



Los primeros intentos en la traducción automática constituyeron uno de los fracasos más espectaculares de la inteligencia artificial pero en la actualidad se están aplicando nuevos enfoques y técnicas a este problema, lo cual permite vislumbrar la aparición de sistemas competentes dentro de algún tiempo.

Los fracasos habidos hasta ahora se deben fundamentalmente a que se construyeron sistemas basándose en la teoría del lenguaje de Chomsky, la cual se queda solamente con la estructura de la frase considerando que a una estructura profunda (un sentido, una idea) le corresponde una estructura superficial (una frase). Esto no tiene en cuenta por tanto, los distintos sentidos que puede tener la frase según el contexto, situación, entonación, creencias, saber compartido, etc.. Como no se tiene en cuenta los distintos sentidos y significados que una frase o, más correctamente, un texto puede tener en el otro idioma. De esta forma los sistemas construidos sólo sirven para traducir frases unívocas. Lo que se pretende en este estudio es indicar cómo se podrían incorporar algunas de estas nuevas variables a un sistema experto para conseguir una mayor aproximación al problema de la traducción automática.

# Consideraciones sobre el desarrollo de un sistema experto para la traducción automatizada

CARLOS GARCÍA-FIGUEROLA  
PANIAGUA, ADOLFO  
DOMÍNGUEZ OLLERO, JOSÉ  
LUIS ALONSO BERROCAL,  
EMILIO RODRÍGUEZ VÁZQUEZ  
DE ALDANA, ÁNGEL  
FRANCISCO ZAZO  
RODRÍGUEZ

*Universidad de Salamanca*

*Early attempts at machine translation were one of the most spectacular failures in artificial intelligence, but today new approaches and techniques are being applied in this area and this may lead to the development of more competent systems in the not too distant future.*

*Earlier failures in machine translation were basically due to the fact that systems were modelled on Chomsky's theory of language, which merely addresses the structure of the sentence and posits that a surface structure (a sentence) is in a one-to-one correspondence with a deep structure (a meaning, an idea). This approach therefore does not take into account the different meanings that a sentence may convey depending on context, situation, intonation, beliefs, shared knowledge, etc. As this approach does not consider variation in meaning it cannot account for the different meanings which a sentence (or rather, a text) may have in the target language. For this reason, such systems serve only to translate invariant sentences.*

*This paper offers some considerations as to how such meaning variables could be incorporated into an expert system in order to achieve more satisfactory results in machine translation.*



## 1 MÉTODO UTILIZADO PARA LA CAPTACIÓN DEL CONOCIMIENTO

El método para la captación del conocimiento es fundamentalmente la entrevista con reorientación de conocimiento, aunque también se observa el trabajo del experto y se le pide que resuelva en voz alta casos prácticos.

Al principio se utiliza un tipo de entrevista informal y una observación del trabajo del experto para centrar, conocer y delimitar más el problema, y después entrevistas dirigidas pero sin limitación de tiempo, aunque a veces con *limitación de conocimiento*.

No se utiliza la entrevista con limitación de tiempos porque el proceso de la traducción escrita (es el que trata el sistema), a diferencia de la traducción oral o interpretación, no es un proceso que dependa demasiado del tiempo y al sistema no se le va a exigir una traducción en tiempo real. Además el sistema está orientado como una ayuda para el traductor, el sistema realiza una traducción pero esta en muchos casos deberá ser revisada. Por todo esto la entrevista con limitación de tiempo no tiene mucho sentido y por lo tanto no se utiliza.

En cuanto a la entrevista con limitación de conocimiento sí nos puede ser útil, pues la información de entrada para el sistema será única y exclusivamente el texto de origen. Todo texto puede ser ambiguo o incompleto en cuanto a que no siempre plasma de una forma clara lo que quiere decir el autor (esto es el llamado aspecto de *inacabamiento del texto*) y no tenemos la posibilidad de preguntar al autor para resolverlo. Esto nos lleva a tener una información incompleta e información con poca certeza. Por otro lado, el conocimiento necesario para traducir engloba todos los cono-

cimientos y experiencias del traductor además de todos los diccionarios y enciclopedias que suele consultar en su trabajo, en definitiva un conocimiento tan extenso que no hay más remedio que limitarlo a la hora de construir el sistema experto.

El método para la captación del conocimiento que más se va a utilizar consiste en realizar la entrevista al experto (sea dirigida o informal), extraer el conocimiento más importante o al menos el que más nos interesa para el sistema, presentarle al experto este conocimiento intentando determinar si es correcto y si es lo suficientemente completo; se realizan los retoques y ampliaciones necesarias y se le vuelve a pasar al experto; el proceso se repite hasta que no se saca nuevo conocimiento del experto y este conocimiento es considerado correcto tanto por el experto como por el ingeniero del conocimiento. Este proceso se tiene que realizar para todos los problemas que se pueden dar, de aquí que las primeras entrevistas sean precisamente para encaminar el problema y para determinar los posibles problemas y factores que se pueden encontrar al traducir.

Se puede plantear la cuestión que para algunos problemas el proceso anteriormente expuesto se haga excesivamente largo, que no quede claro el proceso para resolver el problema o que no se pongan de acuerdo el ingeniero del conocimiento y el experto. En estos casos lo que se hace es que el ingeniero del conocimiento prepara un texto en que se da el problema concreto a resolver y le pide al experto que lo resuelva en voz alta. Esto nos permite clarificar cómo el experto razona para resolver el problema y así poder continuar sin atascarnos en un problema.

## 2 EL PROCESO DE TRADUCIR

### *La traducción del traductor (el experto)*

Traducir puede definirse como hablar para decir lo ya dicho en otra lengua. Como lo dicho se puede decir en otra lengua es que lo dicho es distinto y separable de la lengua. Este dicho es el llamado sentido que no es exactamente el significado sino que simplificándolo se puede decir que es el significado en una situación concreta del habla. De todo esto se puede deducir que el traducir es extraer el sentido de un texto (en la traducción escrita que es la que nos interesa) y reformular ese sentido en el otro idioma. Por esto al tener el sentido del texto nos olvidamos del primer idioma de forma que ya no nos debe influir (siempre que no se esté traduciendo poesía o textos en los que la forma sea importante).

Traducir por tanto no consiste en una operación de transformación de palabras de un idioma al otro, tampoco consiste en traducir trozos sueltos de frases ni formas gramaticales puesto que no es siempre posible encontrar formas lingüísticas equivalentes y la transcodificación total no existe. Por último, traducir no es tampoco comparar campos semánticos puesto que el sentido engloba a los campos semánticos y los supera (entendiendo los campos semánticos como los significados del diccionario). Es cierto que el traductor tiene que operar sobre los campos semánticos, pero no tendrá que traducir palabras aisladas sino situadas en un acto de habla en el que será evidente lo que se quiere decir (este querer decir será muchas veces distinto de la semántica de las palabras).

El traductor tiene que entender lo que dice el hablante, por qué lo dice, por qué elige unas palabras en lugar de otras en función del con-

texto, todo lo cual requiere un esfuerzo de reconstrucción del pensar del otro.

### *Los pasos para traducir. Comprensión del texto*

Una lectura repetida del texto origen hasta que se tiene una comprensión de este. Esta comprensión requiere comprender y conocer la lengua pero también otras variables extralingüísticas como lo gestual, la mímica, pulsaciones del habla (en la traducción oral), la expresividad del hablante que condiciona el ritmo, la prosodia, el tono, el registro y los aspectos emocionales. Pero además toda una serie de creencias, prejuicios, ideas recibidas, la tradición, prácticas sociales o modos de hacer, situaciones en diacronía, normas de comportamiento sociales, conocimientos sobre la vida en general, saberes enciclopédicos, esquemas de interpretación de la realidad, modelos del mundo y sobre todo conocimientos temáticos sobre el asunto del que se va a traducir. Esto es la llamada precomprensión y está constituida por todo el conocimiento que se necesita tener y compartir con el autor del texto que se traduce.

### *Aspectos que hay que tener en cuenta para captar el sentido*

a) La unidad de verbalización que se refiere a una cadena de signos lingüísticos empleados en un acto de habla (no es lo mismo que una frase gramatical). Es una unidad con relación al sentido.

b) La estructura lingüística de superficie que representa la estructura fono-morfo-léxico-semántica de la unidad de verbalización. Es la llamada estructura superficial de la frase según Chomsky pero puede coincidir con una frase, conjunto de frases, fragmento de frase o una cláusula, párrafo o período. Representa la cadena de signos supuestamente estructurada.





c) El potencial semántico que representa la significación de la estructura lingüística en el diccionario. No sólo el significado del diccionario de las palabras sino también el significado del diccionario de las frases, considerado por tanto fuera de todo contexto, al margen de toda situación, en abstracto. Sería la estructura profunda según Chomsky.

d) Estructura rítmica que representa los fenómenos prosódicos que se manifiestan mediante peculiaridades léxicas o sintácticas en la lengua escrita. Estos fenómenos son objeto de la rítmica. La entonación puede transformar el sentido de un enunciado y es por lo tanto un medio para la comunicación.

e) El lenguaje del orador o lengua del texto original con sus sistemas, signos, reglas, sistema estructurado semántico (estructura semántica de la lengua).

f) La lengua del traductor, que se refiere a la lengua del texto destino o final también con sus signos, reglas y sistema estructurado semántico.

g) El sistema rítmico, que representa un sistema de los fenómenos prosódicos de la entonación que tiene lugar en el plano lingüístico (funciones distintivas, integradoras, delimitadoras), en el plano sociolingüístico (antropológicas, sociológicas, idiomáticas) y en el plano expresivo (lúdicas, poéticas y emotivas). Se puede pensar que hace referencia de una manera casi exclusiva a la traducción literaria y más concretamente a la traducción poética pero en realidad también afecta a todas las manifestaciones de la lengua.

h) Sentido, intentado que se puede definir como la percepción que tiene el autor del texto origen. Es lo que quiere transmitir el autor del texto.

i) Los conocimientos, saber enciclopédico o cultura general, es decir, los conocimientos que

son conscientes o concienciables y en particular los saberes sobre el tema de traducción.

j) Las creencias, las normas y las prácticas: por una parte tenemos un conjunto de creencias, normas y prácticas sociales que actúan sobre el autor del texto y del que traduce (sin que estos sean necesariamente conscientes de ellas).

k) La situación social, que es un conjunto de relaciones sociales, la situación es el espacio y el tiempo, y contiene todos los elementos extralingüísticos que funcionan como variables de la comprensión. La situación no es sólo el lugar o el «decorado» o el conjunto de personas y cosas que se describen (referentes inmediatos) sino sobre todo la red de relaciones e intereses sociales.

l) El «mundo» o «mundillo» social al que pertenece la situación. Toda situación pertenece a un conjunto de prácticas sociales que es el mundo.

m) El plano del tiempo histórico en el que se sitúa el texto.

### **Reconstrucción del sentido**

Una vez que se ha comprendido el texto y se tiene el sentido de éste, sus ideas, hay que expresar este sentido en el otro idioma procurando que no nos influya el idioma de partida. Así se consigue que el texto resultante no parezca una traducción sino un texto escrito directamente en ese idioma.

Tendremos que tener en cuenta todos o casi todos los aspectos que se tenían en cuenta para captar el sentido pero aplicándolos ahora al segundo idioma y a su cultura. Aunque realmente en el texto se reflejara la cultura, creencias, situación, etc., del traductor.

### ***Esquema de razonamiento en la traducción escrita y diferencias con la interpretación***

En la traducción del texto los intentos de aprehensión del significado, de un análisis de un sistema por otro, de transposición por asociación y de comprensión del sentido seguida de reexpresión, están tan imbricados y son tan solidarios unos de otros que de hecho es imposible atribuirles un orden fijo. Las interferencias a todos los niveles constituyen un estadio obligado de toda traducción, un componente esencial del proceso mismo, representan la fase durante la cual el mensaje pierde su envoltura original para endosar la nueva.

A pesar de esto se puede sacar un orden más o menos preciso del razonamiento que se sigue para la traducción escrita (no está nada claro para la traducción oral). Este razonamiento comienza con una primera tendencia a sustituir por calco los elementos léxicos y las estructuras sintácticas (por este orden) del texto original por los del terminal y en una segunda fase se procedería a revisar y corregir esos calcos desde la apreciación global del texto de llegada. En cualquier caso no parece que la sustitución de todos los componentes del significado puedan realizarse al mismo tiempo y así se hace necesario el mantenimiento de una parte del enunciado original como soporte de operaciones ulteriores.

Los intérpretes profesionales podrían objetar que el modo de proceder tiene como modelo de desarrollo «la unidad de sentido» que puede definirse como un segmento del texto que es capaz de suscitar una comprensión básica y coherente. El caso de la traducción oral consecutiva constituye una excepción, por cuanto en el momento de reformular el mensaje las palabras han desaparecido de la memoria activa o memoria de trabajo.

Parece no obstante, que el proceso mimético, la traducción por simple sustitución, constituye la primera tendencia de todo traductor. Lo que ocurre es que el profesional de la traducción, que conoce de antemano lo equivocado de esta vía, está ya en la predisposición de olvidar las palabras del original para dedicarse por entero a la reformulación del texto o discurso terminal, contando sólo con el sentido que de aquellas palabras ya desechadas se ha producido. Éste es el mejor medio, quizás el único, de evitar la tendencia espontánea y fácil al calco y a la atracción, lógica, de las interferencias. El mantenimiento en memoria verbal o desverbalizada de las partes sucesivas del texto discurso en cuanto a forma y contenido es una operación necesaria para que se cree el *continuum* de sentido, se eviten las desviaciones y se progrese con acierto.

La diferencia con la traducción oral, ya sea simultánea o consecutiva, es que el intérprete no puede detenerse. En simultánea tiene éste que desverbalizar el contenido de las unidades del discurso porque las palabras se desvanecen, a lo sumo puede retener en su memoria inmediata o por escrito los términos que son simplemente objetos de saber y no de interpretación, con objeto de incluirlos en la equivalencia contextual. En consecuencia el intérprete sólo suele disponer del tiempo para retener los contenidos puramente transcódificables por su univocidad de significación y poner en símbolo un discurso también necesariamente desverbalizado, mediante figuraciones o resúmenes que le permitan reactivar el sentido en un discurso terminal ya bosquejado, ideogramado óptimamente.

Las técnicas, como se ve, son distintas; el fenómeno procesual sigue siendo el mismo. El precepto como resultado de una necesaria y





urgente desverbalización entra en memoria, pero la memoria que lo capta necesita a corto o medio plazo de las muletas del símbolo (el símbolo se entiende como la estructuración de signos), porque caduca (en el sentido de intermitente), lo mismo que el precepto mental.

En la traducción de la escritura se le libera al intérprete de esas dos trabas: el tiempo y la evanescencia del signo. Ambas libertades pueden permitir una entrada en lo mimético del calco, una contemplación de las interferencias y el mantenimiento a voluntad de partes problemáticas o privilegiadas del original, pues a pesar de que el sentido entra en la memoria equivocadamente o que no se recorten adecuadamente las unidades textuales, siempre cabe la posibilidad de corregir, de irse liberando poco a poco de la hipnosis del texto de partida mediante desverbalizaciones intermitentes sin fin y reformulaciones cada vez más precisas.

El traductor cuenta con la vuelta perenne del texto, y en este sentido su tarea de mantener el precepto captado por desverbalizaciones seguidas es siempre una ventaja. El lado negativo lo constituyen los parámetros que definen el texto con ausencia del autor y su diálogo (si algo no está claro no tenemos al autor para preguntárselo). Este aspecto de inacabamiento, de tarea incompleta sobrepasa en dificultad traductora la premura y la aventura en las que se realiza la traducción oral.

También parece claro que el traductor oral puede empezar por el registro formal que prefiera. Sin duda este registro debería ser el propio de la función o funciones que privilegia el propio texto en su carácter «genérico» y su penetración exegética debería recorrerlo de principio a fin, ya que el texto se le ofrece al traductor como un todo cuya unidad es el mismo texto de manera que pueda luego esta-

blecer las subunidades textuales y resolver todos los problemas que la lectura profunda le haya ido suscitando.

Mediante un rodeo largo, mediante reconstrucciones de parámetros que no están en la forma, podrá captar un sentido que será la suma desigual de un conjunto de preceptos y apreciaciones.

Nuestro sistema solamente va a encargarse de la traducción escrita y el proceso que sigue el traductor sería el de la figura 1.

### *Una posible clasificación de las traducciones*

Teniendo en cuenta los aspectos de la lengua que se contemplan en la traducción (situación, creencias, cultura, mundillo,...) se puede establecer una primera clasificación de las traducciones:

a) Traducción de nivel 0: Es aquella en la que las palabras tienen un significado y sentido unívoco y por tanto es una traducción literal. Existen palabras que tienen siempre el mismo significado y también el mismo sentido como pueden ser los números, palabras técnicas, etc. La traducción de estas palabras no requiere más que una transcodificación pero se pueden considerar excepciones dentro de la lengua.

b) Traducción de nivel 1: Es aquella en la que se maneja la semántica de las palabras (significados del diccionario) pero dentro de un tema concreto. Sería una traducción mejor que la de nivel cero porque tenemos varias posibilidades para una palabra según el tema. Además la traducción semántica es la traducción de expresiones enteras del idioma cuya traducción difiere mucho de la literal. Muchas veces la traducción de una palabra no se ajusta al significado del diccionario, ni siquiera en un tema delimitado, porque el autor la utiliza con un sentido distinto. Algunas veces ese sentido queda



1.º Lectura repetida del texto, en la que aparece siempre la tendencia inicial al calco y en la que el traductor ve los problemas. Estos problemas son sobre todo de la semántica de las palabras. (Este paso 1º se repite tantas veces como sea necesario).



2.º Búsqueda de las palabras desconocidas en cuanto a su traducción (búsqueda en los diccionarios bilingües).



3.º Búsqueda de las palabras desconocidas en cuanto a su significado. Son casi siempre palabras técnicas que el traductor desconoce (búsqueda en diccionarios unilingües, enciclopedias, o incluso preguntar a expertos en el tema concreto objeto de la traducción).

4.º Se realiza una primera traducción (mental) con una tendencia literal. En realidad esto se hace a la vez que se lee el texto y se ven los problemas, pero este paso se termina cuando se tienen resueltos dichos problemas.



5.º Normalización sucesiva en el idioma de llegada de la primera traducción. Se intenta que el texto de llegada sea realmente español, francés, etcétera (que parezca un texto escrito directamente en español, francés, etcétera). (Este paso 5º se repite tantas veces como sea necesario).



6.º Se escribe el texto en el idioma de llegada.



7.º Se compara el texto final con el de partida hasta que se está seguro de que el sentido es el mismo y que la traducción es correcta. (Este paso 7º se repite tantas veces como sea necesario).

encerrado en la traducción porque queda implícito el contexto pero otras muchas veces esto no es así y la traducción queda mal.

c) Traducción de nivel 2: Es aquella en la que se maneja directamente el sentido del texto. Es por tanto la traducción del traductor, del experto. Tampoco es una traducción perfecta porque distintos traductores traducen un mismo texto de formas diferentes, con distinto vocabulario e incluso con distintas connotaciones, y esto se debe a que la traducción depende de la cultura, conocimiento, etc., del traductor, que siempre influyen.

### 3 TRADUCCIÓN AUTOMATIZADA

#### *La traducción del ordenador*

A primera vista se puede ver que por ordenador, al menos en la actualidad, no se puede conseguir una traducción como la del traductor. Para que fuera posible una traducción de este nivel (nivel 3) se tendría que representar de una forma unívoca todo lo que es y representa el sentido, es decir se tendría que representar de forma unívoca todos los conocimientos, ideas, conceptos, creencias, vivencias, etc., que el experto tiene en su cabeza. Por si esto fuera poco se tendría que establecer una forma de relacionar las dos lenguas en las que se traduce con este sentido. Como esto es imposible (o al menos un trabajo enorme) lo único que se puede conseguir es una traducción más o menos aproximada. Sería un sistema experto que ayudara al traductor, realizando una primera traducción que muchas veces sería correcta pero no siempre, por lo que requeriría una revisión por parte de éste. Como cualquier sistema experto, no sustituye al experto (que es lo que se ha pretendido muchas veces en la traducción

Figura 1. Pocosos seguido por el traductor en la traducción escrita



automática) sino que le ayuda en su trabajo y podría servir como sustituto eventual para traducciones muy simples.

### *Hasta dónde podemos llegar*

Las traducciones automáticas que se realizan actualmente se puede considerar del nivel 1. Tratan la semántica de las palabras aisladas (dentro de un tema) y la semántica de algunas frases o expresiones idiomáticas especiales. Por otro lado, manejan las reglas gramaticales para resolver problemas del tipo de: si una palabra es un verbo o un nombre, construcción de tiempos verbales y para la reconstrucción de las frases del texto en el segundo idioma (el del texto destino). Realmente no se puede llegar mucho más lejos, porque cómo se podría representar con los métodos actuales de representación del conocimiento (redes semánticas, marcos, guiones o reglas) todo el conocimiento, experiencia, creencias, etc., del traductor y además siendo este conocimiento y experiencia continuamente cambiante. Estos métodos de representación son sólo viables para porciones bien delimitadas de conocimiento y así se puede ver que los sistemas expertos funcionan mejor cuanto mayor sea su especialización.

Por todo esto podemos hacer una base de conocimiento para una traducción de nivel 1 y añadir algunas de las variables que intervienen en la captación del sentido del texto para conseguir una traducción más correcta pero quedándonos en la semántica, sin llegar al sentido.

### *Variables que se pueden tener en cuenta*

a) La unidad de verbalización que se refiere a una cadena de signos lingüísticos empleados en un acto de habla. Es una unidad con relación al sentido por lo que no se puede utilizar (como ya hemos dicho no podemos manejar el

sentido, sólo el significado, la semántica). Por esto en lugar de la unidad de la verbalización utilizaremos como unidad la oración.

b) La estructura lingüística de superficie que representa la estructura fono-morfo-léxico-semántica de la unidad de verbalización. No podemos manejarla como una unidad de verbalización sino como una unidad de frase. Representaría la estructura gramatical de la frase (en cuanto a sintagmas).

c) El potencial semántico que representa la significación de la estructura lingüística en el diccionario. No sólo los significados temáticos del diccionario para las palabras sino también el significado del diccionario de las frases y expresiones aunque consideradas fuera de todo contexto. Tendremos varios significados para una misma palabra dentro de un tema y esto vendrá dado por los sinónimos del significado de la palabra en el idioma.

d) Estructura rítmica, que representa los fenómenos prosódicos que se manifiestan mediante peculiaridades léxicas o sintácticas en la lengua escrita. Aunque la entonación puede transformar el sentido de un enunciado, como sólo tenemos en cuenta el significado tampoco tendremos en cuenta la estructura rítmica, ni el sistema ritmo.

e) El lenguaje del orador o lengua del texto original con sus sistemas de signos, reglas, sistema estructurado semántico (estructura semántica de la lengua).

f) La lengua del traductor que se refiere a la lengua del texto destino o final también con sus signos, reglas y sistema estructurado semántico.

g) Sentido pretendido, que se puede definir como la percepción que tiene el autor del texto origen. Es lo que quiere transmitir el autor del texto. Tampoco lo podemos tener en cuenta.

i) Los conocimientos, saber enciclopédico o



cultura general, es decir, los conocimientos que son conscientes o concienciables y en particular los saberes sobre el tema de traducción. Esto nos queda reducido en la base de conocimientos al significado de las palabras dentro de los distintos temas. Pero aun así es de gran utilidad en el resultado final porque el traductor no puede saber de todos los temas y siempre tiene que consultar enciclopedias u otros medios para conocer el tema de traducción y esto se lo resuelve el sistema (si está actualizado y tiene una amplia base de conocimientos).

*j)* Las creencias, las normas y las prácticas: Las creencias no se pueden incluir dentro de la base de conocimiento pero sí es posible incluir las normas y las prácticas que se verán reflejadas por un lado en expresiones hechas (ej., cómo se empieza un cuento «érase una vez ...») y por otro lado en significados de palabras (ej., cuando se habla de día de mala suerte en España se utiliza el martes 13 y en otros países el viernes 13).

*k)* La situación social que es un conjunto de relaciones sociales, la red de relaciones e intereses sociales sería también muy difícil de incluir.

*l)* El «mundo» o «mundillo» social al que pertenece la situación, se puede incluir en la base de conocimientos a diferentes argots.

*m)* El plano del tiempo histórico en el que se sitúa el texto. Podría incluirse pero sólo en el sentido de palabras que se han dejado de utilizar o palabras que van apareciendo. En definitiva las lenguas son cambiantes y el sistema tendría que irse ampliando en cuanto a nuevo vocabulario. Por otro lado, tendría que relacionar palabras ya no utilizadas en una lengua con sus nuevos equivalentes y no utilizar en la lengua destino términos en desuso aunque sí reconocerlos en la lengua origen.

### *Qué información vamos a introducir*

Por un lado tendremos una base de hechos que incluirá para cada palabra en cada idioma por separado:

*a)* Los temas en los que se encuentra y su significado asociados según estos temas.

*b)* Las frases hechas, metáforas y expresiones (en general locuciones) que incluyen la palabra y el significado asociado que tendrán de forma íntegra en el otro idioma.

*c)* El mundillo al que pertenece la palabra o si es una palabra general (de cualquier mundillo). Esto lo vamos a incluir dentro del tema no porque sea lo mismo sino porque, al igual que el tema, el significado nos cambia según este mundillo y nos va ser más fácil manejarlo así que incluirlo como una nueva variable aparte.

*d)* Habrá dentro de un tema distintos significados con distintos matices debido a los sinónimos de la significación de la palabra en el otro idioma, por lo que tendremos los sinónimos para cada palabra y habrá que indicar cuáles se usan más que otros (cuáles son más normales teniendo además en cuenta la variable tiempo).

*e)* Las clases sintácticas a las que puede pertenecer una palabra. Si puede actuar como verbo, nombre, adjetivo, etc.

Por otro lado, tenemos la base de reglas en la que se cuenta una serie de grupos de reglas totalmente independientes. Primero aplicamos reglas de un grupo y luego de otro sin que nos influyan de forma directa las anteriores. Por esto lo mejor es hacer distintas bases de reglas que el motor de inferencia evaluará de forma sucesiva y de forma independiente. Estas bases de reglas serían:

*a)* Reglas para extraer la información relacionada directa o indirectamente con el texto





Palabra	Factor de Utilización	Clase Sintáctica	Tema	Significado de la Palabra en el Tema	Locución / Traducción	Sinónimo
Pega	Nº	Nombre	Pájaros	Pie	—	—
Pega	Nº	Nombre	Peces	Rémora	—	—
Pega	Nº	Nombre	General	—	—	Crítica
Pega	Nº	Nombre	General	—	—	Chasco
Pega	Nº	Nombre	General	—	—	Pegamento
Pega	Nº	Nombre	General	—	—	Dificultad
Pega	Nº	Nombre	General	—	—	Engorro
Pega	Nº	Nombre	General	—	Poner una pega / Passer une colle	—
Pega	Nº	Nombre	General	—	Hay una pega / Il y a un os	—
Pega	Nº	Nombre	General	—	Asunto lleno de pegas / Affaire pleine d'anicroches	—
Pega	Nº	Nombre	General	—	De pega / À la gomme	—
Pega	Nº	Nombre	General	—	Esto es la pega / C'est ça le hic	—
Pega	Nº	Nombre	General	—	—	—
Pega	Nº	Verbo	General	—	—	Pegar
Pega	Nº	Verbo	General	—	Hay que ver la vida que se pega / Il faut voir la vie qu'il mène	—
Pega	Nº	Verbo	General	—	—	—

Figura 2: Ejemplo para la palabra «Pega».

que se pretende traducir. Esta información extraída será la base de hechos sobre la que se aplicarán las siguientes reglas (la base de hechos de trabajo).

b) Reglas para determinar la clase sintáctica a la que pertenece cada palabra. Reglas para ver si una palabra actúa como verbo, sujeto, complemento directo, complemento indirecto, etc. Reglas para dejar en la base de hechos de trabajo sólo la información relativa a las clases sintácticas de las palabras. Si, por ejemplo, determinamos que una palabra es un nombre, eliminamos de la base de hechos de trabajo todos los significados que están relacionados con esa palabra y son de otro tipo.

#### **Cómo se va representar esa información. Representación de la base de hechos**

La base de hechos se puede representar de distintas formas:

a) Una primera representación sería: *Palabra, Factor de Utilización, Clase Sintáctica, Tema, Significado de la Palabra en el Tema,*

*Locución, Traducción al francés de la locución, Sinónimo.* Entendiendo la clase sintáctica como por ejemplo nombre, verbo, adjetivo, determinante, etc.; factor de utilización como un valor que indica si la palabra se utiliza mucho, poco o si esta en desuso, y sinónimo no como sinónimo de la traducción sino como otra palabra equivalente en el mismo idioma que tendrá a su vez sus temas, locuciones, significados, etc. Esta forma de representación presenta la desventaja de que hay mucha información redundante y además nos será difícil extraer la información sobre el texto.

#### **Representación de los procedimientos o procesos de producción**

También existen varias formas de representar los procedimientos de producción para nuestro sistema experto de traducción:

a) Se pueden utilizar distintos grupos de reglas de producción que nos permiten ir eliminando la ambigüedad inicial de las palabras o, visto de otra forma, buscar la traducción más

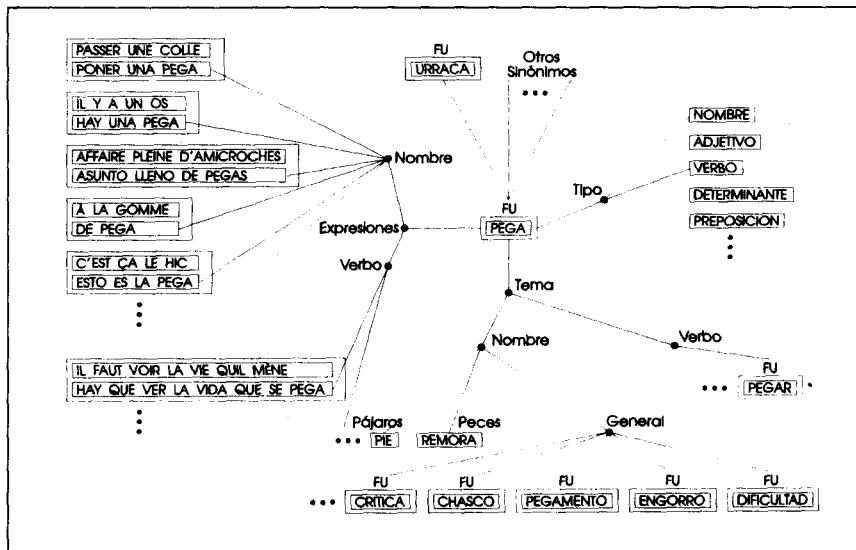


Figura 3. Ejemplo de representación utilizando una red semántica para la palabra «pega»

adecuada de las palabras (o expresiones y frases) dentro de las posibles que tenemos en la base de hechos. También tendríamos reglas para la reconstrucción del texto en el segundo idioma.

b) Se podrían también utilizar marcos que incluirán tanto a los hechos como a los procedimientos de producción. Pero no nos permitirían una división como la que haríamos por grupos de reglas, ni ir teniendo en cuenta sólo la información de la base de hechos que nos interesa (la del texto); habría que evaluarla en conjunto y esto sería menos eficiente. Por eso utilizaremos las reglas de producción.

#### Cómo se va a manejar la información

El proceso va a ser en un principio como una búsqueda del significado más adecuado, para cada palabra del texto, en la que se van aplicando las reglas de producción para ir eliminando significados hasta quedarnos con uno o varios

cuando la palabra se pueda dar en cualquier tema. No siempre es posible elegir un significado dentro de un conjunto de sinónimos para una palabra de manera unívoca, por lo que elegiremos en un primer lugar la palabra con un mayor factor de utilización pero guardando el resto e indicándole al traductor la posible sinonimia. Así permitimos al traductor elegir otra palabra, otro significado, si el que hemos puesto no es del todo correcto. Después, una vez que tenemos todos los significados, se aplica otro grupo de reglas de producción que permiten construir, por frases, el texto en el segundo idioma. Son los dos pasos generales de la traducción: captación del sentido (en nuestro caso del significado) y reconstrucción del texto.

Por todo esto, vamos a tener una memoria de trabajo en la que incluiremos en un principio toda la información de nuestra base de hechos relativa al texto y luego vamos a ir reduciendo la información de la base de hechos de



trabajo hasta no tener ambigüedades. Así vamos a manejar la base de hechos de trabajo para ir reduciéndola lo más posible y hacer que el proceso sea más eficaz.

Por otro lado las reglas de producción también estarán divididas. Tenemos una serie de reglas de producción que son totalmente independientes por grupos y así por ejemplo tendremos reglas para determinar el tema y reglas para la construcción de verbos que no dependen de las primeras. Como son independientes lo mejor es tenerlas separadas reduciendo así el número de reglas que el motor de inferencia tendrá que evaluar en un momento dado y conseguir así una mayor eficiencia.

#### ***Una posible introducción al manejo y validación de los factores de utilización mediante redes neuronales***

Los factores de utilización de las palabras presentan el inconveniente de que al igual que la lengua cambian constantemente y por otro lado es muy difícil determinar, así de entrada, cuál será su valor más adecuado para conseguir que el sistema funcione correctamente.

Una posible forma de manejar estos factores de utilización sería mediante una serie de redes neuronales controladas de forma dinámica. No se trata de una red permanente y para todas las palabras, pues sería inviable.

Lo que se pretende es que cada vez que el sistema traduzca se construya una red neuronal de forma dinámica para que ajuste los factores de utilización teniendo en cuenta los posibles retoques (de palabras nunca de locuciones) que haga el traductor en la revisión del texto. Esta red neuronal sólo tendrá en cuenta la subred (semántica) obtenida para el texto. Esta división de la información (teniendo en cuenta la subred) es correcta puesto que en

realidad cada factor de utilización de una palabra sólo depende directamente de la diferencia de utilización respecto a las palabras que son sus sinónimos para un contexto concreto. El factor de utilización mide por tanto lo que se utiliza una palabra en relación a sus sinónimos directos para todos los temas (su valor dependerá de los temas más frecuentes que traduzca el sistema).

Las características de esta red en cuanto a sus nodos de entrada, salida y enlaces depende en cada caso del tamaño del texto y de su riqueza léxica, en definitiva de la subred semántica que hemos obtenido para dicho texto. Así, la entrada estaría constituida por todas las palabras que el sistema tiene almacenadas en la subred (palabras relacionada directamente con el texto).

Las salidas serían las palabras que constituyen el texto final traducido. Así, la red neuronal cambiará sus pesos, que en este caso son los propios factores de utilización para hacer que los sinónimos escogidos sean los elegidos por el traductor. Hay que dejar claro que la red neuronal no traduce en absoluto, no es que se le dé el texto de entrada y nos proporcione la traducción; lo único que realiza es ajustar los factores de utilización para la elección de la traducción correcta en función de las posibles traducciones según los distintos sinónimos (que en definitiva es la información que tenemos en la subred semántica). Por tanto, la red sólo trabaja con los posibles sinónimos como entrada y los sinónimos más adecuados como salida.

El objetivo de la red es doble, por un lado, mediante un entrenamiento, conseguir que se ajusten los factores de utilización de la forma más adecuada posible y, por otro lado, en la primera programación hacia adelante, antes de



empezar el posible entrenamiento, conseguir la salida de la parte del sistema que trata los sinónimos.

La red, aunque se crea de una forma dinámica, en realidad está ya almacenada en cuanto a que tenemos los pesos (factores de utilización). Esto nos lleva a un tipo de red con dos capas: una de entrada con sus pesos asociados y una capa de salida sin pesos. La red tiene que ser así porque, si queremos añadir por ejemplo una capa intermedia, tendremos también que añadir nuevos pesos no existentes cuyos valores iniciales nos influirán en el resultado y, de todas formas, sería muy difícil almacenar sus valores y utilizarlos posteriormente para otras redes distintas en otras traducciones en las que deberían también influir.

La utilización de esta red neuronal nos lleva a un sistema que depende de su experiencia, de forma que traducirá mejor cuanto más haya traducido y que incluso se puede ir especializando en un tipo de traducción concreta (en cuanto a los temas). Además este sistema tiene la ventaja de que la validación del factor de utilización es mucho más sencilla que si tuviéramos que revisar y escribir de forma manual todos los factores de utilización.

#### 4 CONSIDERACIONES FINALES SOBRE EL SISTEMA EXPERTO

a) Se tratará de un sistema experto en el que se tiene una entrada al comienzo y no se establece un diálogo continuo con el usuario. Se tiene que capturar toda la información del texto de entrada. Se trata por tanto de un tipo de sistema experto denominado «sistema experto en modo *bath*». Aunque se comprobarán los resultados y se irán añadiendo nuevos términos o expresiones al final del proceso

puediéndose establecer entonces un diálogo (pero al final del proceso).

De este tipo de sistema (sistemas en modo *bath*), aunque no es muy infrecuente, hay poca información pero en la práctica los problemas asociados a su construcción son un subconjunto de los que aparecen en los sistemas expertos interactivos.

b) Se trata de un sistema con una base de conocimiento fragmentada pues tanto la base de hechos como la base de reglas están divididas. En cuanto a la base de hechos, lo primero que se hace es dividirla para tener en cuenta sólo la información sobre el texto. Si esto no fuera así el problema se haría intratable. La base de reglas también se divide en grupos de reglas y esto es así porque aunque en principio se podrían tratar todas las reglas en conjunto (no son tantas), éstas son independientes por grupos y siempre es más eficiente el separarlas.

c) El sistema maneja factores de certeza tanto en las reglas como en los hechos. A estos últimos los hemos llamado factor de utilización de las palabras y se manejan mediante una red neuronal.

d) La base de hechos que manejaremos está formada como ya hemos visto por una red semántica que contiene la información sobre el texto.

e) Los procedimientos de producción vienen dados por reglas de producción.

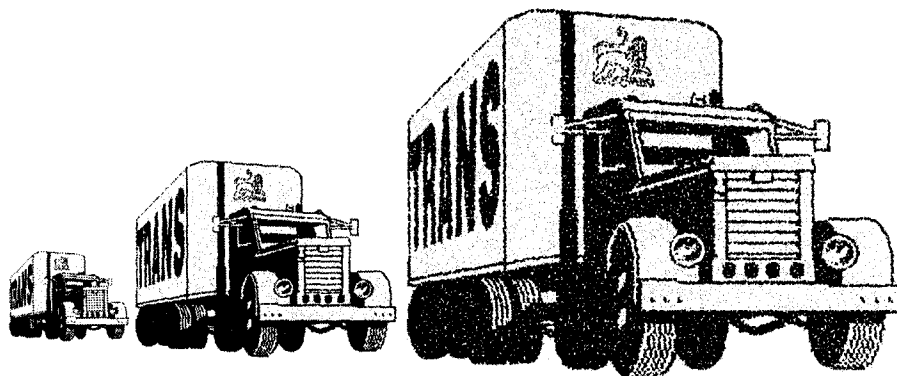
#### 5 FUENTES BIBLIOGRÁFICAS

- Alty, J. L. y M. J. Coombs: *Sistemas expertos: Conceptos y ejemplos*, Madrid: Díaz de Santos, 1986.
- Belkin, N. J., H. M. Brooks y P. J. Daniels: *Knowledge Elicitation Using Discourse Analysis*, Oxford: Research Studies Press, 1988.



- Freeman, James A. y David M. Skapura: *Redes neuronales: Algoritmos, aplicaciones y técnicas de programación*, Madrid: Addison-Wesley/Díaz de Santos, 1993.
- Frost, R.: *Bases de datos y sistemas expertos: Ingeniería del conocimiento*, Madrid: Díaz de Santos, 1984.
- García, M. y F. Mas Martín: *Elementos de sintaxis comparada francesa y española*, Sucesores de Nogués, 1986.
- García Landa, M. y T. Séz Hermosilla: *Teoría de la traducción*, Salamanca, Bruselas, 1994.
- García Yebra, V.: *Teoría y práctica de la traducción*, Madrid: Gredos, 1982.
- Garg-Janrdan, Chaya y Gariel Salvendy: *A Structured Knowledge Elicitation Methodology for Building Expert Systems*, Oxford: Research Studies Press, 1990.
- Gruber, Thomas R.: *Acquiring Strategic Knowledge for Experts*, Oxford: Research Studies Press, 1991.
- Harmon, P. y D. King: *Sistemas expertos: Aplicaciones de la inteligencia artificial en la actividad empresarial*, Madrid: Díaz de Santos, 1988.
- Klinger, G., S. Genetet y J. MacDermot: *Knowledge Acquisition for Evaluating Systems*, Oxford: Research Studies Press, 1988.
- Martínez, A.: *Inteligencia artificial (La gran guía): De los sistemas expertos a las redes neuronales*, Madrid: Jackson, 1991.
- Mira, J.: *Approches to Knowledge Representation: An introduction*, Oxford: Research Studies Press, 1988.
- Nilsson, Nils J.: *Principios de inteligencia artificial*, Madrid: Díaz de Santos, 1987.
- Rich, E.: *Inteligencia artificial*, Barcelona: Gustavo Gili, 1988.
- Roca Pons, J.: *Introducción a la gramática*, Hospitalet (Barcelona): Teide, 1970.
- Sáez Hermosilla, T.: *El sentido de la traducción: Reflexión y crítica*, Salamanca: Universidad de Salamanca y Universidad de León, 1994.
- Witten, Ian H. y Bruce A. MacDonald: *Using Concept Learning for Knowledge Acquisition*, Oxford: Research Studies Press, 1990.

RECIBIDO EN MARZO DE 1996





TRANS · Nº 1 · 1996

This article offers an overview of the range of factors and characteristics which must be considered in the analysis of the expression of intersentential relations in authentic texts. The study is based on two exhaustive corpus surveys of newspaper editorials, one in English and one in Spanish (although references within the text itself are to the English corpus alone). A preliminary discussion is dedicated to general considerations in the analysis and definition of intersentential relations. This is followed by an outline of the different classes of relations which are involved.

# The explicit signalling of intersentential relations in authentic texts: An overview

JOHN HYDE  
*Universidad de Salamanca*

*Este artículo pasa revista a los factores y características que deben ser tomados en cuenta en el análisis de la expresión de la relaciones intraoracionales en textos auténticos. El estudio se basa en dos corpus exhaustivos de editoriales de periódicos, uno en inglés y otro en español (aunque las referencias en el texto se hacen sólo al corpus en inglés). En primer lugar se hacen consideraciones generales sobre el análisis y definición de las relaciones intraoracionales, y a continuación se exponen los distintos tipos de relaciones que se manejan.*