

Reflexión y crítica

Dificultades para una traducción automática: el no-signo

Julián Sanz Pascual

En un reportaje periodístico sobre la feria japonesa de Tsukuba (verano 1985) donde se ofrecía el último grito en electrónica, el autor del mismo se atreve a pronosticar que en un futuro próximo «unos sofisticados computadores portátiles podrán explicar y traducir simultáneamente al turista y al hombre de negocios cualquier indicación del japonés a los principales idiomas»¹. Esta atrevida afirmación no es más que un ejemplo de la superficialidad y de la ligereza con que se suele tratar un tema tan complejo y difícil como es la traducción automática. La historia real del asunto es mucho menos brillante.

«La posibilidad de la traducción automática por un ordenador —explica Terry Winograd— surgió mucho antes de la fabricación comercial de éste. En 1949, cuando los contados ordenadores que estaban en funcionamiento se encontraban en instalaciones militares, Warren Weaver, un matemático pionero en la teoría de la comunicación, señaló que las técnicas desarrolladas para descifrar códigos podían aplicarse a la traducción mecánica.

»En principio la tarea parece sencilla. Dada una oración en un lenguaje fuente, dos operaciones básicas proporcionan la correspondiente en el lenguaje destinatario. Primero se sustituyen las palabras componentes por su traducción; a continuación, se reordenan las palabras de la traducción y se ajustan los detalles...

»Los programas —concluye— daban unas traducciones tan malas que resultaban ininteligibles. El problema estaba en que los lenguajes naturales expresan el significado de una manera distinta de como los códigos criptográficos expre-

¹ EL PAIS SEMANAL, n.º 431, 14 de julio de 1985.

san los mensajes. El significado de una frase en el lenguaje no depende sólo de la forma de la oración, sino también de su contexto»².

Mas, a pesar de todas estas dificultades, el autor se va a referir más adelante a logros importantes que la informática ha tenido en este campo. «Aunque no es posible una traducción mecánica de alta calidad y totalmente automática —dice—, existe soporte lógico que facilita la traducción. Tomemos un ejemplo de ello en la informatización de los medios auxiliares: diccionarios y libros de frases, que van desde los sistemas complejos para los traductores técnicos, en los que la función de buscar una palabra es parte de un programa multilingüe de tratamientos de textos, hasta las "bibliotecas" de frases de bolsillo para turistas». Sin embargo, a pesar de éstos y de otros logros que señala el autor, dice finalmente: «prosigue el trabajo teórico sobre la sintaxis y el significado, pero no se ha conseguido ningún adelanto decisivo en el campo de la traducción por ordenador»³.

Parece claro, entendemos nosotros, de acuerdo con esta exposición de Winograd, que las perspectivas de la traducción automática no son tan halagüeñas como suponía el autor del reportaje sobre la feria japonesa de Tsukuba. Quizá este error de perspectiva, bastante generalizado, nazca de la fascinación que ejerce hoy el campo de la electrónica, al menos por aquello de las tecnologías de punta o tecnologías del futuro. A esto habría que añadir la desmesurada audacia de muchos de los que tratan de aplicar la electrónica a objetivos como éste de la traducción automática. Encerrados, acaso, en una excesiva especialización, se han olvidado de que, detrás del lenguaje, existe toda una larguísima tradición filosófica de estudio, al menos desde los sofistas (s. V a. C.). Sin duda el más genial y el más adelantado de todos los filósofos antiguos en este terreno fue Aristóteles, el creador de la lógica. Pero no hay que olvidar a cuantos retóricos, gramáticos y lingüistas han tratado de afrontar el conocimiento y el dominio del lenguaje. Y la verdad es que, si en el conocimiento siempre se ha avanzado, no ha ocurrido lo mismo en el dominio. Y esta dificultad de dominio nace de que el lenguaje natural es algo vivo, y lo es porque, desde sus más oscuros orígenes, los signos empezaron a significar no de forma analítica o acumulativa, sino de forma sintética o relacional. Y ahí es donde se ha estrellado la lógica, la gramática, la lingüística y hoy se estrella la informática. Quizá la lingüística, al menos como concepción más moderna de la ciencia del lenguaje, se salve en cierto modo, al menos en tanto en cuanto que coloca el hecho lingüístico, el «habla» en palabras de Saussure, por encima de cualquier formalización. La informática, por el contrario, se ve obligada a formalizar el lenguaje para poderlo dominar, pero es a costa de desvirtuarlo. Algo similar le ocurre a la lógica, muy especialmente a la lógica matemática. Tanto la una como la otra pueden llegar a dominar el lenguaje, pero siempre a costa de desnaturalizarlo y

² WINOGRAD, T., «Programación y tratamiento de lenguajes», en «Investigación y Ciencia», noviembre 1984, p. 70.

³ Ibid., pp. 72 y 73.

empobrecerlo. Y sólo en tanto en cuanto que el lenguaje responde a las formulaciones de la lógica y de la informática, podrá ser objeto de la traducción mecánica, pero no más allá. Lo mismo le ocurre a la geometría con respecto a la realidad del mundo físico: sus formulaciones jamás podrán agotar esa realidad, que siempre es mucho más rica y compleja.

Y éste es el muro contra el que se estrella el informático a la hora de intentar hacerse con el lenguaje: éste a primera vista parece lógico, pero después resulta que no lo es. Por poner un ejemplo muy sencillo: si yo voy a una cafetería y pido una cucharilla, me traerán una cuchara pequeña, pero, si pido una manzanilla, no me traerán una manzana pequeña, como sería lo lógico, sino otra cosa. Pero la dificultad más genérica que presenta el lenguaje es la homonimia. Desde el artículo hasta la interjección, pasando por el sustantivo y las demás partes de la oración, prácticamente no hay una sola palabra que no tenga dos o más significaciones distintas. Los artículos determinados todos pueden ser también pronombres personales. Incluso la modesta preposición es polisignificativa. Podemos ilustrarlo con un conocido chiste. Uno pregunta: «¿A que no sabes por qué me duran tanto las gafas?». La respuesta es: «Porque miro mucho *por* ellas». El efecto estético del chiste se produce porque la preposición «por» está tomada de forma biunívoca: como «a favor de» y como «a través de». Pero aún son posibles otros significados como «lugar por donde», «a causa de» y otros.

En cuanto a los sustantivos, podemos ilustrarlo también con otro conocido chiste. Dos amigos se encuentran y uno le dice al otro: «¿No sabes? Me he casado con una *muda*». Y el otro contesta: «Pues yo me he casado con lo puesto». En cuanto a los verbos, es muy conocido también aquél en que uno propone a otro: «Te vendo el coche». Y el otro contesta: «¿Para qué quiero un coche vendido?».

Esta polisignificación pone dificultades para que funcionen ciertos principios lógicos, muy especialmente el de tercero excluido, al menos de forma estricta, que es como exige la informática para poder tratar el lenguaje con un ordenador, para poder entenderse con él mediante reglas. Y es que, en el lenguaje natural, de lo reglado a lo excepcional no hay más que un paso; es más, a veces la excepción es la regla. La diferencia entre la máquina y el hombre ante estas anomalías es que la máquina se descompone, y el hombre también se descompone, pero es de risa.

La única alternativa que le queda al informático ante esta falta de rigor significativo de los términos es la de afrontar el problema mediante listas o paradigmas de posibles significaciones. Teóricamente esto es posible siempre que las listas sean limitadas, pero ocurre que los demás términos, al menos mientras se trate de lenguas vivas, están siempre abiertos a nuevas significaciones. Mas las dificultades no terminan ahí, sino que es necesario identificar a cada término cuando tiene un significado o cuando tiene otro distinto. ¿Y cómo la máquina puede hacer esta identificación sin un signo material que los distinga?

En el lenguaje natural el hablante hace esta identificación mediante lo que podemos llamar de forma genérica el *no-signo*. Al *no-signo* lo podíamos definir

como a un signo que significa sin estar materialmente presente. Esto, en estricto rigor lógico, es absurdo, y no es de recibo para aplicarlo a una máquina. Sin embargo, en los hablantes humanos, funciona perfectamente. Es que el humano, además de su capacidad para captar la presencia material de los signos y para identificarlos, tiene la capacidad de captar otras formas significativas no materiales, como son los silencios en el lenguaje hablado, el tiempo puro, los espacios vacíos en el lenguaje escrito, el espacio puro. En general, es capaz de captar el significado de la ausencia de los signos, los *no-signos*. Dicho de otra manera: el hablante expresa más de lo que expresa y percibe más de lo que sus sentidos le ofrecen. Se puede decir que, además de los sentidos corporales, tiene otro sentido más profundo y complejo, el sentido estético, aunque sea una redundancia.

No obstante, hay no-signos que de alguna manera se materializan en la escritura: tales pueden ser los signos de puntuación, el acento diacrítico y otros. Pero estos signos sólo recogen una pequeñísima parte de las posibilidades de significación de los signos sobre el «vacío» temporal en el lenguaje hablado y sobre el «vacío» espacial en el lenguaje escrito. No obstante, la gran dificultad de la máquina se debe a que ésta sólo es sensible a las relaciones espaciales, no a las temporales. Es decir, la máquina es sensible en mayor o menor medida a lo plástico, a la letra, no lo es a lo acústico, a la música. Nosotros entendemos aquí la música en su sentido más puro, el silencio o no-sonido. Pero ni en lo plástico la sensibilidad de la máquina puede aproximarse siquiera a la del ser humano. El hombre tampoco, por otra parte, ha sido capaz de adecuar y ajustar el lenguaje escrito al lenguaje hablado de una forma perfecta. Es decir, su sensibilidad plástica y su sensibilidad acústica nunca acaban de ajustarse, y eso a pesar de la vieja historia del fonetismo.

El ejemplo más claro y más sencillo de la materialización del no-signo nos lo ofrece la aritmética con el *cero*. Este signo pasó a ocupar el lugar vacío que era necesario dejar para que los signos pudiesen significar por el lugar que ocupaban en la línea. Esta forma de significar por relación hizo posible las operaciones mecánicas de la aritmética, cosa que en otras numeraciones —para nosotros la más conocida es la romana— es imposible. El cero en la aritmética tiene una significación muy precisa, sin embargo también plantea problemas en ciertas operaciones. Por apuntar una muy conocida, $1^0 = 1$. Si $1^1 = 1$, podíamos deducir que $1 = 0$.

Como no podía ser menos, en el lenguaje natural, el no-signo plantea problemas mucho más complicados. En primer lugar, no hay un sólo no-signo, ni siquiera una sola clase de no-signos; tampoco podemos estar seguros de que exista un número limitado de clases de no-signos. No olvidemos que en el campo de lo negativo, que es donde está el no-signo, es imposible hacer ciencia definitiva. Digamos que en el lenguaje natural los valores significativos que desencadena el no-signo son de tal complejidad que desbordan cualquier previsión. Es más, a medida que entramos en este campo, lo que vamos descubriendo

son nuevas negaciones y nuevas posibilidades de negaciones. No obstante, vamos a intentar una clasificación de los no-signos más utilizados en el lenguaje natural, y un breve estudio sobre cada uno de ellos. Entendemos que hay tres clases de no-signos:

1.º El no-signo significa por oposición a un signo que se omite. Por oposición a «s» o a «es», significa singular: mesa(), mesa(s); león(), leon(es). Por oposición a «a» puede significar masculino: león(), león(a). Por oposición a «s» puede significar también la primera y la tercera persona del singular de un determinado tiempo verbal: corría(), corría(s), etc.

Cabe señalar en este punto dos hechos del máximo interés:

a) Que cualquiera de estas reglas está plagada de excepciones: plan(), plan(a), pis(), pis(a).

b) Que en la escritura el vacío y la extensión de cada signo significativo que significa por oposición no se corresponden, sino que el vacío o espacio libre se comparte con el que hay que dejar entre palabras. Ejemplo: «La(s) mesa(s) está(n) deteriorada(s)». «La() mesa() está() deteriorada()». En la escritura real la segunda frase tiene menos longitud que la primera, con lo que el hueco del no-signo desaparece, lo que añade una dificultad más para la mecanización del lenguaje.

2.º El no-signo como pausa significativa. La pausa oral o el espacio vacío entre signos tiene una significación por cuanto que modifica las relaciones de los signos significativos. En el lenguaje hablado, las pausas pueden ir acompañadas de una entonación especial, pero, generalmente, no tienen más signo que el no-signo o *silencio* más o menos *pronunciado*; en el escrito, las separaciones se marcan a veces mediante diferentes signos materiales, que pretenden distinguir las diferentes clases de separaciones o no-signos. Estos son la coma, el punto, el punto y coma, los dos puntos, los puntos suspensivos, el punto y aparte, que no es un signo distinto del punto y seguido. Podíamos añadir la interrogación, la admiración, las comillas y algún otro como el paréntesis y el guión. Hay que advertir que el uso de estos signos no responde a una reglamentación rigurosa. Hasta tal punto es así que la forma de puntuación suele ser un rasgo distintivo de ciertos autores, lo es también de las diferentes épocas de un idioma. Incluso se ha ensayado escribir sin ningún signo de puntuación. A veces los criterios de puntuación no son mantenidos por un mismo autor a lo largo de una misma obra. Se produce aquí un fenómeno similar a la homonimia de que hemos hablado, lo que introduce elementos de dificultad para cualquier normalización.

Y si esta dificultad se da en el lenguaje escrito, que, por su propia naturaleza, se presta más a la normalización, imaginemos lo que ocurrirá en el lenguaje hablado, en el que lo espontáneo de cada hablante está mucho más patente: sus condiciones de todo tipo, grado de cultura, situación social, familiar, emocional, psicológica, etc. Todos estos factores suelen ser determinantes del ritmo y, en general, de la forma de dicción. Claro está que en el lenguaje escrito ocurre algo

similar, pues en él también hay una cierta espontaneidad, muy especialmente en los signos de puntuación.

La enorme importancia de la pausa, tanto la oral como la escrita, para el contenido del mensaje lingüístico, no es necesario encarecerla. Una coma cambiada de lugar puede cambiar todo el mensaje. La razón ya está dicha, porque cambia las relaciones de los términos. Un viejo chiste puede dar idea de lo que queremos decir. En un periódico apareció un anuncio que decía: «Se necesita camarero, inútil presentarse sin referencias». Un lector poco avezado a la lectura, o quizá malicioso, lo leyó así: «Se necesita camarero inútil, presentarse sin referencias». Es claro que, gracias al no-signo como pausa o como silencio, con la misma letra y aún en el mismo orden, se pueden hacer muchas músicas.

3.º El no-signo como orden. Una de las mayores fuentes de simplificación del sistema lingüístico está en el hecho de haber descubierto que el orden también puede significar. Así, con los mismos signos, ordenados de forma distinta, se pueden producir signos distintos. Esto ocurre en los tres niveles básicos del lenguaje: la sílaba, la palabra y la proposición. Sin embargo, tanto en la sílaba como en la palabra, el problema que se plantearía a la traducción automática está prácticamente resuelto, pues en uno y otro caso se reduce a listas limitadas que ya están fijadas. En el caso de las palabras, la lista la constituyen todos los términos del diccionario. Por otra parte, el campo de la creatividad lingüística se reduce prácticamente a la proposición: rarísimamente se da en la sílaba, de forma muy esporádica en la palabra. Es, pues, en la proposición donde se plantean los problemas.

Aclarado esto, es necesario que aceptemos la «existencia» de un fondo temporal para el lenguaje hablado y de un fondo espacial para el lenguaje escrito, una especie de no-ser genérico sobre el cual el ser *es*. Sin ese fondo, no existirían otros órdenes como posibilidad y, entonces, cualquier orden sería único y no significaría nada. El no-ser como fondo, bien espacial bien temporal, hace posibles los diferentes órdenes y, por consiguiente, las diferentes posibilidades de significación de los mismos elementos primarios. Esto es innegable como hecho: «En este bosque no hay un solo hombre», «En este bosque no hay un hombre sólo».

¿Por qué el distinto orden produce distinta significación con los mismos signos? La pregunta, que a primera vista parece ingenua, resulta muy difícil de contestar. Parece claro que, siguiendo un principio de economía que está presente en todos los pasos de la creación lingüística, la utilización de los diferentes órdenes como signos simplifica el lenguaje de forma asombrosa. Pero esto desencadena una segunda cuestión: ¿el hecho de que un orden cargue a una palabra de un determinado contenido conceptual es natural o es convencional? Veamos este sencillo ejemplo: «Esta es gente menuda», «Menuda gente es ésta». ¿Por qué el hecho de colocar delante del sustantivo el adjetivo «menuda» cambia tan radicalmente el contenido conceptual que tenía en el orden inverso? En el estudio comparativo de los diferentes idiomas, nos encontramos con

algunos en los que es inaceptable que el adjetivo calificativo vaya detrás del sustantivo. Esto debe considerarse, creo yo, como una convención, puesto que en castellano no ocurre así. ¿Qué es lo que ha ocurrido en castellano? El buen sentido de estos hablantes, aplicando el principio de economía, les ha llevado a utilizar este orden diferente para expresar un mensaje diferente. Ahora bien, ¿por qué la anteposición significa de forma más figurada y la posposición significa de forma más estricta? A la vista del estudio comparativo con otros idiomas, parece claro que se trata de una convención, pero que después, andando el tiempo, se ha fijado, dando lugar a esta distinción general: delante del sustantivo el adjetivo se hace explicativo, detrás se hace especificativo.

Hay casos, sin embargo, en que sería dudoso el apelativo de convencional. Por ejemplo: «La respuesta inglesa a la cuestión», «La respuesta a la cuestión inglesa». ¿Es natural o es convencional que en el primer caso el término «inglesa» califique a «respuesta» y en el segundo califique a «cuestión»? Parece que el factor proximidad es decisivo, lo que nos permitiría entenderlo como significación natural. Sin embargo esta frase, que está tomada de un Telediario, fue dicha de la forma segunda, pero con el sentido de la primera, pues se refería a la respuesta inglesa a la cuestión de Gibraltar. Parece, por tanto, que el contexto general en que está dicha una frase marca más a los términos que su posición dentro de la propia frase.

De todas maneras hay que decir que el orden como signo depende mucho de cada idioma, de la sensibilidad o estética en que se haya desarrollado. Y este desarrollo depende a su vez de otros elementos sintácticos de que disponga el idioma, como puedan ser las desinencias de caso, que en las lenguas romances han desaparecido prácticamente. Esto suele hacer inaceptable en el orden originario la traducción del latín al castellano. Por poner un ejemplo muy sencillo: la célebre obra de Lucrecio, *De rerum natura*, para traducir en el mismo orden su título al castellano, habría de quedar de esta manera: «De las cosas la naturaleza». Lo que resulta extrañísimo.

Lo que se puede decir, entiendo yo, es que no hay un orden universal para todos los idiomas, pero que cada idioma responde a un sentido del orden, bien que este sentido nunca suele ser cerrado, sino abierto, al menos en lo que se refiere a los hablantes de base. Podemos decir que, originariamente, en todo idioma hay una cierta convención, pero después el efecto significativo del orden de alguna manera se naturaliza. Si nos atenemos a lo que el orden introduce en las cosas físicas como causa de ellas, parece claro que el hecho de que un orden produzca un determinado efecto y otro orden produzca otro será natural y no convencional. Otra cosa es que la diferente estructura que se atribuya a los isómeros químicos, por ejemplo, para explicar sus diferencias cualitativas, responda a la realidad o sólo sea una convención para facilitarnos a nosotros la necesaria explicación. Hoy, sin embargo, hay una ciencia y una técnica que se llama Ingeniería Química, que es capaz de fabricar sustancias que respondan a propiedades previamente establecidas, y esto lo hace fundamentalmente a base

de modificar estructuras macromoleculares. La estructura, pues, como causa de las diferencias cualitativas, no puede ser convencional, sino que ha de ser de naturaleza. Lo que ocurre es que este criterio no es tan fácil aplicarlo a las isomerías lingüísticas, pues el efecto, el resultado efectivo y comunicador, nunca se podrá objetivar a la manera como se puede objetivar en la ciencia química.

Digamos finalmente que los problemas del orden como signo son tan importantes para el lenguaje, introducen tales trastornos que no han faltado lingüistas tan ilustres como Ferdinand Saussure que han cuestionado si la palabra es la unidad básica del lenguaje⁴. Esto crea una inseguridad tal en el traductor que le obliga a replantearse toda la gramática a fin de encontrar un terreno sólido en el que moverse. Y pasa de la traducción de palabras a la traducción de frases sin darse cuenta de que, así, se le escapan los fundamentos mismos del idioma. Lo que ocurre en el fondo es que el lógico, el informático, el lingüista, el gramático, todos viven encerrados en una mentalidad analítica, y el orden precisamente y todos los problemas que plantea desbordan ese análisis.

* * *

Pero las dificultades para la traducción automática no acaban ahí, en todo lo que hemos apuntado sobre el no-signo, sino que todo hecho lingüístico es complejísimo, tan complejo que, a veces, están siendo significativos elementos de nuestra memoria histórica, bien personal bien colectiva, de los que no tenemos conciencia, elementos físicos presentes captados por nuestros sentidos sin que nosotros tengamos tampoco clara conciencia de ello. También se puede decir que todo signo lingüístico, por el hecho de no estar presente, significa de una forma negativa, aunque en algunos casos significa de forma positiva, cuando está elíptico. Ahora bien, hace falta que este término esté sugerido, que se sobreentienda, o lo que es lo mismo, que esté presente sin estarlo materialmente. Esta manera negativa y sugerente de significar abre un campo inmenso de posibilidades al lenguaje, muy especialmente al poético. Así San Juan de la Cruz nos produce un delicioso escalofrío cuando leemos en sus versos: «y déxame muriendo / un no sé qué que quedan balbuciendo»⁵. O este verso del Poema del Cid: «¡Fabla, pero mudo, —varón que tanto callas!»⁶. O esta frase un tanto teatral de «¡Interpreta mi silencio!».

El silencio, lo negativo, el no-signo, la metafísica del lenguaje en una palabra, la parte «espiritual», ¿será capaz de captarla la máquina? Se trata nada más y nada menos que del campo de libertad en el que el ser humano se mueve

⁴ SAUSSURE, F., *Curso de lingüística general*. Akal Editor. Madrid 1980, parte II, cap. II, epígrafe 3

⁵ «Se ha discutido mucho sobre la naturaleza de la palabra, y reflexionando un poco se ve que lo que se entiende por eso es incompatible con nuestra noción de unidad concreta.»

⁶ SAN JUAN DE LA CRUZ, *Cántico espiritual entre el alma y Cristo*

⁶ *Poema del Cid*, Canto 142.

cuando usa del lenguaje, pero sobre todo cuando lo crea. Es que el lenguaje, además de la función comunicadora, tiene también la función opuesta, la de no comunicar, la de defenderse del efecto comunicador. Hasta tal punto es así que, si la máquina o cualquier ser, humano o humanoide, lograra hacerse con todos los secretos del lenguaje llamado natural, nos veríamos obligados a enmudecer. Afortunadamente no hay tal lenguaje natural en sentido estricto, sino que siempre es posible introducir una nueva convención para defenderse, pues el campo de lo negativo no tiene límites. Es la mejor defensa contra las máquinas, contra los lingüistas, lógicos, gramáticos, informáticos, retóricos y otras gentes que, desde tiempos muy remotos, llevan tratando de fijar el lenguaje. ¡Afortunadamente sin conseguirlo! Pienso que las modernas máquinas, incluso aunque sean japonesas, no van a tener más éxito. Pienso también, eso sí, que, si las adiestramos convenientemente, podemos convertirlas en magníficos auxiliares capaces de liberar al hombre de las tareas mecánicas, que son las más penosas, las menos humanas.