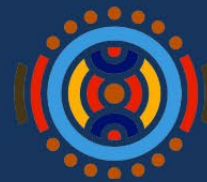


SERIE DE CERO A DIGITAL:

GUÍA DE TÉRMINOS

UNA GUÍA PARA LLEVAR SU IDIOMA A INTERNET

**TRANSLATION
COMMONS**



2022-2032 | INTERNATIONAL DECADE OF
Indigenous Languages

Guía de términos

Autores: Sue Ellen Wright, Akil Iyer, Shuto Kato
Reseñadores y colaboradores: Craig Cornelius,
Alaina Brandt, Julie Anderson, Tex Texin
Mercadotecnia: Leonidas Pappas
Coordinación de proyecto: Paula Cirilo

Agradecemos los comentarios que nos ayuden a mejorar
nuestras pautas.

Escríbanos a krista@translationcommons.org

Esta obra está bajo licencia internacional de Creative
Commons Attribution 4.0.

<https://creativecommons.org/licenses/by/4.0/>

1. INTRODUCCIÓN	4
2. ACERCA DE ESTE DOCUMENTO	5
3. ¿QUÉ ES LA TERMINOLOGÍA Y EL TRABAJO TERMINOLÓGICO?	6
4. FLUJO IDEAL DE TRABAJO	8
4.1 ¿Estás listo para comenzar tu trabajo de terminología?	8
4.2 ¿Ha reunido recursos humanos para hacer el trabajo?	9
4.3 Crear textos y recopilar términos existentes en tu idioma.	10
4.4 Crear (acuñar) términos para ideas y conceptos nuevos faltantes.	10
4.5 Documentar y mantener la terminología en glosarios compartidos.	11
5. APLICAR DE LAS DIRECTRICES	11
5.1 Uso del diagrama de flujo	11
5.2 Lexicografía (diccionarios familiares) y terminología (terminografía)	14
5.3 ¿Cómo identificamos la terminología y por qué es importante el trabajo de la terminología?	15
6. DOCUMENTAR SUS TÉRMINOS	19
7. OTRAS ACTIVIDADES	23
7.1 Crear un corpus	23
7.2 Agregar traducciones para crear un corpus paralelo	24
7.3 Crear sistemas de conceptos genéricos	25
7.4 Sistemas de conceptos parte-todo	26
8. COMPARTIR Y PUBLICAR SUS RECURSOS TÉRMINOS	27
REFERENCIAS	29

1. INTRODUCCIÓN

[Translation Commons](#) es una comunidad de voluntarios sin fines de lucro que apoya la digitalización de idiomas, asesora a profesionales y brinda cursos y recursos para el sector lingüístico.

Uno de los principales programas de Translation Commons es la Iniciativa de digitalización de idiomas (LDI), que busca acercar las capacidades digitales a las comunidades lingüísticas que deseen beneficiarse de ellas. Existen cerca de 6000 idiomas en todo el mundo que tienen una presencia digital pequeña o inexistente. La LDI proporciona una guía para que una comunidad puede seguir con el fin de digitalizar su idioma.

Translation Commons se asoció con el [Año Internacional de las Lenguas Indígenas del 2019](#), una iniciativa de la UNESCO, para centrar más la atención en las comunidades indígenas y la digitalización de sus idiomas. Como parte de su misión, la LDI apoya el acceso digital equitativo a los idiomas indígenas y de otras minorías con el fin de garantizar que estas comunidades lingüísticas puedan participar en actividades globales en línea y, a su vez, beneficiarse de todas las aplicaciones informáticas modernas en su idioma nativo. La creación de pautas que equipen a las comunidades con las herramientas y el entendimiento para digitalizar los escritos y publicarlos en la web les otorga el conocimiento para facilitar el proceso sin perder la autonomía. Además de las pautas, Translation Commons ofrece tutoriales, talleres y ayuda a las comunidades con la digitalización del idioma, poniéndolos en contacto con expertos de la industria que los guiarán a través del proceso de normalización.

Este documento forma parte de una serie de pautas titulada *De cero a digital*, que aborda de manera integral las prácticas de digitalización del lenguaje. Los autores de las pautas son expertos en lingüística y tecnología del lenguaje. El público objetivo es cualquier comunidad lingüística que desee tener la capacidad de usar su idioma en sistemas digitales.

La digitalización amplifica las vías de comunicación de una comunidad lingüística. Ver los [Beneficios de la digitalización de idiomas](#) para obtener más detalles sobre cómo beneficia, tanto a las comunidades indígenas, como al mundo en general.

Para obtener más información sobre el proceso, consulte [De cero a digital: una guía para poner su idioma en línea](#). En la página de [Recursos](#) del sitio web de Translation Commons encontrará información adicional y relacionada con la LDI, entre las que se incluye guías, presentaciones, videos y otros documentos.

2. ACERCA DE ESTE DOCUMENTO

Este documento describe cómo buscar evidencia sobre la terminología en un idioma existente, no documentado o poco documentado. Describe procedimientos para registrar palabras existentes y para crear nueva terminología que apoye a la comunicación sobre temas especializados, para los que las lenguas indígenas pueden no tener palabras particulares. Esta metodología recomendada brinda orientación sobre varios flujos de trabajo para crear o actualizar terminología, incluida la planificación, la recopilación de recursos, el análisis terminológico y la documentación. Este proceso proporciona un medio para que las comunidades indígenas puedan acuñar términos a medida que surja la necesidad, por ejemplo, para temas pertenecientes a campos tecnológicos, médicos, políticos, legales u otros que de algún modo podrían estar fuera del ámbito histórico de su cultura o idioma. Ampliar y actualizar este tipo de vocabulario ayudará a una comunidad indígena a acceder y utilizar recursos para facilitar su inclusión en un entorno digital.

Algunas lenguas en peligro de extinción tienen una larga tradición de escritura y otras son estrictamente orales. Algunas son vistas como "idiomas familiares" que no se comparten con extraños. Los defensores del idioma necesitan documentar las palabras que ya existen y crear nuevas palabras para representar objetos, conceptos e ideas que la gente necesitará si quiere digitalizar su idioma.

La audiencia para las que se ha destinada estas directrices son las comunidades indígenas que desean documentar su vocabulario, los expertos en la materia que apoyan el trabajo de terminología en uno o más idiomas y las organizaciones que apoyan a las comunidades lingüísticas en el proceso de digitalización. El objetivo es poder utilizar la colección de terminología resultante para recursos orales, escritos y, eventualmente, digitales.

Los flujos de trabajo descritos se pueden aplicar al trabajo de terminología en cualquier dominio o área temática. Los ejemplos que se dan aquí se basan en gran medida en el idioma inglés, pero ciertas reglas y modelos generales pueden modificarse para aplicarse a diferentes idiomas y situaciones, aunque cierta información no se aplica a todos los idiomas. Aun así, los ejemplos deben brindar a los lectores un buen punto de partida para identificar los parámetros que deben tenerse en cuenta en su idioma al realizar el trabajo de terminología.

Este documento describe áreas fundamentales de conocimiento, mejores prácticas y recursos (tanto tecnológicos como de otro tipo) que ayudarán a las comunidades indígenas a medida que realizan trabajos de terminología en sus idiomas para apoyar los esfuerzos de preservación, revitalización y digitalización de idiomas.

3. ¿QUÉ ES LA TERMINOLOGÍA Y EL TRABAJO TERMINOLÓGICO?

Las palabras, abreviaturas o incluso frases que se utilizan para fines especiales en los idiomas se consideran parte de la terminología lingüística especial del mismo. Ciertamente, la terminología de la ciencia y tecnología es ampliamente reconocida, pero todas las actividades humanas especializadas, en todas las culturas, utilizan una terminología especial ya sea para cocinar, cazar o expresar valores culturales. El trabajo de terminología se lleva a cabo con dos propósitos en mente:

- Recopilación de palabras y términos: se recopila y documenta palabras y términos existentes. En lenguas indígenas e indocumentadas, esta actividad es parte de la documentación y planificación lingüística y se recopilan todas las palabras y términos que podamos. En este entorno, no hay que limitarse a los términos lingüísticos especiales de por sí, sino que se asignan conceptos a los campos temáticos donde sea relevante.
- Acuñar nuevos términos para conceptos que faltan: cuando es posible, se usan los recursos existentes del idioma que se han registrado para crear nuevos términos cuando encontramos que hay conceptos sobre los que necesitamos hablar que aún no tienen término en el idioma. Los terminólogos a menudo identifican estos conceptos comparando los similares y reconociendo que uno que conocen en un idioma no ha sido creado o acuñado en el otro. La designación adecuada para estos espacios vacíos es "términos que faltan". A veces, los idiomas llenan estos espacios con préstamos. El inglés tiene palabras como *tsunami* y *Schadenfreude* que se tomaron prestadas de esta manera. Con los lenguajes en desarrollo, generalmente se considera prudente acuñar nuevos términos utilizando los recursos existentes en lugar de tomarlos prestados.

El trabajo de terminología es importante para la preservación cultural y la promoción de la igualdad. Cuando los idiomas están en peligro, el trabajo de terminología se puede utilizar para crear un registro de un idioma y revivir su uso. El galés y el hebreo moderno son ejemplos de idiomas que estaban en peligro de desaparecer pero que se han convertido en importantes idiomas nacionales, gracias en parte al trabajo terminológico sistemático.

El objetivo del trabajo de terminología es recopilar las palabras que las personas realmente usan y registrar la comprensión compartida de sus significados. Una colección de tales palabras relacionadas con las diversas áreas de actividad especializada en una comunidad proporciona una base para que los guardianes de las lenguas indígenas participen activamente en el proceso de documentación de su idioma, preservando la

precisión de los términos y sus usos apropiados. Como se señaló anteriormente, el trabajo de terminología incluye la acuñación de nuevos términos cuando se adopta una nueva tecnología, que será el caso de muchos idiomas indígenas que se están digitalizando o preparando para su uso en dispositivos digitales. Los términos digitales como *navegador* y *herramienta*, pueden tomarse prestados de palabras cotidianas, pero términos como *menú* y *fuentes* e incluso *correo electrónico* requerirán atención porque su significado puede resultar desconocido.

4. FLUJO IDEAL DE TRABAJO

El siguiente esquema presenta un flujo ideal de trabajo para terminología relacionado con la digitalización de idiomas. Todos los pasos son importantes para completar el proyecto, pero puede que no sea factible seguirlos en el orden exacto en que se enumeran. Esta lista es una guía aproximada para su trabajo.

4.1 ¿Estás listo para comenzar tu trabajo de terminología?

- ¿Tu idioma tiene las características que necesitas para la digitalización? Por ejemplo:
- ¿Su idioma tiene un sistema de escritura?
- ¿Está disponible su sistema de escritura en el [estándar Unicode](#) para registrar su terminología utilizando aplicaciones informáticas?
- ¿Se ha configurado su sistema de escritura para la entrada y visualización de datos del teclado en computadoras y dispositivos digitales?
- ¿Alguien ha elaborado reglas gramaticales básicas para su idioma?

Si la respuesta a cualquiera de estas preguntas es "NO", consulte la pestaña [Recursos](#) de Translation Commons para obtener más información sobre cómo abordar estos problemas. Pero no se detenga aquí, incluso si algunos de estos pasos están en progreso, puede haber muchas otras cosas que puede hacer mientras tanto. Algunas de las sugerencias de estas directrices pueden resultar repetitivas, pero se

centran en el proceso de identificación y registro de términos, así como en la creación de otros nuevos. Es útil tener en cuenta que se puede comenzar a realizar grabaciones de audio mucho antes de que se haya finalizado un guion o un sistema de escritura y se puede comenzar a utilizar un guion en evolución para registrar información manualmente antes de que se haya completado el proceso de digitalización.

4.2 ¿Se han reunido recursos humanos para hacer el trabajo?

- ¿Se han reunido representantes de la comunidad para trabajar en el proyecto?
- ¿Se ha obtenido el apoyo de los ancianos y otros líderes de grupo?
- ¿Cuáles hablantes del idioma conocen sus temas? ¿Quién es un gran cocinero, cazador u otro experto? Ponerse en contacto con ese experto cocinero o cazador y animarlo a que hable de su trabajo.
- ¿Se ha encontrado algunas personas que puedan brindar soporte técnico para manejar las tareas informáticas involucradas en el proyecto?
- ¿Saben cómo reconocer y recopilar palabras y términos?

- Reunir un equipo puede tomar tiempo, especialmente si se congrega expertos en diferentes [campos temáticos](#) para ayudar a todos a comprender los objetivos. Esto puede ser un proceso continuo a lo largo del tiempo. Se puede usar tutoriales y otros materiales para capacitar a las personas en reconocer términos y registrar sus significados. Si se está desarrollando la gramática y estandarizando la ortografía, por ejemplo y las personas necesitan aprender a aplicar reglas gramaticales básicas, agregue esta información a su programa de capacitación.

A medida que se avanza, hay que asegurarse de verificar el trabajo con los miembros de la comunidad. ¿Están de acuerdo con las decisiones? ¿Hay diferencias de opinión sobre la pronunciación o la forma como se deben representar las palabras en el guion?

Se puede encontrar que los dialectos emergen a medida que se avanza en el trabajo. Todos estos temas deben ser considerados para llegar a un consenso entre el grupo.

4.3 Crear textos y recopilar términos existentes en su idioma.

- Se trabaja preferiblemente en un [tema](#) específico a la vez, pero hay que prepararse para documentar cualquier cosa nueva que se presente e indicar el campo del tema a medida que se avanza.
- Indicar que se tiene listas de palabras, diccionarios o textos en el idioma que traten el tema y que puedan usarse para comenzar una colección de términos.
- Hablar con los expertos que haya identificado y hacer grabaciones de audio de sus historias y comentarios.
- Transcribir sus grabaciones, empezando por escribir a mano si es necesario.
- Identificar palabras y términos importantes y empezar a hacer listas y a escribir [definiciones](#) o explicaciones.
- Si se documenta términos en el área donde se vive y se desconoce los términos de algunos de los objetos que se ven, se debe intentar encontrar esos términos para poder incluirlos en las listas.
- ¿Los términos que se han escogido, son precisos? Verificar con varios hablantes, ¿hay formas variantes? ¿Hay dialectos?
- Identificar términos relacionados mientras se avanza.
- Almacenar textos y listas de palabras juntos para crear un [corpus](#).
- Cuando se tiene un corpus grande, se pueden utilizar [herramientas](#) informáticas para organizar el corpus o para documentar los términos.

4.4 Crear (acuñar) términos para ideas y conceptos nuevos faltantes.

- ¿Hay términos que se necesitan y faltan en el idioma?
- Consulta con varios o incluso muchos hablantes para saber si los términos acuñados tienen sentido.

4.5 Documentar y mantener la terminología en glosarios compartidos.

- ¿Se ha decidido cómo hacer un seguimiento de los términos y sus significados?
- ¿Se ha decidido los tipos de datos que documentarán cada término?
- ¿Se ha decidido el sistema de manejo de terminología y el [modelo de datos](#) que se utilizará para documentar los términos?
- ¿Se ha documentado cada concepto importante en una colección terminológica, que se ha organizado en una base de datos terminológica?
- ¿Se ha verificado la calidad de cada entrada de conceptos?
- ¿Se ha compartido la base de datos terminológica con los usuarios que la necesitan?
- ¿La base de datos terminológica describe lo que se ha encontrado en la investigación? Esto se llama *terminología descriptiva*.
- ¿La base de datos terminológica les dice a las personas cómo deben usar los términos? Esto se llama *terminología prescriptiva*.
- ¿Las comunidades relevantes de hablantes son referidas a los conceptos en la base de datos terminológica de manera consistente?
- ¿La terminología muestra una descripción general de éstas preguntas. Hay que recordar que esta es una vista simplificada, se puede hacer partes del proyecto en un orden diferente o en paralelo. Cada proyecto es diferente. Se tiene que ser flexible y creativo mientras se actualiza con el tiempo?

5. APLICAR LAS DIRECTRICES

5.1 Uso de diagramas de flujo

El diagrama de flujo se muestra en la Figura 1

Hay que tener en cuenta que en el segundo elemento en la columna de la derecha se prefiere las escrituras alfabéticas, donde los símbolos en la escritura representan

sonidos, sin referencia al significado. Una opción es trabajar con una escritura logográfica, donde los caracteres del idioma representan un [concepto](#) o idea sin referencia a la forma en que se pronuncia la palabra o término respectivo. El inglés, el francés y el árabe son ejemplos de escrituras alfabéticas; Los kanji chinos y japoneses son ejemplos de escrituras logográficas.

Reunir a las personas para que participen sin duda debe ser una de las primeras actividades y una vez que se hayan identificado, se puede empezar trabajar. A medida que se habla con la gente para ganar su confianza y cooperación se puede comenzar a realizar grabaciones de audio mucho antes de finalizar un guion o un sistema de escritura. Se puede comenzar a utilizar el guion en evolución para registrar información manualmente antes de que se haya completado el proceso de digitalización. Con el tiempo, a medida que más personas se involucran y usan teléfonos celulares y otras herramientas digitales, se puede usar los medios digitales para recopilar información de un grupo cada vez más amplio de personas que conocen el idioma.

Estos esfuerzos ahorrarán un tiempo valioso y proporcionarán información inicial que se puede utilizar para crear modelos para actividades futuras, a medida que el proyecto madura. Este trabajo inicial también proporcionará ejemplos verificables de texto que se necesitará más adelante, cuando se solicite el estado de idioma ISO 639, el registro de guiones o la certificación Unicode. A medida que el proyecto avanza, se vuelve a consultar el diagrama de flujo para ver si hay que ponerse al día con algunos pasos que no se pudieron hacer antes.

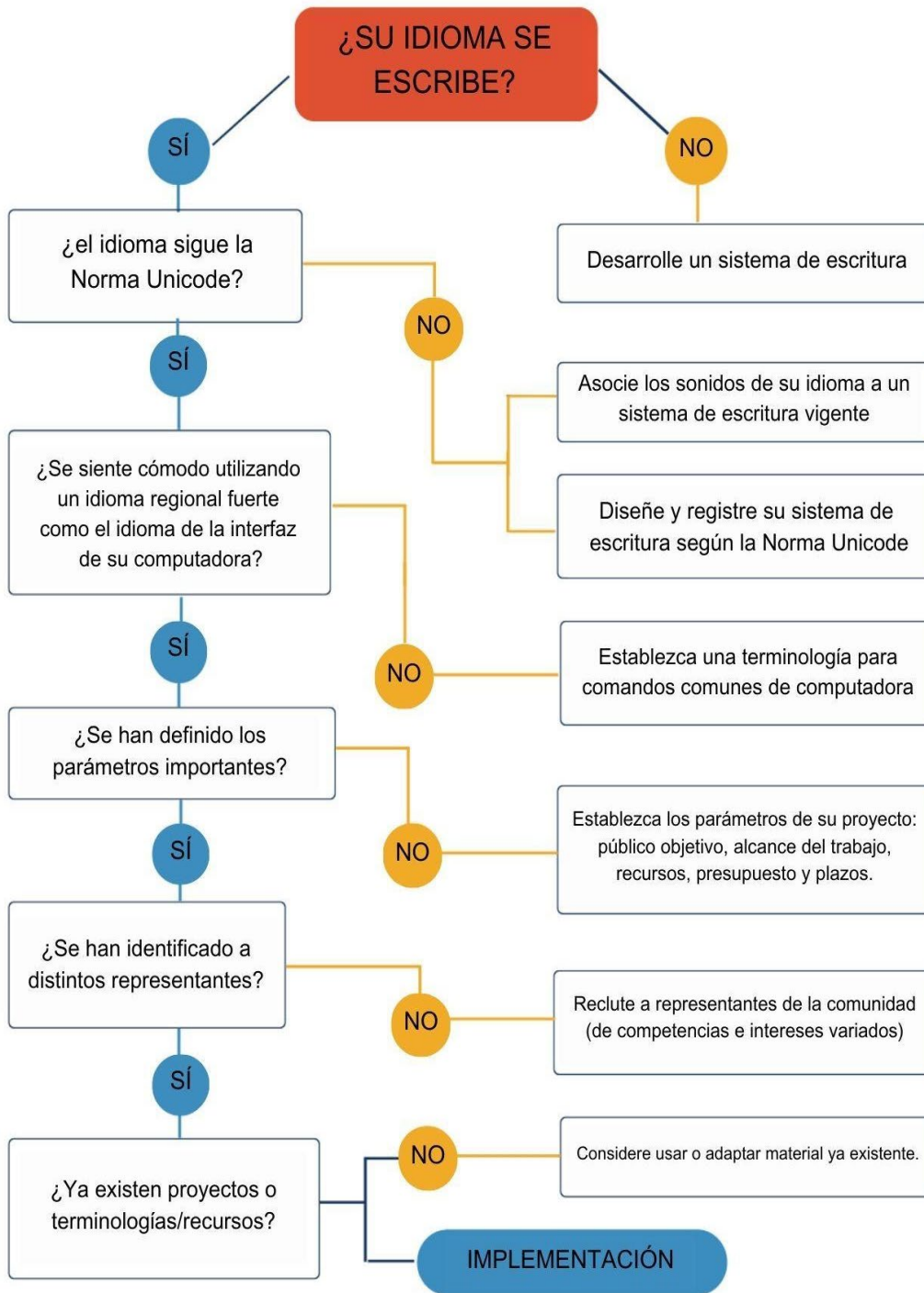


Figura 1: Diagrama de flujo de preparación del proyecto

5.2 Lexicografía (diccionarios familiares) y terminología (terminografía)

Los diccionarios tradicionales enumeran palabras en entradas donde todos los significados ([definiciones](#)) asociados están incluidos en la misma entrada, como se muestra en la Figura 2.

Los diccionarios lexicográficos son muy útiles cuando se trabaja en un solo idioma, porque se puede registrar toda la información sobre una sola palabra en un solo lugar. También se puede incluir información gramatical y usarse para documentar muchas palabras pequeñas (a veces llamadas palabras de sonido en las colecciones de corpus) como artículos, preposiciones, conjunciones y otras partículas del lenguaje que se usan tanto en forma cotidiana como en el especializada.

rattle v.intr.

1. (traquetear) Hacer una rápida sucesión de sonidos cortos
2. (traquetear) Moverse haciendo estos sonidos (un carro traquetea sobre las piedras)
3. (chacharear) Hablar de forma rápida e incesante

--v.t.

1. (hacer ruido) Provocar que algo haga ruido (hacer ruido con el pomo de una puerta)
2. decir o recitar algo rápidamente.
3. (desconcertar) confundir o molestar (abuchear a un orador, por ejemplo).

--s.

1. (traqueteo) Sucesión rápida de sonidos cortos
2. (estertor) Ruido de una respiración ronca al pasar el aire a través de la mucosidad en una garganta parcialmente cerrada. ESTERTOR DE LA MUERTE
3. (ruido) alboroto, griterío
4. (cascabel) anillos rugosos en la punta de la cola de la serpiente de cascabel con los que produce su sonido de cascabeleo
5. (sonajero) utensilio que sirve de juguete infantil o instrumento musical y produce ruido al agitarse

Colocación: deambular en una casa que es demasiado grande para uno

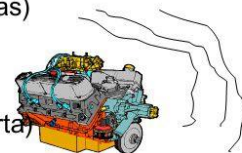


Figura 2: Entrada de diccionario (lexicográfico) para la palabra *traqueteo*

Sin embargo, las entradas lexicográficas se vuelven problemáticas cuando se intentan usar como base para diccionarios bilingües o multilingües o si desea documentar muchas palabras dialectales o sinónimas. Cada uno de los

significados en la entrada tiene el potencial de ser una palabra diferente cuando busca equivalentes en otro idioma y diferentes ortografías o sinónimos terminarán en múltiples entradas. Esta es la razón por la que, hoy en día, la mayoría de los traductores y otros autores e investigadores bilingües, así como los planificadores lingüísticos, utilizan lo que se denomina el “enfoque terminológico” para registrar la información léxica.

5.3 ¿Cómo se identifica la terminología y por qué es importante ese trabajo?

Usando el enfoque terminológico de la documentación del lenguaje, se crea una entrada separada para cada uno de los diferentes significados de una palabra. Eso puede sonar complicado porque la única entrada en la Figura 2 se convierte en once entradas si se crea entradas de terminología. Sin embargo, si se hace eso, es fácil asignar un término, sus sinónimos y equivalentes en múltiples idiomas o dialectos a cada uno de los significados, es decir, asignar equivalentes precisos para cada idea o concepto que representa. Este enfoque es esencial no solo para una traducción precisa, sino también, para navegar por la acuñación de nuevos términos para los conceptos que se necesitan en el proceso de digitalización o para otros conceptos faltantes (como en medicina).

Para visualizar la relación entre una idea o concepto determinado y el término que lo representa, a menudo dibujamos un triángulo que muestra (1) el *concepto* que el término crea en nuestra mente, (2) el *objeto* del término al que se refiere (a veces llamado el *referente* o el *objeto en el mundo real*) y (3) el término mismo. Si un término “funciona”, podemos usarlo cuando hablamos o escribimos y el oyente o lector puede formarse inmediatamente la misma idea en su mente. Cuando nos comunicamos con más de un idioma, es importante que el objeto (el referente) sea el mismo o muy similar y que todos, en ambos idiomas, entiendan el mismo concepto que se tiene en mente. Solo entonces se puede decidir que un término en el idioma A es equivalente a un término en el idioma B. En la Figura 3 se muestra un ejemplo de un triángulo semántico, usado para representar el significado.

La disposición de los triángulos se vuelve muy útil cuando se compara idiomas. Se empieza por reconocer que una palabra representa un concepto. Luego se busca el concepto equivalente en el idioma de destino. Si no lo hay, tenemos una situación típica de una palabra "faltante" y podemos crear una nueva o tomar una prestada. Idealmente, el trabajo de terminología se propone identificar los términos existentes en una comunidad lingüística en lugar de tomar prestadas palabras de idiomas externos. El término *árbol* es un buen ejemplo porque el idioma nativo americano Cheroqui tiene una palabra para *árbol*: ᏍᏉᏉ (quigvi [kʷigəi]). Sin embargo, hay muchas otras palabras en inglés para las que originalmente no había un equivalente cheroqui. Supongamos que se quiere acuñar un buen término para el concepto representado por el término inglés *email*. Si se quiere evitar los términos prestados, no nos gustaría simplemente deletrear una palabra en Cheroqui que suene como "E-MAIL".

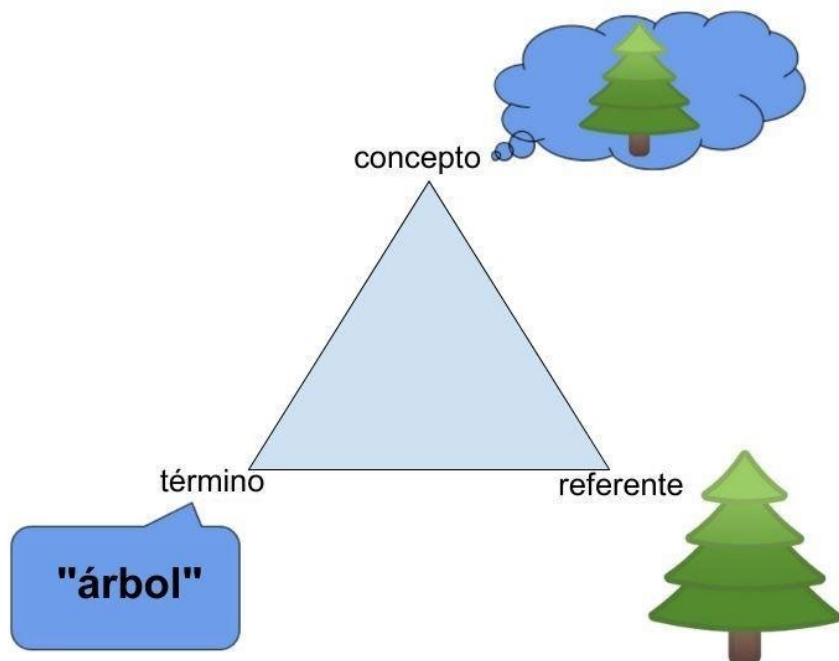


Figura 3: “Triángulo semántico”, donde el término “árbol” coincide con el objeto (referente) árbol, lo que desencadena la idea o concepto del árbol en la mente de la persona que escucha o lee el término.

En el trabajo de terminología, consideramos eso que a pesar de si el hablante de inglés y el hablante de cheroqui han tenido la misma experiencia con los correos electrónicos

en el momento en que se crearon los términos, el *email en sí* es la misma cosa para ambos. El objeto o *correo electrónico* de referencia nunca cambia en esta ecuación. A veces, los objetos conocidos, como los árboles, se verán un poco diferentes para los dos hablantes (por ejemplo, un roble se ve diferente de un baobab), pero hay que fijarse en las características comunes de los dos objetos cuando se los nombra o define. Es, principalmente, el término nodo el que puede ser diferente y reflejar las diferentes experiencias de los hablantes, lectores y escritores. Se puede comparar la palabra en inglés *email*, que es una abreviación de *electronic mail*, y la palabra cheroqui ᏓᏍᏁᏃᏅᏍᏉ ᏐᏍᏁᏃᏅᏍᏉ (*anagalisgv goweli* [anagalisgã goujeli]) que se traduce directamente como *papel relámpago*. Los hablantes de inglés, aunque han visto relámpagos, lidian a diario con la electricidad que fluye a través de un sistema eléctrico (representado aquí por un enchufe) y saben que así es como les llega el correo electrónico. Por supuesto, los hablantes de modernos de cheroqui comparten esta experiencia, pero el idioma en sí no participó en la misma evolución. Aquí se tiene la convergencia del concepto de electricidad que viaja a través de un cable y electricidad en forma de rayo, junto con la velocidad y las propiedades mágicas del correo electrónico. Y los cheroquis eligen usar una palabra natural para electricidad en este término, por lo tanto, *papel relámpago*.

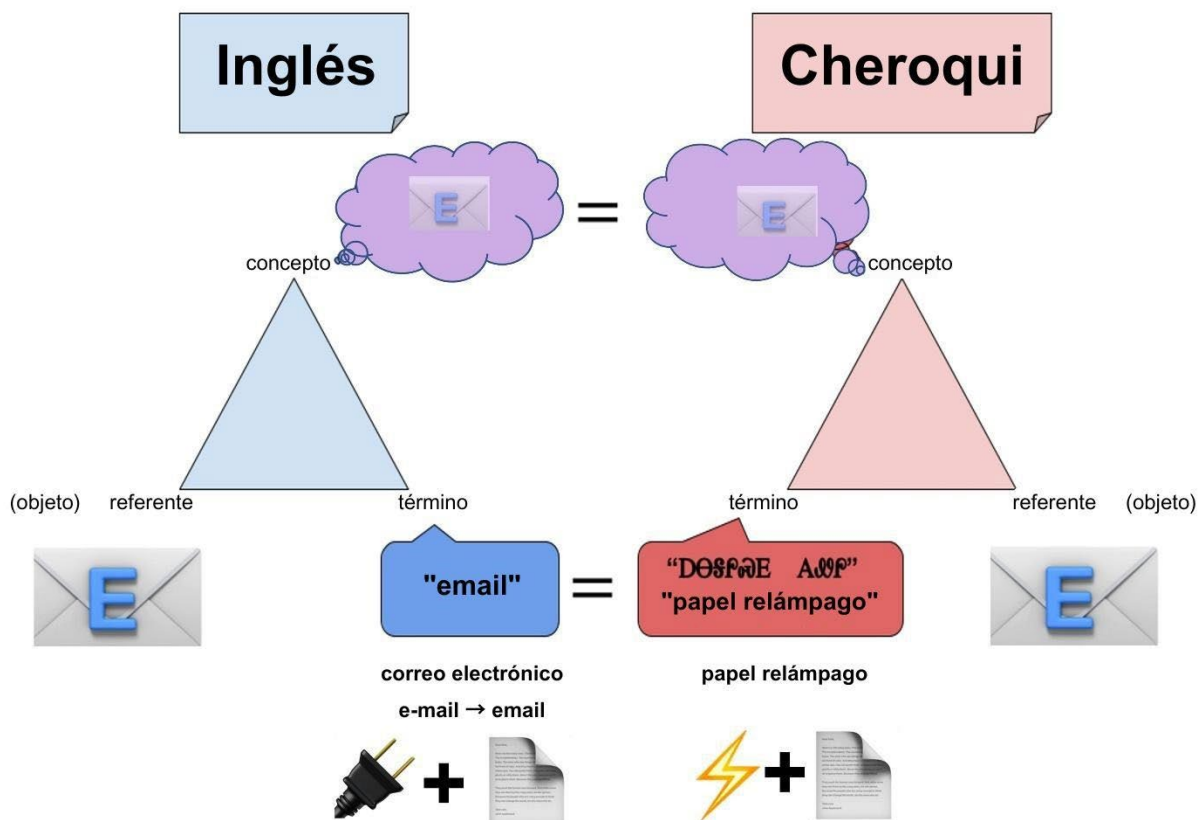


Figura 4: Triángulo que muestra la relación entre conceptos en dos idiomas, donde los conceptos coincidentes admiten términos equivalentes para el mismo objeto de referencia.

Este enfoque para hacer coincidir término + idea (concepto) es importante porque, como se señaló, algunos términos significan más de una cosa y algunas ideas pueden representarse con más de un término, lo que hace que sea peligroso tratar de hacer coincidir término con término de forma independiente. Este enfoque facilita la combinación de conceptos en dos o más idiomas, acomoda diferentes formas en múltiples dialectos e informa ortografías alternativas.

Una de las principales razones para hablar de entradas de terminología bilingüe es que los términos indígenas a menudo se documentan junto con sus equivalentes en un segundo idioma, como el cheroqui y el inglés. Aunque, generalmente, se evita tomar prestados, si algunos términos ya se han tomado prestados y están bien establecidos, es posible que desee mantenerlos en la lista de terminología o agregarlos más tarde a la

base de datos terminológica. Sin embargo, si es posible, se aconseja crear términos propios utilizando palabras familiares en lugar de tomarlas prestadas. Este enfoque ayuda a capacitar a sus hablantes para que entiendan y tomen el control de la digitalización de su idioma. Los beneficios también se extienden a otras áreas a medida que se crean nuevos materiales. Por ejemplo, en el campo de la medicina, el uso de palabras familiares ayuda a los trabajadores de la salud a acceder a la información y responder y comunicarse rápidamente y respalda los resultados exitosos.

6. DOCUMENTAR LOS TÉRMINOS

Las listas de términos son útiles para empezar, pero también se necesitan definiciones, a veces varias, para los términos y, a medida que crece su lista o listas, se vuelve cada vez más difícil encontrar información en ellas cuando la necesita. Con el tiempo se querrá crear *entradas de terminología*, especialmente si se desea documentar los términos junto con sus equivalentes en más de un idioma. Una simple entrada completa (*modelo de entrada de términos*) puede parecerse a la Figura 5.

La información típica que se incluye en una entrada de término incluye los siguientes datos, generalmente llamados *categorías de datos*:

- **Campo de asunto**
- **Definición** del concepto más una fuente, si tiene una
- **Término**
- **Parte de al oración**
- **Contexto** en una oración de muestra donde el término se usa de manera típica, más un origen.

Tema:	comunicación digital
Término en cheroqui:	DƏSPƏDE AƏP
Categoría gramatical:	sustantivo
Definición:	DƏSPƏDY AƏP: ƏZPƏDY DcƏV.əcəW0ə DƏSPƏDY DhYcə.İcəDY EW0əY ƏcəDY Ɔcə E.İcəDY DSVi.əcə.İ DcəDLhA.İcəDY hԛԛԛԛԛ ƆT Ɔcə Dδ ƆhA.İ Ԝh.İcəDY ƆtT T.ƆP ƆSԜİP Fuente: https://language.cherokee.org
Contexto:	Ɔ Dԛԛԛ DƏSPƏDY AƏP ƆS0əİB ƆPƆT ƆZPƏDY Ɔ'ƆPcəԛԛBhRY. Fuente: Roy Boney <roy-boney@cherokee.org>
Término en inglés:	Email
Categoría gramatical:	sustantivo
Definición:	mensaje enviado por vía electrónica por un usuario de computadora a uno o más destinatarios a través de una red de internet Fuente: https://www.google.com/search?q=define+email&rlz=1C1CHBF_enUS866US866&oq=define+mail&aqs=chrome..69j57j0i512l9.2230j0j15&sourceid=chrome&ie=UTF-8
Contexto:	Puedo acceder a mi correo electrónico desde mi celular o desde mi computadora. Fuente: Translation Commons

Figura 5: Diseño típico de entrada de términos

Se puede comenzar ingresando la información en un archivo de texto o en un archivo de procesamiento de texto, pero es más útil usar algún tipo de programa de hoja de cálculo. Un archivo de hoja de cálculo podría parecerse a la Figura 6. Las dos imágenes de esta figura son en realidad una fila larga en dicha hoja. Hay formas de crear hojas de cálculo más estrechas, pero colocar cada concepto para ambos (o todos) los idiomas en una sola fila hace que sea muy fácil exportar los datos a otros sistemas más adelante, si se decide crear una base de datos. Las hojas de cálculo son buenas para recopilar cantidades pequeñas de términos, pero también se vuelven engorrosas a medida que crece la colección. Si se diseña para que coincida con un modelo de base de datos terminológica, será relativamente fácil exportar e importar datos.

Cherokee term	PoS	chr-Definition	chr-Def-Source	chr-Context	chr-Ctx-Source
DƏSŋə A.əŋŋ	noun	DƏSŋə A.əŋŋ: əZŋə DəV.əWŋə DƏSŋə DhYəJə Ewŋə əəV Ɂə E.əəV DSV.əəJ DəIhAJə hLGLəŋə əT Ɂə Də ŋhAJ Jh.əəV əT TJP əSLŋŋ	https://language.chokeee.org	ə DŋC DƏSŋə A.əŋŋ əSŋə əŋT əZŋə əŋəəLhRY.	Roy Boney <roy-boney@chokeee.org>
English term	PoS	en-Definition	en-Def-Source	en-Context	en-Ctx-Source
email	noun	message distributed by electronic means from one computer user to one or more recipients via a network	https://www.google.com/search?q=what+is+an+email&rlz=1C1CHBF_enUS866US866&oq=what+is+an+email&aqs=chrome..69j57j0i512j9.3696j0j15&sourcejd=chrome&ie=UTF-8	I can access my email on my phone or from my computer.	Translation Commons

Figura 6: Entrada de término creada como una sola fila en un programa de hoja de cálculo.

Campo de asunto: la categoría de campo de asunto en la entrada ayuda a organizar los términos según el tema. Siempre se debe usar valores estables (por ejemplo: *Médico*, no variar entre *Médico* y *Medicina*). Si se es constante, es fácil filtrar el recurso para todos los términos médicos.

Definición: Las buenas definiciones suelen ser lo más breves posible, pero tan completas como sea necesario. Si el término nombra una cosa, la definición generalmente indica que el término "es un" concepto más amplio con características distintivas (por ejemplo, *lobo* (es un) *gran animal carnívoro de la familia de los perros*). Para algunos otros términos, se puede indicar "es parte de" un concepto más amplio (por ejemplo, *hoja* (es una) *parte de una planta que está unida a esta y la sostiene a través de un proceso llamado fotosíntesis*). Los verbos a menudo se definen con una frase verbal explicativa (por ejemplo, *caminar: moverse a un ritmo regular levantando cada pie por turno*). Las secciones 7.3 y 7.4 analizan la creación de diagramas de sistemas conceptuales. Una vez se ha recopilado una cantidad considerable de términos, puede comenzar a representar las relaciones entre esos utilizando ese tipo de gráficos. Este proceso puede llevar a afinar las definiciones para reflejar las relaciones entre sus conceptos.

Término: al comienzo del proyecto, es posible que se desee comenzar a registrar los términos apenas se escuchan. A medida que se aprende más sobre la gramática, se debe averiguar si hay, algo así como, una *forma base* para los términos. En muchos idiomas, es el singular la forma más común de un sustantivo, pero el plural puede ser diferente y en algunos idiomas muchas funcionan de diferentes maneras en la oración. Si hay muchas formas, hay que decidir cuáles son los mejores procedimientos para registrarlas. La forma base simple a menudo se llama *lema*. Ejemplos de lemas en español incluyen "corre" (no "corres" o "corriendo"), "libro" (no "libros" o "libra"), "rojo" (no "enrojecer" o "enrojecimiento", que son en realidad conceptos diferentes).

Parte de la oración: la discusión en el último párrafo solo hablaba de sustantivos y verbos, y existen pequeñas palabras funcionales como preposiciones y conjunciones, adjetivos que generalmente describen sustantivos, adverbios que describen verbos, palabras y frases que no son sustantivos. Estas categorías se denominan “partes de la oración”. Esta situación varía de un idioma a otro, por lo que el análisis gramatical del idioma cubrirá estas variaciones y otras. La parte de la oración de un término y otras notas gramaticales se registran mejor en un diccionario monolingüe, mientras que los términos mismos, sin importar la parte de la oración, se registrarán en la base de datos terminológica. Sin embargo, las palabras únicas, en particular los adjetivos y los adverbios, se pueden registrar en el recurso terminológico.

Hay que tener en cuenta que a veces los sustantivos en un idioma pueden ser verbos en otro o viceversa. El inglés a menudo convierte los sustantivos verbales en verbos de acción y otros idiomas (potawatomi, un idioma nativo americano, por ejemplo) pueden tratar los sustantivos animados, que siempre están en movimiento, como si fueran verbos (por ejemplo, la palabra para *río*). Puede que haya que pensar en cómo funciona *la parte de la oración* en el idioma propio y ajustar la forma para registrar la información en la base de datos terminológica.

Contexto: Si se tiene una buena oración de ejemplo con el término, se registra en este campo. Es especialmente útil encontrar contextos que indiquen o, al menos, insinúen la definición del concepto y que usen modismos o giros de frases que se usan

comúnmente con este. Usando el término "libro", por ejemplo, la oración "Juan compra libros cada vez que viene a la ciudad." no insinúa el significado de "libro", así como "Se sentaron juntos y se turnaron para leer en voz alta un libro de historias sobre los pescadores del Mar de Java Meridional".

Nota: La entrada de ejemplo no muestra un campo de *nota*, pero puede agregar una si tiene información adicional sobre cualquier parte de la entrada.

7. OTRAS ACTIVIDADES

7.1 Crear un corpus

¿Su idioma tiene colecciones de textos existentes? Además, tienen escrituras antiguas que solo ahora se están digitalizando, lo que los hará accesibles y los preservará. Otras comunidades tienen una rica tradición de "literatura" oral (cuentos, historias, música y poesía) que se han transmitido oralmente a lo largo de los siglos, pero no están escritas y pueden estar en peligro de perderse si el número de hablantes disminuye y las generaciones más jóvenes ya no memorizan los textos. Como se señaló anteriormente en este documento, se puede comenzar grabando en audio este tipo de "textos" y luego transcribirlos para su digitalización. En cualquier caso, se recopilará gradualmente un cuerpo de textos mediante el registro y la recopilación de información oral y escrita. Tomada como un todo, su colección puede comprender un *corpus*. Hay información sobre la recopilación de corpus, la preservación de los derechos de autor y la creación de archivos en Translation Commons [Resources](#).

Existe una gran variedad de herramientas informáticas diseñadas para manipular y trabajar con corpus, incluida la extracción de términos. Las [herramientas de extracción de términos](#) están diseñadas especialmente para identificar términos candidatos y posibles definiciones o contextos, trabajando con los principales idiomas. Si ha acumulado un corpus significativo, es posible que pueda obtener ayuda de programación para personalizar una o más de estas herramientas para usarlas con su idioma. Cuando

se trabaja con textos que representan el patrimonio cultural de su grupo, siempre es importante proteger y respetar las creencias y tradiciones. Hay que trabajar cuidadosamente con el equipo de colaboradores para garantizar que los materiales sagrados o confidenciales estén protegidos incluso mientras se trabaja para preservarlos para las generaciones futuras.

7.2 Agregar traducciones para crear un corpus paralelo

El valor del corpus se amplifica significativamente si se puede agregar una capa bilingüe porque esto permite expandir los recursos de terminología a través de la traducción humana y automática (TA). La creación de esta capa bilingüe, específicamente mediante la traducción de texto oración por oración a otro idioma, de modo que haya una combinación estricta de oraciones traducidas y fragmentos de oraciones, da como resultado la creación de una memoria de traducción que consta de los llamados *segmentos de texto*. Los textos resultantes componen entonces un “corpus paralelo”. Para agregar TA a su conjunto de recursos digitales, primero es esencial crear un conjunto sustancial de documentos traducidos que se puedan usar para “entrenar” su sistema automatizado para descubrir y extraer términos y segmentos de manera consistente. (Consulte Pautas de traducción automática de cero a digital).

El primer paso importante para implementar soluciones de traducción es adaptar las herramientas de traducción asistida por computadora (TAO) existentes para procesar su guion y mostrar correctamente su idioma. Una vez que los traductores humanos puedan usar estas herramientas para producir una memoria de traducción (MT), se puede comenzar a guardar segmentos coincidentes que consisten en oraciones en el idioma de origen o frases traducidas a otro idioma. Puede utilizar estos segmentos para guardar o reconstruir documentos completos, lo que le permitirá crear un corpus paralelo de textos traducidos coincidentes. A medida que crece el corpus, puede solicitarse la asistencia de los principales proveedores de TA para agregar su idioma a los idiomas cubiertos por los sistemas informáticos. Las MT significativas también se pueden usar para extraer no solo

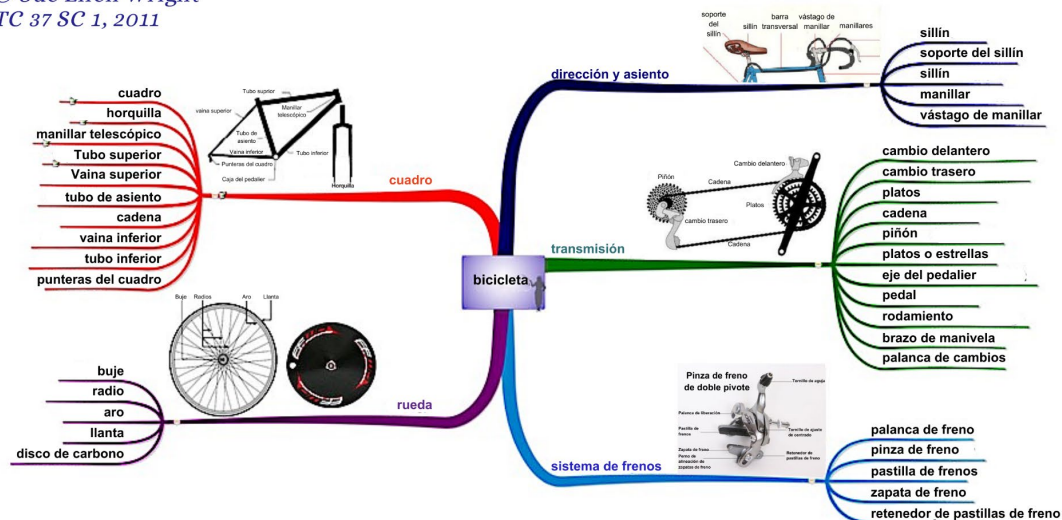
listas de términos, sino también gran parte de los datos que se necesitan para entradas terminológicas completas y significativas.

7.3 Crear sistemas de conceptos genéricos

Los terminólogos recomiendan encarecidamente que organizar los términos a medida que se avanza en los sistemas de conceptos. Al describir las definiciones, hablamos de las relaciones "es un" y "es parte de". Pregúntese: ¿qué es el término A? Si la respuesta es que el *Término A es una especie de Término B*, puede comenzar a construir un sistema de concepto genérico. Por ejemplo, muchas personas están familiarizadas con las bicicletas. Si tuviera que crear un sistema conceptual que muestre todos los tipos de ciclos que se le ocurran, podría parecerse a la figura 7. Este tipo de sistema de conceptos es una buena manera de aprender a definir. Si miras el dibujo, puedes empezar una definición escribiendo: *una bicicleta es una especie de vehículo terrestre de dos ruedas que frecuentemente es accionado por humanos mediante pedales*.

Bicicletas: partes y juegos de parte

© Sue Ellen Wright
TC 37 SC 1, 2011



Un juego de ruedas es parte de una bicicleta.
Una rueda es parte de un juego de ruedas.
Un radio es parte de una rueda.

Figura 8: Sistemas de conceptos parte-todo

Dibujar sistemas de conceptos también es una buena manera de identificar los términos que faltan cuando recopila o crea términos equivalentes. Los sistemas de conceptos también suelen revelar que en realidad hay múltiples significados para un término, en los que no se había pensado inicialmente.

8. COMPARTIR Y PUBLICAR SUS RECURSOS TÉRMINOS

Como se indicó anteriormente, la mejor manera de hacer crecer una colección terminológica y hacerla accesible a un público más amplio es archivar sus datos utilizando un sistema de manejo de terminología que está disponible en la web, por ejemplo, el software [Terminologue](#), que originalmente fue desarrollado por Fiontar & Scoil na Gaeilge para [Foras na Gaeilge](#) para administrar *Téarma*, la base de datos terminológica nacional para irlandés ([téarma.ie](#)). El sitio web de Terminologue proporciona información tutorial sobre cómo configurar bases de datos terminológicas y comenzar a ingresar datos.

Coordine con su grupo para decidir quién será el principal responsable de editar y administrar su base de datos terminológica. Ponerla a disposición de los usuarios interesados, para acceder y leer a través de computadoras, teléfonos y otros dispositivos digitales. Proporcionar medios para que los invitados y las partes interesadas proporcionen sugerencias para nuevas entradas y ediciones para las existentes, pero proteja los datos de ediciones no deseadas o cambios no intencionados.

REFERENCIAS

Proyecto de ejemplo:

Patyegarang, el sitio web de educación de idiomas indígenas australianos

<http://www.indigoz.com.au/language/gaps.html>

Esta página web proporciona excelente información sobre cómo acuñar términos, basándose en la experiencia de una comunidad australiana.

Herramientas:

Herramientas para la gestión de corpus

<https://www.Corpus-Analysis.com> proporciona una lista completa de 261 herramientas para usar en el análisis de corpus.

<https://tesolpeter.wordpress.com/a-brief-guide-to-corpus-analysis-tools/>

<http://inmyownterms.com/readings-tools-and-useful-links-for-corpus-analysis/>

Hay numerosas colecciones de herramientas y la selección cambia constantemente. Muchos están disponibles de forma gratuita y algunos pueden hacer concesiones especiales para proyectos como estos.

Manejo de terminología:

Terminologue es una herramienta de gestión terminológica de código abierto. El software es desarrollado y mantenido por el [grupo de investigación Gaois](#) en [Fiontar & Scoil na Gaeilge, Universidad de la ciudad de Dublín](#). El software tiene derechos de autor de la [Universidad de la ciudad de Dublín](#) y está disponible bajo la [licencia MIT](#) de código abierto. El desarrollador principal es [Michal Boleslav Měchura](#). Descargue el software de [este](#) sitio web si desea instalar su propia instancia de *Terminologue*.

<https://www.terminologue.org/docs/info.cs/>



Herramientas de extracción de términos:



<https://termcoord.eu/free-term-extractors/>

Categorías de datos terminológicos:

Hay muchas más categorías de datos que se pueden utilizar en las entradas de terminología. Puede encontrar una colección de ellos, junto con definiciones y ejemplos de uso en <https://www.datcatinfo.net>.

Ayudas informáticas para traductores (Herramientas TAO), tanto gratuitas como pagadas:

<https://www.marstranlation.com/blog/top-free-and-paid-cat-tools>

Buenas agencias. "Los 10 mejores programas de traducción asistida por computadora gratuitos y de código abierto" <https://www.goodfirms.co/blog/the-top-10-free-and-open-source-computer-assisted-translation-software>

Translate5, Herramienta de traducción de código abierto.

<https://www.translate5.net/en/translate5-open-source-translation-system-2>

Lingüistas:

Maria Fica

Claudia Botero

Silvia Ortega Lorenzo

John Brown McKeller

Gerente de Proyectos de Localización:

Chen Yao, Yutong Du,

Sophie Liu, Han Liu,

Liuyi Yang, Jiaqian Wang,

Di Wan, Yiwei Wu

Ingenieros de Localización:

Luiyi Yang, Jiaqian Wang,

Yiwen Wang