

APROXIMACION FUNCIONAL A DIAC: DICCIONARIO INTELIGENTE DE AYUDA A LA COMPRESION

X. Arregi, X. Artola, A. Díaz de Ilarraza, F. Evrard (*), K. Sarasola

Lengoaia eta Sistema Informatikoak (UPV-EHU)
Informatika Fakultatea, p.k. 649. 20080 DONOSTIA
e-mail: xabier@gorria.if.ehu.es

(*) ENSEEIHT (Toulouse)

Abstract

DIAC, diccionario inteligente de ayuda a la comprensión, se ha concebido como un sistema computacional para facilitar y enriquecer la utilización por parte de un usuario humano de la información contenida en un diccionario monolingüe. El propio sistema está dotado de capacidad de deducción, de forma que pueda responder a solicitudes no triviales del usuario.

DIAC se está desarrollando a partir de una Base de Datos Diccional (BDD), cuyo contenido se ha analizado para estructurar una Base de Conocimientos (BCD), mediante la especificación de la semántica asociada a las diversas estructuras definitorias. Esta BCD es el soporte sobre el que estamos desarrollando el sistema de explotación "inteligente". Para ello es preciso determinar y especificar cuáles son las necesidades del usuario a las que habrá que responder. Desde este punto de vista la BCD se concibe como una estructura de Tipos Abstractos de Datos caracterizada por un conjunto de operaciones.

En el presente artículo se hace un breve resumen de los pasos seguidos hasta ahora en el proceso de extracción de conocimiento semántico a partir de un diccionario monolingüe convencional (2) para a continuación presentar someramente un esquema de la representación que para la base de conocimiento diccional se propone (3).

El objetivo principal de la comunicación es sin embargo presentar el aspecto funcional del sistema (4) que denominamos diccionario inteligente de ayuda a la comprensión.

Areas: Lexicografía computacional. Representación del conocimiento. Sistemas de ayuda inteligentes.

APROXIMACION FUNCIONAL A DIAC: DICCIONARIO INTELIGENTE DE AYUDA A LA COMPRESION

1.- Introducción.

El Diccionario Inteligente de Ayuda a la Comprensión (DIAC) se concibe como un sistema diccionarioal monolingüe (explicativo) para un usuario humano. Es precisamente este enfoque de usuario humano frente a sistema de procesamiento de lenguaje natural el que diferencia este trabajo de otros en la misma línea de adquisición de conocimiento semántico a partir de diccionarios convencionales. En este sentido, el sistema ha de proveer de posibilidades diversas de acceso a la información haciendo factible la obtención de conocimiento implícito que se supone reside en un diccionario. Se trata, en suma, de dotar al sistema de mecanismos de razonamiento que hagan que su comportamiento sea el de un sistema inteligente, análogo en cierta manera al uso que los humanos hacemos de los diccionarios cuando recurrimos a ellos.

Entre los objetivos del trabajo se encuentra la propuesta de representación de conocimiento para un sistema de estas características y la implantación en el mismo de mecanismos de razonamiento adecuados que hagan posible la respuesta a diferentes peticiones del usuario.

Antes de pasar a presentar este último aspecto de la funcionalidad a nivel del usuario, haremos un breve resumen de los pasos seguidos hasta ahora en la construcción de la Base de Conocimiento Diccionarioal (BCD) soporte de DIAC y de la propuesta de representación de conocimiento.

2.- Del diccionario a la BCD.

Para el presente trabajo se ha partido de un pequeño diccionario monolingüe de francés (*Le Plus Petit Larousse*, Librairie Larousse: Paris, 1980) que, a falta de una versión en soporte magnético, fue almacenado en una base de datos relacional convencional. Este diccionario ha servido de soporte a los estudios empíricos realizados que nos han llevado a la propuesta de representación que hacemos.

Le Plus Petit Larousse (LPPL) consta de cerca de 23000 acepciones correspondientes a casi 16000 entradas. Contiene para cada entrada información referente principalmente a categoría gramatical, definiciones (cortas, 3.27 palabras/definición como media) o referencias a sinónimos de las diferentes acepciones, marcas de uso y ejemplos en un número reducido de casos, etc.

El estudio empírico se ha realizado fundamentalmente sobre el campo de definición de cada concepto en cuanto que, en el diccionario, la definición es el factor más relevante en lo que a conocimiento semántico sobre el concepto se refiere. Este estudio ha tenido como objetivo principal la caracterización del metalenguaje lexicográfico utilizado en la redacción de las definiciones con vistas a:

- (1) especificar la semántica de las diferentes estructuras sintácticas utilizadas en las definiciones, y
- (2) obtención de elementos distinguidos susceptibles de funcionar como relatores específicos (entre conceptos) (Vossen et al., 1989,90)

Primeramente se ha realizado –para cada categoría de entradas– un estudio frecuencial de palabras en las definiciones. Se ha procedido después al etiquetado o "tagging" de las apariciones de palabras en definiciones para lo que se ha hecho uso fundamentalmente del propio diccionario; ha sido éste un proceso semiautomático donde ciertas ambigüedades han sido resueltas por procedimientos manuales y donde posteriores estudios estadísticos sobre secuencias de categorías que ocurren en las definiciones de conceptos han servido como realimentación al propio proceso. Los resultados de este proceso nos han llevado a una tasa superior al 85% de apariciones etiquetadas y un 70% de definiciones en las que todas las palabras han sido etiquetadas.

El estudio frecuencial de secuencias de categorías que ocurren en las definiciones ha servido como base a un proceso de análisis bottom-up cuyo resultado ha sido un recuento de categorías sintácticas de orden superior para las definiciones de sustantivos, verbos y adjetivos. Estos resultados junto con estudios particularizados sobre palabras de frecuencia elevada susceptibles de convertirse en relatores específicos constituyen la base para la última fase de análisis de las definiciones.

La última fase de análisis de las definiciones ha consistido en un análisis basado en pattern-matching jerárquico parcial (Alshawi 87, 89). En esta fase se asocia a cada pattern de análisis una interpretación semántica expresada en términos de la representación propuesta.

Esta última fase de análisis ha sido aplicada ya a las definiciones de sustantivos, verbos y adjetivos de LPPL. Para los sustantivos se ha definido un total de 65 patterns que han dado como resultado un 92% de definiciones analizadas con "éxito". Con esta técnica de análisis parcial, consideramos éxito para las definiciones de sustantivos aquellos casos en los que se ha logrado reconocer al menos un primer grupo nominal en el texto de la definición, en el cual se supone se encontrará generalmente el concepto hiperónimo del definido. Para el caso de los verbos una jerarquía formada por 53 reglas ha servido para analizar con éxito el 98% de las definiciones elegidas. Las definiciones de adjetivos han sido analizadas mediante 45 patterns con un porcentaje de éxito del 84%.

No hay que olvidar que con esta técnica se obtiene frecuentemente conocimiento incompleto al analizarse parcialmente las estructuras sintácticas de las definiciones. Es con este conocimiento incompleto con el que se pretende trabajar al responder a las peticiones del usuario.

En la figura 1 se puede ver un ejemplo de pattern utilizado para el análisis de definiciones de sustantivos y la interpretación semántica asociada. Esta interpretación semántica consiste en dos partes bien diferenciadas: la especificación de las unidades a crear –en la red de conceptos interrelacionados en la que se basará la representación –, en el caso del ejemplo el concepto definido (DEFINIENDUM) y los dos conceptos referenciados en su definición (REF1 y REF2), y la asignación de valores en diferentes slots de las unidades creadas (en este caso, en el slot DEF-SYNONYME de la unidad que representa al concepto definido se añaden los dos conceptos referenciados).

```

;;      Pattern sintáctico RN1102 (definiciones de sustantivos)
          ((M+ NOM 1) ", " (M+ NOM 2))
;;      Descripción semántica asociada:
          `(RN1102
;;      unidades a crear
          ((DEFINIENDUM)
           (REF1 REF , (CAN '(NOM 1)) (ENTITES) NOM)
           (REF2 REF , (CAN '(NOM 2)) (ENTITES) NOM))
;;      asignación de valores
          ((,DEFINIENDUM DEF-SYNONYME ,REF1)
           (,DEFINIENDUM DEF-SYNONYME ,REF2))

```

Fig.1 Ejemplo de regla semántica asociada al pattern RN1102

En la elección de los conceptos que van a formar un primer prototipo de la BCD –por razones de implementación–, se han tenido en cuenta dos criterios básicos: un criterio lexicográfico por el cual se seleccionan aquellos conceptos que son entrada en un diccionario básico del francés (G. Gougenheim, *Français Fondamental*. Didier: Paris, 1958) y un criterio gramatical por el cual se recogen solamente sustantivos, adjetivos y verbos como conceptos "definidos" para esta primera versión del sistema. Esto nos lleva a la construcción de un prototipo que va a tener como mínimo unos 2600 conceptos representados, aparte de aquellos conceptos referenciados eventualmente en las partes analizadas de las definiciones de los primeros. El proyecto se encuentra actualmente en la fase de construcción inicial de la BCD por medio de la aplicación de las reglas semánticas asociadas a los diferentes patterns.

3.- Representación del conocimiento y razonamiento.

Los procesos explicados brevemente en el punto anterior han tenido como resultado la extracción de diferentes relaciones léxico-semánticas subyacentes en la definiciones de un diccionario y la constatación de que dicha extracción puede ser realizada utilizando técnicas de análisis más o menos convencionales. Relaciones como la sinonimia, taxonómicas –hiperonimia, hponimia, taxonimia propiamente dicha–, meronómicas (parte, elemento, conjunto, miembro de), causales, etc., son relaciones que implícita o explícitamente se encuentran en los diccionarios, y particularmente, en las definiciones que los mismos dan a las diferentes entradas.

Son estas relaciones las que la representación del conocimiento a proponer debería reflejar y poner a disposición de ser explotadas por las diferentes operaciones de utilización que se van a definir más adelante.

La implementación de la BCD se realiza sobre una máquina de arquitectura simbólica, utilizándose para ello el entorno de ingeniería del conocimiento KEE. KEE es un entorno que proporciona, además de un lenguaje de representación basado en frames, diversas herramientas de razonamiento que van desde reglas –encadenamiento hacia adelante y hacia atrás–, hasta programación orientada a objetos, valores activos y razonamiento en diferentes "mundos" o contextos.

La BCD se compone de tres bases de conocimiento. BC-THESAURUS es la representación del diccionario a modo de red semántica de frames, donde cada frame o unidad representa una acepción o concepto y las unidades están relacionadas entre sí por medio de slots que representan relaciones léxico-semánticas. BC-STRUCTURES contiene meta-conocimiento sobre las unidades pertenecientes a las otras dos bases de conocimiento: en esta BC se definen las diferentes estructuras que pertenecen a la BCD, especificándose los slots o atributos que les corresponden, describiéndose a su vez los slots por medio de facetas que especifican sus rangos de valores, modos de herencia, etc. Por fin, BC-DICTIONNAIRE nos permite acceder del nivel de palabra al nivel de concepto o acepción que constituye la unidad básica de la BCD.

La representación en BC-THESAURUS se basa en diferentes tipos de unidades –se distingue básicamente entre conceptos-tipo y configuraciones (o representación de estructuras sintagmáticas asociadas a la aparición de conceptos-tipo en definiciones)– relacionados entre sí por medio de slots de naturaleza diversa. Se distinguen slots de información general, definitorios —representación de relaciones explícitas extraídas del análisis de las definiciones—, sintagmáticos, metalingüísticos y relacionales.

Desde el punto de vista de la funcionalidad, son estos slots relacionales los que nos dan una visión de la BCD como red semántica de conceptos interrelacionados. Entre estos slots se encuentran aquellos que son representación de las relaciones léxico-semánticas

mencionadas anteriormente, junto con los correspondientes a los relatores específicos obtenidos de los estudios empíricos de los que se ha hablado más arriba. Es de destacar que las relaciones taxonómicas de hiponimia e hiperonimia se representan aprovechando la relación SUBCLASS.OF (y su inversa) que el entorno nos ofrece, lo que facilita en gran manera la posterior explotación de las posibilidades de razonamiento y deducción derivadas de las mismas.

Para terminar con este aspecto de la representación, simplemente dar unas pautas de lo que consideramos va a ser la primera expresión del razonamiento en la BCD tras su construcción. Esta BCD inicial, resultado de la operación CONS-INICIAL representada en la figura 2, resultará enriquecida a continuación por la operación CONS-RELACIONAL.

En la BCD inicial construída tras la aplicación de la semántica asociada a los patterns de análisis a las definiciones provenientes de la BDD, nos encontramos con unos conceptos interrelacionados por medio de slots fundamentalmente definitorios y eventualmente sintagmáticos y/o metalingüísticos (en términos de los cuales se expresan básicamente estas reglas semánticas). La operación CONS-RELACIONAL incluye procesos de transferencia de conocimiento de los slots definitorios a los relacionales —implementados por medio de valores activos o demons, y entre los que merece destacar el proceso de construcción de taxonomías—, desambiguación de conceptos para la que se utiliza dinámicamente la propia BCD, explotación de propiedades de ciertas relaciones, etc.

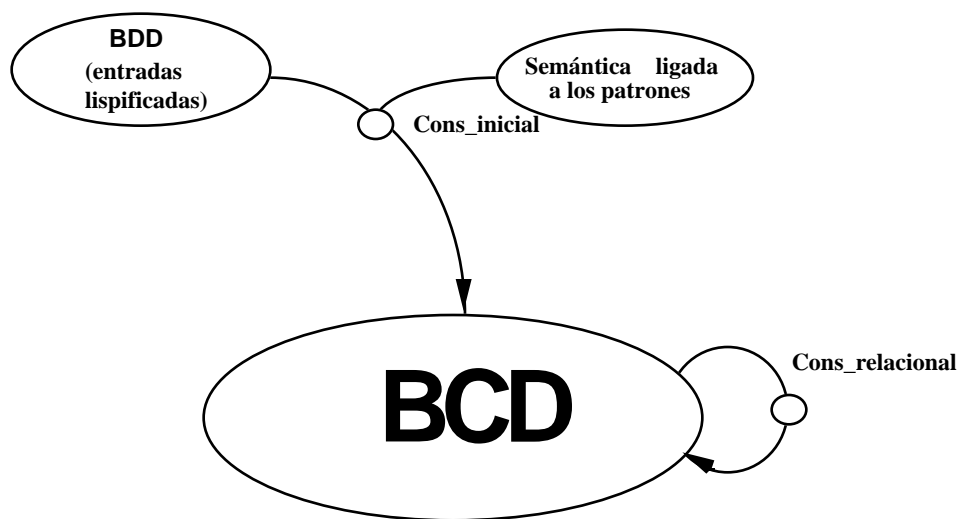


Fig. 2 Construcción de la BCD

Operación CONS_INICIAL

Datos: Entradas lispificadas (BDD)
Semántica ligada a los patterns de análisis

Resultados: Base de Conocimiento Inicial

Proceso: Para cada entrada de la BDD, aplicar la semántica asociada al pattern sintáctico correspondiente a la misma

Operación CONS_RELACIONAL

Datos: Base de Conocimientos Inicial (representación definitoria)

Resultados: Base de Conocimientos Diccionalial (representación relacional)

Proceso: Transferencia de conocimiento de slots definitorios a relacionales, construcción taxonómica, explotación de propiedades de ciertas relaciones

Mediante estas operaciones constructoras se configura un modelo de datos basado en frames. Este modelo tiene asociadas unas operaciones primitivas (Reimer et al., 83), que van a utilizarse como soporte de operaciones más complejas.

4.- Análisis de funcionalidad

Al margen de las tareas de construcción y mantenimiento de la BCD, propias de un usuario especializado, el sistema DIAC debe responder a los requerimientos de aquellos usuarios que accedan a él para recibir, verificar o aclarar algún tipo de información léxica. En principio, el usuario se expresaría en lenguaje natural sirviéndose para ello de Palabras, Conceptos, Definiciones, Atributos (propiedades genéricas de los conceptos), Valores (de los atributos), Restricciones (pueden ser tanto semánticas como sintagmáticas), Diferencias (valores, constituyentes o partes de definiciones) y Categorías (morfosintácticas) a la hora de formular sus consultas. Estas consultas serían sometidas a un procesamiento previo (pre-proceso o fase de comprensión) que las traduciría a una notación más cercana a las características de representación diccionarial. En esta notación las definiciones, restricciones y diferencias se representan mediante listas de pares atributo-valor y/o tripletas concepto-relación-concepto. En la fase previa tendría lugar, por ejemplo, la conversión de "palabras" a "conceptos" o "sentidos", unidad de información básica en la BCD. Del mismo modo, las respuestas del sistema se procesarían (post-procesamiento o fase de generación) con el fin de facilitar al usuario su lectura y comprensión.

En este análisis de funcionalidad no vamos a centrarnos en los aspectos relacionados con el pre- y post-procesamiento, que por otra parte no son objetivos del presente proyecto, y pasamos por tanto a describir la clasificación y especificación de las consultas.

Previamente es interesante considerar la definición de algunas funciones básicas:

. Definición de conceptos:

def: CONCEPTO \rightarrow DEFINICION
C \rightarrow def (C) = definición del concepto C en DIAC

. Parentesco taxonómico:

Δ : CONCEPTO x CONCEPTO \rightarrow BOOLEAN \perp
C1 , C2 \rightarrow C1 Δ C2 = T sii los dos conceptos pertenecen a la misma taxonomía.

. Proximidad semántica:

\cong : DEFINICION x DEFINICION \rightarrow BOOLEAN \perp
D1 , D2 \rightarrow D1 \cong D2 = T sii las dos definiciones son iguales o semánticamente "próximas".

Se deben establecer criterios de proximidad entre definiciones. Algunas condiciones básicas para que D1 \cong D2:

- . genus (D1) Δ genus (D2), en definiciones clásicas (genus-differentia)
- . características o atributos de genus(D1) compatibles con las de genus(D2). En una primera aproximación las características serán compatibles si no son antónimas.

. Relación atributo-valor:

σ : ATRIBUTO x CONCEPTO \rightarrow VALOR
A , C \rightarrow A σ C = V, tal que V es el valor del atributo A del concepto C. A partir de ahora denotaremos esta relación como A(C) = V

. Subsumisión taxonómica:

\sqsubseteq : TAXONOMIA x TAXONOMIA \rightarrow BOOLEAN \perp
T1 , T2 \rightarrow T1 \sqsubseteq T2 = T sii todos los conceptos de T1 están en T2.

. Satisfacibilidad de restricciones:

\models : RESTRICCION x CONCEPTO \rightarrow BOOLEAN \perp
R1 , C1 \rightarrow R1 \models C1 = T sii def(C1) cumple la restricción R1.

Nota.- BOOLEAN \perp : {true, false, \perp }, tal que el resultado de cualquier función será \perp (indefinido) si alguno de los argumentos es indefinido o el proceso de resolución no llega a buen término. Esta consideración es válida en todas las funciones booleanas que se especifiquen.

4.1.- Funciones sobre DIAC

El siguiente diagrama recoge gráficamente las operaciones que constituyen el entorno funcional de DIAC, así como los dominios y rangos que corresponden a cada una de ellas:

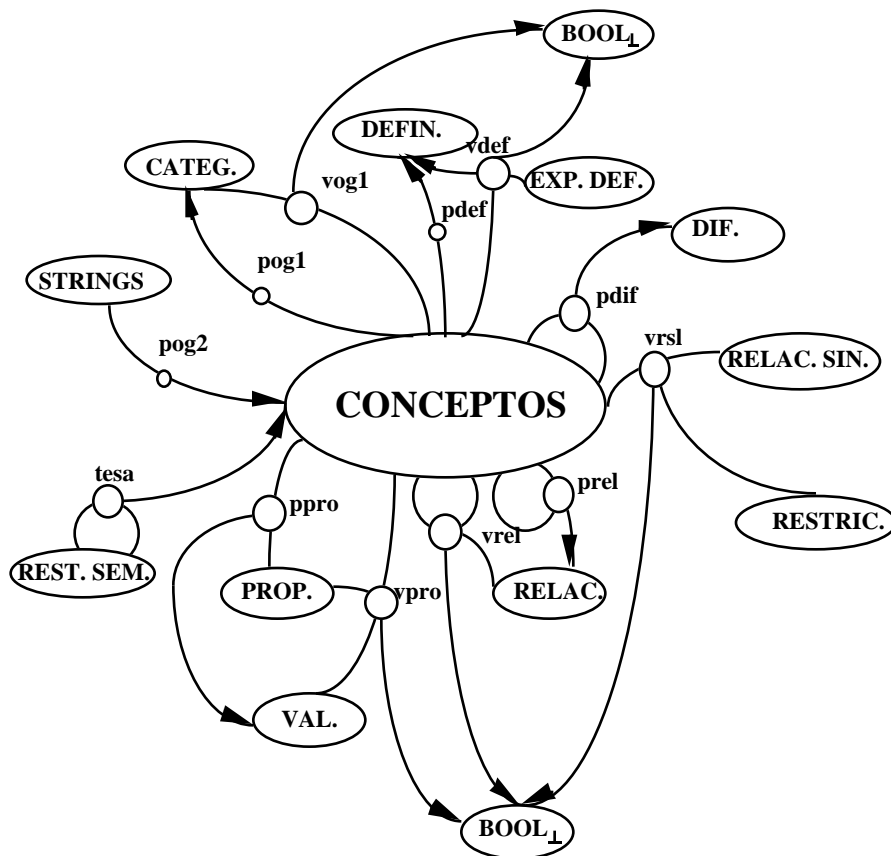


Fig. 3 Entorno funcional de DIAC

Vamos a especificar, a continuación, algunas de las operaciones más relevantes. Todos los ejemplos han sido obtenidos de LPPL.

Petición de definición/significado (PDEF):

Datos: C ∈ CONCEPTO

Resultados: D ∈ DEFINICION, tal que def (C) = D

Proceso: Consulta de significado en la BCD.

Ejemplo:

- U.- PDEF (guêpe I I).
- S.- HYPERONYME (guêpe I I) = insecte I I
- CARACTERISTIQUE (guêpe I I) = hyménoptère I I
- POSSESSION (guêpe I I) = aiguillon I I

Verificación de definición/significado (VDEF):

Datos: C ∈ CONCEPTO

D ∈ EXPR. DEFINITORIA

Resultados: true \leftrightarrow def(C) \cong D

Proceso:

- . Verificar condiciones de proximidad entre def(C) y D.
- . Generar definiciones D', tal que D' \cong def(C) y verificar condiciones de proximidad entre D' y D.

La generación de definiciones se lleva a cabo mediante estrategias de sinonimia y herencia aplicadas a los constituyentes de def(C).

Ejemplo:

- U.- **VDEF** (**HYPERONYME** (*guêpe* I 1) = *insecte* I 1
 POSSESSION (*guêpe* I 1) = *aiguillon* I 1)
- S.- Oui, **HYPERONYME** (*guêpe* I 1) = *insecte* I 1
 CARACTERISTIQUE (*guêpe* I 1) = *hyménoptère* I 1
 POSSESSION (*guêpe* I 1) = *aiguillon* I 1

Reformulación de una definición (RDEF):

Datos: D \in DEFINICION

Resultados: D' \in DEFINICION, tal que D' \cong D

Proceso:

Generación de D' mediante:

- . Sinonimia: Sustitución de los conceptos más relevantes de D por sus respectivos sinónimos.
- . Expansión: Sustitución de **HYPERONYME**(D) por def(**HYPERONYME**(D)). En lugar del genus se puede expandir algún otro constituyente básico.
- . Reformulación según el contexto: Sean D=(c1(a), c2(b), c3(d)) y def(e)=(c1(a), c2(b)), la reformulación generaría D'=(def(e), c3(d)).

Ejemplo:

- U.- **PDEF** (*email* I 1)
- S.- **HYPERONYME** (*email* I 1) = *enduit* I 1
 CARACTERISTIQUE (*email* I 1) = *vitriifiable* I 1
 POSSESSEUR (*email* I 1) = *faïence* I 1
- U.- **RDEF** (*def(email* I 1))
- S.- **HYPERONYME** (*email* I 1) = *substance* I 1
 CARACTERISTIQUE (*email* I 1) = *liquide* I 2
 CARACTERISTIQUE (*email* I 1) = *vitriifiable* I 1
 POSSESSEUR (*email* I 1) = *faïence* I 1
(mediante la consulta de *enduit* I 1: *substance liquide étendue sur*)

Petición de características/propiedades de un concepto (PPRO):

Datos: C \in CONCEPTO

A \in ATRIBUTO

Resultados: V \in VALOR, tal que A(C) = V

Proceso:

Básicamente se trata de encontrar alguna referencia al atributo A en def(C) o en alguno de sus "semejantes" y obtener el valor correspondiente. Las estrategias que se siguen para la resolución de esta operación son:

- . Recurrir a def(C) y ver si A se indica explícitamente.
- . Recurrir a def(A) y obtener sinónimos o hiperónimos de A, dado que alguno de ellos podría aparecer explícitamente en def(C).
- . Examinar los hiperónimos de C y las propiedades heredadas.
- . Analizar los valores que aparecen en def(C) y comprobar si son adecuados respecto de A.
- . Tratar los sinónimos de C.

Ejemplo:

U.-	PPRO (<i>couleur (ambre I 1)</i>)
S.-	<i>jaune I 1</i>

U.-	PPRO (<i>synonyme (grand I 1)</i>)
S.-	<i>gros I 1</i>

Verificación de características/propiedades de un concepto (VPRO):

Datos: C ∈ CONCEPTO
A ∈ ATRIBUTO
V ∈ VALOR

Resultados: true ↔ A(C) = V

Ejemplo:

U.-	VPRO (<i>couleur (or I 1) = rouge I 2</i>)
S.-	<i>faux</i>

U.-	VPRO (<i>synonyme (grand I 1) = gros I 1</i>)
S.-	<i>vrai</i>

U.-	VPRO (<i>classe (souris I 1) = rongeur I 2</i>)
S.-	<i>vrai</i>

Proceso:

Consiste en detectar referencias al atributo A o al valor V y verificar si se cumple la propiedad A(C)=V. Las estrategias seguidas son similares a las de la petición de propiedades.

Búsqueda de relaciones entre dos conceptos (PREL):

Datos: C1, C2 ∈ CONCEPTO

Resultados: R ∈ RELACION, tal que C1 R C2

Proceso:

Se debe encontrar algún camino en la red semántica partiendo de C1 y de C2. Este tipo de recorridos originan un costo enorme (difícilmente implementable), lo que hace imprescindible la utilización de heurísticos que reduzcan el espacio de búsqueda. Estos heurísticos deben posibilitar un análisis eficiente de las relaciones, en particular, de las taxonómicas y

sinonímicas. Es importante considerar aquí la participación interactiva del usuario como guía en el proceso de búsqueda.

Ejemplo:

U.- **PREL** (*guêpe I 1* , *miel I 1*)
 S.- RESULTAT-INV (*miel I 1*) = *produire I 1\$18*
 PREDICATION (*abeille I 1*) = *produire I 1\$18*
 (*produire I 1\$18*: config. que representa "produire le miel")
 HYPERONYME (*abeille I 1*) = *insecte I 1*
 CARACTERISTIQUE (*abeille I 1*) = *hyménoptère I 1*
 HYPERONYME (*guêpe I 1*) = *insecte I 1*
 CARACTERISTIQUE (*guêpe I 1*) = *hyménoptère I 1*

Verificación de relaciones entre dos conceptos (VREL):

Datos: C1, C2 ∈ CONCEPTO
 R ∈ RELACION

Resultados: true ↔ C1 R C2

Proceso:

Similar a la búsqueda de relaciones pero con la ventaja de que el recorrido por la red, en este caso, está dirigido por la relación.

Ejemplo:

U.- **VREL** (*abeille I 1* , *miel I 1* ,
 <PREDICATION (*abeille I 1*) = *produire I 1*
 RESULTAT (*produire I 1*) = *miel I 1*>)
 S.- Oui, PREDICATION (*abeille I 1*) = *produire I 1\$18*
 RESULTAT (*produire I 1\$18*) = *miel I 1*

Búsqueda de diferencias entre dos conceptos (PDIF):

Datos: C1, C2 ∈ CONCEPTO

Resultados: DIF1, DIF2 ∈ DIFERENCIA, tal que def(C1)-DIF1 ≅ def(C2)-DIF2

Precondición: C1 Δ C2

Proceso:

Se examinan las diferencias existentes a partir del hiperónimo común o, cuando se trata de sinonimia parcial, en las acepciones no-sinonímicas. De alguna forma se debe determinar la parte común de las definiciones y el resto es la diferencia.

Ejemplo:

U.- **PDIF** (*bicyclette I 1* , *tricycle I 1*)
 S.- POSSESSION (*bicyclette I 1*) = *roue I 1\$3*
 (*roue I 1\$3*: config. que representa "deux roues")
 POSSESSION (*tricycle I 1*) = *roue I 1\$5*
 (*roue I 1\$5*: config. que representa "trois roues")

 U.- **PDIF** (*congre I 1* , *perche I 1*)
 S.- LOCATIF (*congre I 1*) = *mer I 1*
 LOCATIF (*perche I 1*) = *eau I 2\$4*
 (*eau I 2\$4*: config. que representa "eau douce")

 U.- **PDIF** (*laver I 1* , *balayer I 1*)

- S.- INSTRUMENT (*laver I 1*) = *liquide I 1*
 INSTRUMENT (*balayer I 1*) = *balai I 1*

Consulta tesaúrica (THES):

Datos: LR=<R1, R2, ..., Rn>, tal que todos los Ri ∈ RESTRICCION

Resultados: LC=<C1, C2, ..., Cm>, tal que para cualquier Cj se cumple: R1 ∧ ... ∧ Rn ⊨ Cj

Proceso:

- . Determinación de la taxonomía T1, de forma que $\forall C \in T1$ se verifique $R1 \models C$.
- . A continuación detección de la taxonomía T2, tal que $T2 \subseteq T1$ y $\forall C \in T2: R1 \wedge R2 \models C$.
- . Así sucesivamente hasta que $\forall C \in Tn: R1 \wedge \dots \wedge Rn \models C$. Eventualmente Tn puede ser un único concepto o, incluso, la taxonomía vacía.
- . Todo concepto que pertenece a Tn está en LC.
- . Debe tenerse en cuenta que LR puede ser una definición D, donde los constituyentes de D son implícitamente restricciones de selección.

Ejemplo:

- U.- **THES** (<HYPERONYME (X) = *poisson I 1* >)
 S.- *anguille I 1, bar I 2, barbu I 2, brochet I 1, congre I 1, ...*
- U.- **THES** (<HYPERONYME (X) = *instrument I 1*,
 OBJECTIF (X) = *mesurer I 1*>)
 S.- *baromètre I 1, dynamomètre I 1, télémètre I 1, ...*

Nota.- Se permite un uso análogo de variables en VREL (para la expresión de relaciones).

Otras utilidades:

En el diagrama funcional se representan, además de las que ya se han especificado, otras funciones que vamos a reseñar brevemente:

- . Consulta de restricciones de selección (VRSL): Verifica si se puede dar una determinada relación sintagmática entre un concepto y los que cumplen una determinada restricción.
- . Ayuda a la paráfrasis: Reformula una frase dada por el usuario. Es muy parecida a la reformulación de definiciones, pero en este caso la frase reformulada no tiene por qué ser una definición.
- . Funcionalidad ortogramatical: Se incluyen en este grupo todas las funciones que hacen referencia a aspectos gramaticales y de ortografía. Aunque no son complejas, tienen especial interés en la composición de funciones. A modo de ejemplo, podrían utilizarse junto con la consulta tesaúrica para la resolución de crucigramas.
- . Composición de funciones: La posibilidad de combinar distintas funciones amplía considerablemente la capacidad del sistema. Además, la composición puede ser enriquecida mediante la utilización de estructuras condicionales o incluso iterativas. Es decir, se trataría de un lenguaje de consulta de la BCD, lo que supondría un cambio cualitativo en el modo de utilización de los diccionarios.

5.- Conclusiones

Se ha descrito el proceso de extracción de conocimiento semántico de un diccionario monolingüe convencional. Como resultado de este proceso se ha llegado a la especificación de diferentes relaciones léxico-semánticas subyacentes principalmente en la componente definitoria del diccionario. Estas relaciones constituyen el fundamento de la propuesta de representación de conocimiento que se hace para la Base de Conocimiento Diccional. A diferencia de la mayoría de sistemas en el área en los que el objetivo es más la construcción de la componente léxica de un entorno de procesamiento de lenguaje natural, DIAC se concibe como herramienta de ayuda para un usuario humano.

Bibliografía

- Alshawi, H. "Processing dictionary definitions with Phrasal Pattern Hierarchies", *Computational Linguistics* 13, 3-4, 195-202. 1987.
- Alshawi, H. "Analysing dictionary definitions" in B. Boguraev, T. Briscoe eds., 153-169, *Computational Lexicography for Natural Language Processing*. New York: Longman, 1989.
- Alterman, R. "A dictionary based on concept coherence", *Artificial Intelligence*, 25, 153-186. 1985.
- Amsler, R.A. *The Structure of The Merriam-Webster Pocket Dictionary*. Ph. D. Dissertation, Computer Science. University of Texas, Austin. 1980.
- Amsler, R.A. "A Taxonomy for English Nouns and Verbs", *Proc. 19th Annual Meeting ACL*, 133-138. 1981.
- Amsler, R.A. "Lexical Knowledge Bases", *Proc. COLING (Stanford Univ.)*, 458-459. 1984.
- Arango Gaviría, G. *Une approche pour amorcer le processus de compréhension et d'utilisation du sens des mots en langage naturel*. Thèse de 3^e cycle (Paris VI). Publications du Groupe de Recherche Claude François Picard. 1983.
- Artola, X., Evrard, F. "Dictionnaire intelligente d'aide à la compréhension". Com. seleccionada para publicación en Actas EURALEX (Benalmádena), 1990.
- Binot, J.L., Jensen, K. "A semantic expert using an on-line standard dictionary", *Proc. IJCAI*, 709-714. 1987.
- Boguraev, B., Briscoe, T. eds. *Computational Lexicography for Natural Language Processing*. New York: Longman, 1989.
- Brustkern, J., Hess, K.D. "The BONNLEX lexicon system" in J. Goetschalckx, L. Rolling eds., 33-40, *Lexicography in the electronic age*. Luxembourg: North-Holland, 1982.
- Byrd, R.J., Calzolari, N., Chodorow, M.S., Klavans, J.L., Neff, M.S., Rizk, O.A. "Tools and Methods for Computational Lexicography" *Computational Linguistics* 13, 3-4, 219-240. 1987.
- Calzolari, N., Pecchia, L., Zampolli, A. "Working on the italian machine dictionary: a semantic approach" in A. Zampolli, N. Calzolari eds. *Proceedings of the International Conference on Computational Linguistics* (Pisa, 1973). Vol. 37, 49-69. 1980.
- Calzolari, N. "Machine-readable dictionaries, lexical data bases and the lexical system", *Proc. COLING (Stanford Univ.)*, p. 460. 1984a.
- Calzolari, N. "Detecting patterns in a lexical data base", *Proc. COLING (Stanford Univ.)*, 170-173. 1984b.

- Calzolari, N., Picchi, E. "Acquisition of semantic information from an on-line dictionary", *Proc. COLING* (Budapest), 87-92. 1988.
- Chodorow, M.S., Byrd, R.J. "Extracting semantic hierarchies from a large on-line dictionary", *Proc. ACL*, 299-304. 1985.
- Chodorow, M.S., Ravin, Y., Sachar, H.E. "A tool for investigating the synonymy relation in a sense disambiguated thesaurus", *Proc. 2nd conference on Applied Natural Language Processing* (Austin), 144-153. 1988.
- Chouraqui, E., Godbert, E. "Représentation des descriptions définies dans un réseau sémantique", *Actes 7ème Congrès Reconnaissance des Formes et Intelligence Artificielle* (AFCET-INRIA, Paris), 855-868. 1989.
- Copestake, A. "An approach to building the hierarchical element of a lexical knowledge base from a machine-readable dictionary", com. presentada en *First Int. Workshop on Inheritance in NLP* (Tilburg). 1990.
- Ferrari, G. "Dictionnaire automatique et dictionnaire-machine: une hypothèse" in A. Zampolli, N. Calzolari eds. *Proceedings of the International Conference on Computational Linguistics* (Pisa, 1973). Vol. 36, 257-262. 1980.
- Gosselin, M. *Le bon usage des dictionnaires*, Les éditions la liberté, Sainte-Foy (Québec), 1989.
- van den Hurk, I., Meijs, W. "The dictionary as a corpus: analyzing LDOCE's definition-language", *Corpus Linguistics II*, 99-125.
- Kay, M. "The dictionary server", *Proc. COLING* (Stanford Univ.). 1984.
- Litkowsky, K. C. "Models of the semantic structure of dictionaries", *American Journal of Computational Linguistics*, Mf. 81, 25-74. 1978.
- Litkowsky, K. C. "Requirements of text processing lexicons", *Proc. 18th Annual Conference of the ACL* (Philadelphia, Pennsylvania), 153-154. 1980.
- Markowitz, J., Ahlswede, T., Evens, M. "Semantically significant patterns in dictionary definitions", *Proc. 24th Annual Meeting ACL* (New York), 112-119. 1986.
- Michiels, A., Noël, J. "Approaches to thesaurus production" in J. Horecky ed., 227-232, *COLING 82*. Academia: North-Holland, 1982a.
- Michiels, A., Mullenders, J., Noël, J. "The LONGMAN-LIEGE project" in J. Goetschalckx, L. Rolling eds., 201-210, *Lexicography in the electronic age*. Luxembourg: North-Holland, 1982b.
- Nagao, M., Tsujii, J., Ueda, Y. et Takiyama, M. "An attempt to computerize dictionary data bases" in J. Goetschalckx, L. Rolling eds., 51-73, *Lexicography in the electronic age*. Luxembourg: North-Holland, 1982.
- Papegaaij, B.C., Sadler, V., Witkam, A.P.M. "Experiments with an MT-directed lexical knowledge bank", *Proc. COLING* (Bonn), 432-434. 1986.
- Pazienza, M.T., Velardi, P. "A structured representation of word-senses for semantic analysis", *Proc. 3rd European Conference ACL* (Copenhague), 249-257. 1987.
- Quillian, M.R. "Semantic Memory" in M. Minsky ed., 227-270, *Semantic Information Processing*. Cambridge (Mass.): MIT Press, 1968.
- Reimer, U., Hahn, U. "A formal approach to the semantics of a frame data model" *Procs. of IJCAI'83*, 337-339, 1983.
- Tsurumaru, H., Hitaka, T., Yoshida, S. "An attempt to automatic thesaurus construction from an ordinary japanese language dictionary", *Proc. COLING* (Bonn), 445-447. 1986.

- Vossen, P., Meijs W., den Broeder, M. "Meaning and structure in dictionary definitions" in B. Boguraev, T. Briscoe eds., 171-192, *Computational Lexicography for Natural Language Processing*. New York: Longman, 1989.
- Vossen, P., Serail, I. *DEVIL: A taxonomy-browser for decomposition via the lexicon*. Techn. Report. Faculty of Arts, Univ. of Amsterdam, 1990.
- Webber, H. R. "Machine-readable components in variety of information-systems applications" *Proc. COLING'84* (Stanford). 1984.
- Winston, M., Chaffin, R., Herrmann, D. "A taxonomy of part-role relation", *Cognitive Science*, vol. 11, 417-444. 1987.
- Yoshida, S., Tsurumaru, H., Hitaka, T. "Man-assisted machine construction of a semantic dictionary for natural language processing" in J. Horecky ed., 419-424, *COLING 82*. Academia: North-Holland, 1982.