

SISTEMA INTÉRPRETE PERSONAL Y AUTOMÁTICO

Álvaro de AGUSTÍN VÁZQUEZ



RETENDO en estas páginas dar a conocer un sistema de traducción simultáneo, tanto oral como escrito, que está en estudio en las FAS norteamericanas. Las operaciones en Afganistán e Irak han demostrado, de nuevo, la dificultad que supone el idioma para trabajar con aliados en operaciones combinadas y también la conveniencia de poder entenderse con la sociedad del entorno, ya sean combatientes enemigos o no. Sólo tres ejemplos:

En Irak un informe destaca que fue necesaria más de media hora para que un mando polaco entendiera unas instrucciones que el coronel norteamericano, autor del informe, considera se podrían haber comunicado en un texto de tres líneas.

También en Irak se enviaron unas muestras de sangre de un herido para ser analizadas en Estados Unidos, se tardaron tres días en tener los resultados. Sin embargo, en el mismo hospital de Bagdad se disponía del equipo y personal necesario para hacerlo, pero claro no hablaba inglés, no se enteró de la necesidad y no pudo informar lo fácil y rápido que se podría haber hecho el análisis.

Por último, en Afganistán se llegó a la conclusión de que el tiempo empleado en atender a cada herido/enfermo afgano era más del doble del necesario en casos donde no hay problemas de idioma.

El *Sequoyah Foreign Language Translation System* (S-FLTS), que es como se denomina el proyecto, lo he conocido gracias a mi actual destino en Norfolk, en el Cuartel General Aliado del Mando de Transformación de la OTAN, conocido por SACT. Con la llegada de la transformación a la antigua SACLANT se intensificó un inicial enlace que ya había con el mando conjunto de las FAS norteamericanas, el llamado JFCOM, mando que en ese país se

dedica, como una de sus prioridades, a la transformación de las FAS. Pues bien, como resultado de la mencionada intensificación de contactos se crearon grupos de trabajo para coordinar esfuerzos en diferentes áreas, el relativo a CIS y Sistemas C4 fue donde yo me integré y de esa forma cayeron en mis manos las primeras noticias del proyecto S-FLTS que se está estudiando en el JFCOM.

El S-FLTS nació como consecuencia de la búsqueda de superioridad en sistemas de información que el documento *Joint Vision 2020* establece como elemento clave para garantizar una posición dominante de las FAS norteamericanas en conflictos venideros. Se pretende un sistema que ofrezca una traducción automática oral y escrita en ambos sentidos. Es decir, en la parte oral, que creo es la más novedosa e interesante, se persigue que yo pueda hablar en español y la máquina «le hable» en inglés a mi interlocutor, traduciendo en tiempo real mi discurso. De la misma manera al oír a un americano, incluso de Minnesota, en su idioma local, la máquina me lo diría a mí en castellano. He puesto el ejemplo del inglés por razones personales, ya que mi actual trabajo hace que mi pelea con este endemoniado idioma sea diaria, dura y de resultado incierto, como por otra parte y honradamente creo le pasa a la inmensa mayoría de los que estamos destinados en la OTAN. Referente a la parte escrita, el requisito mayor, nacido de las necesidades vistas en las últimas operaciones, es que el S-FLTS permita la traducción inmediata de material encontrado o requisado en el campo de batalla.

Como muestra del interés norteamericano en esta herramienta, en un memorando fechado el 24 de septiembre pasado, el secretario de Defensa de este país colocaba la mejora de los sistemas de traducción automática como la sexta prioridad de su departamento para el año fiscal 2005.

Situación actual

Los diversos departamentos del Gobierno norteamericano disponen ya de varias herramientas para traducción automática de textos. No es muy difícil que cualquiera de nosotros dispongamos de algún programa de este tipo. Por ejemplo, en el buscador *Google* de Internet, cuando la página *web* encontrada está escrita en inglés te ofrece traducirla, lo cual hace en un tiempo de segundos. El problema es que lo que lees es un castellano sin apenas sentido. También es cierto que con una relativamente baja inversión se consiguen mejores herramientas que estas gratuitas, pero el resultado aún deja que desear.

Recuerdo cómo en mi curso de Guerra Naval tuve «la fortuna» de que el libro que me tocó para su recensión era de economía y en inglés. Muy decidido me dirigí a un amigo, gran aficionado a la informática, que me prestó tres

traductores, puedo asegurar que tras varios intentos y alguna tarde perdida no saqué el menor provecho de estos programas pseudotraductores.

Pero bueno, volviendo al tiempo real y dejando atrás el año 97, hoy en día en el JFCOM trabajan con programas, unos comerciales y otros del Gobierno, que disponen de enormes bases de datos, con miles y miles de palabras y frases y hacen unas casi aceptables traducciones de material escrito. Digo «casi» por que lo pueden ser para temas generales, pero aún no dan la fiabilidad necesaria en temas de Gobierno, ni de negocios donde se tomen decisiones con cantidades de muchos ceros, y mucho menos en operaciones militares donde las decisiones pueden poner en riesgo vidas humanas.

En definitiva la situación actual es que se dispone de programas que efectúan una relativa traducción de textos, que necesitan equipos grandes donde correr estos programas y que la traducción, si se quiere que sea aceptable, no la hacen en tiempo real. También se dispone de programas para traducción oral simultánea, pero los resultados son pobres y con fiabilidad aún menor que en el caso de traducción de textos.

Sistemas que ofrecen soluciones parciales

Personal traductor

Disponer de personal, bien con la lengua que necesitamos traducir como lengua materna o con profundos conocimientos de ella, ha sido el sistema tradicional utilizado. Por ejemplo, España contrató personal de este tipo para sus unidades durante las operaciones de paz en Bosnia o en Albania. Sin embargo, son varios los problemas que conlleva el uso de estas personas, yo destacaría como más importantes:

- La dificultad para disponer de traductores en número suficiente, particularmente para idiomas no muy extendidos. Hay que tener en cuenta que actualmente se hablan en el mundo 6.700 lenguas y unos 37.000 dialectos.
- El personal traductor puede dominar muy bien la lengua que necesitamos para una determinada operación, pero nos encontraremos en muchos casos que la que no domina es nuestra propia lengua, con lo que continúa nuestro problema.
- La fiabilidad del trabajo de estos traductores es escasa en la mayoría de los casos, ya sea por su propio desacuerdo con nuestras operaciones, por temor a futuras represalias del enemigo (para él puede ser su vecino de años) o en general por no ser personal profesional y acostumbrado a trabajar en el ámbito militar.

Doces TSS

Sistema usado actualmente por el Ejército norteamericano, basado en programas informáticos comerciales COTS y en desarrollos propios del Gobierno GOTS. Tiene una cierta capacidad para identificar el idioma, traducir textos, caso de que esté en la lista de los idiomas que maneja, priorizar su importancia, basándose en la aparición de ciertas palabras, y gestionar la documentación escrita enemiga que se obtenga durante las propias operaciones. Dispone también de una herramienta más limitada para reconocimiento óptico de escritos manuales OCR y para identificar y traducir información oral, presentando la traducción en forma de escrito.

Coalition Chat Line Multi Lingual Capability

Esta capacidad se está utilizando actualmente en Irak, funciona de forma similar a un «chat» de Internet, donde cada uno usa su propio idioma, tanto para enviar comentarios como para leer los del resto de participantes del «chat», que, por supuesto, los habrán introducido en su lengua. A día de hoy se puede trabajar con los idiomas inglés, español, ruso, polaco y ucraniano. En un futuro próximo se añadirá el árabe y el turco. Para ello y para la mejora general del sistema hay aprobado un presupuesto de 2,2 millones de dólares para este año, como se ve no es una herramienta de desarrollo muy caro (compárese con los más de 800 millones de euros que anualmente gasta la Unión Europea en traducciones). Los informes que he leído de este proyecto son favorables, pero insisten en una fiabilidad limitada, particularmente si los usuarios usan frases hechas o lenguaje coloquial. El aumentar las bases de datos con este tipo de frases es en general lo que añade calidad al sistema, claro que manejarlas y discernir entre una gran cantidad de frases y expresiones coloquiales es lo que complicó el diseño del programa informático.

Trim

Sistema similar al anterior, presentado por Corea durante el último ejercicio JWID 03. La demostración se hizo entre coreano e inglés y consistió en un «chat» con hasta 25 clientes de ambos idiomas y con un servidor donde corría la aplicación, facilitando la traducción a todos ellos en lo que se puede considerar tiempo real (traduce 150.000 palabras por hora). Si bien he leído buenas críticas de esta herramienta, de nuevo se considera que debe mejorar para ofrecer una mayor fiabilidad.

Cybertrans

También usado por el Gobierno norteamericano, traduce al inglés documentos escritos en 57 idiomas de los considerados poco extendidos (la mayoría de uso en pequeñas áreas de Asia y África). Sólo en nueve de los 57 idiomas este sistema permite la traducción inversa. Es una herramienta útil y facilita el trabajo de inteligencia, pero no se puede considerar un traductor de textos definitivo.

S-Minds

«Speaking Minds» es un programa que pretende permitir una conversación entre dos personas con traducción inmediata, es decir, justo lo que buscamos. La ha desarrollado el Departamento de Defensa con la intención de que sirva para interrogatorios a detenidos. Por ahora dispone de traducción de español y serbocroata en doble sentido y de inglés a chino y al árabe en un solo sentido. El «soft» desarrollado usa Windows como plataforma y aunque ya se está utilizando no ofrece aún la fiabilidad buscada.

Tanto en el área comercial como en la gubernamental existen más herramientas, las presentadas son una muestra de las más avanzadas. Como conclusión se puede decir que ninguno de estos productos da una solución definitiva. Si bien en la traducción de textos se está muy cerca de disponer de máquinas fiables y útiles, en lo que se refiere a la traducción oral simultánea y automática será necesario esperar un tiempo. De esta carencia nace el proyecto S-FLTS del que a continuación paso a hablar.

El proyecto S-FLTS

Primero es necesario aclarar que lo que busca el Departamento de Defensa con este proyecto no es mejorar la intercomunicación de los pueblos con lenguas diferentes. Sin embargo, una vez más creo que nos encontramos ante el caso de que las necesidades militares puedan desarrollar una herramienta muy útil para la sociedad civil.

Lo que se pretende en el ámbito militar es que las unidades que tengan que desplegarse en territorio hostil, donde el inglés no sea la lengua de uso diario, dispongan de un traductor oral/escrito automático y en tiempo real. La razón es triple:

- Poder comunicarse de forma inmediata y fiable con la población que habite en el área de operaciones.

- Obtener información del enemigo capturado o de la documentación requisada, ya sea ésta confidencial o abierta.
- Facilitar la colaboración con aliados de la operación, pues ya se sabe que la inmensa mayoría de las operaciones militares futuras serán conjuntas y combinadas.

Además de las capacidades ya vistas en estos sistemas de traducción oral y escrita, en tiempo real y en ambos sentidos, el sistema S-FLTS también dispondrá de la posibilidad de traducción de textos a vía oral y viceversa. Su capacidad de reconocimiento óptico de caracteres permitirá que información «escaneada» o filmada pueda ser traducida. Hablamos, por ejemplo, de pintadas en una pared, señales de tráfico en carreteras o papel manuscrito, en este último caso dependerá, como es lógico, de que la letra sea inteligible para el equipo OCR.

El programa que resulte de este proyecto ha de ser interoperable con el sistema de mando y control conjunto norteamericano, el cual, a su vez, deberá ser interoperable con el futuro Bi-SC AIS, sistema conjunto aliado de la OTAN.

Permítanme aquí una corta reflexión sobre el futuro sistema C4I conjunto para nuestro Ministerio de Defensa. Sé que este sistema C4I conjunto es una de sus preocupaciones y un trabajo pendiente en el EMACON, considerando que el actual sistema aéreo —ICC— y el marítimo —MCCIS— son ambos sistemas aliados y nacionales, y que como tales sistemas aliados van a converger en el mencionado Bi-SC AIS, parece lógico considerar la posibilidad de que nuestro futuro sistema nacional conjunto sea el propio Bi-SC AIS aliado, lo cual, a mi modo de ver, sería una forma de asegurar su correcto funcionamiento y la interoperabilidad con los aliados, por otra parte con toda seguridad su desarrollo saldría bastante más barato.

Volviendo al proyecto S-FLTS, éste se ha dividido en tres fases:

- Fase I, del 2007 al 2011.
- Fase II, del 2012 al 2016.
- Fase III, del 2017 al 2020.

Al final de la primera fase se dispondrá de un sistema que facilitará traducciones oral y escrita de nivel básico, con módulos orientados a misiones específicas (operaciones de paz, apoyo sanitario, operaciones especiales, interrogatorios, etc.). El decir que se está a un nivel básico no es muy concreto, el documento que define las capacidades de esta nueva herramienta especifica que el resultado de esta fase debe reflejar un nivel mínimo aceptable para que sea operativamente útil en el desarrollo de operaciones militares.

Veamos los números que se manejan en la definición de capacidades para intentar concretar algo más. Las traducciones tendrán un 70 por 100 de fiabili-

dad, para hacernos una idea podemos suponerlo como si el sistema tradujera o entendiera al mismo nivel que uno de nosotros que tenga un «posee de fortuna». Los estudios hechos confirman que con este 70 por 100 se puede entender un artículo de un periódico generalista.

También nos puede servir como referencia que los diferentes módulos tendrán diccionarios de entre 2.000 y 3.000 palabras en inglés y una cantidad similar del idioma de que se trate. Esto lo podemos comparar con las 600 frases fijas de que disponen los sistemas de traducción simultánea oral que utiliza actualmente la administración norteamericana. Como ejemplo, uno de los módulos será el denominado «core» empleado para saludos iniciales, identificación y fórmulas básicas de comunicación. Podemos concluir que con 2.000 palabras las necesidades iniciales de comunicación pueden ser cubiertas.

En lo referente al diccionario del traductor de textos tendrá en esta primera fase unas 50.000 entradas en inglés y otro tanto en el idioma de que se trate.

En la fase II se aumentará el número de idiomas cubiertos por el sistema, se añadirán nuevos módulos específicos para más tipos de operaciones y se aumentará la fiabilidad de las traducciones hasta alcanzar un mínimo del 80 por 100, con un objetivo del 90 por 100. Podemos considerar que la máquina responde a las capacidades de un «posee auténtico» (con algún tres y muchos cuatros). Al final de esta fase se dispondrán de módulos en 125 idiomas, priorizados por el Departamento de Defensa. El vocabulario de los módulos de traducción oral será de entre 5.000 y 10.000 palabras y el diccionario del traductor de textos se ampliará hasta las 75.000 entradas en cada idioma.

En la fase III y última la fiabilidad de las traducciones será tal que incluso los modismos, bromas o expresiones coloquiales serán traducidos sin problemas, dando la posibilidad a una traducción oral en la que ambos lados empleen su forma natural de expresarse. Para terminar las comparaciones digamos que ahora la máquina tiene los «cuatro cincos». Respecto a números se estará en una fiabilidad de entre el 90 y el 98 por 100, con 225 idiomas disponibles. No se han considerado máximos para el vocabulario que tratarán los traductores oral y de textos, se supone que la tecnología del momento (hablamos del final de la segunda década de este siglo) permitirá la alta fiabilidad, objetivo de esta última fase.

El material obtenido al final debe ser:

- Los módulos correspondientes a cada lengua, con un total ya mencionado de 225 diferentes. Estos módulos serán siempre mejorables y sus diccionarios ampliables.
- El programa informático, común para todos los idiomas, y que, sin duda, se debe considerar la parte más complicada del proyecto.
- Los equipos donde corra el programa. Naturalmente habrá diferentes opciones según se vaya a trabajar en un Estado Mayor o lo lleve un

teniente al mando de un pelotón, correspondiendo equipos de diferentes tamaños y diferentes prestaciones. Por otro lado, las comunicaciones vía satélite permitirán a pequeñas unidades, o incluso a soldados aislados, acceder a grandes sistemas y bases de datos situadas en el cuartel general.

Por último mencionar que, como siempre, el enemigo no descansa e indudablemente la «nueva arma» tiene sus peligros. Por ejemplo, el enemigo buscará cómo perturbar el programa de forma que la traducción que obtengamos sea totalmente engañosa, o si se utiliza el traductor para dar órdenes a fuerzas aliadas el enemigo intentará interferir y que las órdenes que le lleguen sean diferentes, etc. Una vez más, con toda nueva capacidad aparecerá la contramedida correspondiente.

Conclusión

La información aquí ofrecida es escasa y, sin duda, tendrá errores; sin embargo, creo que el proyecto es interesante y que es bueno darlo a conocer. Por otro lado sería bueno que la Armada, o posiblemente mejor dicho en este caso, el Ministerio de Defensa, pueda situarse en la vanguardia de la investigación en este interesante campo. Una vía puede ser seguir de cerca las herramientas/sistemas de este tipo que se prueben en el ejercicio JWID de cada año.

Considero por último que la idea extendida durante años de que el inglés sería la única lengua del futuro, debe darse por olvidada. Buena prueba de ello es la fuerza del español en los Estados Unidos. Donde año tras año aumenta su influencia y el número de personas que lo usan. No quiere esto decir que el español vaya a sustituir al inglés, más bien creo que en el futuro los pueblos van a seguir usando sus propias lenguas, las únicas en que realmente se encuentran cómodos para relacionarse, y por esta razón las «maravillosas máquinas» de que trata este artículo tendrán un rápido y extenso desarrollo. De esta forma no se podrá ya decir que con inglés (algunos añaden «y español») se pueden hacer negocios o turismo por toda la tierra. Lo que será necesario es llevar encima una SIPA, palabreja que me acabo de inventar, sin ánimo ni esperanza de que haga fortuna, y con la única intención de que el amable lector que haya llegado hasta aquí termine el artículo con una sonrisa.