

## REPRESENTACION INTERMEDIA DE ESTRUCTURAS NOMINALES COMPLEJAS EN SISTEMAS DE T.A. BASADOS EN TRANSFER.

Juan Alberto Alonso (METAL - Siemens S.A.)  
Marta Carulla Pahl (EUROTRA - España)

### Resumen

El artículo trata de la representación intermedia de estructuras nominales en sistemas de traducción automática multilingües basados en "transfer". Se introduce primero el concepto de representación intermedia y la necesidad de contar con ella. A continuación se plantean los problemas que supone utilizar una representación de este tipo y se proponen una serie de métodos que pueden ayudar a solventar tales problemas. Finalmente, presentamos varios ejemplos centrados en construcciones nominales compuestas en alemán que ilustran la utilización del método seguido.

### 1.- Planteamiento del Entorno.

Todo lo presentado en este artículo está pensado para su aplicación en un sistema de traducción automática que cuente con las siguientes características:

- Basado en "Transfer".
- Multilingüe: el sistema no está específicamente pensado para un par de lenguas determinado, sino que está diseñado para poder traducir varios pares de lenguas diferentes.
- Ampliable: la inclusión de un nuevo par de lenguas debe ser posible y no debe afectar al núcleo del sistema. Para incluir un nuevo par de lenguas X-Y se deben codificar los léxicos monolingües X e Y (si es que no existen ya), el léxico bilingüe X-Y, la gramática de análisis de la lengua X (si no existe ya) y la gramática de síntesis de la lengua Y (si no existe ya). En cualquier caso, tanto los léxicos monolingües como las gramáticas de análisis y síntesis de una lengua deben ser aprovechables para nuevos pares de lenguas.

## 2.- El concepto de Representación Intermedia en sistemas de T.A. basados en Transfer.

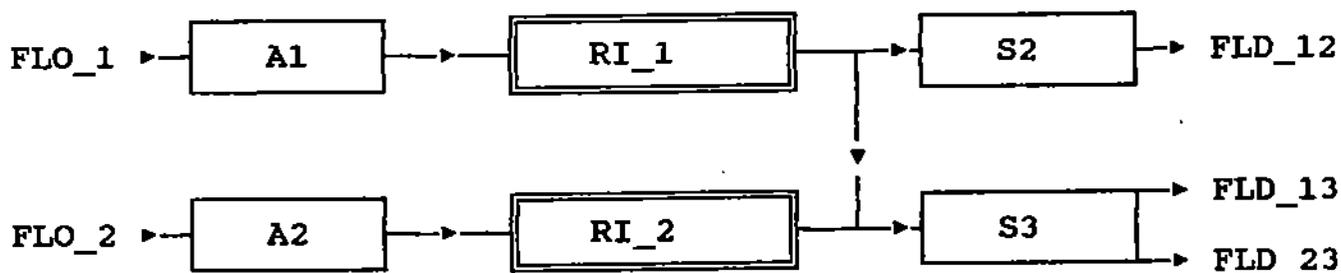
### 2.1.- Objetivos.

La utilización de una representación sintáctica pivote universal (o Representación Intermedia: RI) entre las fases de análisis y de generación de un sistema de T.A. basado en "Transfer" tiene como objetivo tres puntos básicos:

- Reducir el módulo de transfer para cada par de lenguas a, básicamente, un diccionario bilingüe. Las entradas del diccionario bilingüe pueden incluir información para realizar transformaciones estructurales (adición, borrado o cambio de material léxico) en la fase de generación que sean dependientes de la entrada en cuestión.
- Cada lengua de origen tendrá un único módulo de análisis que será independiente de la lengua de destino. Cada módulo de análisis debe construir árboles bien formados según las especificaciones de la RI.
- Cada lengua de destino tendrá un único módulo de síntesis que será independiente de la lengua de origen. Cada módulo de síntesis tomará árboles RI como entrada.

### 2.2.- Requisitos y definiciones aplicables a árboles RI.

Dada la siguiente figura:



donde

- FLO = Frase Lengua Origen,
- FLD = Frase Lengua Destino,
- RI = Arbol en Representación Intermedia
- A = Módulo de Análisis,
- S = Módulo de Síntesis,

la estructura e información que conforman los árboles de la RI deben cumplir los siguientes requisitos:

### Requisito de Equivalencia Débil de Análisis.

Dos módulos de análisis diferentes para dos lenguas origen diferentes deben dar como resultado dos árboles RI que sean débilmente equivalentes para dos frases de entrada que tengan el mismo significado en los dos idiomas.

$(FLO_1 \Leftrightarrow FLO_2) \Rightarrow (RI_1 \succ\prec RI_2)$  en la figura.

### Definición de Equivalencia Débil.

Dos árboles RI T1 y T2 resultantes de los módulos de análisis para las lenguas L1 y L2, respectivamente, son débilmente equivalentes entre sí  $(T1 \succ\prec T2)$  si,

- o T1' es fuertemente equivalente a T2, o
- T2' es fuertemente equivalente a T1,

donde T1' y T2' son el resultado de someter T1 y T2, respectivamente, a las transformaciones estructurales especificadas en las entradas léxicas de transfer correspondientes al material léxico contenido en los árboles.

### Definición de Equivalencia Fuerte

Dos árboles RI T1 y T2 son fuertemente equivalentes  $(T1 \leftrightarrow T2)$  si para cada nodo N1 perteneciente a T1 hay un nodo correspondiente N2 en T2 que tenga el mismo camino (lista de nodos que hay que recorrer desde el nodo raíz hasta N), etiqueta (categoría sintáctica de N) y secuencia (número de orden de izquierda a derecha en caso de que haya nodos hermanos con igual camino y etiqueta que N) que N1 y el mismo conjunto de rasgos y valores que N1 (sin tener en cuenta rasgos léxicos).

### Requisito de Síntesis para Árboles Débilmente Equivalentes

Cada módulo de síntesis debe generar frases con igual significado cuando tengan como entrada árboles RI que sean débilmente equivalentes.

$(RI_1 \succ\prec RI_2) \Rightarrow (FLD_{13} \Leftrightarrow FLD_{23})$  en la figura.

(' $\Leftrightarrow$ ') significa "tienen igual significado")

### Requisito de Conservación de Significado

Dos módulos de síntesis diferentes para dos lenguas de destino diferentes deben generar dos frases con el mismo significado si, o bien tiene el mismo árbol RI como entrada, o bien tienen dos árboles RI débilmente equivalentes como entrada.

$(RI_1 \succ RI_2) \implies (FLD_{12} \iff FLD_{13})$   
 $(FLD_{12} \iff FLD_{23})$  en la figura.

### 2.3.- Diseño e implementación de la RI.

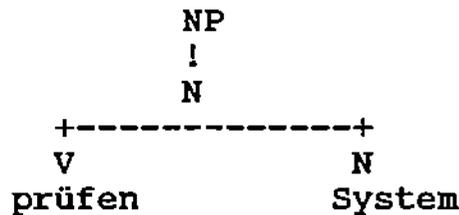
En el momento de diseñar e implementar una representación en interlingua sintáctica, hay tres puntos a tener en cuenta:

- Es preferible representar la información en forma de rasgos que en forma de estructura. Muchos nodos presentes en el árbol de "parsing" se pueden eliminar en el árbol RI correspondiente, expresándose su información en forma de pares rasgo-valor contenidos en nodos superiores. Esto suele ser el caso con los afijos, partículas marcadoras de caso, algunos determinantes, verbos modales y auxiliares, algunos adverbios con valor aspectual, etc.
- La información contenida en los pares rasgo-valor no debe referirse a fenómenos superficiales de la lengua de origen, sino a información de nivel más profundo común a todas las lenguas en juego. Los árboles RI no deben tener información sobre género gramatical, sino sobre género natural, no sobre caso sino sobre rol o función, no sobre tiempo gramatical sino sobre tiempo "cronológico", etc.
- El árbol en RI no debe contener información que sea específica de una lengua en particular y que no aporte nada al significado de la frase de cara a su traducción a las otras lenguas en juego.

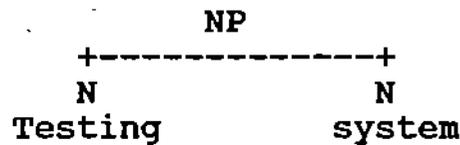
### 3.- El problema de la representación de las expresiones nominales complejas.

El problema básico planteado a la hora de utilizar una RI como la anteriormente presentada es la dificultad de encontrar una representación que se adapte a las distintas estructuras de análisis que diferentes idiomas presentan para frases con idéntico significado. Veamos unos ejemplos:

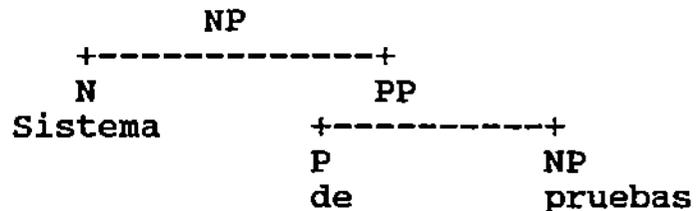
De: Prüfsystem:



En: Testing system:



Es: sistema de pruebas:



Esto es uno de los casos "fáciles". Frases alemanas del tipo "Das auszugebende Geld" corresponden en castellano a estructuras con una frase de relativo "El dinero que se debe gastar".

Si en un sistema de T.A. como el que nos ocupa se quiere utilizar una única RI como estructura pivote, se debe definir una arquitectura determinada de los árboles RI para cubrir casos como estos, en los que diferentes lenguas presentan estructuras superficiales diferentes.

#### 4.- Arquitectura e información necesaria de la RI para estructuras nominales.

Nuestra propuesta incluye la anotación de los nodos con rasgos que den información sobre las relaciones funcionales entre los distintos constituyentes así como de sus roles temáticos (en caso de tenerlos). Al mismo tiempo proponemos una arquitectura de RI para estructuras nominales consistente en un aplanamiento máximo de los árboles de análisis.

##### 4.1.- Relaciones funcionales.

Cada uno de los constituyentes desempeña una función en relación con los demás. Tal función puede ser la de núcleo, especificador, modificador o complemento. Así por ejemplo, en la frase:

"La presentación de nuestro artículo en S. Sebastián"

el núcleo de toda la estructura es "presentación", "artículo" desempeña una función de complemento con respecto a su núcleo, que es "presentación", "nuestro" desempeña una función de especificador con respecto a su núcleo que es "artículo" y "en S. Sebastián" desempeña una función de modificador con respecto a su núcleo que es "presentación".

En el punto 5 se presentarán ejemplos concretos que demuestran la necesidad de que esta información esté presente en la RI.

#### 4.2.- Roles temáticos.

La existencia de información sobre roles temáticos en aquellos constituyentes que los puedan llevar es igualmente necesaria para la traducción y debe estar presente en la RI (ver punto 5).

La diferencia entre un modificador y un complemento es que un núcleo puede asignar un determinado rol temático a sus complementos pero no a sus modificadores.

En nuestra propuesta, consideraremos que sólo los núcleos que sean nombres deverbales pueden tener complementos y por lo tanto asignar roles. Por "nombres deverbales", entendemos no solo los nombres directamente derivables de un verbo, sino también todos aquellos que puedan formar parte de construcciones verbales fijas, (tipo "tener miedo a/de").

Los roles temáticos que utilizaremos son: Agente, Tema, Lugar, Origen, Destino (cf. Cook89), aunque se pueden añadir nuevos roles (Experimentador, Paciente, Resultado, etc.) en caso de que interese tener una distinción de Tema más fina.

El criterio para decidir si un constituyente es modificador o complemento con respecto a su núcleo es que dicho constituyente pueda llenar uno de los roles exigidos por el marco temático del verbo subyacente en el núcleo. Esto significa que aquellos nombres deverbales que subcategorizan roles temáticos heredan el marco de subcategorización de su correspondiente verbo (cf Randall83).

Según todo esto, en nuestro anterior, el sintagma "de nuestro artículo" sería un complemento del núcleo "la presentación", ya que "artículo" puede ocupar la posición de Tema en el marco argumental del verbo "presentar", y por lo tanto, del deverbale "presentación".

#### 4.3.- Aplanamiento de árboles

De cara a conseguir estructuras similares para diferentes idiomas, consideramos que una buena técnica sería la de aplanar los árboles de análisis lo más posible. Las relaciones de dominancia y precedencia que estaban expresadas en el árbol de análisis de

forma estructural, las representaremos ahora en forma de rasgos. Cada nodo llevará información sobre qué constituyente es su núcleo. Igualmente, interesa aplanar los sintagmas preposicionales, manteniendo la información sobre la preposición de la lengua de origen en forma de rasgo. Así, la RI para el ejemplo que nos ocupa sería:

NP			
NUC	COMPL	ESP	MOD
presentación CAT_Es=N	artículo CAT_Es=N NUC=presentación ROL=Tema PREP_LO=de	nuestro CAT_Es=DET NUC=artículo	S. Sebastián CAT_Es=N NUC=presentación PREP_LO=en

#### 5.- Ejemplos de representación en RI para compuestos nominales.

En las lenguas germánicas, en menor medida en griego y en vasco, y en mucha menor medida en las lenguas románicas, se da el fenómeno de los compuestos nominales. Se trata en todo caso de compuestos morfológicos:

Alemán: Bushaltestelle, Handbremsvorrichtung.  
 Inglés: handbag.  
 Griego: Odhondoyatrós, Psarotaverna.  
 Vasco: Postakutxa.  
 Castellano: matamoscas.

La inmensa mayoría de compuestos morfológicos tienen su contrapartida como estructuras NP AP/PP en las lenguas románicas

Bushaltestelle = parada de autobús.  
 Handbremsvorrichtung = dispositivo de frenado manual.  
 Handbag = bolso.  
 Psarotaverna = taberna de pescado.  
 Postakutxa = apartado de correos.

Como ejemplo concreto presentaremos el tratamiento de los compuestos nominales no-lexicalizados en alemán y su correspondencia en inglés y en castellano.

Ya hemos visto que las estructuras nominales se representan superficialmente de maneras muy distintas y que el objetivo fundamental de la representación interlingual es su independencia de esta representación superficial. De cara a una representación más profunda tiene interés también representar de igual forma construcciones débilmente equivalentes dentro de una misma lengua

p.ej. en alemán "Computerverkauf" y "Verkauf von Computern" o en castellano "país comunitario" y "país de la Comunidad".

Hemos hablado de la información que debe contener la RI y en el siguiente apartado vamos a representar algunos ejemplos concretos de cada uno de los tipos de compuestos nominales en alemán y a la vez mencionaremos algunas generalidades que sirven de ayuda a los componentes de análisis y de generación que utilizan la RI como estructura pivote.

En alemán se distinguen principalmente tres mecanismos de composición nominal:

- la composición de dos o más nombres
- la composición de raíz verbal y nombre
- la composición de adjetivo y nombre

#### 5.1. La composición de dos o más nombres

Ejemplos:

i)

Busfahrer	bus driver	conductor de autobús
Programmentwicklung	programm development	desarrollo de programas
Datenschutz	data protection	protección de datos
Problemlösung	problem solving	resolución de problemas

ii)

Landhaus	country house	casa de campo
Fabrikarbeiter	factory worker	trabajador de fábrica
Nordseeoel	North Sea oil	petróleo del Mar del Norte
Metallindustrie	metal industry	industria metalúrgica
Hauptaufgabe	main task	tarea principal
Grundfähigkeit	basic capacity	capacidad básica
Endprodukt	final product	producto final
Schlüsselwort	key word	palabra clave
Mitgliedsstaat	member state	estado miembro

Como hemos dicho toda estructura nominal tiene su núcleo; en los compuestos nominales el núcleo suele ser el nombre que aparece más a la derecha. Así el núcleo de "Landhaus" es "Haus" (casa) y de "Autozubehörindustrie" (industria de accesorios de coche) es "Industrie".

El núcleo del compuesto puede subcategorizar o no a los demás elementos del compuesto. Si el núcleo no tiene marco de subcategorización el elemento no-núcleo es siempre modificador. En cambio, si el núcleo tiene marco de subcategorización el elemento

no núcleo puede ser modificador o complemento.

El caso en que el no-núcleo toma un papel argumental ha sido especialmente estudiado en el campo de los compuestos, seguramente por su gran regularidad. Todo nombre derivado de un verbo que subcategoriza más de un rol temático puede formar un compuesto con su tema. Este tipo de compuestos raras veces está lexicalizado, suele permitir la paráfrasis y suele ser regular también de cara a la traducción.

En alemán se utiliza el compuesto preferentemente cuando el complemento es genérico: la lectura preferida de "Problemlösung" es "resolución de problemas"; para expresar la lectura con el complemento definido "la resolución del problema" se prefiere la construcción "die Lösung des Problems".

Las diferentes generaciones de las lenguas destino tendrán sus propios mecanismos de construcción de estructuras superficiales a partir de la RI. Así, para compuestos cuyo no-núcleo sea un complemento (los ejemplos anteriores encuadrados en (i)), el castellano asignará estructuras tipo NP[núcleo] + PP["de" + complemento]. Así:

Busfahrer (Bus es un complemento de Fahrer)

"conductor de autobús"

Programmentwicklung (Programm es complemento de Entwicklung)

"desarrollo de programas"

Para aquellos compuestos cuyo no-núcleo tiene función de modificador (los ejemplos anteriores encuadrados en (ii)), no es predecible a priori la estructura en la lengua de destino, siendo esta normalmente dependiente del material léxico. La relación entre el núcleo y su modificador está subespecificada semánticamente y por tanto en las demás lenguas estos compuestos se traducirán por estructuras que reflejen esta subespecificación. El contexto es de vital importancia para la comprensión de este tipo de compuestos y es también el contexto el que determinará la lectura genérica o definida del modificador puesto que no existirá ninguna preferencia explícita.

A continuación podemos ver algunos ejemplos con su respectiva estructura RI:

De:

NP	
+	-----+
NUC	COMPL
Fahrer	Bus
CAT-De=N	CAT-De=N
COMP-De=sí	COMP-De=sí
	NUC=Fahrer
	ROL=Tema

En:

NP

NUC driver CAT-En=N	COMPL bus CAT-En=N NUC=driver ROL=Tema

Es:

NP

NUC conductor CAT-Es=N	COMPL autobús CAT-Es=N NUC=conductor ROL=Tema PREP-Es=de

### 5.2. La composición de raíz verbal y nombre

Ejemplos:

i)		
Prüfverfahren	testing procedure	proceso de inspección
Badeanzug	bathing suite	traje de baño
ii)		
Schwimmkran	floating crane	grua flotante
Klappstuhl	collapsible chair	silla plegable

En este tipo de compuestos alemanes, el elemento no nuclear siempre es un modificador. De hecho, en estos compuestos es el núcleo el que a veces ocupa un rol temático del marco argumental del modificador. Por ejemplo, en "Schwimmkran" ("grúa flotante"), "Kran" ("grúa") es el Tema de "schwimmen" ("flotar"). No ocurre así en el caso de "Prüfverfahren" ("proceso de inspección") o de "Schwimmwettbewerb" ("concurso de natación"), donde el núcleo no ocupa ningún papel temático del verbo que hace de modificador del compuesto.

Como criterios de ayuda a la generación se podría decir, que la tendencia general es que la forma verbal tenga su correspondencia en el nombre derivado de este verbo si el nombre no ocupa un papel temático del verbo (ejemplos i)). Así:

Schwimmlehrer: "profesor de natación"  
Malwettbewerb: "concurso de pintura"

Si el no-núcleo ocupa el rol de Tema del verbo que actúa como modificador dentro del compuesto, la correspondiente versión en castellano suele presentar un participio o un adjetivo derivado

del verbo correspondiente. Por ejemplo:

Hängebrücke: "puente colgante"  
Waschkleid: "vestido lavable"

He aquí algunos ejemplos de representación en RI para estos compuestos:

De:

```

                        NP
          +-----+
          NUC                      MOD
          Kran                      schwimmen
          CAT-De=N                  CAT-De=V
          COMP-De=sí                COMP-De=sí
                                   NUC=Kran
                                   ROL-NUC=Tema
  
```

En:

```

                        NP
          +-----+
          NUC                      MOD
          crane                      floating
          CAT-En=N                  CAT-En=A
                                   NUC=crane
                                   ROL-NUC=Tema
  
```

Es:

```

                        NP
          +-----+
          NUC                      MOD
          grúa                      flotante
          CAT-Es=N                  CAT-Es=A
                                   NUC=grúa
                                   ROL-NUC=Tema
  
```

### 5.3. La composición de adjetivo y nombre

Gesamtausgabe	complete edition	edición completa
Hoechstgeschwindigkeit	maximum speed	velocidad máxima
Zentraleinheit	central unit	unidad central
Privatbereich	private area	ámbito privado

El adjetivo tiene siempre la función de modificador del núcleo en este tipo de compuestos. Estos compuestos no ofrecen ningún tipo de dificultad estructural a la hora de su traducción. Como ejemplo de representación RI podemos poner:

De: NP

+-----+	
NUC	MOD
Einheit	zentral
CAT-De=N	CAT-De=A
COMP=Sí	COMP=Sí
	NUC=Einheit

En: NP

+-----+	
NUC	MOD
unit	central
CAT-En=N	CAT-En=A
	NUC=unit

Es: NP

+-----+	
NUC	MOD
unidad	central
CAT-Es=N	CAT-Es=A
	NUC=unidad

## 6.- Otras estructuras nominales complejas.

Aparte de los compuestos nominales presentados en la sección anterior, existen más tipos de estructuras nominales complejas que son susceptibles de ser representadas de manera similar para diferentes idiomas siguiendo nuestra propuesta. He aquí algunos ejemplos de no compuestos con modificadores verbales:

- Das gelesene Buch: "el libro leído"

De: NP

+-----+	
NUC	MOD
Buch	lesen
CAT-De=N	CAT-De=V
	NUC=Buch
	VOZ=pasiva <sup>1</sup>
	ASPECTO=perfectivo
	ROL-NUC=Tema

<sup>1</sup> Aunque esto no es tema del presente artículo, se podrían representar los participios verbales mediante rasgos interlinguales (en vez de con información morfológica de la lengua en cuestión, como es VFORM), como serían "VOZ" y "ASPECTO". Así "gelesen" ("leído") iría marcado con [VOZ=pasiva] y [ASPECTO=perfectivo], mientras que "schwimmend" ("que está nadando") iría marcado como [VOZ=activa] y [ASPECTO=progresivo].

En: NP

+-----+														
<table border="0"> <tr> <td style="text-align: center;">NUC</td> <td style="text-align: center;">MOD</td> </tr> <tr> <td>book</td> <td>read</td> </tr> <tr> <td>CAT-En=N</td> <td>CAT-En=V</td> </tr> <tr> <td></td> <td>NUC=book</td> </tr> <tr> <td></td> <td>VOZ=pasiva</td> </tr> <tr> <td></td> <td>ASPECTO=perfectivo</td> </tr> <tr> <td></td> <td>ROL-NUC=Tema</td> </tr> </table>	NUC	MOD	book	read	CAT-En=N	CAT-En=V		NUC=book		VOZ=pasiva		ASPECTO=perfectivo		ROL-NUC=Tema
NUC	MOD													
book	read													
CAT-En=N	CAT-En=V													
	NUC=book													
	VOZ=pasiva													
	ASPECTO=perfectivo													
	ROL-NUC=Tema													

Es: NP

+-----+														
<table border="0"> <tr> <td style="text-align: center;">NUC</td> <td style="text-align: center;">MOD</td> </tr> <tr> <td>libro</td> <td>leer</td> </tr> <tr> <td>CAT-Es=N</td> <td>CAT-Es=V</td> </tr> <tr> <td></td> <td>NUC=libro</td> </tr> <tr> <td></td> <td>VOZ=pasiva</td> </tr> <tr> <td></td> <td>ASPECTO=perfectivo</td> </tr> <tr> <td></td> <td>ROL-NUC=Tema</td> </tr> </table>	NUC	MOD	libro	leer	CAT-Es=N	CAT-Es=V		NUC=libro		VOZ=pasiva		ASPECTO=perfectivo		ROL-NUC=Tema
NUC	MOD													
libro	leer													
CAT-Es=N	CAT-Es=V													
	NUC=libro													
	VOZ=pasiva													
	ASPECTO=perfectivo													
	ROL-NUC=Tema													

- Das schwimmende Kind: "el niño que está nadando"

De: NP

+-----+														
<table border="0"> <tr> <td style="text-align: center;">NUC</td> <td style="text-align: center;">MOD</td> </tr> <tr> <td>Kind</td> <td>schwimmen</td> </tr> <tr> <td>CAT-De=N</td> <td>CAT-De=V</td> </tr> <tr> <td></td> <td>NUC=Kind</td> </tr> <tr> <td></td> <td>VOZ=activa</td> </tr> <tr> <td></td> <td>ASPECTO=progresivo</td> </tr> <tr> <td></td> <td>ROL-NUC=Tema</td> </tr> </table>	NUC	MOD	Kind	schwimmen	CAT-De=N	CAT-De=V		NUC=Kind		VOZ=activa		ASPECTO=progresivo		ROL-NUC=Tema
NUC	MOD													
Kind	schwimmen													
CAT-De=N	CAT-De=V													
	NUC=Kind													
	VOZ=activa													
	ASPECTO=progresivo													
	ROL-NUC=Tema													

En: NP

+-----+														
<table border="0"> <tr> <td style="text-align: center;">NUC</td> <td style="text-align: center;">MOD</td> </tr> <tr> <td>child</td> <td>swim</td> </tr> <tr> <td>CAT-En=N</td> <td>CAT-En=V</td> </tr> <tr> <td></td> <td>NUC=child</td> </tr> <tr> <td></td> <td>VOZ=activa</td> </tr> <tr> <td></td> <td>ASPECTO=progresivo</td> </tr> <tr> <td></td> <td>ROL-NUC=Tema</td> </tr> </table>	NUC	MOD	child	swim	CAT-En=N	CAT-En=V		NUC=child		VOZ=activa		ASPECTO=progresivo		ROL-NUC=Tema
NUC	MOD													
child	swim													
CAT-En=N	CAT-En=V													
	NUC=child													
	VOZ=activa													
	ASPECTO=progresivo													
	ROL-NUC=Tema													

Es: NP

+-----+														
<table border="0"> <tr> <td style="text-align: center;">NUC</td> <td style="text-align: center;">MOD</td> </tr> <tr> <td>niño</td> <td>nadar</td> </tr> <tr> <td>CAT-Es=N</td> <td>CAT-Es=V</td> </tr> <tr> <td></td> <td>NUC=niño</td> </tr> <tr> <td></td> <td>VOZ=activa</td> </tr> <tr> <td></td> <td>ASPECTO=progresivo</td> </tr> <tr> <td></td> <td>ROL-NUC=Tema</td> </tr> </table>	NUC	MOD	niño	nadar	CAT-Es=N	CAT-Es=V		NUC=niño		VOZ=activa		ASPECTO=progresivo		ROL-NUC=Tema
NUC	MOD													
niño	nadar													
CAT-Es=N	CAT-Es=V													
	NUC=niño													
	VOZ=activa													
	ASPECTO=progresivo													
	ROL-NUC=Tema													

- Das auszuführende Programm: "el programa que se debe ejecutar"

De:

NP	
+-----+	+-----+
NUC Programm CAT-Es=N	MOD ausführen CAT-Es=V NUC=Programm VOZ=pasiva ASPECTO=perfectivo MODAL=obligatorio ROL-NUC=Tema

En:

NP	
+-----+	+-----+
NUC program CAT-En=N	MOD execute CAT-En=V NUC=program VOZ=pasiva ASPECTO=perfectivo MODAL=obligatorio ROL-NUC=Tema

Es:

NP	
+-----+	+-----+
NUC programa CAT-Es=N	MOD ejecutar CAT-Es=V NUC=programa VOZ=pasiva ASPECTO=perfectivo MODAL=obligatorio ROL-NUC=Tema

En principio, esta información basta para la RI. Las generaciones respectivas tendrán suficiente información como para saber que, por ejemplo, un modificador verbal dentro de un NP con marcas de voz pasiva, aspecto perfectivo y modalidad obligatoria se realiza como una cláusula de relativo en castellano y en inglés ("que se debe ejecutar", "which has to be executed") y como un participio presente con "zu" en alemán ("auszuführend-").

## 7.- Conclusión.

Hemos presentado una propuesta de representación intermedia que elimina gran parte de los casos en los que los árboles sintácticos provenientes de los diferentes análisis de diferentes lenguas eran débilmente equivalentes<sup>2</sup>.

Nuestro enfoque supone que todos los constituyentes de un sintagma nominal están al mismo nivel en el árbol, que sus relaciones de dominancia y precedencia se expresan en forma de rasgos y que están anotados con su función y su rol (en caso de tenerlo). Dado que las relaciones funcionales (núcleo-modificador-complemento) se conservan en la gran mayoría de lenguas (a diferencia de la categoría), la mayoría de estructuras nominales se pueden representar de forma que sus árboles RI sean fuertemente equivalentes.

Sólo en aquellos casos en que la lexicalización o la descomposición se da de una lengua a otra, suponiendo una adición o una eliminación de un elemento al pasar de una lengua a otra (por ejemplo: "Straßenlampe" -> "farola", o "Chemikalien" -> "compuestos químicos"), los respectivos árboles RI serán débilmente equivalentes.

A lo largo de este artículo se ha supuesto que los módulos de análisis tienen la información necesaria para generar estructuras RI con la arquitectura y la información propuestas a partir de la estructura superficial de la lengua de origen y que los módulos de generación son capaces de generar estructuras superficiales correctas en la lengua de destino a partir de la información del árbol RI. Obviamente, existen numerosos problemas, sobre todo en la fase de análisis, que hacen que estas suposiciones no sean triviales: adjunción de modificadores y de complementos a sus núcleos, detección de complementos y de sus roles temáticos, etc.

Por último, falta por extender este enfoque a otros tipos de constituyentes no nominales (sintagmas adjetivales, verbales y diferentes tipos de cláusulas).

<sup>2</sup>

Siempre que en la definición de "relación de equivalencia fuerte" (punto 2.2) se considere la etiqueta de los nodos como su relación funcional, y no como su categoría.

**péndice:** Lista de rasgos usados en los árboles RI

SPECTO	aspecto verbal
AT-Lg	categoría sintáctica del constituyente en la lengua Lg. (Lg= 'De': alemán, 'En': inglés, 'Es': español)
OMP	el constituyente forma parte de un compuesto morfológico.
OMPL	complemento
SP	especificador
OD	modificador
ODAL	modalidad verbal
UC	núcleo
REP-Lg	preposición del PP en la lengua Lg.
OL	rol temático del complemento
OL-NUC	rol temático del núcleo
OZ	voz verbal

**ibliografía.**

- Alonso90]: ALONSO, J.A., Transfer InterStructure: Designing an 'Interlingua' for Transfer-based MT Systems. Artículo invitado en "The Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages", LRC, Austin, Texas, EE.UU., Junio 1990.
- Bourbeau88]: BOURBEAU, L & LEHRBERGER, J. Machine Translation: Linguistic Characteristics of MT System and General Methodology of Evaluation. John Benjamins Publishing Company. 1988 Amsterdam/Philadelphia.
- Cook89]: COOK, W.A. Case Grammar Theory. Georgetown University Press. 1989 Georgetown, EE.UU.
- Hutchins86]: HUTCHINS, W.J.; Machine Translation: past, present and future. Ellis Horwood Ltd. 1986 Chichester, England.
- Nirenburg87]: NIRENBURG, S. Ed. Machine Translation. Cambridge University Press. 1987 Cambridge, UK.
- Randall83]: RANDALL, Thematic Structure and Inheritance (sin publicar), 1983.