

# LAS EMPRESAS COMO USUARIAS DE PRODUCTOS DE LA LENGUA, A TRAVES DE LOS MEDIOS TELEMATICOS

CARMEN DIEZ CARRERA

## INTRODUCCION

El tema abordado en este artículo se centra en los diversos productos de la lengua, fruto de las actividades que las empresas dedicadas a la lingüística informática ofrecen a otras empresas, instituciones, universidades y usuarios en general, difundidos a través de los medios telemáticos, entre los cuales cabe citar el Ibertex en España. Tanto las empresas como los usuarios en general necesitan de estos productos, en algunos casos, como medios de obtención de información y como fin en sí mismos, en otros.

## NECESIDAD DE ESTOS PRODUCTOS

Cada día se hace más evidente en nuestra sociedad la necesidad del uso de productos de la lengua. Las primeras tentativas parecen surgir del deseo de conseguir una comunicación universal entre los hombres: los tres mil idiomas y seis mil dialectos que existen en el mundo ocasionan barreras lingüísticas que se intentan paliar con la creación de lenguas artificiales, pensemos en el esperanto; y, en este siglo, con el desarrollo de la traducción automática. La investigación en este campo concreto nos da el primer producto de la lengua, que surge para romper la incomunicación entre los hombres. Éste es uno de los que más auge está adquiriendo y resulta esperanzador observar el aumento de gastos por tal concepto. El

mayor número de textos para traducir proceden del inglés, -con un aumento día a día considerable del japonés-, lo que pone de manifiesto, como se ha dicho reiteradamente, que el uso y difusión de una lengua depende del nivel de desarrollo tecnológico del país donde se habla, más que del número de hablantes. Como consecuencia de esto, el mayor desarrollo de los productos de la lengua, lo encontramos en los países anglosajones.

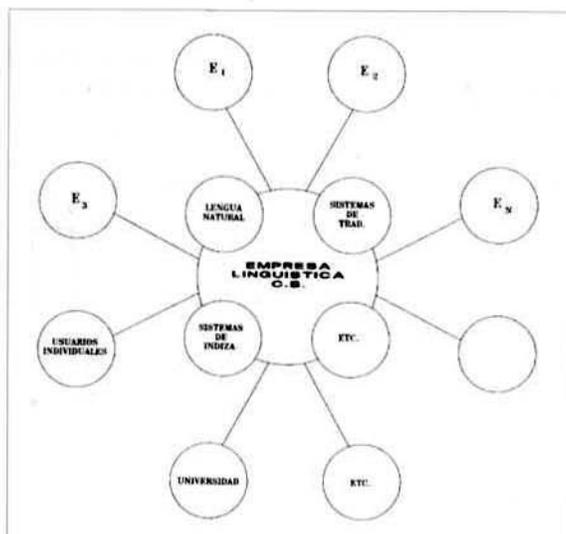
Si el punto de partida fue romper las barreras lingüísticas entre los hombres, (estamos haciendo una justificación desde el punto de vista lingüístico, pues en el fondo sabemos que siempre existen móviles económicos y de progreso), la investigación continua con un deseo de romper también otras fronteras: las que separan a los hombres y a las máquinas. De nuevo, la lengua es la protagonista y las investigaciones se orientan a alcanzar un medio más eficaz para obtener un trabajo más rápido y eficiente a través de las máquinas.

De todo ello se desprende que las empresas necesitan de estos productos para hacer frente al rápido desarrollo de la tecnología y para poder acceder a la ingente información que necesitan. Esto queda reflejado en la constante creación de bancos de datos, en la proliferación de nuevos y más sofisticados productos informáticos, etc. que encuentran en la lengua la herramienta más eficaz para normalizar y facilitar su uso.

## PRODUCTOS DE LA LENGUA Y MEDIOS TELEMATICOS

Hecha esta justificación de la necesidad de los productos de la lengua, podemos situar el entorno de estudio, dentro de la Lingüística Informática (tecnología lingüística) que, a su vez, se integra en las denominadas Industrias de la Lengua. Pero, ¿qué son las industrias de la lengua?: son el resultado de informatizar sus elementos y las reglas que rigen su funcionamiento. Con esto se consigue un análisis automático de la lengua, no aún con la profundidad deseada ni en todos los niveles lingüísticos, pero útil para elaborar productos (sistemas de indización, sistemas terminológicos, sistemas de traducción, etc.) que ayudan a realizar trabajos en los que la lengua es protagonista, de una manera más rápida y precisa. Al automatizarse los elementos de la lengua, surge un mercado potencial de productos lingüísticos demandados por el mercado, que deriva a su industrialización. Una industria que requiere múltiples saberes y, por tanto, necesita equipos interdisciplinarios para poderse desarrollar: la lingüística, las comunicaciones, la informática, la psicología, la fisiología... etc., están imbricadas, girando en torno a la primera.

En este artículo, nos estamos centrandó en la parte lingüística de las industrias de la lengua y, recalcando el objetivo, vamos a analizar algunos de los productos que pueden ser usados de forma compartida por las distintas empresas, a través de los medios telemáticos, como mostramos en el gráfico siguiente:



C.S.: Centro Servidor.

E: Empresa.

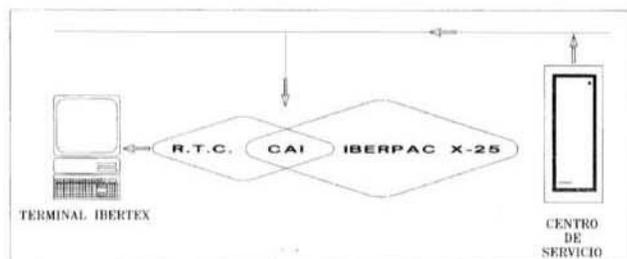
ETC.: otros productos / otros usuarios, según el caso.

Según observamos, los protagonistas de la creación de productos que permiten el acceso a la información, y del manejo de la información son empresas: empresas lingüísticas, como creadoras de los productos de la lengua, y empresas de cualquier otro tipo como usuarias y consumidoras de esos productos.

Hoy en día, las empresas dedicadas a la lingüística informática están adquiriendo gran notoriedad y relevancia, -un ejemplo de ellas es la Sociedad francesa ERLI-; éstas cuentan con un campo de aplicación de gran incidencia en las empresas demandantes en cuanto que se encargan de desarrollar productos útiles para usuarios que no pueden soportar, o bien no es de su competencia, el gasto de investigación y desarrollo que conlleva. Ahora bien, uno de los problemas que se plantean radica en la distribución de sus productos, es decir, su comercialización y una de las soluciones que contemplamos es el uso de los servicios telemáticos, mediante los cuales las empresas lingüísticas pueden facilitar sus productos a través de un centro servidor de la red Ibertex española, o bien crear su propio centro.

Los productos lingüísticos constituyen una gama de servicios que facilitan el acceso a la información, haciendo uso de tecnologías como la informática y las

redes de comunicación. El videotex es el soporte propuesto como herramienta compartida por las empresas, tanto para el acceso a la información como para el uso de los productos de la lengua. Veamos de manera gráfica su funcionamiento<sup>1</sup>:



Aunque este servicio telemático es cada día más conocido, resumamos los requisitos para su uso. Se necesita un terminal específico o un Pc y una tarjeta modem conectada a Ibertex, que es la red formada por la red telefónica conmutada de uso habitual y la red Iberpac de transmisión de datos; a través de un 'interface' sencillo entre el hombre y la máquina, compuesto de comandos de función, el videotex nos proporciona una serie de opciones que van desde la consulta de información distribuida en bases de datos, al acceso a aplicaciones teleinformáticas, pasando por la comunicación entre terminales.

Así, una empresa de lingüística se encargará de desarrollar productos que sirvan, según se ha dicho al principio, como medios de acceso a la información: el tratamiento de la lengua natural, por ejemplo, en el manejo de las bases de datos, y como fin: en aquellas aplicaciones en las que la lengua trata sobre la lengua -la metalengua-, como en el caso de la traducción automática. Por tanto, la lengua natural o humana es la base de la comunicación, no sólo ya entre hombre y hombre, también y es a lo que se aspira, entre el hombre y la máquina y las máquinas entre sí. La lengua es, por consiguiente, un elemento fundamental en el uso y difusión de la información y, por tanto, de gran importancia ya que la información es pieza clave en nuestra sociedad, nadie duda de ello, e incluso se tiene conciencia de que la información aumenta la productividad. La lengua natural, dentro del planteamiento que estamos haciendo, y el soporte videotex son dos 'interfaces' de comunicación interactiva.

Una empresa lingüística se encarga de elaborar, entre otros, los siguientes productos:

1. Tomado de FUINCA: Los servicios de información electrónica. Bases de datos y videotex. Madrid: Fundesco, 1990, p. 146.

- Productos referidos a los sistemas de análisis de la lengua natural, presentes y necesarios en el resto de los productos.
- Productos referidos a los sistemas de información y bases de datos.
- Y otro tipo de productos para ayudar a la comunicación de aquellas personas con discapacidades o para el aprendizaje de una lengua extranjera, o el perfeccionamiento de la lengua nativa.

Con ello las empresas aminoran los costes económicos de I + D referidos a productos como la traducción automática, la indización y recuperación del contenido de los documentos mediante lenguajes de indización ya elaborados, productos referidos al acceso a bases de datos con una comunicación interactiva en lengua natural... etc., todo esto tiene una repercusión en las empresas de rapidez, competitividad y eficacia en el acceso a la información, que se traduce en una reducción de los recursos económicos y humanos empleados en ellos.

Pasemos a analizar, de un modo didáctico, algunas de estas aplicaciones:

**I.- Productos para el análisis de la lengua.** El tratamiento de la lengua natural, oral y escrita, es el pilar central que sirve de intermediaria entre los usuarios y los productos lingüísticos e informáticos. Para ello, es necesario realizar investigaciones que abarquen cada una de las áreas tradicionales de estudio lingüístico: fonética-fonología, morfología, sintaxis, semántica y pragmática. El primer área es fundamental para el tratamiento de la lengua natural oral; en este campo necesitamos realizar especificaciones concernientes a los fonemas y sonidos de una lengua, además de tener en cuenta los otros niveles de estudio, para conseguir la aplicación de generación y síntesis de voz, que nos permite hablar con el ordenador. Las investigaciones se orientan en la actualidad al reconocimiento del habla continua, no sólo de palabras aisladas, con independencia del hablante, no sólo los que han sido registrados, y con léxicos muy amplios. En cuanto a los sistemas de síntesis, se pretende que los futuros sistemas de salida del habla se acerquen a la voz humana, con sus características prosódicas, y en los que la identidad de cualquier hablante humano pueda ser simulada.

Con un planteamiento informático del estudio de la morfología de la lengua se consigue elaborar analizadores morfológicos, que no tienen existencia por sí mismos, salvo para un conocimiento más sistematizado de la lengua en general; sin embargo, es uno de los elementos que está presente en cada uno de los productos de la lengua. Para la construcción del analizador debemos dar al ordenador todas las especificaciones concernientes a la forma de las palabras y a sus posibles transformaciones: cambio de raíz, por ejemplo, como en dormir, duermo, intercambio de letras como en luz-luces, etc. Con éste la máquina podrá segmentar las palabras en sus componentes y buscar las combinaciones correctas.

Con la sintaxis informática, se consigue el desarrollo del analizador sintáctico, que como en el caso anterior no tiene por sí mismo una aplicación, pero sí forma

parte de los productos finales de las industrias de la lengua. Este analizador sintáctico sirve para asignar a cualquier oración de una lengua dada, una estructura, a partir de un plan de instrucciones o algoritmos llamados 'parser', que tienen como misión identificar los elementos de la oración que hay que analizar y especificar la relación existente entre esos elementos.

Pero la comunicación que parece tan sencilla se complica con el problema de la significación, para ello, se intenta con la semántica y la pragmática que las máquinas comprendan el significado de lo que escribimos o de lo que decimos. Se necesita elaborar una teoría semántica que explique el conocimiento que un hablante competente tiene del significado de los enunciados, algo difícil de sistematizar al ordenador, pues el significado de una oración no es simplemente la suma de los significados de cada una de las palabras que la constituyen, sino que viene dado por otros elementos como el hablante, el oyente, el momento de la locución, el fondo sociocultural y el contexto. Todos estos aspectos no son fáciles de programar y en ellos investiga la Inteligencia Artificial. Entonces hablamos del analizador semántico, siguiendo a Meya y Huber<sup>2</sup>, como "un sistema que permite al ordenador aplicar procedimientos de razonamiento y extraer conclusiones". Entre los aspectos que debe solucionar el analizador encontramos la ambigüedad, la anáfora, la elipsis, etc.

Otro producto sería un diccionario aglutinador o enciclopédico que denominamos 'macrodiccionario'<sup>3</sup>. Este diccionario incluye la máxima información lingüística y extralingüística, y constituye la base lexical y de conocimiento gramatical de donde el resto de las aplicaciones de la lingüística informática extraen información.

Toda esta investigación y desarrollo de productos conforma el espacio de actividad propio de las empresas lingüísticas, como base para conseguir la comunicación con las máquinas en lengua natural y el desarrollo de los productos lingüísticos que veremos a continuación, entre otros.

## II.- Productos referidos a sistemas de información y bases de datos.

En éstos, están presentes las líneas de investigación anteriormente expuestas: por un lado, el empleo de la lengua natural, como medio para interrogar a las bases de datos y como procedimiento óptimo que facilita el manejo y difusión de éstas; por otro lado, la necesidad de analizadores morfológicos, sintácticos y semánticos para conseguir productos que nos permitan un mejor manejo de la información. Dentro de éstos, teniendo como llave de entrada la lengua natural, encontramos, entre otros, *los sistemas de indización* que sirven para describir y después recu-

2. Montserrat Meya Llopart, Wolfgang Huber: *Lingüística computacional*. Barcelona: Teide, 1986.

3. Carmen Diez Carrera. *Industrias de la lengua. Planteamiento de elaboración de un 'macrodiccionario' aplicado al léxico de Tierra de Campos*. Tesis doctoral. Madrid, 1992.

perar el contenido de los documentos; generalmente estos lenguajes están formados por términos de la lengua humana y se organizan en bases de datos en las que se especifican relaciones lingüísticas entre ellos. Aquí también incluimos los *sistemas terminológicos*, encargados de registrar y hacer un seguimiento de los términos correspondientes a los distintos saberes del conocimiento humano. Otro producto relacionado con éste, del cual se nutre, son los *sistemas de traducción automática* que nos permite la transferencia mecánica del texto de una lengua natural (llamada lengua fuente) al texto de otra lengua natural (traducción bilingüe) u otras (traducción multilingüe) (llamadas lengua objeto o de destino).

Este servicio lo encontramos ya en Francia, país en el que más desarrollo ha adquirido el videotex, mediante el sistema SYSTRAM -uno de los sistemas pioneros- el usuario francés, a través del MINITEL, puede obtener traducciones de una página por minuto con 8 pares de lenguas (este servicio se denomina TGV por analogía con los trenes de alta velocidad).

Los usuarios de este tipo de productos no necesitan realizar dichas investigaciones, ni comprar paquetes de 'software', sino utilizarlos de manera compartida a través de las redes telemáticas. Y para nutrir éstas hace falta fomentar la creación de equipos interdisciplinarios volcados en la elaboración de productos derivados de la lengua natural.

En síntesis, la idea que hemos querido transmitir es la función preminente que desempeñan los productos lingüísticos en la gestión de la información, y las ventajas que supone su uso compartido por las empresas y usuarios en general, derivadas del empleo de las redes telemáticas para su explotación.