

LA LINGÜÍSTICA COMPUTACIONAL Y LA IMPORTANCIA DE LA HERRAMIENTA “ANTCONC” PARA EL ANÁLISIS DE LA LENGUA CASTELLANA

Lic. Franz Canaza Aduviri¹

Resumen

El presente artículo enfatiza la importancia de la lingüística computacional en el análisis lingüístico en Bolivia, donde se hablan más de 36 lenguas indígenas además del castellano. La lingüística computacional permite el procesamiento de grandes cantidades de datos lingüísticos en un corto período de tiempo, es fundamental en el análisis de las diferentes variedades del castellano en el país. El desarrollo de corpus lingüísticos y el apoyo de la herramienta AntConc resulta trascendental en el análisis de una lengua. Sin embargo, Bolivia enfrenta desafíos para la aplicación de una metodología adecuada en la recolección de datos representativos, además del acceso limitado a recursos y tecnologías específicas.

Palabras clave: corpus lingüístico, tecnología lingüística, variación lingüística, dialectos, Antconc

Bolivia cuenta con una amplia variedad de lenguas, ya que en su territorio se hablan más de 36 lenguas indígenas, además del castellano, que es el idioma oficial del país. El análisis del castellano es objeto de estudio en distintas áreas, como la lingüística, educación y comunicación. Asimismo, la lingüística computacional ofrece herramientas relevantes para analizar las distintas variedades del castellano, lo que permite identificar patrones lingüísticos de manera efectiva.

La lingüística computacional es una disciplina que ha ido ganando importancia en el campo de la lingüística en las últimas décadas. Esta disciplina combina los conocimientos de la lingüística y la informática para el desarrollo de aplicaciones que permitan el análisis y la comprensión de los lenguajes humanos. A través de la utilización de técnicas y herramientas computacionales, la LC permite el procesamiento de grandes cantidades de datos lingüísticos en un corto período de tiempo y con alta precisión.

¹ Franz Canaza Aduviri es Licenciado en Lingüística e idiomas, y, actualmente, docente académico de la carrera de Lingüística e Idiomas de la Universidad Mayor de San Andrés.

En el caso de Bolivia, fue utilizada para el análisis del castellano que es una lengua pluricéntrica, lo que significa que presenta variaciones lingüísticas en diferentes regiones del mundo, el castellano presenta características lingüísticas propias, que reflejan la influencia de las lenguas indígenas y la historia social y política del país. Por esta razón, es un tema de gran interés para los investigadores de la lingüística en el país y en el extranjero.

El desarrollo de esta disciplina surge como una necesidad para analizar las distintas variedades del castellano y paliar la falta de recursos y herramientas lingüísticas adecuadas para ese fin. En este sentido, el tratamiento de corpus lingüísticos es un paso fundamental en el análisis de una lengua porque facilita el análisis de grandes cantidades de datos lingüísticos y permite identificar patrones lingüísticos, lo que resulta relevante para el estudio de las diferentes variedades de la lengua castellana.

Entre las herramientas más utilizadas para el desarrollo de corpus lingüístico se encuentra AntConc, que es un software libre y gratuito que permite el análisis de textos lingüísticos a través de la identificación de patrones y la realización de búsquedas específicas. Este programa es una herramienta valiosa para la lingüística computacional, dado que permite la identificación de modelos lingüísticos.

Sin embargo, el desarrollo de esta disciplina en Bolivia aún enfrenta desafíos importantes. Uno de estos desafíos es la falta de una metodología adecuada que permita la recolección de datos representativos de todas las variedades lingüísticas del país. Otro desafío es el procesamiento de datos en lenguas indígenas, lo que requiere de herramientas y recursos específicos para el análisis lingüístico de estas lenguas. Además, el acceso limitado a tecnologías y recursos puede restringir el desarrollo de la lingüística computacional en el país. Esto incluye la falta de conocimiento a software y recursos lingüísticos específicos, así como la falta de capacitación en lingüística computacional para los investigadores y estudiantes.

Desarrollo

En este apartado se abordan diversos conceptos clave relacionados con la lingüística computacional. Se explora el uso de AntConc, una herramienta ampliamente utilizada para el análisis lingüístico, que permite examinar de manera detallada y precisa

diferentes aspectos del lenguaje. Además, se presentan trabajos destacados tanto a nivel nacional como internacional que han contribuido significativamente al desarrollo de este campo de estudio. En resumen, este apartado proporciona una perspectiva sólida sobre la lingüística computacional, destacando su importancia, las herramientas clave y los avances más recientes en este emocionante campo interdisciplinario.

Lingüística computacional y lingüística del corpus

En primer lugar, es importante aclarar que la lingüística de corpus (LC) no es lo mismo que la lingüística computacional. El objetivo principal de la lingüística computacional es diseñar herramientas informáticas que permitan un procesamiento óptimo, es decir, la comprensión y generación del lenguaje natural (Moreno, 1998, p. 13). Esto posibilita un análisis confiable, eficiente y, sobre todo, automático. Con este fin, se han desarrollado diversos programas que se basan en modelos estadísticos o reglas, los cuales permiten, por ejemplo, la caracterización y búsqueda de categorías léxicas o la combinación de palabras en textos, así como la descripción y búsqueda de estructuras sintácticas específicas.

En palabras de Bolaños (2015):

La lingüística de corpus centra su interés en la utilización de los corpus disponibles o en la confección de corpus mediante el empleo de las herramientas que han diseñado para este fin los lingüistas computacionales. El lingüista de corpus no es necesariamente un programador como sí lo es el lingüista computacional. Se trata, pues, de actividades estrechamente ligadas y con linderos borrosos pero, en principio, podemos señalar que tienen objetivos distintos. (p.34)

La lingüística computacional se encarga de desarrollar software especializado para el procesamiento automático eficiente del lenguaje natural. Por otro lado, la lingüística de corpus se dedica a realizar análisis lingüísticos de aspectos léxicos, gramaticales, semánticos y pragmáticos (discursivos) utilizando las herramientas informáticas creadas con ese propósito.

De la misma manera, la lingüística computacional es una disciplina que se encarga de la intersección entre la lingüística y la tecnología. Según Jurafsky y Martin (2020) se puede definir como “el

estudio de la computación de lenguajes naturales y la creación de sistemas que pueden procesar o ‘comprender’ lenguajes naturales” (p. 3). Esto implica que se utilizan técnicas computacionales para analizar y entender el lenguaje humano.

La elaboración de corpus lingüísticos es una parte fundamental de la lingüística computacional. Según Bolaños (2015), analiza la definición del concepto de corpus, sus diferentes tipos y las características más destacadas. Además, se resalta el valor metodológico que la lingüística de corpus tiene para la investigación en diversas áreas de la lingüística moderna, entre muchas otras. De este modo, el autor destaca la importancia de contar con corpus lingüísticos de alta calidad y diversidad para realizar investigaciones y análisis precisos y fiables. Además, es importante contar con herramientas y software especializado para la creación, organización y análisis de los corpus.

En definitiva, la lingüística computacional y la lingüística del corpus son disciplinas interrelacionadas que tienen un gran impacto en el estudio y comprensión del lenguaje humano. La utilización de técnicas computacionales para analizar el lenguaje permite una mayor precisión y objetividad en los análisis, mientras que los corpus lingüísticos ofrecen una base de datos fiable y diversa para la investigación en lingüística.

Importancia de la lingüística computacional y lingüística del corpus

La lingüística computacional es parte de la lingüística que se encarga del estudio de la aplicación de la tecnología a la comprensión del lenguaje humano. Bolivia, como país multilingüe y diverso en lenguas y dialectos, puede beneficiarse enormemente de las herramientas y técnicas de la lingüística computacional para el análisis y comprensión de su rica variedad lingüística. Además, tiene una gran importancia en el desarrollo de tecnologías de la información y la comunicación, como la traducción automática, el reconocimiento de voz y la creación de chatbots². Como señala Martínez-Cámara (2019), es una herramienta para trabajar en distintos campos del lenguaje natural.

Por otra parte, la lingüística del corpus ayuda a identificar patrones y tendencias en el uso del lenguaje, lo que resulta útil tanto

² Un chatbot es un programa de computadora que simula conversaciones humanas. Utiliza inteligencia artificial para comprender y responder preguntas de manera automatizada.

para la investigación lingüística como para el desarrollo de tecnologías. Como señala Sinclair (2004) en su libro “Introducción a la lingüística de corpus” establece que la LC ofrece una completa introducción al análisis de extensas colecciones de textos y cómo se puede emplear esta metodología para investigar el lenguaje, además aborda aspectos fundamentales como la creación de corpus, el etiquetado de partes del discurso y el análisis de concordancias, entre otros temas-

La elaboración de corpus lingüísticos es relevante para el estudio de las lenguas y dialectos que se hablan en el país. Bazzo (2013) plantea que la lingüística del corpus se basa en el uso de amplias colecciones de datos lingüísticos con el fin de investigar y comprender el lenguaje. Además, es crucial desarrollar materiales lingüísticos para el estudio y la preservación de las lenguas, especialmente aquellas en riesgo de desaparecer. Los corpus lingüísticos proporcionan una documentación detallada del uso del lenguaje en diversos contextos, lo que permite un análisis exhaustivo de sus características.

En otras palabras, la lingüística computacional y la elaboración de corpus lingüísticos son importantes para el estudio y la comprensión de la diversidad lingüística en Bolivia. Su uso puede contribuir a la investigación lingüística, la preservación de las lenguas y dialectos, así como al desarrollo de tecnologías de la información y la comunicación.

Conceptos clave para la elaboración de corpus lingüístico

La elaboración de corpus lingüísticos es una tarea fundamental en el estudio de la lingüística computacional, y requiere de la comprensión de ciertos conceptos clave para su correcta elaboración y análisis. A continuación, se presenta los siguientes:

1. Selección de textos: Es un proceso clave que se basa en criterios claros y precisos. Los textos seleccionados deben ser representativos del lenguaje que se quiere estudiar y deben cubrir una amplia gama de géneros y estilos.
2. Etiquetado lingüístico: Es el proceso de marcar y etiquetar los elementos lingüísticos de un texto. Esta anotación es necesaria para el análisis del corpus, y debe ser realizada con cuidado y precisión.
3. Normalización: Es el proceso de estandarizar la ortografía y la gramática de los textos en un corpus lingüístico. Esto es importante para evitar ambigüedades y errores en el análisis computacional.

4. Metadatos: Son información adicional que se añade a cada texto en un corpus lingüístico, como el autor, la fecha de publicación, el género, etc. Estos metadatos son útiles para el análisis y la clasificación del corpus.
5. Tamaño del corpus: Es un factor importante en su análisis computacional. Un corpus demasiado pequeño puede no ser representativo del lenguaje en cuestión, mientras que uno demasiado grande puede ser difícil de analizar de manera eficiente.
6. Acceso y distribución: Deben ser cuidadosamente planificados para garantizar su disponibilidad y uso por parte de la comunidad científica. Esto incluye la elección de formatos de archivo compatibles con los programas de análisis lingüístico y la creación de herramientas de búsqueda y visualización para el corpus.

La elaboración de un corpus lingüístico requiere de la comprensión de conceptos clave como la selección de textos, la anotación lingüística, la normalización, los metadatos, el tamaño del corpus y el acceso y distribución del mismo. Es importante tener en cuenta estos conceptos para garantizar la calidad y eficiencia en el análisis computacional del corpus.

Herramientas para análisis lingüístico

La creación de corpus lingüísticos permite la recolección de datos lingüísticos en diferentes contextos, lo que es para la investigación lingüística y la enseñanza del castellano. En este sentido, el uso de herramientas como AntConc permiten analizar patrones lingüísticos en diferentes variedades de una lengua, como el castellano andino, el castellano de los valles y el castellano amazónico.

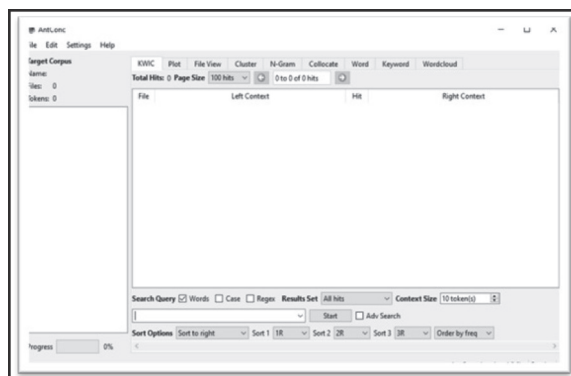


Ilustración 1 Software de análisis Lingüístico ANTCONC

La utilización de corpus lingüísticos es fundamental para llevar a cabo análisis lingüísticos del castellano. En este sentido, según López (2018) es un conjunto de textos de un mismo origen y que tiene por función recopilar un conjunto de documentos tales como ensayos, obras de teatro, transcripciones entre otros con el fin de reunir en una misma base de datos o programa el uso de un término de la lengua en un momento dado.

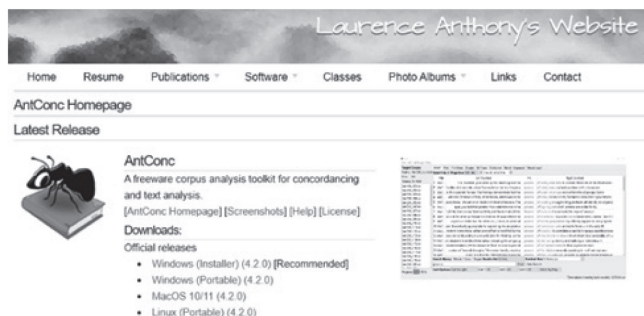
Por otro lado, AntConc es una herramienta de gran utilidad para el análisis de textos lingüísticos, como señala González (2019). Sin embargo, aún existen desafíos importantes en el acceso a tecnologías y recursos lingüísticos específicos para la lingüística computacional en Bolivia. López (2020) destaca que la falta de acceso a estos recursos sigue siendo uno de los principales desafíos en el campo de la lingüística computacional en Bolivia.

Como se aprecia, ambos autores destacan la utilidad de las herramientas lingüísticas en el análisis del castellano. Sin embargo, es importante tener en cuenta que la lingüística computacional en Bolivia todavía enfrenta desafíos importantes. Como señala López (2020), la falta de acceso a recursos específicos sigue siendo un obstáculo a superar.

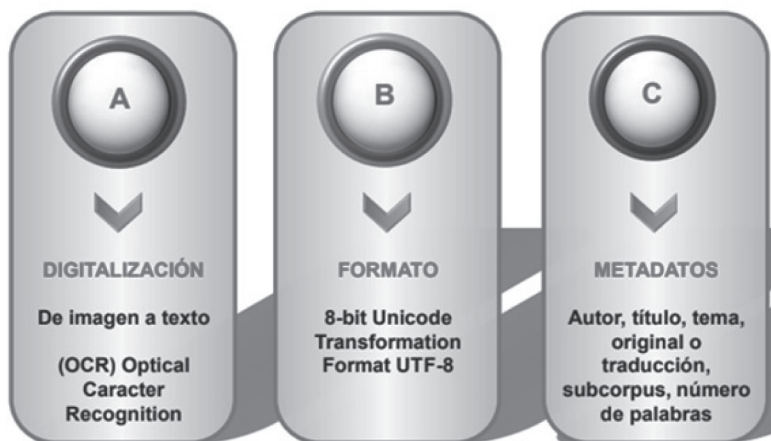
Por otro lado, la lingüística computacional también permitió la identificación de patrones lingüísticos en diferentes géneros textuales en lengua castellana. En este sentido, se desarrollaron herramientas para el análisis de géneros textuales específicos, como la prensa escrita, la literatura y el discurso político.

Pasos para la elaboración de corpus con ANTCONC

1. Descargar e instalar Antconc: Lo primero que se debe hacer es descargar el programa Antconc desde su página web oficial y seguir los pasos de instalación.



1. **Obtener los textos:** Se deben recopilar los textos que se desean utilizar para la elaboración del corpus. Estos pueden ser textos ya existentes, como libros, artículos, periódicos, entre otros. También se pueden utilizar transcripciones de entrevistas o grabaciones de audio.
2. **Preparar los textos:** Es importante tener en cuenta que los textos deben estar en formato de texto plano (.txt) además de tener una codificación en (UTF-8) y que deben estar limpios de cualquier formato o etiqueta que no sea texto. Si los textos tienen formato Word, por ejemplo, se deben guardar como texto plano para utilizar en Antconc.
3. **Cargar los textos en Antconc:** Una vez que se tienen los textos preparados y el archivo de configuración listo, se pueden cargar los textos en Antconc. Esto se hace mediante la opción “Cargar Archivos” en el menú “Archivo”.
4. **Analizar el corpus:** Con los textos cargados, se puede proceder a realizar análisis lingüísticos mediante las herramientas que ofrece Antconc, como la concordancia, el análisis de frecuencia, entre otros.
5. **Exportar resultados:** Finalmente, se pueden exportar los resultados obtenidos para su posterior análisis y visualización en otras herramientas o programas.



Es importante tener en cuenta que estos son los pasos generales para la elaboración de un corpus con Antconc y que pueden variar dependiendo de las necesidades específicas de cada proyecto.

Análisis cuantitativo de corpus con ANTCONC

Este punto se centra en la interpretación de los resultados obtenidos del análisis lingüístico realizado con AntConc. Una vez que se haya explorado y analizado el corpus utilizando las diversas herramientas y funcionalidades, es importante interpretar los hallazgos y extraer conclusiones significativas. En este apartado se presentan algunos aspectos claves en cómo abordar un análisis:

- a) **Identificación de patrones y tendencias lingüísticas:** Analiza los resultados obtenidos a partir de las herramientas de AntConc para identificar patrones y tendencias en el uso del lenguaje en un corpus. Se puede examinar las palabras más frecuentes, las colocaciones más comunes, las estructuras sintácticas predominantes, entre otros aspectos relevantes. Esto permitirá obtener una comprensión más profunda de cómo se utiliza el lenguaje en el dominio específico de una investigación.
- b) **Análisis de concordancias y KWIC (*Key Words in Context*):** Revisa las concordancias y los resultados de KWIC para examinar el contexto en el que se utilizan ciertas palabras o frases clave en el corpus. Esto proporciona información sobre los diferentes usos y significados de las palabras, así como las asociaciones y coocurrencias más comunes. Puedes realizar comparaciones entre diferentes contextos o subcorpus para obtener conocimientos o percepciones adicionales.
- c) **Análisis de coocurrencia y redes léxicas:** Explora las herramientas de AntConc que permiten analizar la coocurrencia de palabras y construir redes léxicas. Estas herramientas coadyuvan en la identificación de las palabras que tienden a aparecer juntas en el corpus y a visualizar las relaciones semánticas entre ellas. Esto puede ser especialmente útil para detectar términos clave en el dominio de una investigación y comprender las relaciones entre ellos.
- d) **Relación de los resultados con el objetivo de investigación:** Asegúrate de relacionar los resultados obtenidos con el objetivo de tu investigación y las preguntas de investigación planteadas al comienzo. Evalúa si los hallazgos respaldan o refutan tus hipótesis iniciales, y si proporcionan insights relevantes para tu campo de estudio. Puedes discutir las implicaciones teóricas, prácticas o metodológicas de tus resultados y cómo contribuyen al conocimiento existente.

- e) Limitaciones y futuras investigaciones: Reconoce las limitaciones de tu estudio y discute las áreas que podrían ser objeto de futuras investigaciones. Puedes mencionar aspectos del análisis que podrían mejorarse, limitaciones del corpus utilizado, o preguntas adicionales que surjan a partir de tus resultados. Esto abrirá la puerta a futuras investigaciones y proporcionará oportunidades para el desarrollo de nuevos enfoques o metodologías.

Es importante para cerrar el estudio y proporcionar una síntesis de los resultados obtenidos a través del análisis lingüístico realizado con AntConc. Se debe presentar conclusiones de manera clara y coherente, respaldándolas con evidencias concretas extraídas de un corpus y análisis.

Análisis cualitativo de corpus con ANTCONC

En este punto, se abordará la discusión de los resultados obtenidos a partir del análisis lingüístico con AntConc y su relevancia dentro del contexto más amplio de la lingüística computacional y los estudios de corpus. Aquí hay algunas consideraciones importantes que se pueden incluir:

- a) Comparación con estudios anteriores: Discute cómo los resultados se comparan con los hallazgos de estudios anteriores que abordan temáticas similares. Se puede resaltar las similitudes y diferencias entre los resultados de otros investigadores, y analizar las posibles razones detrás de estas disparidades. Esto contribuirá a la comprensión del estado actual del conocimiento en tu área de investigación y demostrará la relevancia de tus hallazgos.
- b) Reflexión sobre la metodología utilizada: Evalúa críticamente la metodología utilizada en tu estudio y reflexiona sobre sus fortalezas y limitaciones. Discute los desafíos encontrados durante el proceso de elaboración del corpus, el análisis con AntConc y la interpretación de los resultados. Proporciona sugerencias para mejorar la metodología en futuros estudios o abordar posibles obstáculos.

Es necesario mencionar que la característica principal de AntConc es trabajar con un vasto corpus para tener resultados más objetivos.

Estudios realizados con el uso de procedimientos computacionales

Desde el ámbito nacional, cabe destacar la obra “Tópicos e imágenes de sí y del otro en la política boliviana. Análisis de los discursos presidenciales de Carlos Mesa” escrita por Juan Marcelo Columba Fernández (2021). En esta obra, se examinan más de 200 discursos presidenciales bolivianos del periodo 2003-2005, utilizando métodos de estadística léxica asistida por computadora. El autor se centra tanto en las representaciones verbales de los interlocutores políticos como en la evolución de los temas en la agenda gubernamental.

El libro ofrece un detallado análisis de cómo se construyen y proponen imágenes del emisor y del receptor en el discurso político. Mediante la combinación de enfoques cualitativos y cuantitativos para el análisis del discurso, la investigación explora nuevas formas de estudiar y analizar el lenguaje político en Bolivia. Esto representa una valiosa contribución intelectual en el campo de las ciencias del lenguaje y la comunicación política en el país

Asimismo, en la OpenInfo 2020 de la Carrera de Informática de la UMSA, evento realizado en La Paz – Bolivia, Mauro Costantino (2020) presenta una visión general de cómo la Lingüística Computacional y la Lingüística de Corpora deben colaborar con la Informática para abordar numerosos proyectos que aún están pendientes en nuestro país. Tanto en el ámbito del Castellano Boliviano como en el de las Lenguas Nativas, existen amplias oportunidades y una gran necesidad de investigación, incluyendo la realización de proyectos, tesis y trabajos académicos.

Ahora bien, en la carrera de Lingüística e Idiomas se llevaron a cabo varios proyectos de investigación que utilizaron la herramienta AntConc como base. Algunos ejemplos de estos proyectos incluyen: un análisis del dialecto en películas bolivianas utilizando la herramienta AntConc; un análisis comparativo, léxico y pragmático sobre la frecuencia de aparición en el álbum de Morat y Karol G; un análisis semántico y etimológico de la discografía de la cantautora boliviana Matilde Casazola; un análisis del léxico y los aspectos contextuales de los libros de Edgar Allan Poe; un análisis léxico de la frecuencia de palabras en los poemas de Pablo Neruda; un estudio de las diferencias en el léxico entre los géneros del trap y el folclore desde la perspectiva de la lingüística de corpus; un análisis comparativo léxico-semántico del Salmo 119 en las traducciones RVR 1960, NTV, NVI y TLA de la Biblia;

el uso de conectores en la literatura boliviana; y un análisis léxico de los sufijos utilizados en los discursos de los expresidentes de Bolivia.

Además de los proyectos mencionados, estos son solo ejemplos de la amplia gama de investigaciones llevadas a cabo por los estudiantes de quinto año de la carrera de Lingüística e Idiomas durante el semestre I/2023. Estos proyectos demuestran la versatilidad y la aplicabilidad de la herramienta AntConc en el análisis lingüístico en diferentes áreas y contextos.

El uso de AntConc permitió a los estudiantes explorar y profundizar en diversos temas lingüísticos, desde el análisis comparativo de la música popular hasta el estudio de los aspectos semánticos y etimológicos de la discografía de artistas locales. Asimismo, examinaron el impacto del dialecto en la producción cinematográfica nacional, analizar el léxico y los aspectos contextuales en obras literarias reconocidas y explorar las diferencias y similitudes en el uso del lenguaje en distintos géneros musicales.

Estas investigaciones no sólo enriquecieron el conocimiento lingüístico de los estudiantes, sino que también al campo académico y científico, aportando nuevos hallazgos y perspectivas en el estudio del lenguaje. Al brindar las herramientas necesarias y fomentar la investigación en la carrera de Lingüística e Idiomas, se logró promover un ambiente de creatividad y excelencia académica, permitiendo a los estudiantes desarrollar sus habilidades analíticas y su capacidad para abordar cuestiones lingüísticas complejas de manera rigurosa y fundamentada.

Por otra parte, en un contexto internacional, el uso de herramientas computacionales para el análisis de corpus lingüísticos se convirtió en una práctica común y ampliamente empleada en diversos países. Sin embargo, en el caso de Bolivia, esta tendencia no se ha desarrollado de manera tan extendida debido a ciertas limitaciones y desafíos que existen en el ámbito académico y educativo del país.

Una de las razones principales es la falta de actualización de las mallas curriculares en las instituciones educativas bolivianas. El campo de la lingüística computacional y el análisis de corpus ha experimentado avances significativos en los últimos años, con el desarrollo de herramientas y técnicas cada vez más sofisticadas. Sin embargo, las estructuras curriculares en muchas instituciones

aún no incorporaron estos avances, lo que limita la capacidad de los estudiantes y profesionales para utilizar estas herramientas en su trabajo de investigación y análisis lingüístico.

Además, la falta de acceso a recursos tecnológicos y a una infraestructura adecuada también juega un papel importante en la limitada adopción de herramientas computacionales en el análisis de corpus en Bolivia. El acceso a computadoras, software especializado y conexión a Internet estable puede ser un desafío para muchos estudiantes y académicos, lo que dificulta su capacidad para llevar a cabo investigaciones y proyectos que requieren el uso de estas herramientas.

No obstante, es importante destacar que hay un creciente interés y conciencia sobre la importancia de la lingüística computacional y el análisis de corpus en Bolivia. Cada vez más profesionales y académicos reconocen el valor y el potencial que estas herramientas aportan al estudio y comprensión de las lenguas y el lenguaje en nuestro contexto específico.

En este sentido, es necesario promover la actualización de las mallas curriculares en las instituciones educativas, incorporando la enseñanza y el uso de herramientas computacionales para el análisis de corpus lingüísticos. Asimismo, se requiere mejorar el acceso a recursos tecnológicos y establecer colaboraciones y proyectos de investigación que impulsen el desarrollo de la lingüística computacional en Bolivia.

De esta manera, se podrá superar los desafíos actuales y aprovechar el potencial de las herramientas computacionales en el análisis de corpus, permitiendo a los estudiantes y profesionales bolivianos ampliar sus conocimientos y contribuir de manera significativa al avance de la lingüística en nuestro país.

Perspectivas futuras de la lingüística computacional en Bolivia

A pesar de los avances en la lingüística computacional, aún existen desafíos en la recolección y procesamiento de datos lingüísticos en el contexto boliviano. Uno de los principales desafíos es la recopilación de información en las variedades del castellano que se hablan, lo que requiere de una estrategia de muestreo adecuada que permita la recolección de datos representativos de todas las variedades lingüísticas. Otro desafío es el procesamiento de datos

en lenguas indígenas, lo que requiere de herramientas y recursos específicos para el análisis de estas lenguas.

También, se necesitan más esfuerzos para mejorar la accesibilidad de la lingüística computacional a la comunidad de hablantes de lenguas indígenas. Esto podría incluir la creación de aplicaciones móviles de aprendizaje de idiomas y la promoción de la enseñanza de la lingüística en las comunidades indígenas. Para lograrlo, se debe tener en cuenta el acceso limitado a tecnologías y recursos, lo cual puede obstaculizar el desarrollo de la lingüística; además, es necesario contar con software y recursos lingüísticos específicos, así como brindar capacitación en lingüística computacional a los investigadores y estudiantes.

En ese contexto, la herramienta Antconc desempeña un papel fundamental al permitir análisis exhaustivos y obtener resultados precisos y objetivos. Facilita el procesamiento y manipulación de grandes cantidades de datos textuales, esencial para investigaciones lingüísticas rigurosas. Sus funcionalidades avanzadas, como el análisis de frecuencia y la generación de concordancias, proporcionan una visión detallada de los textos analizados, agilizando la investigación y garantizando resultados confiables.

Por último, para lograr avances significativos en la investigación, es esencial fomentar la colaboración entre lingüistas, científicos de la computación y hablantes de lenguas indígenas y del castellano. La creación de redes de colaboración y la participación activa de la comunidad en los proyectos de lingüística computacional contribuirán a desarrollar soluciones efectivas y relevantes para las necesidades de la comunidad.

Conclusiones

La lingüística computacional es una disciplina importante que, a través de la herramienta Antconc, permite desarrollar distintos análisis del castellano y lenguas indígenas en Bolivia, permitiendo la identificación de patrones lingüísticos en diferentes variedades lingüísticas y la identificación automática de errores gramaticales y ortográficos. Sin embargo, aún existen desafíos en la recolección y procesamiento de datos, como la falta de acceso a tecnologías y recursos lingüísticos específicos y la necesidad de capacitar a investigadores y estudiantes en lingüística computacional.

Asimismo, esta disciplina es fundamental para el estudio y análisis del lenguaje en la actualidad, permitiendo la elaboración de corpus lingüísticos que facilitan la comprensión de las características y variaciones de una lengua en un determinado contexto. Gracias a la tecnología, se pueden realizar análisis más precisos y rápidos, lo que permite avances significativos en diferentes campos, como la traducción automática, la minería de textos, la inteligencia artificial y el procesamiento del lenguaje natural.

En este artículo, se presentó una forma distinta de abordar a la ciencia lingüística empleando herramientas computacionales para la elaboración de un corpus lingüístico en nuestro contexto científico, además, de conocer la importancia que tiene la lingüística computacional en investigaciones sobre las lenguas en Bolivia, especialmente el castellano porque permite entender las particularidades y variaciones de esta lengua en un contexto sociocultural específico.

Además, se abordó la importancia de la lingüística de corpus y el empleo de la herramienta Antconc en el estudio del lenguaje y su relación con la tecnología, así como la necesidad de contar con corpus lingüísticos representativos y actualizados para realizar análisis más precisos y relevantes. Asimismo, se destaca la importancia de la colaboración interdisciplinaria para la elaboración de corpus lingüísticos y su análisis, involucrando a expertos en lingüística, informática y otras disciplinas afines.

Es importante destacar que la elaboración de corpus lingüísticos no es una tarea sencilla y requiere de un trabajo riguroso y detallado, pero los resultados permiten un mayor entendimiento del lenguaje y su relación con el contexto sociocultural en el que se desenvuelve. Además, estos pueden ser utilizados en diversas áreas, como la educación, la literatura, la comunicación, la política y la cultura en general.

En resumen, la lingüística computacional y el análisis de corpus lingüísticos con la herramienta Antconc se presentan como una oportunidad para profundizar en el conocimiento de las lenguas y su relación con el contexto sociocultural en el que se desarrollan.

BIBLIOGRAFÍA

- Baker, P. (2006). *Uso de corpus en el análisis del discurso*. Continuum.
- Bazzo, W. A. (2013). *Fundamentos de Lingüística de Corpus*. Parábola Editorial.
- Bernal, E. (2005). *Análisis de corpus: aplicaciones al español*. En J. Gutiérrez-Rexach y L. Silva-Villar (Eds.), *La lingüística del siglo XXI. Reorientaciones y perspectivas*. Trotta.
- Biber, D. et al. (1998). *Lingüística de corpus: Investigando la estructura y el uso del lenguaje*. Cambridge University Press.
- Columba, J. (2021). *Tópicos e imágenes de sí y del otro en la política boliviana. Análisis de los discursos presidenciales de Carlos Mesa*. CIDES-UMSA.
- Costantino, M. (2020). *Lingüística Computacional y de Corpora - OpenInfo2020*. Recuperado de <https://www.audacy.com/podcast/mauro-in-bolivia-957aa/episodos/openinfo2020-ling%C3%BC%C3%ADstica-computacional-y-de-corpora-679db>
- González, R. (2019). *AntConc: una herramienta informática para el análisis de corpus en lingüística hispánica*. Lingüística y Literatura.
- González, R. (2019). *Herramientas digitales para el análisis de corpus lingüísticos en Bolivia*. Boletín de Lingüística.
- Jurafsky, D. y J. Martin. (2020). *Introducción al procesamiento del lenguaje natural, Lingüística computacional y reconocimiento de voz* (2nd ed.) Prentice Hall.
- López, F. A. (2018). *¿Qué es un corpus lingüístico y cuál es su uso?* Recuperado el 1 de julio de 2023, de <https://www.weluntranslations.com/post/2018/02/12/-qué-es-un-corpus-lingüístico-y-cuál-es-su-uso>
- López, J. (2020). *Análisis de un corpus de blogs bolivianos para el estudio del lenguaje juvenil*. Revista Iberoamericana de Educación a Distancia.
- Martínez-Cámara, E. (2019). *Lingüística de corpus para el análisis gramatical: una guía para la investigación*. Routledge.
- Moreno, A. (1998). *La lingüística computacional. Introducción a los modelos simbólicos, estadísticos y biológicos*. Editorial Síntesis.
- Sinclair, J. (1991). *Corpus, concordancia, colocación*. Oxford University Press.
- Sinclair, J. (2004). *Introducción a la lingüística de corpus*. Routledge.
- Villavicencio, A. y Bondi, M. (2008). *Lingüística de corpus. Teoría y práctica para el análisis automático del lenguaje natural*. Universidad Nacional de Córdoba.