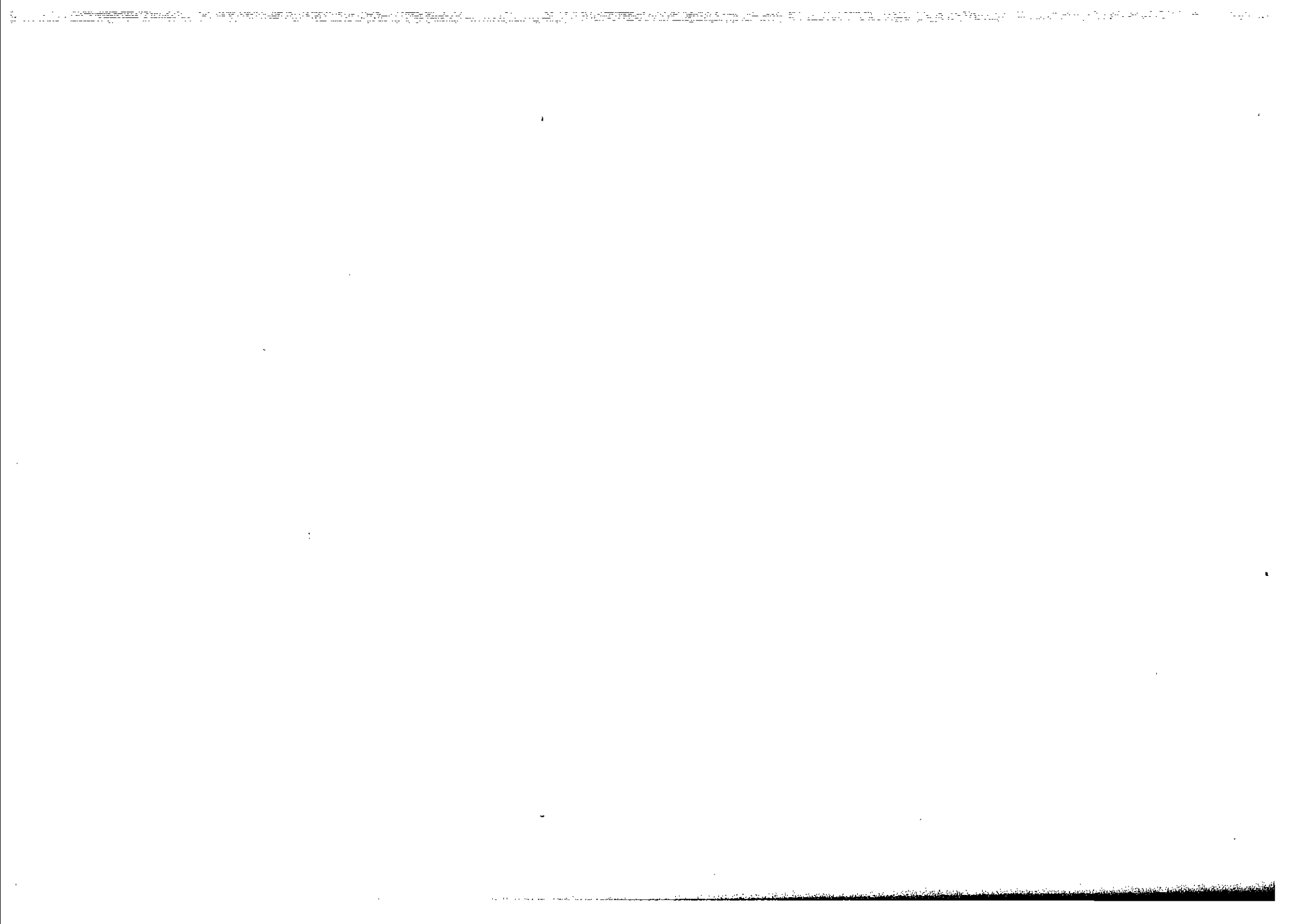


5 - TRADUCCIÓN AUTOMÁTICA



JULIETTA: UN PROTOTIPO DE TA BASADO EN LFG

Gabriel Amores

Departamento de Lengua Inglesa
Universidad de Sevilla

1. Introducción

Se presenta aquí una fase más avanzada del sistema descrito anteriormente en Amores (1991). En aquel trabajo estudiamos la viabilidad de usar la Gramática Léxico Funcional (LFG) de Bresnan y Kaplan (Bresnan ed. 1982) en el contexto de la traducción automática (TA).

La traducción automática se ha convertido en un área prioritaria de investigación y discusión en los últimos cuatro decenios. La tarea de crear sistemas para mecanizar el proceso de traducción se ha perseguido desde una amplia variedad de enfoques, a veces casi contradictorios entre sí. Encontramos así grupos de investigación que tratan de desarrollar sistemas basados en enfoques estadísticos, sin hacer uso de información lingüística (Brown et al. 1990), frente a otros que buscan una *interlingua* universal (Goodman y Nirenburg eds. 1991).

Sin embargo, un análisis más general de los últimos adelantos apunta hacia una clasificación de los sistemas de TA en dos categorías, según las motivaciones que condujeron a su creación.

Por un lado, encontramos sistemas creados con el fin de satisfacer necesidades específicas de traducción. La experiencia indica que cuando se definen claramente el entorno de uso y los criterios de diseño que de ellos se derivan, la TA puede tener gran éxito. Podrían citarse como ejemplos el sistema ENGSPAN/SPANAM de la Organización Panamericana de la Salud, en Washington, D.C. (Vasconcellos y León 1988) o Meteo (Chandioux 1988).

Este enfoque demuestra que la TA es útil, que puede ser efectiva en función de los costes, y que puede producir resultados más que aceptables que satisfacen a los clientes en general. Sin embargo, y a pesar de logros tan sobresalientes, los sistemas que pertenecen a esta categoría pueden ser objeto de crítica en dos niveles: tienden a presentar una falta de modularidad en la separación de los algoritmos y los datos lingüísticos, y no están motivados lingüísticamente.¹

Por otro lado, encontramos sistemas motivados lingüísticamente, primordialmente dirigidos a la resolución elegante de problemas lingüísticos y de traducción. EUROTRA (Allegranza et al. 1991) y sus derivados CAT y MiMo (Sharp 1988 y van Noord et al. 1991) y UNITRAN (Dorr 1990) podrían citarse como ejemplos de sistemas pertenecientes a esta categoría. Estos sistemas proporcionan soluciones eficaces y elegantes a muchos problemas difíciles de traducción, pero se presume que no pueden llegar a ser operativos.

El prototipo de TA que presentamos aquí intenta demostrar que la adopción de una teoría lingüística sólida y con una orientación computacional, concretamente LFG, ofrece un punto de partida excelente para el desarrollo de sistemas de TA que sean operativos y estén motivados lingüísticamente.

Este trabajo va más allá de propuestas anteriores que persiguieron la misma idea. Frey y Reyle (1983), Eisele y Dörre (1986), Kudo y Nomura (1986), Netter y Wedekind (1986), Momma y Dörre (1987), Kaplan et al. (1989), Sadler et al. (1990), Zajac (1990) y Ruiz et al. (1991) implementaron prototipos de limitada cobertura lingüística para demostrar la viabilidad del empleo de LFG en el contexto del procesamiento del lenguaje natural o TA.

¹ "Motivados lingüísticamente" se refiere aquí a sistemas que adoptan una teoría lingüística específica.

Sus trabajos dejaron abierta la cuestión fundamental de si un sistema basado en LFG respondería igual de bien al tratar con textos reales.

2. Descripción general del sistema

JULIETTA sigue un enfoque de transferencia, el cual encaja perfectamente con los dos niveles de representación propuestos en LFG (Netter 1986 y contra Zajac 1990).

Mientras que la estructura-c incluye información específica al idioma, y es, por tanto, descartada durante la transferencia, las relaciones gramaticales que encontramos en la estructura-f proporcionan una información más abstracta, ideal para ser manipulada durante la transferencia.

El sistema genera estructuras-c y -f para cada oración en lengua fuente. La transferencia se efectúa desde la estructura-f de la lengua fuente, y produce una estructura-f en lengua destino, desde la cual se genera su correspondiente estructura-c. Este enfoque sigue las líneas propuestas por Kudo y Nomura (1986) y contra Kaplan et al. (1989).

Como ejemplo ilustrativo, presentamos a continuación el resultado de los tres módulos del sistema para una frase simple.

English Input

The boy ate a cake.

English c-structure

```
s(clh(cl(np(detp(det2(det(the))),
          n2(n1(n(boy))))),
      vph(vp(vg(v(ate))),
          np(detp(det2(det(a))),
              n2(n1(n(cake))))))))
```

English f-structure

```
pred:eat ([subj,obj])
num:sing
vtype:action
tense:past
obj: pred:cake
      cake isa edible
      spec:a
      count:yes
      role:theme
      agr:num:sing
subj:  pred:boy
      boy isa animate
      spec:the
      count:yes
      role:ag
      agr:num:sing
```

Spanish f-structure

```
pred:comer ([subj,obj])
tense:past
vtype:action
num:sing
refl:yes
subj:  pred:niño
      spec:el
      count:yes
      agr: gen:masc
      num:sing
obj:  pred:pastel
      spec:un
      count:yes
      agr: gen:masc
      num:sing
```

Spanish c-structure

```
o(prop(snh(sn(sres(res(el)),
              stbar(st(niño))))),
   sv(vbg(clt(se))
       vb(comió)),
   snh(sn(sres(res(un)),
          stbar(st(pastel))))))
```

JULIETTA: UN PROTOTIPO DE TA BASADO EN LFG

Gabriel Amores

Departamento de Lengua Inglesa
Universidad de Sevilla

1. Introducción

Se presenta aquí una fase más avanzada del sistema descrito anteriormente en Amores (1991). En aquel trabajo estudiamos la viabilidad de usar la Gramática Léxico Funcional (LFG) de Bresnan y Kaplan (Bresnan ed. 1982) en el contexto de la traducción automática (TA).

La traducción automática se ha convertido en un área prioritaria de investigación y discusión en los últimos cuatro decenios. La tarea de crear sistemas para mecanizar el proceso de traducción se ha perseguido desde una amplia variedad de enfoques, a veces casi contradictorios entre sí. Encontramos así grupos de investigación que tratan de desarrollar sistemas basados en enfoques estadísticos, sin hacer uso de información lingüística (Brown et al. 1990), frente a otros que buscan una *interlingua* universal (Goodman y Nirenburg eds. 1991).

Sin embargo, un análisis más general de los últimos adelantos apunta hacia una clasificación de los sistemas de TA en dos categorías, según las motivaciones que condujeron a su creación.

Por un lado, encontramos sistemas creados con el fin de satisfacer necesidades específicas de traducción. La experiencia indica que cuando se definen claramente el entorno de uso y los criterios de diseño que de ellos se derivan, la TA puede tener gran éxito. Podrían citarse como ejemplos el sistema ENGSPAN/SPANAM de la Organización Panamericana de la Salud, en Washington, D.C. (Vasconcellos y León 1988) o Meteo (Chandioux 1988).

Este enfoque demuestra que la TA es útil, que puede ser efectiva en función de los costes, y que puede producir resultados más que aceptables que satisfacen a los clientes en general. Sin embargo, y a pesar de logros tan sobresalientes, los sistemas que pertenecen a esta categoría pueden ser objeto de crítica en dos niveles: tienden a presentar una falta de modularidad en la separación de los algoritmos y los datos lingüísticos, y no están motivados lingüísticamente.¹

Por otro lado, encontramos sistemas motivados lingüísticamente, primordialmente dirigidos a la resolución elegante de problemas lingüísticos y de traducción. EUROTRA (Allegranza et al. 1991) y sus derivados CAT y MiMo (Sharp 1988 y van Noord et al. 1991) y UNITRAN (Dorr 1990) podrían citarse como ejemplos de sistemas pertenecientes a esta categoría. Estos sistemas proporcionan soluciones eficaces y elegantes a muchos problemas difíciles de traducción, pero se presume que no pueden llegar a ser operativos.

El prototipo de TA que presentamos aquí intenta demostrar que la adopción de una teoría lingüística sólida y con una orientación computacional, concretamente LFG, ofrece un punto de partida excelente para el desarrollo de sistemas de TA que sean operativos y estén motivados lingüísticamente.

Este trabajo va más allá de propuestas anteriores que persiguieron la misma idea. Frey y Reyle (1983), Eisele y Dörre (1986), Kudo y Nomura (1986), Netter y Wedekind (1986), Momma y Dörre (1987), Kaplan et al. (1989), Sadler et al. (1990), Zajac (1990) y Ruiz et al. (1991) implementaron prototipos de limitada cobertura lingüística para demostrar la viabilidad del empleo de LFG en el contexto del procesamiento del lenguaje natural o TA.

¹ "Motivados lingüísticamente" se refiere aquí a sistemas que adoptan una teoría lingüística específica.

Sus trabajos dejaron abierta la cuestión fundamental de si un sistema basado en LFG respondería igual de bien al tratar con textos reales.

2. Descripción general del sistema

JULIETTA sigue un enfoque de transferencia, el cual encaja perfectamente con los dos niveles de representación propuestos en LFG (Netter 1986 y contra Zajac 1990).

Mientras que la estructura-c incluye información específica al idioma, y es, por tanto, descartada durante la transferencia, las relaciones gramaticales que encontramos en la estructura-f proporcionan una información más abstracta, ideal para ser manipulada durante la transferencia.

El sistema genera estructuras-c y -f para cada oración en lengua fuente. La transferencia se efectúa desde la estructura-f de la lengua fuente, y produce una estructura-f en lengua destino, desde la cual se genera su correspondiente estructura-c. Este enfoque sigue las líneas propuestas por Kudo y Nomura (1986) y contra Kaplan et al. (1989).

Como ejemplo ilustrativo, presentamos a continuación el resultado de los tres módulos del sistema para una frase simple.

English Input

The boy ate a cake.

English c-structure

```
s(clh(cl(np(detp(det2(det(the))),
          n2(n1(n(boy))))),
      vph(vp(vg(v(ate)),
          np(detp(det2(det(a))),
              n2(n1(n(cake))))))))))
```

English f-structure

```
pred:eat([subj,obj])
num:sing
vtype:action
tense:past
obj: pred:cake
      cake isa edible
      spec:a
      count:yes
      role:theme
      agr:num:sing
subj:  pred:boy
      boy isa animate
      spec:the
      count:yes
      role:ag
      agr:num:sing
```

Spanish f-structure

```
pred:comer([subj,obj])
tense:past
vtype:action
num:sing
refl:yes
subj:  pred:niño
      spec:el
      count:yes
      agr: gen:masc
      num:sing
obj:  pred:pastel
      spec:un
      count:yes
      agr: gen:masc
      num:sing
```

Spanish c-structure

```
o(prop(snh(sn(sres(res(el)),
              stbar(st(niño))))),
   sv(vbg(clt(se))
      vb(comió)),
   snh(sn(sres(res(un)),
         stbar(st(pastel))))))
```

Spanish output

El niño se comió un pastel.

3. Presupuestos teóricos

La principal motivación de esta fase del trabajo es estudiar los límites de la aplicación de la teoría LFG en la traducción automática de textos reales. Describimos aquí los resultados obtenidos con resúmenes de medicina. Estos textos son similares a los que se envían habitualmente a ser traducidos mecánicamente y presentan cierta consistencia estructural, lo que permite centrarse en cuestiones de representación y calidad de la traducción.

Asimismo, pretendemos demostrar que la adopción de un dominio semántico restringido como es la medicina no debe limitar la posible aplicación del sistema a otros campos semánticos. Para ello sugerimos una nueva estrategia de TA que está a medio camino entre la basada en el conocimiento (Goodman y Nirenburg eds. 1991) y la sintáctico-semántica, propia de los actuales sistemas de 2ª generación (Slocum ed. 1988).

4. Estrategia de TA

Es bien conocido que en la TA se necesita información de tipo extralingüístico, o referente al mundo o al dominio semántico, para desambiguar y definir relaciones de dependencia que han de ser transferidas a la lengua destino.

Estos argumentos han dado lugar a propuestas de representaciones más abstractas, de tipo semántico-conceptual en TA. No obstante, desde un punto de vista práctico y real, casi todos los problemas de traducción entre dos lenguas se pueden resolver siguiendo un enfoque de transferencia.

La estrategia que hemos seguido en JULIETTA pretende integrar ambas propuestas. Mantenemos una representación de tipo sintáctico-funcional, pero hacemos uso de información de tipo extralingüístico para la correcta desambiguación de estructuras.

Este enfoque apunta hacia una concepción multiforme de la TA. Se defiende aquí que un sistema de TA de ámbito general puede constar de un módulo sintáctico principal basado en LFG, enriquecido con redes semánticas satélite que faciliten información específica referente al campo semántico. Esta información semántica está disponible durante el análisis, la transferencia y la generación, con independencia del diccionario, como se detallará en la siguiente sección.

Las diferentes bases de datos semánticas podrían actuar como interruptores que se activarían según el tipo de texto. En el desarrollo de sistemas de TA para dominios semánticos específicos, la separación entre la información sintáctica y semántica proporciona un marco más flexible que el propuesto por sistemas diseñados para sublenguajes específicos (Sager et al. 1987). Así, es posible crear sistemas para otros dominios semánticos en un espacio de tiempo reducido, actualizando o redefiniendo la red semántica y sin necesidad de modificar el diccionario fuente.

La ventaja de este enfoque reside en su modularidad. Hasta ahora, la técnica habitual en TA había sido incluir en la entrada del diccionario marcas semánticas tomadas de un conjunto finito de rasgos (Bennett 1990). Pensamos que ese tipo de información es útil, pero no suficiente, sobre todo en casos de ambigüedad estructural.

5. Interacción sintáctico semántica durante el análisis

No hay espacio aquí para describir con detalle la cobertura lingüística del sistema. Baste decir que la versión actual puede tratar todas las estructuras que se reflejan en la sección de Resultados. Me detendré solamente a continuación en cómo interacciona el diccionario del sistema con la jerarquía semántica.

En JULIETTA el diccionario es independiente del campo semántico al que pertenezca el texto fuente. El componente sintáctico es autónomo, y solicita información del componente semántico en casos de ambigüedad. Este tiene forma de red semántica, que se actualiza paralelamente al diccionario. La estrategia que seguimos consiste en asumir que hay estructuras en inglés que dan lugar a ambigüedad, y que han de ser tratadas teniendo en cuenta el conocimiento extralingüístico.

Consideremos por ejemplo las secuencias Adjetivo + Nombre + Nombre en inglés, como **living donor transplantation** o **systolic blood pressure**, y las siguientes reglas de la gramática para ver un ejemplo de cómo interaccionan los distintos tipos de información.

- (1) N2 → AdjP N1
- (2) AdjP → Adj
- (3) N1 → N N
- (4) N1 → N

El problema estriba en decidir cómo se organiza la premodificación nominal, es decir, si estamos ante **living (donor transplantation)** o **(living donor) transplantation** y **systolic (blood pressure)** o **(systolic blood) pressure**.

Dado que los procesos de retroceso (backtracking) son costosos en Prolog, hemos optado por una estrategia de dilación (delay) con prioridad de extensión (breadth-first). El analizador consume la secuencia completa, que genera por defecto una estructura-f en la que el adjetivo modifica al grupo nominal. A continuación, se consulta la rutina `semantically_related/2`, que se encarga de confirmar la agrupación propuesta o devolver la estructura-f reorganizada.

El predicado opera de izquierda a derecha, buscando relaciones semánticas entre el adjetivo y los sucesivos nombres, o entre el primer nombre y el segundo. En la presente versión se han definido tres posibilidades para dar cuenta del corpus analizado.

```
semantically_related(Item1, Item2) :-
    Item1 property_of Item2;           % blood pressure
    obj_nominalization(Item2, Item1); % health care
    subj_nominalization(Item2, Item1). % cell decomposition
```

Centrándonos en el primer subpredicado (`Item1 property_of Item2`), el programa consulta una versión adaptada de la red semántica UMLS (McCray 1989), que contiene información acerca del dominio de la medicina en forma de jerarquía `isa (es_un)`. La red semántica nos dice que

```
Item1 property_of Item2 :-
    Item1 isa body_attribute,
    Item2 isa organism.
```

Esta relación es verdadera entre **living** (`living isa body_attribute`) y **donor** (`donor isa organism`), pero no entre **donor** y **transplantation** o entre **living** y **transplantation**, con lo que la secuencia queda desambiguada. La misma técnica se utiliza para desambiguar **systolic blood pressure**, ya que **pressure** y **blood** están relacionados semánticamente.

Las estructuras-f resultantes para **living donor transplantation** y **systolic blood pressure** aparecen a continuación. El signo '+' indica que **living** está modificando a **donor** y que **blood** está modificando a **pressure**.

```
pred:transplantation          pred:blood + pressure
vnom:yes                      agr:num:sing
count:yes                     mods:pred:systolic
agr:num:sing                  temp:no
descr:  pred:living + donor
count:yes
agr:num:sing
```


Una vez que la secuencia ha sido desambiguada, el componente de transferencia puede definir reglas para obtener la traducción deseada: 'transplante de donantes vivos' y 'presión sanguínea sistólica'.

Como hemos indicado, la información de tipo semántico se codifica en la jerarquía *isa* paralelamente al diccionario. La única excepción son algunos verbos, que incluyen información acerca del tipo de sujeto y objeto que seleccionan. Pensamos que, excepto en casos de uso figurado del lenguaje, este tipo de información es de tipo más universal e independiente del campo semántico. Esta interacción se consigue mediante la ampliación del proceso de unificación de manera que permita la inclusión de relaciones *isa*. Así, por ejemplo, la entrada del verbo *eat* quedaría como sigue:

```
dic(v, eat, [pred:eat([subj,obj]), vtype:action,
            subj:[pred:Pred1, role:ag, Pred1 isa animate],
            obj:[pred:Pred2, role:theme, Pred2 isa edible]]).
```

Esta entrada indica que el verbo *eat* toma un sujeto animado y un objeto comestible.

6. Algunos problemas de transferencia

6.1 Transferencia léxica accionada por información semántica

El componente de transferencia también hace uso de la red semántica para la selección léxica. Por ejemplo, el verbo *raise* en inglés se transfiere de diversas maneras al español dependiendo unas veces de cuál es el núcleo del objeto (*flag*, *money* o *issue*), o de sus rasgos semánticos (+animado o +vegetal).

<i>raise</i> + <i>flag</i>	=>	izar una bandera
<i>raise</i> + <i>money</i>	=>	conseguir dinero
<i>raise</i> + <i>issue</i>	=>	plantear una cuestión
<i>raise</i> + [+plants] <i>cotton</i>	=>	cultivar algodón
<i>raise</i> + [+animal] <i>dogs</i>	=>	criar perros

Las reglas de transferencia tienen la siguiente forma:

```
([pred:raise([subj,obj]), obj:Obj]
  <> (pred:P, P isa plant)
  => ([pred:cultivar([subj,obj]))).

([pred:raise([subj,obj]), obj:Obj]
  <> (pred:P, P isa animal)
  => ([pred:criar([subj,obj]))).
```

En la versión anterior de este prototipo ya demostré que la inclusión de información de roles temáticos sobre los argumentos del verbo en la estructura-f simplifica la transferencia estructural y contribuye a generar traducciones naturales y elegantes.

6.2 Transferencia de fenómenos de 'head shifting'

El sistema es también capaz de dar cuenta de otros casos difíciles de traducción entre el inglés y el español. Son ejemplos típicos los fenómenos de 'head shifting' como *just* y *usually* donde un núcleo de cláusula en inglés es degradado a XCOMP en español y un ADJUNCT es promovido a núcleo de cláusula. Dichos problemas pueden resolverse por medio de reglas de transferencia estructural que hacen uso de la información disponible en la estructura-f.

Ejemplos

The mail just arrived => El correo acaba de llegar
 The mail has just arrived => El correo acaba de llegar
 The mail had just arrived => El correo acababa de llegar
 The boys usually sleep in the house => Los niños suelen
 dormir en la casa

The mail just arrived vs. El correo acaba de llegar

pred:arrive ([subj])	pred:acabar ([vcomp])
num:sing	vtype:process
tense:past	tense:pres
vtype:process	num:sing
adj:type:subj	vcomp:pred:llegar ([subj])
form:just	inf:de
subj:pred:mail	subj:pred:correo
count:no	count:no
spec:the	spec:el
agr:num:sing	agr:gen:masc
	num:sing

6.3 Transferencia de verbos tipo like

Un segundo ejemplo de traducción difícil lo constituyen aquellos verbos que asignan funciones diversas a los mismos argumentos en inglés y en español, como es el caso de *like/gustar*. La traducción de estos verbos suele plantear problemas a algunos sistemas basados en transferencia, teniendo que recurrir a procedimientos *ad hoc*. JULIETTA basa su estrategia para transferir estos verbos en la propuesta de sujetos no nominativos en LFG (Levin 1985) y en la existencia de una jerarquía de roles temáticos (Kiparsky 1987).

Se defiende aquí que a Pedro en *A Pedro le gustan los pasteles* presenta rasgos de sujeto. Estos rasgos son: 1) admite elevación de sujeto (*A Pedro parece gustarle el vino*) y 2) al serle asignado el rol temático de *experimentador* a *a Pedro* y de tema a los *pasteles*, el *experimentador* tiene prevalencia en la asignación de la función sujeto frente al tema. Por otra parte, interpretar *a Pedro* como *frontalización* del tópico plantea también problemas ya que habría que explicar como anormal una construcción intuitivamente más natural en español que la que no presentara *topicalización* (??*los pasteles le gustan a Pedro*). Obviamente, la falta de concordancia sujeto dativo / verbo principal plantea problemas en principio a nuestra interpretación. Sin embargo, pensamos que la definición de sujeto ha de ser más amplia que la basada *solamente* en la concordancia del sintagma nominal inmediatamente dominado por la oración con el verbo principal. Nuestra explicación es solamente una hipótesis de trabajo, pero ofrece grandes ventajas en el tratamiento de estas construcciones entre inglés y español.

Esta estrategia proporciona una solución elegante, que hace innecesario recurrir a una representación más abstracta, de tipo interlingua² para dar cuenta de tales diferencias. Sólo es necesaria una regla de transferencia como la siguiente

```

- ([pred:like ([subj, obj]), subj:S, num:N]) &&& (subj:S) =>
  ([pred:gustar ([dsubj, obj]),
   dsubj:[pcase:a],
   obj:[def:yes, agr:[num:Num]],
   num:Num, clt_doubling:yes]).

```

² Según defiende Dorr (1990).

La regla dice que el verbo *like* se transfiere como *gustar* en español, pero el sujeto pasa a ser un sujeto dativo en español (dsubj), que será generado como sintagma preposicional en caso de sujeto no pronominal en inglés; el verbo pasa a concordar con el objeto, que ha de ser definido para insertar un artículo en español, y, finalmente, el verbo *gustar* en español presenta 'clitic doubling'. Las estructuras-f para el inglés y el español quedan como sigue

Peter likes cakes vs. A Pedro le gustan los pasteles

```

pred:like([subj,obj])
num:sing
tense:pres
vtype:state
obj:pred:cake
  count:yes
  role:theme
  agr:num:plur
subj:pred:peter
  peter isa animate
  role:exp
  agr:num:sing

```

```

pred:gustar([dsubj,obj])
vtype:state
tense:pres
clt_doubling:yes
num:plur
dsubj:pred:'Pedro'
  pcase:a
  agr:num:sing
obj:pred:pastel
  count:yes
  def:yes
  spec:el
  agr:gen:masc
  num:plur

```

7. Inserción de artículos en español

El inglés y el español difieren en el uso de los artículos. La correcta inserción de artículos suele ser un problema complejo para cualquier sistema de TA. En JULIETTA la decisión sobre inserción de artículos se realiza en la última fase de la transferencia, tras la transferencia estructural y léxica, y se basa en cuatro criterios. Estos criterios son restrictivos, es decir, solamente se insertará un artículo cuando estemos absolutamente seguros de ello, para evitar la sobregeneración.

7.1 Sujetos

Como hipótesis de trabajo estimamos que los sujetos en español han de ser definidos [+definite]. Esta fase dio como resultado las siguientes inserciones.

Replies received from 27 transplant centres representing 17 health regions gave data on more than 1200 transplants from living donors.

Las respuestas recibidas de 27 centros de transplante que representaban 17 regiones de salud ofrecieron datos sobre más de 1200 trasplantes de donantes vivos.

Transplants from living donors accounted for 0.25 % of the total experience of health regions.

Los trasplantes de donantes vivos representaron el 0.25% de la experiencia total de las regiones de salud.

In national samples of 9921 10-year-olds and 3259 adults in Britain systolic blood pressure was inversely related to birth weight.

En muestras nacionales de 9921 niños de 10 años y 3259 adultos en Gran Bretaña la presión sanguínea sistólica estaba inversamente relacionada con el peso al nacer.

7.2 Postmodificadores

La fase de análisis ha determinado una representación distinta para postmodificadores (**pmod**) y para objetos preposicionales (**pobj**). En caso de **pmods**, primero se acude a una rutina para la pluralización del postmodificador consultando una pequeña versión de la jerarquía semántica para el español, como muestra el ejemplo siguiente

Two centres had abandoned living donor transplantation => transplante de donantes vivos.

```
obj:pred:transplantation
  vnom:yes
  role:theme
  agr:num:sing
  descr:pred:living + donor
    donor isa human
    agr:num:sing
```

```
obj:pred:transplante
  vnom:yes
  pmod:pred:donante
    pcase:de
    mod:pred:vivo
  agr:gen:masc
  num:sing
  agr:gen:masc
  num:sing
```

7.3 Inserción idiosincrática

Algunos nombres siempre llevan artículo en español, como es el caso de **por ciento**, que tiene artículo y exige un artículo en el su objeto preposicional

30% of surgeons -> el 38 por ciento de los cirujanos.

7.4 Objetos preposicionales

La mayoría de los **pobjs** tienen que ser [+definite] en español. Por ejemplo,

outcome of transplantation	-> resultado de el transplante
views on expansion	-> opiniones sobre la ampliación
favour the expansion	-> estar a favor de la ampliación
influences blood pressure	-> influye en la presión
predisposes to differences	-> predispone a las diferencias
differences in cardiovascular mortality	-> diferencias en la mortalidad cardiovascular
related to birth weight	-> relacionado con el peso al nacer
independent of gestational age	-> independiente de la edad gestacional
attributed to foetal growth	-> atribuirse a el crecimiento fetal

8. Resultados

Por último, mostramos a continuación los resultados obtenidos de traducir dos **resúmenes** de medicina. Los resúmenes aparecieron en British Medical Journal 298, 04/03/89 p. 564, y 25/02/89, p. 490 respectivamente.

ABSTRACT 1

Title: transplants from living donors in the United Kingdom and Ireland: a centre survey.

A survey was carried out to determine for the first time the extent of transplantation from living donors in the United Kingdom and Republic of Ireland and the views of transplant surgeons regarding future developments. Questionnaires were sent to 32 transplant centres representing 18 health regions and covered their extent of experience of transplantation, sources of donors, ages of donors and recipients, outcome of transplantation, and views on expansion of living donor transplantation services. Replies received from 27 transplant centres representing 17 health regions gave data on more than 1200 transplants from living donors. Transplants from living donors accounted for 0.25% of the total experience of health regions. Two centres had abandoned living donor transplantation. Sixty percent of transplant surgeons favoured expansion of the living donor programme to meet a shortage of kidneys from cadavers and the remainder thought that existing programmes were optimal. Living donor transplantation promises to be an important factor in the future planning of health care resources.

TRANSLATION

Título: Transplantes de donantes vivos en Gran Bretaña e Irlanda: un estudio de centros.

Se llevó a cabo un estudio para determinar por primera vez el alcance de transplante de donantes vivos en Gran Bretaña e Irlanda y las opiniones de cirujanos de transplante respecto a desarrollos futuros. Se enviaron cuestionarios a 32 centros de transplante que representaban 18 regiones de salud y cubrieron su grado de experiencia en transplante, orígenes de donantes, edades de donantes y receptores, resultado del transplante y opiniones sobre la ampliación de servicios de transplante de donantes vivos. Las respuestas recibidas de 27 centros de transplante que representaban 17 regiones de salud aportaron datos sobre más de 1200 transplantes de donantes vivos. Los transplantes de donantes vivos representaron el 0.25% de la experiencia total de las regiones de salud. Dos centros habían abandonado el transplante de donantes vivos. El sesenta por ciento de los cirujanos de transplante estaban a favor de la ampliación del programa de donantes vivos para afrontar una escasez de riñones de cadáveres y el resto pensó que los programas existentes eran inmejorables. El transplante de donantes vivos promete ser un factor importante en la planificación futura de los recursos de atención sanitaria.

ABSTRACT 2

Title: growth in utero, blood pressure in childhood and adult life, and mortality from cardiovascular disease.

In national samples of 9921 10-year-olds and 3259 adults in Britain systolic blood pressure was inversely related to birth weight. The association was independent of gestational age and may therefore be attributed to foetal growth. This suggests that the intrauterine environment influences blood pressure during adult life. It is further evidence that the geographical differences in average blood pressure and mortality from cardiovascular disease in Britain partly reflect past differences in the intrauterine environment. Within England and Wales 10-year-olds living in areas with high cardiovascular mortality were shorter and had higher resting pulse rates than those living in other areas. Their mothers were also shorter and had higher diastolic blood pressures. This suggests that there are persisting geographical differences in the childhood environment that predispose to differences in cardiovascular mortality.

TRANSLATION

Título: crecimiento en útero, presión sanguínea en infancia y vida adulta y mortalidad por enfermedad cardiovascular.

En muestras nacionales de 9921 niños de 10 años y 3259 adultos en Gran Bretaña la presión sanguínea sistólica estaba inversamente relacionada con el peso al nacer. La asociación era independiente de la edad gestacional y puede por consiguiente atribuirse al crecimiento fetal. Esto sugiere que el entorno intrauterino influye en la presión sanguínea durante la vida adulta. Es evidencia adicional que las diferencias geográficas en la presión sanguínea media y mortalidad por enfermedad cardiovascular en Gran Bretaña en parte reflejan diferencias anteriores en el entorno intrauterino.

En Inglaterra y Gales los niños de 10 años que vivían en zonas con mortalidad cardiovascular alta eran más bajos y tenían pulsaciones más altas en reposo que los que vivían en otras zonas. Sus madres eran también más bajas y tenían presiones sanguíneas diastólicas más altas. Esto sugiere que hay diferencias geográficas persistentes en el entorno de infancia que predisponen a las diferencias en la mortalidad cardiovascular.

9. Conclusiones y desarrollos futuros

Pensamos que los resultados obtenidos son de una calidad más que aceptable, y apoyan la estrategia que aquí hemos defendido. Por consiguiente, pensamos que LFG ofrece grandes ventajas para el desarrollo de sistemas de TA que aspiren a ser operativos y estén motivados lingüísticamente. LFG ofrece, además, la ventaja de ser un formalismo, con lo que puede aumentarse o modificarse en función de su aplicación.

Los resultados refuerzan asimismo la idea de que la mayoría de los fenómenos de traducción pueden representarse empleando un sistema de TA de tipo transferencia.

En sucesivas versiones del programa nos proponemos seguir aumentando la cobertura lingüística basándonos en textos reales. Pretendemos aplicar el sistema a otros dominios semánticos siguiendo la misma estrategia. Al mismo tiempo, nos iremos centrando en problemas de traducción inglés-español y en cuestiones de representación lingüística poco tratados hasta ahora en LFG, tales como construcciones comparativas y apositivas. Asimismo, pretendemos estudiar estrategias para el tratamiento de construcciones no gramaticales en un entorno de unificación, que también habrían de ser traducidas en un entorno real.

10. Referencias

- [Allegranza et al. 1991] Allegranza, V., P. Bennett, J. Durand, F. Van Eynde, L. Humphreys, P. Schmidt y E. Steiner. 1991. «Linguistics for Machine Translation: The Eurotra Linguistic Specifications.» En Copeland et al. eds. *Studies in Machine Translation and Natural Language Processing*. Vol 1. Commission of the European Communities, Brussels, pp. 15-125.
- [Amores 1991] Amores, G. 1991. «Un mecanismo de transferencia para LFG en DCG-Prolog.» *Procesamiento del Lenguaje Natural*, 9 pp. 69-80.
- [Bennett 1990] Bennett, W.S. 1990. «How Much Semantics is Necessary for MT Systems?» En *Proceedings of the Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Language*. Austin, TX, pp. 261-271.
- [Bresnan ed. 1982] Bresnan, J., ed. 1982. *The Mental Representation of Grammatical Relations*. Cambridge, MA, MIT Press.
- [Brown et al. 1990] Brown, P.F., J. Cock, S.A. Della Pietra, V.J. Della Pietra, F. Jelinek, J.D. Lafferty, R.L. Mercer y P.S. Roossin. 1990. «A Statistical Approach to Machine Translation.» *Computational Linguistics* 16 (2), pp. 79-85.

- [Chandioux 1988] Chandioux, J. «Meteo™: An Operational Machine Translation System.» Presentation, RIAO 88 (Manchester Institute of Technology, March 1988).
- [Dorr 1990] Dorr, B. 1990. «A Cross-Linguistic Approach to Translation.» Austin, TX: Proceedings of the *Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Language*, pp. 13-33.
- [Eisele y Dörre 1986] Eisele, A. y J. Dörre. 1986. «A Lexical Functional Grammar in Prolog.» En *Coling 86*. Bonn, pp. 551-553.
- [Frey y Reyle 1983] Frey, W. y U. Reyle. 1983. «A Prolog Implementation of Lexical Functional Grammar as a Base for a Natural Language Processing System.» En *Proceedings of the 1st Meeting of the Association for Computational Linguistics*. Pisa.
- [Goodman y Nirenburg eds. 1991] Goodman, K. and S. Nirenburg eds. 1991. *The KBMT Project: A Case Study in Knowledge-based Machine Translation*. San Mateo, CA: Morgan Kaufmann.
- [Kaplan et al. 1989] Kaplan, R., K. Netter, J. Wedekind, and A. Zaenen. 1989. «Translation by Structural Correspondences.» 4th Conference of the European Chapter of the ACL. Manchester, United Kingdom, pp. 272-281.
- [Kiparsky 1987] Kiparsky, P. 1987. «Morphology and Grammar Relations.» Ms., Stanford, CA: Stanford University.
- [Kudo y Nomura 1986] Kudo, I. and H. Nomura. 1986. «Lexical-Functional Transfer: A Transfer Framework in a Machine Translation System based on LFG.» *Coling 86*. Bonn, Germany, pp. 112-115.
- [Levin 1985] Levin, L. 1985. «Identifying non Nominative Subjects in LFG.» *Proceedings of the Eastern State Conference in Linguistics*, pp. 313-324.
- [McCray 1989] McCray, A. 1989. «The UMLS Semantic Network.» En L.C. Kingsland ed. *Proceedings of the 13th Symposium on Computer Applications in Medical Care*. IEEE Computer Society Press, Los Angeles, CA, pp. 503-507.
- [Momma y Dörre 1987] Momma, K. y J. Dörre. 1987. «Generation from f-structures.» En Klein y van Benthem eds. *Categories, Polymorphism and Unification*. Centre for Cognitive Science, University of Edinburgh, Edinburgh, pp. 148-167.
- [Netter 1986] Netter, K. 1986. «Getting Things out of Order: An LFG-Proposal for the Treatment of German Word Order.» *Coling 86*, pp. 494-496.
- [Netter y Wedekind 1986] Netter, K. y J. Wedekind. 1986. «An LFG-based Approach to Machine Translation.» *Proceedings of IAI-MT86*, Didweiler, Germany, pp. 199-209.
- [Ruiz et al. 1991] Ruiz, J.C., J. Abaitua and J.R. Zubizarreta. 1991. «Un Compilador de LFG y su Aplicación al Euskara.» *Procesamiento del Lenguaje Natural*, 9 pp. 177-191.
- [Sadler et al. 1990] Sadler, L., I. Crookston, D. Arnold y A. Way. 1990. «LFG and Translation.» En Proceedings of the *Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Language*. Austin, TX, pp. 121-131.
- [Sager et al. 1987] Sager, N., C. Friedman, y M.S. Lyman. 1987. *Medical Language Processing*. Addison Wesley, Reading, MA.
- [Sharp 1988] Sharp, R. «CAT2 - Implementing a Formalism for Multi-lingual MT.» en Proceedings of the *2nd International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Language*. Pittsburgh, PA.
- [Slocum ed. 1988] Slocum, J. ed. 1988 *Machine Translation Systems*. Cambridge, United Kingdom: Cambridge University Press.
- [Van Noord et al. 1991] Van Noord, G., J. Dorrepaal, P. van der Eijk, M. Florenza, H. Ruessink y L. des Tombe. 1991. «An Overview of MiMo2.» *Machine Translation* 6 (3), pp. 215-228.

- [Vasconcellos y León 1988] Vasconcellos, M. and M. León. 1988. «SPANAM and ENGSPAN: Machine Translation at the Pan American Health Organization.» En: J. Slocum, ed. *Machine Translation Systems*. Cambridge, United Kingdom: Cambridge University Press, pp. 187-235.
- [Zajac 1990] Zajac, R. 1990. «A Relational Approach to Translation.» Proceedings of the *Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Language*. Austin, TX, pp. 235-255.