EL ANALIZADOR MORFOLOGICO DEL ARABE DE UN SISTEMA DE TA ARABE-ESPAÑOL

I INTRODUCCION

Lo que aquí presentamos es un Analizador Morfológico (AM) del Arabe que se ha implementado en el marco de un proyecto de T.A. Arabe-Español. En un sistema de TA en el que intervenga la Lengua Arabe, cobra especial importancia el módulo de Análisis Morfológico por cuanto es una lengua muy flexiva y derivacional siendo necesario descomponer la mayoría de palabras árabes en raíces y patrones (esquemas de vocalización).

Al hablar del Arabe nos referimos al MSA (Modern Standard Arabic), que es un puente entre el clásico del Corán y los dialectos que se hablan en los diferentes países, y que es utilizado en los medios de comunicación en las relaciones internacionales, etc.

CARACTERISTICAS DEL ARABE:

Es una lengua semítica y como tal, en cuanto a la grafía:

a) La escritura es cursiva. No tiene mayúsculas y casi todas las letras tienen cuatro formas diferentes según lugar de la palabra en el que estén (a principio de palabra, a mitad, al final o aisladas).

b) No se vocaliza y se utilizan poco los signos de puntuación.

c) La escritura es de derecha a izquierda.

En cuanto a la morfología:

a) Se declina y tiene duales.

b) Muy derivativa. La mayoría de las palabras derivan de raíces formadas, normalmente, por tres consonantes (radicales) y en algunos casos por dos, cuatro y, muy raramente, por cinco. En Arabe es difícil separar la morfología flexiva de la derivacional que, además, es aparentemente muy regular y consiste básicamente en la aplicación de un conjunto predeterminado de patrones a raíces triconsonánticas con el fin de formar unos lexemas a los que, a su vez, se afijan una serie de prefijos y sufijos, también predeterminados.

Por lo general, cada patrón, además de proporcionar información de tipo morfológico y sintáctico, también nos da información semántica.

Así, por ejemplo, la raíz "KTB" confiere la idea de "escribir" a las palabras en las que puede ser encontrada y la raíz "QTL" la idea de "matar". El 1, 2 y 3 de los esquemas corresponden a la primera, segunda y tercera radical respectivamente.

Esquema: **1a2a3**

"KTB" ---- "**KaTaB**" ---- Trad.: escribir

"QTL" --- -"**QaTaL**" ---- Trad.: matar

Esquema: **1a'2a3**

"KTB" ---- "**Ka'TaB**" ---- Trad.: mantener correspondencia con

"QTL" ---- "**Qa'TaL**" ---- Trad.: combatir con

Esquema: **1i2a'3**

"KTB" ---- "**KiTa'B**" ---- Trad.: libro

"QTL" ---- "**QiTa'L**" ------ Trad.: combate

Esquema: **ma12a3a#**

"KTB" ---- "**maKTaBa#**" ---- Trad.: biblioteca

"QTL" ---- "**maQTaLa#**" ---- Trad.: masacre

Hay, como se ve, una cierta correspondencia con lenguas indoeuropeas como el español, pero aquí la derivación afecta a la forma interna de la raíz, que en muchos casos llega a ser irreconocible y, además, la derivación se hace tanto por medio de prefijos y sufijos como de infijos.

Ejemplo: "TR" --> "**TaRa**" Trad.: tu ves

Raíz: "R?Y"

Forma Vocalizada: **TaRa**"

Prefijo: "ta"

Esquema Regular: **12a3**

Esquema final: **1a**

c) Es muy flexiva. Posee prefijos, sufijos e infijos.

d) Es aglutinante. Una partícula interrogativa, conjunciones, preposiciones, el artículo y los pronombres personales se unen a la palabra.

e) Las radicales a veces se transforman, se pierden e, incluso, se fusionan siguiendo unas normas morfofonológicas.

Y en cuanto a la sintaxis:

a) Respecto al Orden, aunque se considera una lengua VSO en realidad es bastante libre.

b) Las concordancias en género y número presentan muchas particularidades, dependiendo, sobre todo, de orden.

II PROBLEMAS COMPUTACIONALES QUE REPRESENTAN ESTAS CARACTERISTICAS

a) Al no estar vocalizado, la ambigüedad de análisis es enorme. Teniendo como entrada una forma superficial no vocalizada, el Analizador Morfológico debe proporcionar como salida todas las posibles palabras legales que correspondan a dicha forma.

"FHM" ----->	"FaHiMa"	comprender (él comprende) Act.
	"FuHiMa"	ser comprendido (él/ello es comprendido) Pas.
	"FaHMu"	comprensión (Nom)
	"FaHMa"	comprensión (Ac)
	"FaHMi"	comprensión (Gen)
	"FaHiMu"	sagaz (Nom)
	"FaHiMa"	sagaz (Ac)
	"FaHuM"	y ellos (Fa + Hum)
	"FaHiM"	y enamórate (Fa + HiM) Conj+Imperativo, Raiz: HYM

b) La falta de estandarización a nivel grafémico, es decir, el uso libre de algunos signos diacríticos (omisión o cambio de forma) en la escritura del MSA, aumenta la ambigüedad.

c) Para analizar una palabra se ha de reconocer su raíz, lo que a veces es difícil porque se ha perdido o transformado alguna(s) de las radicales.

Esquema: **1a2a3 + tu** (suf. 1ª sg, Perfecto Ind. Activa)

"QWL" --- **"QaWaL" + tu ("QaWaLtu")** ---> **"QuL" + tu ("QuLtu")** Trad.: dije

(Forma "regular" predecible) (Forma real encontrada)

d) Aunque todos los patrones (o esquemas) son potencialmente derivables de cualquier raíz, éstas toman, idiosincráticamente, sólo alguno de ellos, y para saber cuáles son hay que consultar obligatoriamente el diccionario. Hay que tener en cuenta que, entre nominales y verbales, el nº de patrones existentes supera los quinientos. En la práctica, pocas raíces aceptan más de 20 esquemas.

e) El trabajar con caracteres árabes y el que la lengua se escriba de derecha a izquierda presenta graves problemas de incompatibilidad de software.

III ENFOQUES PARA LA RESOLUCION DE ESTOS PROBLEMAS

1) Para analizar morfológicamente una palabra, hay que vocalizarla. Para ello, se necesitan conocimientos de tipo ortográfico, morfológico, sintáctico, semántico e incluso pragmático. Tras el análisis morfológico, permanecen muchas ambigüedades que se irán eliminando en las sucesivas fases del análisis, puesto que, en muchos casos, las vocales que caracterizan el caso, el tiempo y modo verbal y la indeterminación, no pueden decidirse antes del sintáctico.

2) Hay que estandarizar el texto que se quiere analizar: forzar la presencia del 'tashdiid' en formas 2 y 5, unificar las distintas grafías de la letra 'ya', etc...

3) Hay que transliterar y por ello necesitamos un conversor que nos convierta los caracteres árabes e transliterados y viceversa.

4) Como ya hemos dicho, el AM tiene que proporcionar la raíz de la palabra que se quiere analizar y el patr superimpuesto a dicha raíz. Y ello es debido, entre otras razones, a que el léxico monolingüe está organiza por raíces y no por lemas ("stems") y por lo tanto, la clave de acceso al mismo será la raíz y no la for canónica (o el lema) como sucede en otras lenguas.

El léxico árabe de análisis sigue un enfoque basado en raíces debido, entre otras razones, a la dificultad averiguar el singular que corresponde a un determinado patrón de Plural Fracto. El Plural Fracto árabe - diferencia del llamado Plural Sano que consiste en un sufijo específico de plural y que, por consiguiente, reducible a un lema - es un patrón de plural completamente nuevo formado por derivación no lineal a partir las mismas tres o cuatro radicales del singular. Por lo tanto, es casi siempre imposible deducir su singular (e es, su lema) ya que cada patrón de Plural Fracto tiene un número bastante elevado de potenciales singular Así pues, la única forma de encontrar en el léxico Arabe monolingüe una entrada léxica conociendo únicame su patrón de plural es a través de la raíz que nos proporciona el análisis morfológico y que constituye, co ya hemos mencionado, la entrada léxica de este diccionario.

Ejemplo: **"KuTuB"** (plural fracto) ----> **"KiTa'Ba#"**

"KiTa'B"

"KuTTa'B"

"KaTiYB"

"KaTiYBa#"

"Ka'TiB"

"KaTB"

"KiTBa#"...

(singulares reales que podrían tomar ese patrón de plural,

aunque en realidad sólo **"KiTa'B"** lo hace)

IV DESCRIPCION DE UN ANALIZADOR MORFOLOGICO DEL ARABE

A) *Características desde el punto de vista lingüístico:*

Se analizan palabras sin vocalizar introducidas indistintamente en árabe o en transcripción. En este último caso, se efectúa una conversión, previa al análisis, ya que el sistema trabaja internamente con caracteres transliterados. Obtenemos esa misma palabra con sus posibles vocalizaciones, su raíz y un listado de las partes que la componen (afijos) con su información lingüística asociada.

Se han considerado afijos: partícula interrogativa, conjunciones, partículas prefijadas, artículo, prefijos y sufijos verbales, sufijos nominales (plurales sanos y duales) y pronombres sufijados. Los infijos y las aformativas de las formas verbales y de los nombres (en el sentido árabe tradicional del término) forman parte de los patrones.

Ejemplo: **"maKTaBa#"** = ma12a3a#

"taKa'TaB" = ta1a'2a3

A partir del Análisis Morfológico de la palabra:

- Para el Verbo obtenemos, antes de acceder al léxico: tiempo y modo, persona, voz, categoría, raíz y afijos, forma, "bab" (diferentes vocalizaciones de la forma I en perfecto e imperfecto) y la palabra base vocalizada.

- Para el Nombre obtenemos, antes de acceder al léxico: género, número, casos posibles, categoría, subcategoría, raíz y afijos, rasgos de tipo semántico y la palabra base vocalizada.

La información lingüística se encuentra en:

a) Un Léxico Cerrado de Partículas y Palabras Bilíteras con su información asociada.

b) Un Léxico Abierto de Palabras Extranjeras, Nombres Propios y Palabras Quintilíteras (5 radicales).

c) Un Léxico Abierto de Raíces con los patrones nominales y verbales que éstas aceptan y con su información lingüística asociada (Léxico General).

d) Una Lista de Patrones con su información asociada de tipo ortográfico, morfológico, sintáctico y semántico.

PASOS QUE SIGUE EL PROCESO DE ANALISIS:

El proceso no es lineal sino que va alternando la búsqueda a la izquierda y a la derecha de la palabra segu criterio.

1. Comprueba que la palabra no esté en el léxico cerrado de partículas y bilíteras;
2. Mira si la palabra tiene una terminación única de nombre (como es la "ta marbuuta" o el acusati indeterminado con alif de contención);
3. Busca, separa y almacena pronombres sufijados (uno o dos);
4. Identifica y separa: a) partícula interrogativa
   b) conjunciones
   c) partículas prefijadas;
5. Comprueba la existencia del artículo, lo separa y almacena la información obtenida;
6. Busca, separa y almacena la información de sufijos nominales;
7. Busca, separa y almacena la información de prefijos y/o sufijos verbales;
8. Mira y selecciona las posibles vocalizaciones de los afijos;
9. Lo que queda de la palabra, habiendo extraído todo lo anterior, es comparado con una lis esquemas o patrones vocalizados, obteniendo las posibles raíces y patrones que encajan;
10. Comprueba la compatiblidad entre prefijos y sufijos y entre éstos y los patrones identific unificando la información lingüística que les acompaña.
11. Accede al léxico de raíces (mediante una tabla de acceso léxico rápido) para comprobar que esa raíz (o raíces) y, en el caso del nombre, que ésta admite el esquema de nombre identificado el caso del verbo, que la raíz admite la forma y el "bab" obtenidos;
12. Consulta el léxico de palabras quintilíteras y de nombres extranjeros.

EJEMPLO: "FSYKTTBHWH'" Trad.: y se lo hará escribir

PASOS:

1) No está en el léxico de Partículas y Palabras Bilíteras
2) No tiene terminación nominal
3) 1er. Pronombre: "h'", 2º. Pronombre: "hw" sufijados
4) a) No hay partícula interrogativa
   b) Conjunción: "f"
   c) Partícula: "s" (partícula que indica futuro y exige tiempo Impf)
5) No hay artículo
6) No hay sufijo nominal
7) Prefijo Verbal: "y", Sufijo: ""
8) "h'" = "ha'"
   "hw" = "huw"
   "f" = "fa"
   "s" = "sa"
   "y" = "yu", "ya"
   " " = "u", "a", "".
9) "kttb" ---> **1a22i3** ---> **"KaTTiB"** ---> RAIZ: "KTB"
   1u22i3 ---> **"KuTTiB"** ---> RAIZ: "KTB"
   1a22a3 ---> **"KaTTaB"** ---> RAIZ: "KTB"
   1u22a3 ---> **"KuTTaB"** ---> RAIZ: "KTB"
   ...
10) Comprueba la compatiblidad de afijos y patrones y obtiene como resultado lo siguiente:
   "yu"--"**KaTTiB**"--"u" (Forma:II, 3ms Impf. Ind, Act, Raíz:KTB)
   "yu"--"**KaTTiB**"--"a" (Forma:II, 3ms Subjuntivo Act, Raíz:KTB)
   "yu"--"**KaTTiB**"--"" (Forma:II, 3ms Yusivo Act, Raíz:KTB)
   "yu"--"**KaTTaB**"--"u" (Forma:II, 3ms Impf. Ind. Pas, Raíz:KTB)
   "yu"--"**KaTTaB**"--"a" (Forma:II, 3ms Subjuntivo Pas, Raíz:KTB)
   "yu"--"**KaTTaB**"--"" (Forma:II, 3ms Yusivo Pas, Raíz:KTB)

11) Accede al léxico monolingüe donde encuentra que existe la raíz: "KTB" y que además acepta los patrones "1a22i3" y "1a22a3". Recupera la información lingüística asociada a la entrada léxica.
12) No encuentra la palabra base en el léxico de Nombres Extranjeros y Quintilíteras.

OUTPUT1: FORMA VOCALIZADA:"FaSaYuKaTTiBuHuWHa'"
RAIZ: "KTB"
FORM: II
CONJ: "fa"
PART: "sa" (indica: aspecto futuro, verbo finito en imperfecto)
PREF: "yu" (indica: 3ªms imperfecto indicativo)
PALBASE: **"KaTTiB"**
SUF: "u" (indica: 3ªms imperfecto indicativo)
PRON1: "huw" (pronombre personal 3ªms)
PRON2: "ha'" (pronombre personal 3ªfs)
TMOD: Imperfecto indicativo
PERS: 3ªms

ESPECIFICACIONES ADICIONALES:

a) Existen numerosas restricciones de tipo ortográfico, como por ejemplo la de incompatibilidad entre alguna consonantes, entre consonante y vocal, etc.

b) En cada paso existe la posibilidad de "backtracking", es decir, como los diferentes segmentos identificable pueden formar parte de la raíz o incluso del esquema, el sistema proporciona los diferentes análisis en orde de mayor a menor probabilidad.

EJEMPLO: "TRJM"

a) Análisis 1: Forma Vocaliz:"**TaRJuMu**" Trad.: ella apedrea

Raíz: "RJM"

Pref: "ta"

PalBase: "RJM"

Suf: "u"

b) Análisis 2: Forma Vocaliz: "**TaRJaMa**" Trad.: él tradujo

Raíz: "TRJM"

Pref: --

PalBase: "TRJM"

Suf: "a"

c) Los esquemas llevan como información asociada una serie de reglas ortográficas y morfológicas que permiten obtener la raíz, especialmente cuando se trata de raíces enfermas (irregulares). Estas reglas se varios tipos:

- Reglas de instanciación cuando falta o cambia una radical:

Ejemplo: rg1(Nslot):- (es_y(Nslot), es_w(Nslot)).

Dado un slot cualquiera, instanciarlo como "w" o como "y".

Esta regla aplicada:

Esquema ($1a'3$, [rg1(N2)]). --> la Radical 2 puede ser "y" o "w" en los verbos cóncavos.

Ex. "Q'L" puede pertenecer al verbo "QYL" o al verbo "QWL" por lo que "'" se instancia como "y" o como "w".

- Reglas de vocalización (pérdida o substitución de vocales o sukún):

Ejemplo: rg13(Vocal,Nvocal1,Nvocal2):- Nvocal1 =Vocal, pierde_sukun(Nvocal2).

Dada la vocal asociada a un slot, pasa a estar asociada a otro slot, perdiéndose todo signo sobre el primero.

Esta regla aplicada:

Esquema ($12u3$, [rg13(v2,nv1,nv2)]). --> Así, en el caso de los verbos cóncavos, "yaQ*WuL" queda convertido en "yaQuWL".

- Reglas de incompatibilidades (consonante-consonante, vocal-consonante):

Ejemplo: rg11 (Nslot):- no_es_w(Slot).

Dado un slot cualquiera, nunca puede ser "w".

Esta regla aplicada:

Esquema($12i3$, [rg11(slot2)]). La "i" del esquema exige que el slot2 nunca sea "w".

Ex. En el caso de la forma gráfica "yQWL" , cuya 2ª Radical es "W", y que podría tomar los esquemas "12i3" y "12u3", la regla rechazará el primero.

B) *Características desde el punto de vista informático: Arquitectura y Funcionamiento.*

El analizador ha sido escrito en Arity PROLOG (Arity prolog, Arity Co):

Toda la información lingüística necesaria para su funcionamiento (esquemas, información léxica, reglas, etc...) esta almacenada en forma de cláusulas prolog. El espacio requerido para almacenar dichas cláusulas es considerablemente grande. Para su manipulación Arity ofrece:

- Gestión transparente de memoria virtual. El espacio máximo que podemos consumir sólo viene

limitado por la capacidad de nuestro disco duro.

- Formas eficientes para acceder a dicha información (e.g. árboles B+ para el acceso a la informació léxica y tablas de hash para el acceso a los esquemas)

Se trabaja con un Shell del DOS en Arabe (SAKHR, Alalamiah Software Co.) en el cual se ha modificado página de códigos y la estructura del teclado para permitir la introducción indistinta del texto a traducir en ára y transliterado.

Estructura General:

El analizador está dividido en tres módulos, a saber:

- Un primer módulo que incluye el motor del analizador morfológico.

- Un módulo intermedio que ofrece una interficie al módulo motor del analizador para que éste pu acceder inteligentemente (y de forma transparente) a la información lingüística.

- Un último módulo para la gestión de la estructura de análisis morfológico (instanciación y cons de rasgos).

La implementación del algoritmo de análisis de una palabra árabe (módulo motor) es, a grandes rasgos, c sigue:

1) Extraer partículas afijadas a la palabra base.

Una partícula afijada en su forma más general tiene la siguiente estructura:

```
particula_afijada(      Partícula,
                        Lista_rasgos_morfológicos,
                        Lista_tipos_afijos_incompatibles,
                        Vocalización1(
                                Lista_rasgos_morfológicos ),
                        (...)
                        VocalizaciónN(
                                Lista_rasgos_morfológicos )

).
```

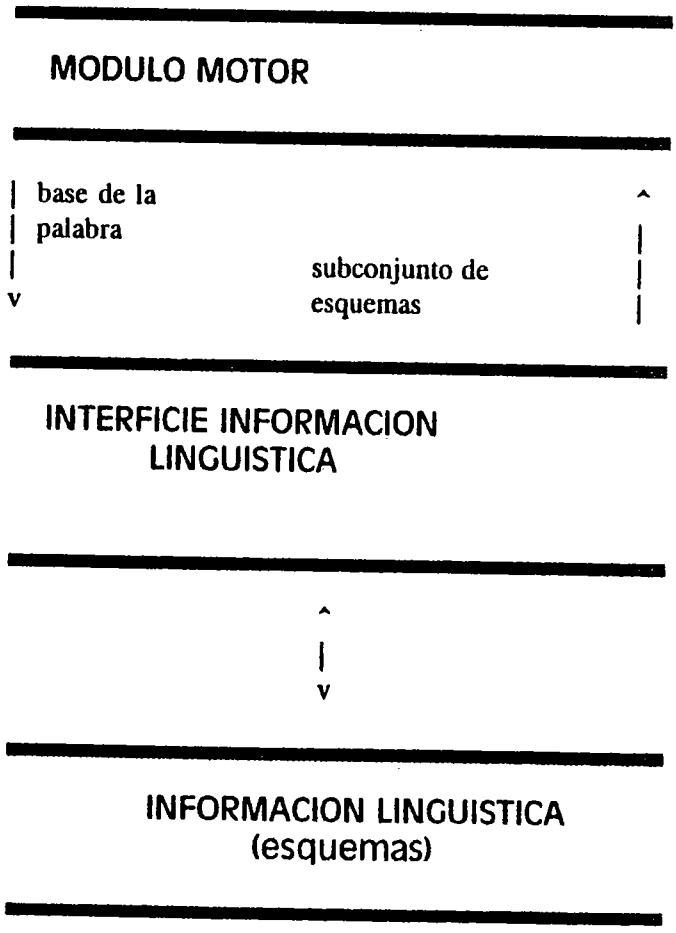
- Se especifica el orden en que debemos extraer los diferentes tipos de afijos (así como cuáles de ellos son optativos).

- Algunas incompatibilidades entre tipos de afijos (combinaciones de afijos no válidas), se indican como información adicional asociada a los mismos afijos, aunque la mayoría de combinaciones inválidas no serán ni tan siquiera consideradas.

- Al extraer un afijo, se comprueba si habíamos extraído previamente un afijo incompatible con él, y se instancian en la estructura de análisis los rasgos morfológicos asociados.

2) Cotejar la base de la palabra con un conjunto de esquemas viables y obtener la(s) vocalizacion(es).

- Los esquemas con los que se efectúe el proceso de cotejado serán únicamente aquéllos que nos sugiera el módulo intermedio (entre los que siempre estarán todos los válidos).



- El acceso a los esquemas se efectúa a través de la interficie de acceso mediante una llamada de la forma:

consultar_esquemas_viables(Base_palabra, Esquemas_viables)

(Indicando la base de la palabra, se obtiene el subconjunto de esquemas potencialmente válidos)

Ejemplo:

Si la base de la palabra es "**K'TB**", el módulo motor recibe como esquemas potencialmente válidos (de entr un número total de esquemas que puede superar los quinientos):

- **1a'2a3**
- **1a'2i3**

En este caso, los dos esquemas obtenidos son efectivamente válidos. En general, muchos de ellos deberán s descartados posteriormente como se explica más abajo.

Un esquema en su forma más general tiene la siguiente estructura:

```
esquema(        Esquema,
                Lista_reglas_morfológicas,
                Lista_rasgos_morfológicos,
                opciónA(
                        Lista_reglas_morfológicas,
                        Lista_rasgos_morfológicos),
                opciónB(
                (...)
).
```

- Las razones por las que pueden descartarse los esquemas obtenidos son varias:

a) El esquema no es cotejable con la base. (e.g. la base de la palabra es: "**K'TB**" y uno de esquemas seleccionados es "**1a2a'3**" . Aunque el esquema tiene el mismo número de elementos que la bas la palabra a analizar, la distinta posición del <'> los hace absolutamente incomparables).

b) No se puede aplicar a la base de la palabra alguna de las reglas morfológicas asociadas. (Ex. nos encontramos ante un esquema regular, y se obliga a que el tercer slot del esquema - 3ª radical de la raíz - sea diferente del segundo)

Ejemplo:

"**MRR**" -- no es cotejable con el patrón --- **1a2a3** ya que, según la regla: 2\==3

"**MRR**" -- es cotejable con el patrón -- **1a22**

c) No se puede instanciar alguno de los rasgos indicados en el esquema. (Ex. se trata de un esquema de 'perfecto' y hemos extraído previamente un afijo que indicaba 'yusivo').

3) Seleccionar una de las posibles vocalizaciones para cada uno de los afijos extraídos previamente (instanciando la información morfológica asociada a la vocalización en particular) y concatenarlos junto con la base vocalizada para obtener la vocalización completa de la palabra.

4) Detectar incompatibilidades ortográficas a partir de la vocalización completa de la palabra y de una lista de reglas ortográficas. En este paso, descartamos análisis espurios motivados por la no inclusión de estas reglas en los propios esquemas (estas reglas ortográficas se aplican independientemente del esquema, por ser comunes a todos ellos).

a) El hecho de que la información lingüística sea completamente independiente de los otros módulos:

- Permite la posibilidad de añadir nueva información y/o modificar la ya existente (p.e. añadir un esquema nuevo, o modificar la definición de una de las reglas morfológicas asociadas a uno antiguo) sin que esto comporte tener que hacer ningún cambio en los otros módulos.

- Libera al lingüista de la obligación de conocer los detalles de implementación del motor del analizador permitiéndole modificar libremente la información lingüística.

b) El acceso del módulo motor a la información lingüística a través de una interfície intermedia y no directamente:

- Permite reestructurar la información lingüística en profundidad, siempre y cuando mantengamos invariable la especificación de dicha interficie.

c) Tener un módulo para la gestión de una estructura de análisis morfológico permite que:

- Modificar la implementación de la estructura no afecte a los otros módulos.

- Añadir nuevos rasgos (p.e. a petición del lingüista) sólo suponga tener que modificar la estructura ligeramente y advertir al módulo motor de la existencia de nuevos predicados de consulta/instanciación de estos nuevos rasgos.

CONCLUSIONES:

Por la propia naturaleza del Arabe, el problema de la vocalización de una palabra va más allá del análisi morfológico de la misma, dado que se necesita información sintáctica, semántica e incluso pragmática par desambiguar totalmente todas las posibles vocalizaciones.

Pero es el Analizador Morfológico el núcleo de este proceso de vocalización y el que se ha presentado es u AM capaz de tratar con éxito tanto la morfología flexiva como la derivacional, extrayendo de una palabra n sólo los prefijos y sufijos, sino también la raíz subyacente y el patrón de vocalización aplicado a dicha raíz

Asimismo, en el diseño de este Analizador Morfológico se ha tenido en cuenta no sólo su utilización en Sistema de T.A. (tanto para analizar como para ayudar en la creación dinámica del léxico), sino también posible uso en un entorno CALL (Computer Assisted Language Learning).

REFERENCIAS

Hasta la fecha, se han realizado diversos AM para el Arabe con el mismo enfoque, generalmente, o basándose en el modelo de Koskenniemi (dos niveles), debido a las características de la Lengua. La implementación se ha hecho, sobre todo, en Prolog y Lisp.

Beesley, K.R. 1990. Finite-State Description of Arabic Morphology, in: *The 2nd Conference and Exhibition on Bilingual Computing in Arabic and English*, University of Cambridge.

Ben Hamadou,A.& Saidi,L. 1986. Analyse Morpho-Syntaxique de la Langue Arabe, in:*8ème Séminaire Tuniso-Français d'Informatique et Intelligence Artificielle*, Tunis 5-7 Mai.

Degachi,A. & Smith,J.CH. 1990. The Morphological Component of an Arabic Parser, in: *The 2nd Conference and Exhibition on Bilingual Computing in Arabic and English*, University of Cambridge.

ElSadany,T. & Hashish,M.A. 1987. Arabic Morphological System, IBM Cairo Scientific Center Report.

Keddag,A. 1992. Arabic Morpho-Syntax and Semantic Parsing, in: *Proceedings of the 3rd International Conference on Multilingual Computing (Arabic and Roman Script)*, University of Durham, 10-12 December.

IBM 1987: Arabic Morphological Analyser, in: *Conference on NLP*, Denmark.

Narayanan,A. & Hashem,L. 1992. A Three-Level Finite-State Model for Arabic Morphology, in: *Proceedings of the 3rd International Conference and Exhibition on Multilingual Computing*, University of Durham, 10-12 December.

Koochnik,P. 1989. ARABTN: A Morpho-Syntactic Parser for Arabic, in: *The First Seminar on Bilingual Computing in Arabic and English*, University of Cambridge.