Sincronía
Revista de Filosofía, Letras y Humanidades
Departamento de Filosofía / Departamento de Letras

e-ISSN: **1562-384X**
Año XXVI, Número 81 **Enero**-Junio 2021
DOI: **10.32870/sincronia.axxiv.n78**

UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE CIENCIAS SOCIALES Y HUMANIDADES

Una aproximación a la lingüística computacional.

An approach to computational linguistics.

DOI: 10.32870/sincronia.axxvi.n81.34a22

Mario Casado-Mancebo
Universidad Nacional de Educación a Distancia. (MÉXICO)
CE: mcasado@flog.uned.es / **ID ORCID:** 0000-0001-6003-5190

Recibido: 31/03/2021
Revisado: 28/04/2021
Aprobado: 14/11/2021

RESUMEN

En este trabajo se realiza una aproximación a los aspectos de la Lingüística computacional que tienen que ver con las tecnologías conversacionales. Con la intención de acercar sus relaciones y semejanzas con la lingüística, se parte de la Filología para explicar el surgimiento de esta especialidad de la Lingüística aplicada y se define su objetivo principal. A continuación, se profundiza en las principales aplicaciones del procesamiento del lenguaje natural en el ámbito mencionado señalando la posición que ocupa la programación en ellos y en el trabajo del lingüista. En el desarrollo de la exposición, se presentan y explican conceptos fundamentales que suponen parte de estos procesos como *modelo de lenguaje*, *interpretación de estructuras* y *reconocimiento de entidades*. Para finalizar, se ilustra todo lo anterior con dos ejemplos de aplicaciones de los procesos tratados muy frecuentes en el día a día: un sistema de reconocimiento de voz telefónico y un asistente de voz.

Palabras clave: Filología. Lingüística. Programación. Lingüística Computacional.

Sincronía
Revista de Filosofía, Letras y Humanidades
Departamento de Filosofía / Departamento de Letras
e-ISSN: **1562-384X**
Año XXVI, Número 81 **Enero**-Junio 2021
DOI: **10.32870/sincronia.axxiv.n78**
UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE CIENCIAS SOCIALES Y HUMANIDADES

ABSTRACT

In this work we make an approach to aspects of Computational linguistics that deal with conversational technologies. To get to know its similarities, we start from Philology and Linguistics to finally present the advent of this branch of Applied linguistics. We also deepen in the main tasks of natural language processing for conversation and how programming relates to them and to linguists. Along the way, we discuss key concepts such as *language model*, *structures interpreting* and *entity recognition*. To finish, we will show how computational linguistics are present in our daily life with two examples: an interactive voice recognizer and a voice assistant.

Keywords: Philology. Linguistics. Programming. Computational linguistics.

Introducción[1]

La *lingüística computacional* (LC) está cada vez más presente en nuestro día a día. Si bien hace algunos años su presencia era poco detectable por no traspasar la pantalla del ordenador, actualmente está en nuestro bolsillo –en el *smartphone*–, en nuestro salón –en el televisor–, en la cocina –en el altavoz inteligente— y, además, realizando tareas muy vistosas como mantener al día nuestra lista de la compra, reproducir la música que nos gusta o evitarnos el tedioso momento de levantarnos del sofá porque nos hemos dejado una luz encendida. En este trabajo nos aproximaremos a la lingüística computacional en lo que atañe a las tecnologías de interacción conversacional con el objetivo de conocer un poco más su funcionamiento. Hemos de tener en cuenta que la LC es una disciplina con una larga trayectoria y, por ello, el contenido presentado en este trabajo solo representa la parte del campo mencionada. Esa larga trayectoria ha estado estrechamente vinculada a la computación, la inteligencia artificial, etc.; sin embargo, en esta ocasión pondremos el foco en sus relaciones con la *filología* y la *lingüística* para mostrar el carácter

[1] El trabajo que se expone a continuación se ha realizado en un proyecto financiado por la Universidad Nacional de Educación a Distancia en el marco de las ayudas para la Formación del Personal Investigador (FPI-UNED 2019).

Considero imprescindible agradecer las observaciones y sugerencias de los profesionales de la lingüística y la lingüística computacional que han revisado este trabajo.

Sincronía
Revista de Filosofía, Letras y Humanidades
Departamento de Filosofía / Departamento de Letras
e-ISSN: **1562-384X**
Año XXVI, Número 81 **Enero**-Junio 2021
DOI: **10.32870/sincronia.axxiv.n78**
UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE CIENCIAS SOCIALES Y HUMANIDADES

vertebral que tienen en los desarrollos conversacionales. También veremos los métodos de trabajo y, finalmente, revisaremos las aplicaciones del procesamiento del lenguaje a una de las tecnologías que actualmente presenta mayor interés: los asistentes de voz.

La lingüística computacional

Comencemos por inspeccionar qué separa a la filología de la LC. El punto de partida de los estudios relacionados con la lengua es la filología, una disciplina que ha evolucionado de gran manera durante los últimos años. Tradicionalmente, la filología se ha concebido como "el estudio de textos en una lengua determinada" (Guzmán Guerra y Tejada Caller, 2000, p.21). Sin embargo, con los postulados del estructuralismo y su búsqueda de la estructura inherente del lenguaje, surge la oposición lengua-habla (De Saussure, 2004, párr. 2d) y la disciplina se distancia de los textos para ser situada en los hablantes de una comunidad. Aunque el término *linguistique* aparece por primera vez en 1812 para designar a la disciplina que se ocupa de la historia de las lenguas y de su comparación (Henry, 1812 *apud* Auroux, 1987) y así fue como llegó y pervivió en España durante mucho tiempo, con la sucesiva profundización, la lingüística se acabó constituyendo como una rama de la filología dedicada exclusivamente al estudio del lenguaje como entidad autónoma, quedando la filología restringida a las más tradicionales (Guzmán Guerra y Tejada Caller, 2000, p.20). Nos encontramos entonces con una nueva disciplina con enfoque propio: la lingüística o "el estudio científico del lenguaje"[2] (Aronoff y Rees-Miller, 2003, p.xiv). Entonces, lo que separa filología y lingüística, además de la filiación que mantiene esta con aquella, es el que la lingüística focaliza su campo de actuación sobre el lenguaje en sí mismo, dejando atrás la necesidad de que este sea un mero derivado de los textos.

Los avances en lingüística comenzaron a dotar a esta disciplina de unas sólidas herramientas de análisis, lo que permitió que surgieran toda una serie de enfoques basados en la aplicación de los sucesivos hallazgos. La *lingüística aplicada* (LA) o "investigación teórica y empírica de los problemas del mundo real que tienen el lenguaje como eje central" (Brumfit, 1997, p.93 citado por Davies y

[2] Salvo que se indique lo contrario, todas las traducciones son del autor de este trabajo.

Sincronía
Revista de Filosofía, Letras y Humanidades
Departamento de Filosofía / Departamento de Letras
e-ISSN: **1562-384X**
Año XXVI, Número 81 **Enero**-Junio 2021
DOI: **10.32870/sincronia.axxiv.n78**
UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE CIENCIAS SOCIALES Y HUMANIDADES

Elder, 2004, p.3) es la rama de la lingüística que los agrupa. En ella priman los enfoques interdisciplinares, pudiendo encontrar a la lingüística en colaboración con la psicología (psicolingüística), el derecho (lingüística forense) o la medicina (lingüística clínica), entre otras muchas opciones. En el seno de la LA es donde surge la LC como disciplina que utiliza los métodos de la computación para procesar el lenguaje. Así, Moreno Sandoval (1998, p.14) la define como «el desarrollo de programas de ordenador que simulan la capacidad lingüística humana».

Puesto que la disciplina compañera de la lingüística, en esta materia, es la computación o, como señalaba Moreno Sandoval (1998) "el desarrollo de programas de ordenador", la LC, requiere trabajos de procesamiento y transmisión de voz, procesamiento y generación de datos y cadenas de texto y también tareas de hardware, puesto que todos los trabajos mencionados necesitan un entorno de máquinas que los alojen. El lingüista se ocupará de las tareas que lidian con la lingüística (elaboración de corpus y modelos lingüísticos, revisión de muestras y datos, etc.), aunque en la mayoría de las ocasiones estas tareas estarán anidadas en productos de la ingeniería: archivos de códigos, bases de datos, servidores físicos y virtuales. El compañero del lingüista, por lo tanto, será un ingeniero.

Procesamiento del lenguaje natural

El modelo de lenguaje

El *procesamiento del lenguaje natural* (PLN) es el procedimiento que tiene la LC para simular la capacidad humana del lenguaje. Como el cerebro humano, un sistema de PLN tiene una entrada de datos sobre los que se aplican una serie de procesos deductivos para sacar de ellos unas conclusiones lingüísticas. Para alcanzar estas conclusiones, normalmente es necesario el apoyo de un *modelo de lenguaje*. El modelado lingüístico es, para el lingüista, el núcleo del PLN: por un lado, es el primer paso en el desarrollo de cualquier sistema de procesamiento y, por otro, es donde se requieren las competencias de un lingüista para ser capaz de captar la muestra de lenguaje adecuada, variada y suficiente. Para elaborar un modelo de lenguaje de forma efectiva en el ámbito

Sincronía
Revista de Filosofía, Letras y Humanidades
Departamento de Filosofía / Departamento de Letras
e-ISSN: **1562-384X**
Año XXVI, Número 81 **Enero**-Junio 2021
DOI: **10.32870/sincronia.axxiv.n78**
UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE CIENCIAS SOCIALES Y HUMANIDADES

de las tecnologías conversacionales es fundamental definir previamente algunos aspectos de nuestro proyecto:

a) Qué tipo de procesamiento quiero realizar.
b) Qué tipo de datos necesito para representarlo.
c) De dónde puedo recuperarlos.
d) Cuál es la mejor forma de organizarlos.

La pregunta a) tiene frecuentemente dos posibles orientaciones: la competencia fónica–procesamientos y síntesis de voz para emular la capacidad de los hablantes tanto para interpretar señales acústicas como para generarlas y codificar mensajes en ellas, y la competencia gramatical–procesamiento y generación de cadenas de texto. Más adelante profundizaremos en las posibilidades de estos análisis. Con la respuesta a la pregunta a), sabremos si necesitamos datos lingüísticos de voz o de texto. Sin embargo, quedaría por definir en qué tipo de elocuciones estoy interesado: voces individuales, diálogos, con/sin ruido; textos formales, informales, que reflejen la oralidad, etc. y, con ello, un aspecto fundamental: dónde puedo conseguir esas muestras (pregunta c)). Las muestras de voz son más comunes en forma individual; es decir: grabaciones de un solo individuo. Respecto a las muestras de texto, una fuente importante de testimonios escritos formales son los periódicos en línea. La mayoría de ellos tienen una sección en la que actualizan periódicamente el contenido más reciente en formato XML para optimizar la distribución unificada del contenido (el XML solo contiene la información estructurada: el formato, estilo o visualización lo ponen los programas que se encarguen de mostrar esos contenidos). Esta sección se conoce como *feed RSS* (del inglés *Really Simple Syndication*). Por lo tanto, es un tipo de fuente de información que nos permite acceder cómodamente a las distintas partes de una noticia y recoger las que necesitemos. Si, por el contrario, necesitamos unas muestras de lenguaje más espontáneas, una gran fuente de testimonios son las redes sociales, que en la mayoría de los casos también cuentan con accesos a la fuente de información estructurada.

Sincronía
Revista de Filosofía, Letras y Humanidades
Departamento de Filosofía / Departamento de Letras
e-ISSN: **1562-384X**
Año XXVI, Número 81 **Enero**-Junio 2021
DOI: **10.32870/sincronia.axxiv.n78**
UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE CIENCIAS SOCIALES Y HUMANIDADES

Una vez que hemos encontrado la mejor fuente para obtener los datos que necesitamos, solo nos queda plantear una organización óptima. Esta parte del proceso es fundamental. Los datos que vayamos recogiendo cada vez que hagamos búsquedas irán configurando un corpus, definidos estos como "colecciones extensas de datos lingüísticos orales o escritos" (Moreno Sandoval, 1998, p.26). Los corpus bien diseñados son aquellos que se pueden utilizar para propósitos, proyectos o experimentos muy diferentes. Por este motivo, es fundamental responder a la cuestión d) desde el comienzo de nuestro planteamiento. Un buen diseño de corpus permitirá que en futuras ocasiones podamos acceder cómodamente a la totalidad o al subconjunto que necesitemos de los datos que hemos recogido.

Los corpus en LC se organizan convencionalmente en listas de *objetos*. Un objeto es un elemento abstracto que representa unos comportamientos y propiedades comunes (Phillips, 2010, p.7). Esto quiere decir que nuestros datos deben estar organizados en agrupaciones de elementos con las mismas propiedades. Las RSS de un periódico, retomando el ejemplo anterior, se organizan en objetos que representan noticias. De este modo, cada objeto *noticia* tiene las propiedades comunes *titular*, *fecha*, *autor*, etc. La organización de los datos en objetos permite generar archivos de *datos estructurados*. Este tipo de archivos se caracteriza, como indica su nombre, por tener una estructura interna: la que le hemos dado a nuestros objetos: una estructura basada en las relaciones constantes entre las propiedades que definen nuestros datos. Y el tener nuestros corpus organizados de esta manera es lo que nos permite una gestión y acceso óptimos en cualquier momento que necesitemos recuperar un elemento recóndito del corpus.

Una vez tenemos los datos y están organizados, podemos pasar al modelo de lenguaje. Actualmente existen modelos muy variados que se valen de diferente manera de los avances en computación. Un tipo muy frecuente es el estadístico, que se elabora a partir de corpus de datos lingüísticos lo suficientemente amplios y variados como para servir de representación de lo que el sistema de PLN puede esperar y lo que debe hacer con ellos; es decir, con este modelo de lenguaje, enseñamos al procesador qué cosas esperaremos que entienda y qué debe hacer con ellas. Está formado, por lo tanto, por datos lingüísticos y, en general, serán datos de voz o datos de texto.

Sincronía
Revista de Filosofía, Letras y Humanidades
Departamento de Filosofía / Departamento de Letras
e-ISSN: **1562-384X**
Año XXVI, Número 81 **Enero**-Junio 2021
DOI: **10.32870/sincronia.axxiv.n78**
UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE CIENCIAS SOCIALES Y HUMANIDADES

Para generar un modelo de estas características a partir de nuestros datos, lo más habitual es recurrir al estudio de *n-gramas*. Este modelo se basa en la asunción de que "solo unas unidades anteriores condicionan la probabilidad de aparición de la siguiente unidad" (Moreno Sandoval, 1998, p.181). En el concepto *n-grama*, la *n* determina la cantidad de unidades contextuales que se toman para el modelo: unigramas, una; bigramas, dos; trigramas, tres; etc. De esta manera, se organiza toda una red de conjuntos ordenados de frecuencias de aparición que permiten estimar la probabilidad de que una forma aparezca detrás de otra u otras dentro del modelo. Como se puede apreciar, este modelo comparte con la gramática la idea de que las unidades se distribuyen en el contexto de manera interdependiente (los adjetivos son clases de palabras que acompañan a nombres, los determinantes a estos, etc.).

El problema de esta forma de análisis es que se requieren corpus de entrenamiento muy grandes para poder dar cabida, de manera fiable, a todos los contextos que queremos modelar. Las lenguas con morfologías muy ricas agravan este problema, ya que, por un lado, cada forma flexiva de una misma forma léxica incrementa el tamaño del modelo; y, por otro, muchas formas no aparecen lo suficientemente representadas en el corpus como para que el modelo pueda predecirlas (Bojanowski, et al., 2017, p.135). Las investigaciones han mostrado que una manera de hacer más eficientes los modelos lingüísticos contextuales es recurrir a la información por debajo del nivel de la palabra, tanto a nivel morfológico como de caracteres (Bojanowski, et al., 2017, sec. 2). Este último caso ha mostrado un mejor aprovechamiento de las estructuras, puesto que las atomiza al máximo, logrando entrenamientos más rápidos con unos corpus menos exigentes (Bojanowski, et al., 2017; Joulin, et al., 2017).

La complejidad que puede llegar a alcanzar solamente el proceso de recolección de datos es elevada. Recoger manualmente los datos o calcular los *n*-gramas de un corpus de manera individual, dificultaría en gran medida la consecución de algún objetivo a corto o medio plazo. La programación es la herramienta que los lingüistas computacionales utilizan para agilizar los procesos: trabajos repetitivos, cálculos de gran magnitud, extracción de datos masivos, etc. Como herramienta, puede presentarse en formas diferentes, pero esencialmente iguales: hay múltiples lenguajes de

Sincronía
Revista de Filosofía, Letras y Humanidades
Departamento de Filosofía / Departamento de Letras
e-ISSN: **1562-384X**
Año XXVI, Número 81 **Enero**-Junio 2021
DOI: **10.32870/sincronia.axxiv.n78**
UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE CIENCIAS SOCIALES Y HUMANIDADES

programación (y periódicamente surgen nuevos), pero los fundamentos de la mayoría de ellos comparten similitudes.

Procesamiento de voz

Uno de los enfoques del PLN, como veíamos arriba, es el que reproduce las competencias fónicas de los humanos. Una de las posibilidades es el procesamiento de voz o ASR (*Automatic speech recognition*), que se ocupa de la interpretación de señal acústica o lo que es lo mismo: la "transformación de una señal acústica en una secuencia de palabras sin que sea necesario comprender el significado o la intención de lo dicho" (Renals y Hain, 2013, p.299). Sin embargo, como los mismos autores señalan, el reconocimiento de voz a texto no siempre es posible sin contar con la semántica, puesto que se pueden presentar secuencias ambiguas en cuanto a la segmentación que solo el contexto semántico puede resolver: *Ven arcos en las cumbres borras cosas* frente a *Ve narcos en las cumbres borrascosas*[3]. Pero la segmentación no es el único problema. La variación intralingüística también implica una gran dificultad de reconocimiento. Por un lado, las emisiones de cada hablante son únicas, puesto que la anatomía, fisiología e idiolecto individuales son únicos. Por otro, las emisiones de una misma secuencia por un mismo hablante también son diferentes en cada ocasión. Aun así, los modelos actuales de procesamiento de voz son capaces de lidiar de una manera eficiente con la variabilidad.

Los motores de ASR toman como entrada la señal acústica; sin embargo, para el reconocimiento no se hace un análisis directo de ella. Un preprocesado la filtra y extrae los rasgos acústicos relevantes (Renals y Hain, 2013, p.304). El proceso de reconocimiento se basa en un modelo de hipótesis de probabilidades apoyado en un modelo de lenguaje. Para los procesadores de voz se requieren dos tipos de modelo: uno acústico, que permite identificar una serie de rasgos de la señal con representaciones lingüísticas, y otro lingüístico, que generalmente se compone de oraciones. Generalmente, el modelo acústico no se basa en fonemas sino en difonemas o

[3] Agradezco estos ejemplos al dr. José María Lahoz Bengoechea de la Universidad Complutense de Madrid y a la dra. María del Carmen Horno Chéliz de la Universidad de Zaragoza.

Sincronía
Revista de Filosofía, Letras y Humanidades
Departamento de Filosofía / Departamento de Letras
e-ISSN: **1562-384X**
Año XXVI, Número 81 **Enero**-Junio 2021
DOI: **10.32870/sincronia.axxiv.n78**
UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE CIENCIAS SOCIALES Y HUMANIDADES

trifonemas: segmentos acústicos que incluyen la parte central del fonema y las transiciones al anterior y/o al siguiente para poder capturar la coarticulación de una forma eficiente y evitar tener que modelar un ejemplar para cada manifestación acústica de un fonema.

Procesamiento de texto

La salida de ASR es una cadena de texto con el reconocimiento de elementos que se han procesado a partir de la señal acústica. Aunque ya tenemos unos datos procesables computacionalmente, no sería eficiente alimentar un procesador de texto con las salidas automáticas del motor de reconocimiento de voz. Habitualmente, la normalización audio-texto genera cadenas totalmente expandidas, ya que, como hemos visto, la unidad de análisis del modelo acústico está relacionada con el fonema. Entonces, el reconocimiento automático devolverá secuencias de texto alineadas fonológicamente con el audio: *Busco un audi a tres*, *Fotos del nokia treinta y tres diez* o *Información del vuelo efe erre tres tres cuatro*. No es habitual que un modelo lingüístico de procesamiento de texto estuviera preparado para estas secuencias, sino para *Busco un Audi A3*, *Fotos del Nokia 3310* e *Información del vuelo FR334*. Por lo tanto, antes de iniciar el procesamiento de texto, se realiza una *normalización inversa* de las cadenas de texto que salen del módulo de audio. La normalización inversa se llama así porque toma como entrada unas secuencias que ya han sido normalizadas de audio a texto y se les aplican unas reglas de formateo de texto. Dado que el objetivo es que las secuencias de entrada al módulo de procesamiento de texto sean las esperadas, se considera otra forma de normalización.

Interpretación de estructuras

Dependiendo de nuestro objetivo podemos emprender diferentes procesos que se pueden encadenar unos con otros para profundizar de diferente manera en el procesamiento del texto. Un primer acercamiento podría ser la interpretación de las cadenas de texto según su estructura. En este proceso, se recurre a la distribución sintáctica lineal de la estructura, prescindiendo de la semántica individual de sus partes, para tratar de asignar un significado global determinado a las

Sincronía
Revista de Filosofía, Letras y Humanidades
Departamento de Filosofía / Departamento de Letras
e-ISSN: **1562-384X**
Año XXVI, Número 81 **Enero**-Junio 2021
DOI: **10.32870/sincronia.axxiv.n78**
UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE CIENCIAS SOCIALES Y HUMANIDADES

estructuras. No se busca atomizar la estructura en busca del detalle semántico, sino especificar unos valores determinados. Podríamos buscar el tipo de escrito y, entonces, recurriríamos a la estructura para determinar si es una carta, un correo electrónico, una novela... O podríamos identificar los sentimientos, asignando a cada cadena el valor *tristeza, felicidad, euforia*, etc.

Sin embargo, este tipo de análisis, al carecer de propiedades semánticas, solo es capaz de decirnos la probabilidad de que una secuencia (del nivel que sea: carácter, palabra, oración, etc.) sea tal por los elementos que la componen. En la mayoría de modelos de PLN una interpretación estructural no es suficiente y se requiere algún acceso a la información semántica (Moreno Sandoval, 1998: 120). Una buena forma de acceder a la semántica es el *reconocimiento de entidades*. Se trata de un proceso que atribuye etiquetas a elementos dentro de las estructuras que procesamos. Por ejemplo, en la secuencia *Busco un Audi A3,* unas posibles entidades serían *marca* (Audi) y *modelo* (A3). Es un tipo de análisis que, en combinación con un modelo de *n-gramas*, resulta muy sólido, puesto que permite determinar las etiquetas o entidades de los elementos mediante su distribución contextual. De igual manera que con el método de análisis anterior, el reconocimiento de entidades requiere un entrenamiento previo con un corpus que contemple de una forma suficiente y variada los elementos que identificaremos en nuestro procesamiento.

Si en el ejemplo anterior propusimos un analizador de estructuras que, a nivel global, nos permitía determinar a través de los *n*-gramas el tipo de texto con el que se podía identificar la secuencia de texto, un modelo de lenguaje que, además, incluyera el reconocimiento de entidades representando las categorías gramaticales añadiría la posibilidad de hacer, por ejemplo, un análisis cuantitativo de las categorías gramaticales en cada tipo de texto. Con este tipo de PLN también podríamos optar por un reconocimiento de entidades más semántico y analizar la presencia de marcas comerciales, animales, títulos cinematográficos... Cualquier noción es susceptible de ser entrenada para analizarla en estos modelos.

Sincronía
Revista de Filosofía, Letras y Humanidades
Departamento de Filosofía / Departamento de Letras
e-ISSN: **1562-384X**
Año XXVI, Número 81 **Enero**-Junio 2021
DOI: **10.32870/sincronia.axxiv.n78**
UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE CIENCIAS SOCIALES Y HUMANIDADES

De la lingüística a la acción

Hasta el momento hemos ejemplificado los modelos de análisis expuestos con casos de aplicación muy concretos. La LC, sin embargo, puede realizar un aprovechamiento mayor de la información contenida en los enunciados. Solo necesitamos subir de nivel lingüístico para comprobar el alcance que tiene un modelo de PLN. La clave reside en la pragmática, desde la que podemos hacer un análisis intencional de las locuciones. Este tipo de análisis es el que está presente en los productos de la LC con los que estamos en contacto frecuente. Con nuestras elocuciones son capaces de determinar nuestra intención y producir una acción en consecuencia.

Cuando llamamos a una compañía telefónica es frecuente que de primeras nos atienda un procesador de lenguaje natural que funciona con un sistema de *reconocimiento de voz interactivo* ('interactive voice recognizer', IVR). Este sistema recibe nuestra elocución como señal de audio y la procesa con un motor de ASR (cf. §

Procesamiento de voz) para obtener una cadena de texto que procesar lingüísticamente. Mediante un análisis de la cadena de texto se determina nuestra intención y se desencadena una acción en consecuencia: trasladar nuestra llamada al departamento que corresponda a la intención reconocida. El análisis, sin embargo, no se queda ahí. Pensemos en una elocución como *Quiero contratar una tarifa de 50 gigas con llamadas ilimitadas*. El análisis global de la secuencia de texto desaprovecha gran parte de la información, por lo que es habitual que el modelo de lenguaje incluya una serie de entidades esperables en el contexto. En este caso, una propuesta de entidades podría ser *datos* y *llamadas* de manera que, al pasar por el procesador, la secuencia podría ser etiquetada de la siguiente manera [*Quiero contratar una tarifa de* [*50 gigas*]$_{\text{datos}}$ *con llamadas* [*ilimitadas*]$_{\text{llamadas}}$]$_{\text{contrataciones}}$. Con este análisis, no solo podríamos desencadenar la acción de pasarnos al departamento de contrataciones, sino que además podríamos presentar al cliente la tarifa más adecuada a su petición.

Otro ejemplo, que además está de moda actualmente, lo forman los asistentes de voz. En forma de aplicación del móvil, altavoz doméstico y un sinfín de variantes, estos sistemas también son producto de LC. El funcionamiento es parecido, pero su repercusión o alcance en nuestro día a día es más vistoso, por lo que han sido los detonantes de la emersión de la LC. El interés principal de

Sincronía
Revista de Filosofía, Letras y Humanidades
Departamento de Filosofía / Departamento de Letras
e-ISSN: **1562-384X**
Año XXVI, Número 81 **Enero**-Junio 2021
DOI: **10.32870/sincronia.axxiv.n78**
UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE CIENCIAS SOCIALES Y HUMANIDADES

los asistentes de voz es que cuentan con acceso a Internet y, en muchos casos, al sistema de aplicaciones de nuestro entorno móvil. Esto permite que la intencionalidad de los usuarios se pueda escalar al infinito: cualquier información presente en internet o cualquier aplicación de nuestro móvil se vuelve susceptible de ser accionada mediante el lenguaje. La complejidad del sistema de PLN, en consecuencia, también aumenta. El entrenamiento de los modelos ya no está acotado a un contexto lingüístico específico (en el ejemplo anterior, la telefonía). En este caso, el procesador de lenguaje natural tendría que estar entrenado para comprender y desencadenar una acción ante todos los contextos que se nos ocurran.

Actualmente no es realista pretender que un ordenador esté preparado para cubrir todos los usos que quepan en la imaginación de sus usuarios. La primera limitación y la fundamental sería el espacio físico. Estos sistemas no dejan de ser ordenadores con una memoria y unos recursos asignados. Como veíamos en el apartado **¡Error! No se encuentra el origen de la referencia.**, la organización previa constituye un escalón ineludible en el proceso de desarrollo. Es necesario definir qué casos se procesarán y qué casos no y, en esta última circunstancia, definir el comportamiento del asistente. De esta manera, cuando el usuario intenta iniciar una interacción que no cabe dentro de los límites del asistente, este no le devuelve un error ni lo contraría, sino que es capaz de dar una respuesta normalizada, dando la sensación de que la interacción ha sido exitosa.

Generación de lenguaje natural

Como acabamos de ver en los ejemplos, una parte fundamental de los sistemas de PLN diseñados para la interacción es la posibilidad de entablar diálogos, de proporcionar una respuesta al usuario en relación con el procesamiento de su elocución. Para poder estructurar correctamente un sistema de diálogo, es necesario haber establecido previamente uno de PLN que se encargue de procesar las locuciones de partida y de desencadenar la acción correspondiente. Una vez hecho esto, podemos establecer un sistema de *generación de lenguaje natural* (GLN) que proporcione una reacción (respuestas, preguntas complementarias...) al usuario. En el apartado referente al PLN, hemos visto

Sincronía
Revista de Filosofía, Letras y Humanidades
Departamento de Filosofía / Departamento de Letras
e-ISSN: **1562-384X**
Año XXVI, Número 81 **Enero**-Junio 2021
DOI: **10.32870/sincronia.axxiv.n78**
UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE CIENCIAS SOCIALES Y HUMANIDADES

los diferentes grados de procesamiento que se le pueden aplicar a los enunciados y, en función de la información que extraigamos, podremos generar unos diálogos más o menos precisos. Si nuestro procesamiento únicamente reconoce estructuras de manera global, solo podremos generar diálogos especializados en cada tipo de estructura. Lo hemos visto en el ejemplo de la compañía telefónica: solo si procesamos las entidades que componen los enunciados podemos dar una respuesta más concreta y personalizada.

Estas respuestas pueden proporcionarse en texto, tal como se generan en nuestro módulo de GLN, o en formato hablado utilizando la síntesis de voz. Los sistemas que toman los diálogos generados y los sintetizan imitando la voz humana se denominan TTS (*Text to speech*). Los motores de TTS tienen un sistema de normalización y un modelo acústico. El modelo acústico, basado en versiones complejas del fonema (difonema, trifonema, etc.) para ser capaz de emular la coarticulación natural del lenguaje, permite convertir las unidades de texto en representaciones acústicas. Al tratar los motores de ASR, comentamos cómo la correspondencia entre representaciones gráficas y representaciones acústicas requiere una alineación fonológica exhaustiva. Por este motivo, se incorpora el sistema de normalización, que se ocupa de expandir al máximo las cadenas de diálogo generadas por el módulo de GLN: *3* se convierte en *tres*, *las 12:45* en *la una menos cuarto* y así sucesivamente. Con todo esto, cuando un usuario dirige una elocución al sistema de LC, no solo se desencadena una acción como consecuencia de la intencionalidad, sino que además se le proporciona una respuesta personalizada.

Conclusiones

Hemos visto que la filología y la LC están separadas por la lingüística. Con la ramificación de la filología, surge la lingüística y, a través de los planteamientos de LA en interdisciplinariedad con la computación, surge la LC, una disciplina cuyo objetivo es la reproducción de diferentes aspectos de la facultad humana del lenguaje utilizando las técnicas de la computación.

En el PLN es donde se ponen en práctica los métodos y modelos de la LC. La base del procesamiento son los modelos de lenguaje: corpus de muestra elaborados por el lingüista para

Sincronía
Revista de Filosofía, Letras y Humanidades
Departamento de Filosofía / Departamento de Letras
e-ISSN: **1562-384X**
Año XXVI, Número 81 **Enero**-Junio 2021
DOI: **10.32870/sincronia.axxiv.n78**
UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE CIENCIAS SOCIALES Y HUMANIDADES

dotar al sistema computacional que frecuentemente consisten en un modelo representativo de las unidades que se espera que interprete y cómo debe interpretarlas. Este tipo de modelos de lenguaje son diferentes para cada tipo de procesamiento: los enfocados al procesamiento de voz se componen de muestras acústicas y los enfocados al procesamiento de texto de muestras semejantes a las unidades esperadas. Es fundamental planificar exhaustivamente la estructuración de los datos, puesto que la reutilización de corpus no solo es habitual, sino también señal de que los datos están seleccionados con buen criterio y organizados de una manera óptima. La programación, por su parte, es una herramienta que nos permite agilizar los procesos de desarrollo en la LC.

Una vez hemos planificado los aspectos teóricos y estructurales de nuestro proyecto y hemos recogido los testimonios necesarios, es el momento de proceder a la aplicación de las técnicas de procesamiento. Si hemos optado por procesar muestras de voz, el primer paso será realizar un reconocimiento automático que transforme la voz en texto. Son útiles los motores de ASR, que, mediante modelos acústicos y lingüísticos, son capaces estimar la transcripción óptima para las señales acústicas. La salida de ASR son textos expandidos que, normalmente, no son aptos para el procesamiento lingüístico de texto. Es necesario aplicar una serie de normalizaciones que conviertan esas secuencias en bruto en textos adecuados para los modelos lingüísticos de PLN.

El procesamiento de texto puede ocurrir en diferentes fases según nuestro objetivo, obteniendo diferentes resultados en cada una. Si nuestras necesidades únicamente incluyen el reconocimiento de tipos de estructuras a nivel global, un análisis de probabilidad combinatoria basado en *n*-gramas podría ser suficiente para determinar la categoría de cada estructura. Sin embargo, si tenemos interés en acceder al contenido interno de las secuencias, necesitaremos profundizar en nuestro análisis. El reconocimiento de entidades es la herramienta que nos permite crear modelos de reconocimiento capaces de identificar etiquetas de tipo semántico dentro de nuestras muestras. Cualquier noción es susceptible de ser entrenada en nuestro modelo: categorías gramaticales, tipos de cosas, marcas, nombres propios... Estas dos formas de enfocar el análisis lingüístico no solo son compatibles, sino que, en conjunto, aportan unos resultados más sólidos y con una mayor eficiencia computacional.

Sincronía
Revista de Filosofía, Letras y Humanidades
Departamento de Filosofía / Departamento de Letras
e-ISSN: **1562-384X**
Año XXVI, Número 81 **Enero**-Junio 2021
DOI: **10.32870/sincronia.axxiv.n78**
UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE CIENCIAS SOCIALES Y HUMANIDADES

El alcance de la LC se puede maximizar tan solo pasando de la sintaxis y la semántica a la pragmática. En una interacción con un hablante, un buen reconocimiento y análisis intencional nos permite comprender el fin de la interacción y modelar acciones que respondan a cada elocución. Dependiendo del contexto de nuestros sistemas, tendremos acceso a unos recursos que nos permitirán manifestar unas consecuencias reales a cada petición. Cuando el alcance de nuestro sistema de PLN es muy amplio, es necesario acotar los casos de uso para poder modelar el comportamiento cuando la interacción va más allá de los límites estimados.

De igual manera, hemos visto que, cuando los sistemas de PLN implican una interacción con usuarios, es una buena práctica generar diálogos que proporcionen una retroalimentación del proceso de computación al usuario. Estos diálogos pueden aparecer en forma de texto parametrizable según las categorías que hayamos establecido en nuestro sistema de procesamiento, pero también en forma sonora mediante la síntesis de voz. Una vez hemos definido unos diálogos para las posibles interacciones, podemos incorporar un modelo de TTS para convertir esas representaciones textuales en señal acústica que imita la voz humana y puede simular un diálogo oral con el interlocutor.

Referencias

Aronoff, M. y Rees-Miller, J. (2003). *The Handbook of Linguistics*. Blackwell Publishers.

Auroux, S. (1987). The first Uses of the French Word "Linguistique" (1812- 1880). En: Aarsleff, H.; Kelly, L.G.; Niederehe, H.J (eds.), *Papers in the History of Linguistics*, John Benjamins, Amsterdam. 447-459.

Bojanowski, P.; Grave, E.; Joulin, A.; Mikolov, T. (2017): Enriching Word Vectors with Subword Information. *Transactions of the Association for Computational Linguistic*, (5) 135–14.

Brumfit, C. (1997). How applied linguistics is the same as any other science. *International Journal of Applied Linguistics*, *7*(1). 86-94.

Davies, A. y Elder, C. (2004). *The Handbook of Applied Linguistics*, Blackwell Publishing Ltd., Massachussets.

Sincronía
Revista de Filosofía, Letras y Humanidades
Departamento de Filosofía / Departamento de Letras
e-ISSN: **1562-384X**
Año XXVI, Número 81 **Enero**-Junio 2021
DOI: **10.32870/sincronia.axxiv.n78**
UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE CIENCIAS SOCIALES Y HUMANIDADES

De Saussure, F. (2004). *Escritos sobre lingüística general* (Simon Bouquet, Rudolf Engler y Antoinette Weil, Eds.; Clara Ubaldina Lorda Mur, Trad.), Gedisa editorial, Barcelona.

Guzmán, A. y Tejada, P. (2000). *¿Cómo estudiar Filología?,* Alianza Editorial, Madrid.

Henry, G. (1812). *Historie de la langue Française*, Leblanc, París.

Joulin, A., Grave, E.; Bojanowski, P. y Mikolov, T. (2017). Bag of Tricks for Efficient Text Classification. *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics.* (2). 427-431.

Moreno, A. (1998). *Lingüística computacional. Introducción a los modelos simbólicos, estadísticos y biológicos*, Síntesis, Madrid.

Phillips, D. (2010). *Python 3 Object Oriented Programming*, Packt Publishing, Birmingham.

Renals, S. y Hain, T. (2013). Speech recognition. En: Clark, A. Fox, C. y Lappin, S. (eds.), *The handbook of computational linguistics and natural language processing*, Wiley-Blackwell, Oxford.