Gracias. UNED, Radio Educativa. Hola, buenas tardes. Reciban un saludo de Beatriz Santa María desde el control y de Miguel Minaya del micrófono en nombre del equipo que realiza este programa desde la Universidad Nacional de Educación a Distancia.

Comienza ahora el programa de la UNED, un tiempo dedicado a la educación y la cultura que llega hasta ustedes a través de Radio 3, de Radio Nacional de España en FM y en el 1359 de la Onda Media. También pueden escucharnos en Internet, en nuestra página web, uned.es. Eso sí, pasada una semana de la fecha de emisión en radio. Hoy, sábado 19 de febrero, les ofrecemos una serie de espacios relacionados con buena parte de las carreras y cursos que pueden seguirse a través de la UNED. Los primeros programas tienen que ver con las facultades de Filología, Geografía, Historia y Ciencias, además del espacio La Voz de la Pizarra, del curso de formación del profesorado. Tras este espacio volveremos a dar un avance de la programación que nos llevará hasta las 12 de la noche, horario peninsular. Y ya sin más, vamos a entrar de lleno en el contenido de nuestros espacios. Comenzamos con Aula de Filología. Aula de Filología.

En Aula de Filología nos acompaña en esta tarde Tim Ripp. Tim, buenas tardes. Hola Miguel. Profesor de la Escuela de Ingeniería de la UNED y Elena Bárcena. Elena, buenas tardes. Hola, buenas tardes Miguel. Que es profesora en la Facultad de Filología de la UNED. Es el segundo espacio de una serie de tres programas que tienen durante este curso y este lleva por título un repaso histórico a la traducción automática. Bueno, la primera pregunta quizás es un tanto obvia, pero yo creo que servirá como introducción. ¿Qué es la traducción automática? ¿Quién recoge la pregunta? Bueno, pues básicamente se puede definir la traducción automática como un proceso automatizado o semi-automatizado. Le voy a expresar un texto fuente en un idioma de destino. Claro, dependiendo del sistema y la tecnología hay varias formas de expresarse.

maneras de hacer traducción y algunos herramientas lo hacen automáticamente y otros son más bien una pequeña ayuda para un traductor humano. Sí, hay distintos niveles de automatización. Lo que pasa es que traducción automática se usa como término tanto para la forma de traducción completamente automatizada como para cualquier tipo de traducción en la que interviene un ordenador. Entonces hay una línea clásica que se suele mostrar para ver los distintos niveles de automatización. Tenemos desde la traducción completamente automatizada, pero para llevar a cabo esta traducción debemos, con cierta garantía de éxito, es decir, para obtener una traducción automática de una calidad.

comparable a la que alcanzaría un traductor humano desempeñando la misma tarea, hay que restringir hoy por hoy los textos de alguna forma. Normalmente, el modo mejor de hacerlo es restringiendo el tema de los textos y así tener un traductor automático completamente automatizado que traduzca solamente informes meteorológicos o solamente recetas de cocina. Luego, a continuación, tendríamos lo que se llama la traducción automática asistida por el hombre. Entonces, podemos ver distintos tipos dentro de este tipo de traducción. Uno de ellos es la preedición interactiva, los sistemas de traducción automática que tienen preedición interactiva, post-edición interactiva o también traducción automática interactiva. La preedición es un paso previo a la traducción en la traducción automática.

en el que el usuario repasa el texto fuente, se supone que conoce bien el sistema de traducción antinomática y entonces puede prever los problemas que va a tener a continuación el sistema traduciéndolo y entonces los elimina. Por ejemplo, una cosa que puede hacer es poner las palabras que el sistema no tiene en su diccionario, ponerlas en mayúscula o ponerle algún tipo de banderita o de sobrayado para que el sistema ni siquiera intente traducirlas y luego en una etapa posterior él ya lo hará. A veces esta predicción puede ser interactiva y entonces el sistema puede hacer, antes de intentar traducir, puede hacer un paso previo que consistiría en escanear el texto buscando las palabras que él mismo sabe que no tiene en su diccionario. En su base de datos, en sus diccionarios.

Y, entonces, bueno, pues es un modo mucho más ágil de hacerlo, ¿no? En vez de tener lo que hacer el usuario, lo hace el usuario en colaboración con el sistema. Así pues, bueno, pues imaginemos que precisamente lo que vamos a hacer es buscar palabras desconocidas para el sistema. Entonces, en el caso de la predicción interactiva, el sistema escanearía todo el texto y cuando encuentre una palabra que no sabe, o bien le pide ayuda al preditor humano que estaría delante de él diciendo, puedes, por favor, solucionarme esta palabra, darme la traducción para que luego yo, en una etapa posterior, pueda llevar a cabo la traducción o la dejo así, bueno, pues como decía antes, en mayúscula o de algún modo. En cuanto a la postedición, pues es un poco el proceso inverso. Se trata de hacer un poco lo mismo, pero con el texto de destino. Es decir, el sistema primero intenta traducir.

el texto y luego, por último, repasa, como si dijéramos, aquella traducción que ha producido, viendo cómo decide, ahora una vez que tiene toda la información contextual y todo traducido, cómo decide entre distintas opciones, por ejemplo, para un caso de ambigüedad léxica o así. También puede ser este proceso interactivo y entonces el sistema le puede pedir ayuda al usuario humano, al posteditor humano y decir, bueno, no he podido traducir esta palabra y además, normalmente, cuando no se puede traducir una palabra, tiene consecuencias con las palabras de alrededor, es decir, si no podemos a lo mejor traducir el verbo de la oración porque no tenemos el significado, pues esto, claro, nos va a condicionar quizá, bueno, pues a algún complemento que haya anexo. Entonces, el sistema pide, solicita ayuda al usuario.

Y luego, por último, está la traducción automática interactiva propiamente dicho. Esto conlleva una serie de peligros tremendos. Bueno, en primer lugar voy a decir que en la traducción automática interactiva lo que sucede es que el sistema en pleno proceso de traducción, es decir, cuando está o bien analizando o dependiendo del tipo de problema que se encuentre, bueno, pues encontrará con el problema en distintos puntos exactos del problema. Entonces, va a ser o bien, bueno, pues cuando está analizando el texto fuente o cuando está haciendo una consulta bilingüe en el diccionario o lo que sea, entonces se da cuenta de que no tiene la información necesaria para el proceso de traducción y le pregunta al usuario.

que tal palabra, la palabra ball, por ejemplo, hay que traducirla como baile, pues después de tres o cuatro veces que le haya transmitido esta información, el sistema asuma que cada vez que aparezca la palabra ball haya que traducirla como baile y puede no ser el caso, con lo cual esto puede ocasionar errores al sistema. Sí, esto es lo que Kieran decía ahora. Y claro, esto es una moneda con dos caras, porque este problema viene directamente de la aplicación del aprendizaje de máquina, porque claro, también conviene mucho que el

sistema de traducción automática va adquiriendo vocabulario mientras que hace el proceso de traducción. Entonces conviene de alguna manera que está aprendiendo las palabras y haciendo los vínculos entre los dos significados. Es cierto. Pero también... Hay que tener cuidado. Sí, sí, el peligro de sobregeneralización que digo...

Pero esto es lo ideal, que un sistema vaya mejorando con el uso y se vaya incorporando más información. De hecho, incluso los sistemas de estos llamados de primera generación permiten la incorporación de palabras, incluso de estructuras al sistema. Bueno, el peligro opuesto también existe y es que el sistema tenga que repetir la misma pregunta todo el tiempo. Entonces, si lamentablemente el sistema no tiene capacidad de tomar sus propias decisiones, pues entonces nos podemos encontrar con la triste realidad de que el sistema nos pregunte continuamente, pues en un texto, imaginémonos, sobre procesamiento del lenguaje natural, cómo ha de traducir la palabra al lenguaje. Entonces, bueno, pues puede ser tremendamente tedioso y al final, bueno, pues el usuario puede acabar desechando la posibilidad de emplear una máquina para su traducción. Y luego, por último, otro peligro que se me ocurre es que el usuario no conozca el idioma. Hay que, tiene que tener muy claro el sistema.

de traducción automática, ¿cuáles son los conocimientos específicos que tiene el usuario? Porque claro, si el usuario no conoce el idioma fuente, bueno, pues entonces no puede hacerle preguntas sobre el texto fuente o el de destino o ambos. Hay varios modos de llevar a cabo esta traducción interactiva, pues por medio de un menú, que sería el modo más sencillo, pero quizá el menos interesante, ¿no? Desde el punto de vista del procesamiento del lenguaje natural, o por lo menos desde el punto de vista de la investigación básica, puede haber una interfaz en lenguaje natural y en cuanto a la iniciativa, bueno, pues la iniciativa puede que la tome el sistema, el usuario, o que sea mixta, que también es la más interesante, ¿no? La más flexible, pero la más ambiciosa también. Sí, y también merece la pena comentar que un problema que ha aparecido a lo largo de la historia de la traducción automática es el tema del traslapo entre el conocimiento del sistema y el conocimiento del usuario.

Quiero decir que, como ha dicho Elena antes, estos sistemas, para que funcionen a algún nivel, tienen que estar restringidos a un dominio bastante reducido. Y claro, lo importante también es que este sistema debería trabajar con un usuario que también controle el mismo ámbito, porque de la misma manera que una máquina puede dedicarse a un ámbito pequeño, pasa lo mismo con un traductor humano. Entonces, a lo largo de que intenta sacar aprovecho de un sistema, tenemos que hacer un humano que, por ejemplo, cuya especialidad es la traducción basada en la intención comunicativa, que está trabajando con un sistema que hace traducciones en el mismo ámbito y no, por ejemplo, que tiene la especialidad de trabajar con sistemas métricos, por ejemplo. Sí, sí. Habiendo dicho esto, pasamos a decir brevemente que la traducción... Que el siguiente tipo de traducción automática basándonos en...

en el nivel de automatización sería la traducción humana asistida por ordenador. Antes habíamos visto la traducción automática asistida por el hombre. Y bueno, con este término nos referimos a todas las herramientas, diccionarios, enciclopedias, glosarios, correctores, que como vimos en el anterior programa, pues pueden ser ortográficos, gramaticales, estilísticos. Un entorno de trabajo interesante en cuanto a la traducción humana asistida por

ordenador es lo que se llama la mesa de trabajo del traductor. En este entorno están todas las herramientas integradas y disponibles online para cuando lo necesite el traductor. Entonces el traductor típicamente empieza a desarrollar su propia traducción, pero tiene todas estas herramientas disponibles. Y accede a ellas de un modo muy ágil mientras traduce. De hecho, por ejemplo, tendría la pantalla.

dividida horizontalmente en dos, de tal modo que tendría el texto fuente en la mitad superior y él empezaría a teclear el texto de destino, la traducción, en la mitad inferior. Y entonces, bueno, pues en el momento que tiene un problema, pues accede a las herramientas por medio de un menú en el que estaría todo puesto de modo muy claro. También, con vistas al siguiente y último programa, cuando vamos a hablar sobre el papel del procesamiento de lenguaje en la web, merece la pena mencionar que también hay muchos programas de este tipo disponibles en la web. Hay diccionarios y diorosóra y muchos programas. Sí, tesauros, sí. Antes de pasar a la historia de la traducción automática, hay que mencionar que con el término texto fuente y texto de destino, espero que haya quedado claro que nos referimos al texto en la lengua fuente, o sea, en la primera lengua de la que... Que se va a traducir.

El texto del destino es el texto en la lengua del destino, que no tienen por qué coincidir con la lengua materna del usuario o con una lengua extranjera en absoluto. Simplemente son términos propios de esta disciplina. Si os habéis incorporado tarde a escuchar este espacio, os comentaré que estamos en Aula de Filología y que están con nosotros Tim Reed y Elena Bárcena, ambos de la Universidad Nacional de Educación a Distancia, y que estamos hablando de un repaso histórico a la traducción automática. Yo, cuando leía el título, Tim y Elena, pues me parecía un tanto paradójico hablar de historia, de la traducción automática, pero supongo que, aunque sea una cosa que suene muy novedosa, tiene una historia, claro está, ¿no? ¿Cómo nace la traducción automática? Pues nace en la Segunda Guerra Mundial, así que, bueno, ya tiene 50 añitos. Desde luego es mayor de edad. Hace poco asistimos precisamente a un simposio, o bueno, una conferencia, supongo, en San Diego, California.

California, en la que se celebraba, nunca mejor dicho, los 50 años de la traducción automática. Fue muy interesante porque allí estaban los que se llamaban los pioneers, los pioneros, que fueron las primeras personas, los primeros investigadores en materia de traducción automática, que por supuesto están vivos y bien vivos, y fue un auténtico placer escucharles todas las anécdotas de cómo se desarrollaba la traducción automática hace 30, 40 años. Como digo, nació en la Segunda Guerra Mundial, porque allá por los años 40, las técnicas de criptografía que se empleaban para enviar y recibir mensajes, sugirieron la idea de que los ordenadores recibieran input textual en un idioma dado y lo tradujeran a otro idioma diferente. Sí, esto se recoge en el memorando de Warren Weaver, que escribió en 1940, y muchos que plantearon una fecha, quizás también merecen...

es la pena comentar algunas fechas muy claves en este campo. Primero la fecha de 1951, en que se podía reconocer el primer investigador de tiempo a tiempo completo, que se llama Bar Hillel, y que trabajaba en la MIT en Estados Unidos. En 1952 tuvo lugar la primera conferencia. En 1954 teníamos la primera demostración pública de un sistema de traducción automática en la Universidad de Georgetown, que tradujera unas 49 frases del ruso al inglés, que tenía unas 250 palabras de vocabulario y seis estructuras gramaticales.

Estaba todo muy restringido, de tal modo que fue un éxito rotundo. Claro, porque no era en realidad un lenguaje natural, sino que teníamos una sección muy pequeña. Sí, sí. De lo que es el inglés y el ruso. Sí.

Por fin podemos identificar entre 64 y 66 el informe ALPAC, el informe tan famoso del campo, el informe sobre The Automatic Language Processing Advisory Committee, que era francamente muy negativo. Trataba tanto de la calidad de traducción como la velocidad, la rentabilidad. Sí, pero me gustaría mencionar que sí que es cierto lo que tú dices, Tim, que fue muy negativo y hablaba de la traducción automática, bueno, pues diciendo que es una calidad muy inferior a la que produce un traductor humano, que es muchísimo más lenta y tal. Pero yo creo que aún peor que el contenido del informe ALPAC fue la interpretación que se le dio porque supongo que, bueno, quizá alguno de nuestros estudiantes haya negado. He leído el famoso informe ALPAC que es como un pequeño folleto así y dice explícitamente que se deberían redirigir los esfuerzos investigadores.

hacia, bueno, pues quizá otros objetivos más modestos. Por supuesto, es muy difícil, como se dice en Alpac, y ha quedado demostrado, traducir toda una lengua viva en otra, pero, bueno, pues quizás se podrían hacer esfuerzos incluso dedicarnos solamente a sublenguajes. De hecho, exactamente lo que dices, los sistemas de traducción en sublenguajes funcionan muy bien. Sí, pero bueno, sí, lo trágico que supongo que era a donde ibas a parar, Tim, cuando hablabas del informe Alpac, es que, así como había habido una gran euforia desde 1950, cuando Bargilel y su equipo empezaron a trabajar en este campo, pues después del informe Alpac se cortaron absolutamente todas las subvenciones oficiales que había en Estados Unidos y, bueno, y también en Canadá.

Porque en Canadá había mucho interés, no tanto político y militarista como en Estados Unidos, sino por ser una comunidad bilingüe, ¿no? Inglés y francés. Entonces, bueno, pues tienen que realizar un esfuerzo de traducción tremendo en todos los ámbitos, siempre para tener toda la documentación en los dos idiomas, por lo cual estaban muy interesados en este campo. Y tras el informe ALPAC, pues la verdad es que presenciamos una década de inactividad prácticamente absoluta, con alguna excepción. Y tenemos que esperar muchos años en que aparezcan los primeros sistemas de traducción automática con éxito, porque, como digo, bueno, pues a pesar de que se cortaran todas aquellas subvenciones fabulosas que había habido al principio, bueno, pues algunos grupos, como el de Pete Thomas, etcétera, pues seguían... Y hay dos sistemas, bueno, pues que conviene mencionar porque fueron de los...

De los primeros, allá por los años 70, y curiosamente prácticamente dos de los sistemas que tienen más aceptación y más éxito hoy en día y que se están utilizando más ahora a puertas del siglo XXI. Y uno de ellos es CISTAN, que curiosamente mencionamos en todos los programas sobre traducción automática, pues fue originalmente desarrollado en la Universidad de Georgetown para la traducción del ruso al inglés, perdón, para las fuerzas aéreas norteamericanas. Bueno, ha dado muchas vueltas este sistema y finalmente lo adquirió la Unión Europea, que se sigue desarrollando. La verdad es que hay dos versiones. Hay una que está con la comunidad europea y hay otra que está con una empresa que se llama CISTAN. Y la diferencia básicamente está en la implementación y el tamaño del vocabulario. Se sigue trabajando en...

en el vocabulario, porque la verdad es que la arquitectura del sistema es muy poco sofisticada. Es un sistema prácticamente de primera generación, con lo cual, para que tenga una aplicabilidad mayor, lo que se hace es alimentar las bases de datos léxicas. Se han añadido muchísimos pares de lenguas. Empezó del ruso al inglés, pero la verdad es que hoy en día cubre cantidad de idiomas prácticamente. Los rusos son nuestros amigos hoy en día. Claro, sí. Ya ha pasado la Guerra Fría y ya no hay tanto interés. Es triste, pero es cierto. Y luego el otro sistema que merece la pena mencionar, porque también ha marcado un hito en la historia de la traducción automática, yo creo que es el sistema al que todo el mundo nos referimos de un modo fácil y cómodo cuando alguien critica el campo de la traducción automática, y es Meteo. Meteo, tenemos como fecha, por ejemplo, 1976, que se produjo la primera versión.

operativa y este sistema fue desarrollado en Canadá porque había un problema de orden práctico y es que tenía que traducir, como decíamos antes, al ser una comunidad bilingüe, pues los informes meteorológicos entre el inglés y el francés diariamente. Entonces, es un trabajo sumamente tedioso para un traductor profesional y lo que sucedía es que después de entrenar a una persona en el sublenguaje de los informes meteorológicos, apenas la media de trabajadores superaba los seis meses antes de presentar su dimisión porque era un trabajo francamente aburrido. Entonces, se empezó a desarrollar Meteo en la Universidad de Montreal. Y, bueno, pues como se trata de un sistema muy restringido, con un número de palabras pequeño y también de estructuras, son apenas unos cientos de palabras y unas decenas de estructuras.

gramaticales, el éxito, el nivel de éxito, o sea, de producción correcta de traducciones de meteo, me parece que es como del 98%, que es prácticamente igual que la de un traductor humano muy profesional. Y los fallos ocasionales se deben simplemente a errores tipográficos, o sea, que eso también puede suceder a traductores humanos. Así pues, es un sistema que sigue operativo hoy en día y ha sido un éxito rotundo. No ha sido fácil repetir la experiencia, porque luego se intentó desarrollar el sistema Aviation, para la traducción de los manuales de instrucciones de vuelo y tal, pero claro, era ya un sustituto. Era un sublenguaje mucho más complejo, entonces no se pudo llegar al mismo punto de...

de éxito. Y además de los problemas lingüísticos deberíamos destacar los problemas con los ordenadores, por ejemplo, porque desde hace algún tiempo los ordenadores eran demasiado lentos, tenían muy poco memoria y poco espacio en sus discos duros y una capacidad gráfica muy reducida. Entonces, claro, también... Sí, sí, qué duda cabe que era un factor importante, sí. Y bueno, también a nivel lingüístico, no solamente a nivel computacional, no se habían desarrollado muchas de las teorías que tenemos hoy en día. No olvidemos que por ejemplo, la misma gramática funcional empezó en el año 78-79 y se desarrolló en base a... parece que son 17 o 19 oraciones simplemente. Bueno, siguiendo un poco con lo que es la historia de la traducción automática, mencionaré Eurotra, que es otro delito.

los grandes proyectos europeos de finales de los 70, que comenzó en finales de los 70. Entonces, se trataba de traducir la documentación de todas las lenguas oficiales de la Unión Europea. Bueno, esto no se consiguió. Y se hizo pensar, de hecho hubo muchísimas publicaciones que se titulaban así, algo parecido a esto. Pero estamos encaminándonos hacia un segundo alpac, porque se pensó que, bueno, también se había creado mucha

expectación y había mucha ilusión por la idea de tener un sistema de traducción automática que tradujera toda la enorme cantidad de documentación que en las nueve, me parece, lenguas oficiales de la Unión Europea en aquel entonces, bueno, pues... Hubo, la verdad que... Yo creo que...

sé que no tenemos mucho tiempo. Yo creo que sería un buen momento, porque tampoco queremos terminar el programa siendo muy negativo, porque, sin lugar a dudas, los sistemas de traducción automática sí tienen un papel muy importante. Me ocurre la anécdota del sistema de defensa de Estados Unidos, porque, claro, reciben mucha información en muchos idiomas y no tienen dinero para pagar a traductores. Entonces, lo que hacen es que tienen sistemas de traducción automática que, en primer lugar, hacen una traducción muy rápida. Entonces, hay gente para leer el texto. Y si creen que merece la pena, entonces pasan el texto a traductores humanos. Entonces, aún nos quedan muchas posibilidades. Sí, incluso para la traducción de baja calidad, sí. Y hablando ya de épocas más cercanas, los 80 y los 90. Sí, en los 80 hubo varios intentos de producir sistemas sofisticados, pero lo que pasa es que fallaron a un nivel operacional. Es decir, que funcionaban bien en el entorno del laboratorio, pero no funcionaban fuera de este entorno y dieron un nivel de frustración. Мууиля.rock.es Discracia de las noticias de la ciudad y

Y luego, por último, en los 90, y ahí es donde estamos situados ahora mismo, a pesar de que sea el año 2000, lo que nos encontramos es cada vez con más sistemas de traducción automática comerciales, con estrategias simples y luego, por otra parte, investigación básica altamente especializada. Una de nuestras sorpresas cuando fuimos hace dos años a este congreso que mencionaba al principio, que celebraba los 50 años de la traducción automática, fue la cantidad de sistemas comerciales que había allá. Y, bueno, pues no son brillantes, pero pueden solucionar problemas de traducción muy específicos, por ejemplo, a nivel de empresa o así, haciendo con glosarios muy especializados. Y esto por una parte. Luego, por otra, lo que nos encontramos es que se está desarrollando una investigación básica altamente especializada.

en la que se han incorporado técnicas y estrategias de inteligencia artificial, luego también los llamados métodos estadísticos, es decir, en vez de llevar a cabo un sistema de traducción automática como toda la vida, con reglas lingüísticas, pues incorporar algoritmos matemáticos y trabajar con córpora textuales muy vastos, de millones de palabras. Luego también hay mucha investigación en el campo de los sublenguajes, es decir, los grupos de investigación parece que se están concentrando muchísimo más en estrategias para poder solucionar el problema fundamental, que es el carácter irregular y cambiante y tan vasto y misterioso en última instancia de los idiomas. Bien.

Pues es una pena porque el tiempo nos da para más. Habría mucho que hablar sobre ello, pero no hay más tiempo. No nos queda más. Solo me queda agradeceros vuestra presencia en este espacio, tanto a Tim Reed, profesor de la Escuela de Ingeniería de la UNED, y a Elena Bárcena, de la Facultad de Filología de la UNED también. Y no os digo adiós porque habrá una próxima intervención, sino solamente hasta pronto y muy buenas tardes. Muchísimas gracias, Miguel. Y muchas gracias a todos. Adiós. A vosotros. Por nuestra parte, nada más. Aula de Filología volverá a estar de nuevo con todos ustedes el próximo sábado a partir de las 7 de la tarde, una hora menos en la Comunidad Canaria. Gracias por la atención que nos han prestado. Y no olviden que la UNED les acompaña hasta las 12 de

la noche, horario preinsular. Ahora es el turno de Revista de Geografía e Historia. Les dejo con nuestra compañera Pilar Carrasco.

Gracias.