... ... UNED, Radio Educativa. Hola, buenas tardes, ¿qué tal? Reciban un saludo de Beatriz Santamaría desde el Control Técnico y de quien les habla, Miguel Minaya, en nombre del equipo que realiza este programa desde la Universidad Nacional de Educación a Distancia. ...

La UNED emite su programación educativa y cultural durante todos los fines de semana del curso lectivo, sábados y domingos de 7 a 12 de la noche, horario peninsular, a través de Radio 3, de Radio Nacional de España, en FM, y en el 1359 de la Onda Media. También pueden seguirnos por Internet acudiendo al portal de nuestra universidad, uned.es. Eso sí, pasada una semana de la fecha de emisión en radio. Nuestro programa está dirigido a todos los alumnos que cursan sus estudios a través de la UNED, la Universidad Pública a Distancia, pero también tenemos en cuenta a todos aquellos que sin estar matriculados están interesados en ampliar su horizonte cultural. Hoy, sábado, en la Universidad de Madrid, en el programa de la Universidad de Madrid,

nuestros espacios se centran en buena parte de los cursos y carreras que pueden seguirse a través de la UNED. En la primera parte de la programación nos vamos a dedicar a las facultades de Filología, Geografía, Historia y Ciencias, además del espacio La Voz de la Pizarra, del programa de formación del profesorado. Tras este espacio volveremos a dar un avance del resto de la programación, pero ahora ya, si están dispuestos a acompañarnos, vamos a dar paso al primer espacio educativo. Se trata de Aula de Filología. En Aula de Filología de esta tarde, 24 de febrero, nos acompañan los profesores Elena Bárcena y Tim Ripp. Elena, buenas tardes. Buenas tardes, Miguel. Tim, buenas tardes. Hola, Miguel. Buenas tardes y bienvenido. Ellos ya son casi asiduos, diría. Sí. De estos programas radiofónicos y la verdad que vienen siempre con ilusión y hablarnos de cosas, cuando menos, novedosas.

El título que me han ofrecido para el espacio de hoy de Aula de Filología, repito, es Relevancia del tipo textual para la traducción automática. Ellos son expertos, han trabajado en traducción automática y los programas que vienen haciendo es una especie, no diría serie, porque tampoco tiene una continuidad como tal, pero sí reinciden en ese tema, ¿verdad, Elena? En traducción automática. Sí, pretendemos formar un pequeño mosaico a base de piezas de lo último que se está haciendo en traducción automática, de dónde estamos y de dónde venimos, estas cuestiones existenciales. Exactamente, y con el papel cada vez más importante de la web hay más oportunidades, entonces también queremos un poco destacar estas posibilidades y cómo se pueden aprovechar. Perfecto. El título, vuelvo a repetir, es Relevancia del tipo textual para la traducción automática. La primera pregunta, aunque sea general. Yo creo que para descender luego a cosas más concretas sería hablar de qué es un tipo textual. Elena, ¿te parece? Sí, un tipo textual es...

es, como la misma palabra lo indica, un tipo de texto. Lo que ocurre es que cuando estamos pensando elaborar un sistema de traducción automática y consideramos cuál va a ser el objeto de la traducción, tenemos que considerar si la traducción se va a dedicar a todo lo que es un idioma en general, es decir, toda la lengua inglesa, toda la lengua castellana, o si por el contrario nos vamos a ceñir a un tipo textual. Entonces, las coordenadas que empleamos para determinar lo que es un tipo textual son el género y la temática, lo que se dice en inglés subject matter y el género genre. Entonces, el género no

hay que pensar en él en un sentido estricto, como por ejemplo en literatura o en estilística, sino simplemente un género sería lo que... Los boletines informáticos.

narrativos, la publicidad, los anuncios publicitarios, los informes, las recetas. En cuanto a temática, tendríamos la medicina, la biología, la mecánica. Entonces, un tipo textual sería la división de textos que nos queda entre estas dos coordenadas, el género y la temática. Sí, y si partimos de otro tipo de división, Elena, se me está ocurriendo no solamente temática, género y propósito, como has comentado, sino una división más general. Por ejemplo, la lengua de la medicina, pongo por caso. Lo que ocurriría si solamente pensáramos diseñar y elaborar un sistema de traducción automática con una de estas divisiones, digamos, temática, la lengua de la medicina, la lengua de la lingüística, o, por ejemplo, por propósito, informar, divulgar.

pues lo que ocurriría sería que no conseguiríamos constreñir, ceñir, poner límites a toda la inmensidad que es un idioma, una lengua natural, para aprovecharnos de la simplificación y de la reducción, tanto cuantitativa como cualitativa, para diseñar un sistema de traducción automática dedicado, digamos, o exprofeso, y no nos podríamos beneficiar en tanta medida. Sí, y lo que puede pasar en la mayoría de estos sistemas de traducción automática es que hay un motor de traducción, digamos, y eso va con un conjunto de reglas generales y también un vocabulario general. Pero además, para cada dominio, digamos, además de tener suprimidos… El propio vocabulario muchas veces viene con reglas originales sobre cómo formar los grupos de palabras, etc.

Pero igual es una cosa que vamos a poder ver un poco más adelante. ¿Y cuáles son los beneficios concretos de los tipos textuales para la traducción automática? Pues el gran beneficio que podemos encontrar al reducir el campo de la aplicabilidad de un sistema de traducción automática a un tipo textual es la simplificación. Lo que decíamos antes, la simplificación y la reducción, tanto cuantitativa como cualitativa. Las lenguas naturales son unos entes monstruosos en cuanto al número de palabras por vocabulario, la enorme cantidad de ambigüedades, tanto léxicas como estructurales. Y luego otros fenómenos, la elipsis, la anáfora, que son estos problemas lingüísticos los que constituyen el... el caballo de batalla de la traducción automática y hace que, bueno, pues hasta la fecha...

los grandes éxitos sean muy limitados y los podamos contar casi con los dedos de las manos. Entonces, al constreñir, al ceñir un sistema de traducción automática a la traducción de un tipo textual, como los que estamos diciendo, por ejemplo, de los campos científico-técnicos, campos profesionales, académicos, es reducir toda la problemática que conlleva el captar y formalizar el conocimiento lingüístico y extralingüístico inherente a las lenguas naturales. También lo que suele ocurrir algunas veces, en un caso en concreto, estoy pensando en la empresa Kodak, lo que hacen, como saben que quieren traducir sus manuales a todas las lenguas, sus escritores, lo que hacen es usar un sobrenombre. Un lenguaje de inglés, digamos, porque escriben en inglés, que saben que se puede traducir muy bien. Entonces, están haciendo el problema mucho más sencillo. Sí.

Sí, no me gusta hacer de abogado del diablo, Elena, pero ¿hay alguna desventaja? Pues sí, ¿para qué negarla? Siempre hay, ¿no? Una de ellas, y supongo que nuestros oyentes habrán pensado inmediatamente en ella, es que al ceñir el vocabulario, o sea, la cantidad de palabras que va a haber en las bases de datos léxicas del sistema, al solamente

centrarnos en las estructuras gramaticales que sean pertinentes al tipo textual que estemos queriendo traducir, etc., la aplicabilidad del sistema va a ser mucho menor. Es decir, si elaboramos un sistema para traducir informes meteorológicos, claro, el sistema de traducción automática solo nos va a servir para eso y no podrá traducir ninguna otra cosa. Y luego hay otra desventaja, que esta es un poco más profunda y tiene que ver con el, el modo en el que se elaboran los sistemas de traducción automática.

que se ciñen en un tipo textual, los que se llaman sistemas de traducción automática basado en sublenguajes. Y es que estos sistemas suelen incorporar, para aprovecharnos de la simplificación del sistema y de la regularidad del campo, suelen incorporar conocimiento semántico. Y también, de nuevo, este conocimiento semántico hace que el sistema esté mucho más especializado. Así que es difícil luego incorporar cambios al sistema. Al intentar aprovecharnos de la regularidad semántica del sistema, intentamos casi siempre en estos tipos de sistemas de traducción automática incorporar conocimiento semántico. Pero luego esto se puede volver en contra del investigador a la hora de intentar, por ejemplo, ampliar.

el sistema para que incorpore otro tipo de textos, porque está todo demasiado adaptado, digamos, a las necesidades del tipo textual concreto de que se trate. Y esto en realidad es todo lo que se me ocurre en cuanto a desventajas. Bien, si se han incorporado tarde a escuchar este espacio, les comentaré que estamos en Aula de Filología y que nos acompañan los profesores Elena Bárcena, que es profesora dentro del Departamento de Filologías Extranjeras y Lingüísticas, inglés, y Tim Reed, que es profesor en la Escuela de Ingenieros Superiores, Ingenieros... Industriales, Departamento de Lenguaje y Sistemas Informáticos. Muy bien, que siempre me confundo cómo es tan largo para decirlo. Estamos hablando de relevancia del tipo textual para la traducción automática. ¿Cuál ha sido la actitud de los investigadores de sistemas en traducción automática? Pues, curiosamente, la idea de reducir el objeto de traducción del sistema, el sistema de traducción automática...

Fue una de las primeras ideas que se les ocurrió a los investigadores en cuanto se dieron cuenta de la inmensidad del problema al que se enfrentaban. Pero luego por cuestiones de aplicabilidad y como en los años 50 y 60 interesaban diversos tipos de textos, la verdad es que a la hora de desarrollar los sistemas no se intentó aprovechar esto. Lo que suele pasar, por desgracia, es que como una parte importante del enfoque científico es reducir el problema hasta alguna cosa que se puede tratar. Pero claro, a la hora de pedir becas también hay que hacer lo que suena muy bien. Entonces, claro, un sistema que se puede usar para traducir cualquier cosa suena mucho más interesante, aunque es mucho más difícil de desarrollar. El propio Bargilel, uno de los grandes pioneros de la traducción automática, dijo allá a principios de la década de los 50 que por qué no nos ceñíamos. Recuerdo la cita era al lenguaje, por ejemplo, de los pilotos. Pero claro, sin embargo... Como bien dice Tim...

las necesidades eran otras y, por ejemplo, al gobierno estadounidense en su momento le interesaba crear sistemas que pudieran traducir del ruso informes de tipo militarista y científico-técnico y entonces, bueno, no se llevó demasiado a cabo. Y hoy en día, claro, hay mucho más interés, por ejemplo, en Japón a un nivel de traducir los manuales, porque ahí están las empresas sacando coches y dispositivos electrónicos y lo que más les interesa es producir manuales en todos los idiomas. Sí, entonces, bueno, ha sido un arma de doble filo. Sin embargo, pasado la década de los 60 y 70, con algunas excepciones que quizá luego

mencionaremos, algunos investigadores se dieron cuenta de que realmente con estos sistemas generales no se podía alcanzar un nivel de calidad suficiente, digamos. Entonces, optaron por dos estrategias. Una de ellas era la interacción. Es decir, que el usuario...

el usuario experto, naturalmente, interactuara con el sistema de traducción, bien antes de traducir, durante el proceso de traducción, o inmediatamente después. Son lo que se llaman los subprocesos de preedición, traducción interactiva propiamente dicha, y postedición, para intentar ayudar al sistema con aquellos problemas de ambigüedad, etcétera, que se fueran encontrando. Eso es lo que se suele llamar traducción semiautomática, asistida por el ordenador. No puedes salir a tomar un café mientras que lo hace todo el ordenador. No, exacto, hay que estar un poco pendiente. Y luego, la segunda estrategia es la de restricción del input, y esto está muy conectado con lo que estamos hablando hoy en el programa, con la cuestión de los tipos textuales. Entonces, bueno, pues pensaron que si restringían el input, pues quizás simplificarán el problema. ¿Y se puede hablar? ¿Y otros tipos de restricción de input?

¿O es mucho decir? No, sí que existen y Tim ha apuntado a uno de ellos al hablar de Kodak. Aparte de intentar coger solamente un corpus de textos o un sublenguaje con estas coordenadas que marcábamos de género y temática, también hay compañías muy concretas como Kodak y sistemas como Susi, etcétera, que lo que hacen es dirigir a los usuarios, a las personas que introducen los textos en el ordenador, que se ciñan a un vocabulario determinado y a una serie de estructuras gramaticales. ¿Qué ocurre? Claro, la aplicabilidad. Estamos de nuevo tentando un poco la aplicabilidad del sistema, porque claro, esto es... Los sistemas son útiles dentro de una empresa porque la verdad es que es costoso y lleva tiempo entrenar a los...

a los humanos que actúen con la máquina a que elaboren este tipo de textos, que compongan estos textos sin salirse de las estructuras que el sistema puede. Y además resulta un poco pesado para los pobres lectores porque una cosa es que quieres enterar de cómo funciona la impresora para cambiar el papel, entonces no te importa mientras que entiendes el texto. Otra es que trata de ser una novela. Claro. El profesor Juan Carlos Segue de UMIST en Manchester estuvo presente en una conferencia mundial sobre Sistran, un sistema de traducción automática, uno de los demás exitos en el mundo, y precisamente allí se empezaba a hablar el cuenta del lenguaje Sistran, porque estaba sucediendo precisamente esto, que ya no estábamos realmente tratando inglés o, por ejemplo, informes meteorológicos en inglés, sino que ya era una lista de vocabulario, una lista de estructura, y yo creo que eso, en cierto modo, es...

Es un poco hacer trampa. No es el objetivo al que debemos aspirar. Muy bien. Precisamente quería hablar un poco ahora de esto que has apuntado, Elena. ¿Es la restricción del objeto de traducción automática una cuestión superficial u opcional, un truco o recurso por venir a enlazar con lo que decías de trampa para evitar problemas lingüísticos? Pues fundamentalmente es eso. Es intentar simplificar el problema. Es muy loable y desde luego siempre tiene que existir grupos de investigación que trabajen con lenguas generales. Pero a la vez también es bueno que con un objetivo más a corto o medio plazo se esté trabajando en estos campos con el objetivo de simplificar el problema, como estamos diciendo. Pero es que a la vez lo que sucede es que en estos tipos textuales,

en estos sublenguajes, hay divergencias, no solamente restricciones. Con respecto al lenguaje natural.

Y voy a poner un par de ejemplos. Por ejemplo, en inglés, en el lenguaje, la lengua general en el inglés, overhead es un adverbio. O sea, diríamos, the plane flew overhead. El avión sobrevoló o voló por encima. Pero en el lenguaje del aéreo, de los pilotos, overhead puede ser una preposición y se podría decir, the plane flew overhead Paris. Y también un sustantivo, por ejemplo, el overhead, hablando de una transparencia. Exactamente. O sea, que hay incluso divergencias. Y entonces, si queremos que un sistema de traducción traduzca para este entorno específico aéreo, pues entonces sí que hay que tenerlo en consideración. O sea, no se podría obviar y quizás un sistema de lenguaje general no contemplaría overhead como una preposición. También en el lenguaje médico.

Nos encontramos con que el verbo to present es un verbo preposicional, es decir, que lleva un complemento preposicional, mientras que en inglés general present es un verbo transitivo que lleva un objeto directo. Diríamos en inglés general, por ejemplo, the patient presented the following symptoms. El paciente presentó los siguientes síntomas. Pero en el lenguaje informático, uy, perdón, informático médico, podríamos decir the patient presented with the following symptoms. Así que todas estas divergencias gramaticales son fundamentales y hace del tipo textual ya algo más que una excusa. Por supuesto. Sí. Precisamente te quería preguntar. ¿Qué pasa si hablamos de consecuencias en el diseño de los sistemas de traducción automática? Pues la consecuencia principal para el diseño es, tal y como se está enfocando hasta ahora, es como apuntaba al principio, el uso del conocimiento semántico. ¿Qué ocurre?

hasta ahora en los sistemas de traducción general, que el conocimiento lingüístico fundamental que incorporan es el léxico, es decir, el vocabulario, con un componente morfológico a menudo, sobre todo para idiomas como el alemán, etc. Y un componente, un conocimiento sintáctico. Lo bueno de las reglas sintácticas es su carácter general, su carácter abstracto, que se pueden aplicar sujeto, verbo, predicado, sea la frase que sea, se puede filtrar a través de esta regla. Pero lo que sucede con estos tipos textuales o sublenguajes es que el conocimiento semántico también se puede emplear, incluso el conocimiento pragmático o del mundo real. En cuanto al conocimiento semántico, lo que sucede es que estamos ante un mundo más o menos ordenado, en el que cada entidad tiene su función. Entonces, al tener cada entidad su función,

Esto se suele plasmar lingüísticamente en que cantidad suele verse reflejada en que se ocupa, por ejemplo, se corresponde con una categoría lingüística, digamos nombre, y a la vez suele tener una función gramatical específica. Por ejemplo, el submundo de las recetas médicas suele tener al paciente como protagonista, como actante principal, y esto se plasma lingüísticamente en que the patient o he, she, en fin, hay un número muy pequeño restringido de sintagmas nominales que podrían ser el sujeto de la oración. Y entonces toda esta regularidad de semánticas. Se pueden captar y es lo que nos acerca a una calidad mejor. Y además desde el punto de vista puramente técnico, ¿implica?

una cierta modularidad a los sistemas. Incluso si vamos a la página web de Sistran, donde tiene su sistema, este sistema que está ahí disponible gratuitamente es modular, porque se puede probar usando el motor general y también se puede elegir un módulo de un

sublenguaje y meter la misma frase y ver la diferencia en la traducción que hace. Sí, yo creo que lo explicáis muy claramente. Yo creo que a un nivel que lo puede seguir cualquier alumno. Estoy pensando también en los que no son alumnos de los programas de la UNED, pero que están interesados en temas de ese tipo. Pero quería, Elena, dar una vuelta de tuerca y acercarnos al aspecto pedagógico de este programa. ¿Por qué no nos pones un ejemplo de un sistema que siga el enfoque general y el enfoque de su lenguaje? ¿Te parece? Bien, estupendo, sí. Pues Sistran, como este ejemplo que acabo de dar, es un sistema que da un nivel de traducción general en que podemos meter cualquier texto que tenemos y además existe la posibilidad en la misma página web de elegir un sublenguaje como informática o médico independientemente de qué,. Gracias, Maria. Gracias, quería que hablas, To posteriori… Gracias. Gracias. Gracias a ustedes las preguntas y panelistas les doctora para venir a场, a Children's School la present Austria con lo que tratamos hoy de un pronóstico del tema más profundo y importante, tratando тонoplas, los Bishopram-as engañan, los Brooks Noyan繼續 dicho, los Berktu<|nn|> 주세요,

por ejemplo, y meten un texto un poco más específico, con vocabulario un poco más especial, y ver la diferencia en las traducciones. Sí, de hecho, Sistran es uno de los sistemas que actualmente funcionan con un nivel de éxito mayor en el mundo. Y la página web es, Tim? Es www.sistransoft.com Sistran con Y. Sistransoft.com Sistransoft.com. Merece la pena que jueguen con él, porque hay una demo gratuita, ¿no? Sí, sí, sí. Y en cuanto a un sistema que utiliza el enfoque de sublenguaje, es Meteo. Meteo, además, curiosamente, es un sistema que se desarrolló allá por los años 70, en la Universidad de Montreal, porque lo que sucedía es que, como todos sabemos, Canadá es un país bilingüe, entonces toda la información tenía que estar en inglés y francés, y necesitaban...

necesitaban que alguien o algo tradujera diariamente los informes meteorológicos. Ahora bien, este tipo textual, porque es un tipo textual, parece ser que es bastante aburrido, lo cierto es que es así. Entonces, después de tener un pequeño grupo de traductores profesionales, la mayoría acababan abandonando su puesto de trabajo en un periodo de más o menos de seis meses, con lo cual era costoso entrenarlos y luego había que renovar estos profesionales. Entonces, se desarrolló este sistema y consta solamente de unos cientos de palabras, que son las palabras que aparecen constantemente en los informes meteorológicos, pues lluvia, los nombres de las ciudades de Canadá, etcétera, y unas docenas de estructuras gramaticales. Por supuesto, no hace falta ni interrogativas, ni

negativas ni exclamativas. Bueno, negativas sí, lo siento, pero por ejemplo, oraciones interrogativas y exclamativas no, pasivas, pues está muy restringido gramaticalmente. El resumen es que hay un nivel de éxito del noventa y tantos por ciento tan solo los errores que cometa el sistema, bueno, pues son errores tipográficos que también los puede cometer cualquier ser humano al escribir y está funcionando diariamente desde entonces. Sí, me gustaría que me contarais algo de vuestra experiencia personal, que será lo más próximo y lo más cercano, ¿no? Sí, yo personalmente he trabajado siempre en sublenguajes, es por eso que, bueno, pues quizás estoy llevando un poco mi apasionamiento en este programa. Trabajé en el campo concreto de los anuncios de empleo, los anuncios que aparecen en los periódicos. Se trata de un de un

tipo textual muy regular, porque los tipos de información que aparecen en los anuncios de empleo, pues con muy pocas variaciones, son los mismos. Pues hay una dirección de

contacto, una serie de condiciones, que varía curiosamente entre las distintas culturas. Porque yo trabajé con inglés, francés, castellano y griego. Por ejemplo, en griego eran más importantes algunas condiciones que en España. Era muy interesante. Pero incluso las divergencias eran regulares. O sea, que se podían hacer estudios, o sea, se podían hacer correspondencias. Y luego, bueno, pues en cuanto al tipo de vocabulario, también era bastante constante. No era un sublenguaje completamente cerrado, porque no cerramos tanto el sublenguaje de los anuncios de empleo, digamos, por ejemplo, a un tipo de profesional. ¿No? Pues no. Pues no. Por ejemplo, podríamos habernos ceñido a...

a carreras o a puestos de trabajo científico-técnicos, estaban abiertos y entonces podía aparecer la palabra peluquero como ingeniero, etc. Pero sí, se considera que un sublenguaje está cerrado a partir de cuando empieza a haber un número de palabras constantes a partir de 10.000 palabras. Y entonces, pues sí, en ese sentido sí se podría decir que estaba relativamente cerrado. Bueno, el mejor ejemplo, lo más interesante que he visto en los últimos años fue una vez hace poco en un congreso internacional que llegó un general del éxito americano ahí con una maleta. De la ejército. Sí, todo el mundo pensando, ¿qué es lo que tiene ahí? Y lo que tiene es un nuevo sistema de traducción automática que dan a los soldados porque pueden encontrarse en cualquier lugar y realmente no tienen recursos, no tienen por qué saber el idioma y necesitan saber lo suficiente. Entonces es un buen ejemplo que aquí tenemos un sistema que está restringido. Por el tamaño de la maleta, no podemos literalmente ir metiendo más información porque eso le lleva un... no sé exactamente la capacidad porque tampoco...

estaba muy dispuesto a contar los detalles pero dependiendo muy bien si usted va a Bosnia por ejemplo, bueno metemos este módulo y lo lleva ahí, por ejemplo yo creo que también muy brevemente yo creo que el próximo paso va a ser la integración de todos estos sistemas con entrada de voz porque también lo que realmente les interesa en el mundo real es poder grabar lo que está diciendo alguien y ver la traducción entonces yo creo que también vamos a ver las cosas moviendo en estas líneas. Muy bien pues muy interesante, como siempre que os acerquéis aquí, Etienne no nos queda más tiempo, os tengo que despedir hemos estado hablando, bueno, han estado hablando ellos, para ser exactos, de la relevancia del tipo textual para la traducción automática Elena Bárcena, profesora de la Facultad de Filología y Tim Ripp de la Escuela Superior de Ingenieros Industriales, ambos de la UNED. Muchas gracias por vuestra participación en este espacio, un placer como siempre que venís aquí y hasta una nueva ocasión Buenas tardes. Es un placer para nosotros venir, Miguel. Sí, adiós Gracias Gracias

Termina aquí Aula de Filología. Nuestra sintonía volverá a sonar de nuevo el próximo sábado a partir de las 7 de la tarde, hora peninsular, para dar paso a un nuevo espacio dedicado en esa ocasión a latín. Mirando a las estrellas, visiones del cielo en la literatura antigua, es el tema que desarrollarán los profesores Antonio Moreno y Leticia Carrasco. Por nuestra parte, nada más. Gracias por la atención que nos han prestado. Y recuerden que continúan en compañía de la UNED, la Universidad Pública a Distancia. Ahora tienen ocasión de escuchar Revista de Geografía e Historia. Les dejo con nuestra compañera Pilar Carrasco.