

SEGMORF: UN FORMALISMO PARA ANALIZADORES MORFOLÓGICOS DE DOS NIVELES

Toni Badia, Àngels Egea, Antoni Tuells
Institut Universitari de Lingüística Aplicada
Universitat Pompeu Fabra

La Rambla 30, Barcelona 08002

tbadia@upf.es, ANGELS@slc.ub.es, tuells@upf.es

Abstract

In this paper we show the problems that the present two-level morphological formalisms have in dealing with several morphological phenomena involving interaction between two-level rules (TLR's) and the word grammar (WG). On the one hand, we demonstrate the non-modularity of their WG, since it cannot be designed according to linguistic criteria only. On the other hand, we show how they do not provide a clean and general solution for this kind of morphological phenomena in Romance languages. We therefore propose a new implemented formalism, SEGMORF, that allows the linguist to express in a natural way the interaction between TLR's and the WG, thus making the whole system more modular. SEGMORF is being used in the implementation of a TLM analyzer for Catalan; some results in applying it to Catalan are also showed.

Keywords: morphological analysis

1 MOTIVACIÓN

Para la morfología de dos niveles (TLM, presentada en [Koskenniemi, 1984]), la morfografía se establece mediante reglas de dos niveles (RDN) y la morfotáctica se caracteriza mediante clases de continuación o mediante una gramática de unificación para construir palabras (GP). Separando los dos procesos, se pretende que se puedan expresar de manera natural ambas clases de relaciones. No obstante, como las RDNs intentan encontrar una descomposición léxica de una cadena superficial y como la GP intenta construir una estructura de palabra a partir de aquella descomposición, hay varios fenómenos morfológicos que plantean problemas importantes.

Por un lado, la GP parece requerir información sobre la aplicación de las RDNs, como en los casos (a)-(b) (para seleccionar la descomposición correcta, debería tener información sobre qué regla se ha aplicado).

(a) *puedo* (verbo,1a,sing,presente) \Rightarrow *pod* (raíz_verbal) + *o* (1a,sing,presente)

(b) *pudo* (verbo,3a,sing,pasado) \Rightarrow *pod* (raíz_verbal) + *o* (3a,sing,pasado)

Por el otro, como es bien sabido, hay fenómenos morfológicos, como el Umlautung alemán, que restringen el contexto morfológico en que se aplican (algunas RDNs se aplican sólo en unos procesos concretos de formación de palabra).

Varios fenómenos morfológicos, pues, deberían ser tratados mediante una buena interacción entre las RDNs y la GP. Hasta el momento, la manera habitual de tratar esta interacción es la propuesta por [Trost, 1990], que prevé que:

- la información sobre la regla aplicada tiene que ser traspasada a la GP
- hay que poder restringir la aplicación de algunas RDNs a ciertas clases de morfemas.

En esta propuesta, la transferencia de la información se concibe como la unificación de la estructura de rasgos asociada a la regla con la estructura de rasgos asociada al morfema léxico. Si tiene lugar la unificación, se transfiere la estructura de rasgos resultante a la GP. Aunque esta propuesta funciona, presenta algunos problemas, especialmente en relación al primer requisito:

- la GP no está motivada independientemente: a menudo se la concibe como un proceso que valida la aplicación de las RDNs y que filtra las descomposiciones erróneas. Por lo tanto, su formulación depende de las RDNs.
- la GP es difícil de manejar, puesto que tiene que adaptarse a la interacción de las distintas reglas que seleccionan su propio contexto.

Ambos puntos expresan el mismo hecho: la GP no es modular. Considérese el siguiente fenómeno del catalán:

(c) *espès* (adj,masc,sing)

(d) *espessos* (adj,masc,pl) \Rightarrow *espès* + *s* (pl)

(e) *abús* (nombre,masc,sing)

(f) *abusos* (nombre,masc,pl) \Rightarrow *abús* + *s* (pl)

Evidentemente, las reglas que introducen el acento gráfico o duplican la "s" están motivadas morfológicamente. Según la propuesta de Trost la GP debería tratar la interacción entre reglas, o sea debería tener en cuenta si se han aplicado las reglas que introducen el acento (o duplican la "s"). Y como esta validación tiene que hacerse para varios procesos de formación de palabra, la complejidad de esta tarea puede aumentar enormemente.

Es importante notar que [Karttunen *et al.*, 1992] no resuelve este problema. El hecho de establecer niveles intermedios para relacionar la cadena superficial y la forma léxica, no influye en ningún sentido en la interacción entre las RDNs (o de tres o más niveles) y la GP.

2 NUESTRA PROPUESTA

Para superar las dificultades anteriores, el tratamiento de los fenómenos morfológicos en que interviene la interacción entre los componentes morfografémico y morfotáctico en el marco de la morfología de dos niveles tiene que cumplir con los siguientes requisitos:

1. la GP debería ser lo más modular (independiente) posible
2. hay que poder restringir la aplicación de las RDNs a ciertas clases de morfemas (segundo punto señalado por Trost).
3. debería ser posible especificar RDNs de manera que el contexto morfológico sea tenido en cuenta (junto al morfografémico).

Las RDNs en SEGMORF tienen en cuenta estos requisitos. En lo que sigue nos centramos en los puntos 1 y 3 (el punto 2 ya ha sido tratado anteriormente: [Bear, 1988], [Trost, 1990]). A continuación, mostramos algunas de las reglas para tratar los casos anteriores (a)-(f), que mostrarán la expresividad de las RDNs de SEGMORF.

Desde un punto de vista procedural, se necesitan las siguientes estructuras de datos:

- contextos morfografémicos superficiales izquierdo y derecho
- contextos morfografémicos léxicos izquierdo y derecho
- contextos morfológicos izquierdo y derecho
- contexto de aplicación (una estructura de rasgos que controla la aplicación de las reglas. Por ejemplo [umlautung:y]). Esta estructura de rasgos debe unificar con la estructura de rasgos asociada con cada morfema extraído del léxico.

La aplicación de una regla se valida de una manera homogénea en relación con todos los contextos. Nótese que, aunque se necesitan más reglas, son más fáciles de especificar y que el contexto morfológico ayuda a eliminar antes la aplicación de las reglas no aplicables, evitando así sobregeneración. Asimismo la GP no tiene que tratar la aplicación de las RDNs. Su labor principal es la de validar las descomposiciones léxicas; no se necesita información sobre las reglas, cosa que conlleva una gramática más modular (simple). Para tratar los casos de "pudo" - "puedo" se necesitan las reglas siguientes:

```
Rule diftongacion: {
    ue <=> o
    morphemes_to_be_found = (raiz_verbal +
                             (sufijo_verbal, presente))
}
```

```
Rule cierre: {
    u <=> o
    morphemes_to_be_found = (raiz_verbal +
                             (sufijo_verbal, pasado))
}
```

Nótese que son las propias reglas las que seleccionan los morfemas adecuados; la desambiguación se hace lo más pronto posible (de hecho, al consultar el diccionario); la GP no tiene que comprobar qué regla se ha aplicado.

Finalmente, la siguiente regla da cuenta de los casos (c)-(f), en los que la presencia del acento no depende del contexto superficial, sino del proceso de formación de palabra: aparece siempre que se añade un sufijo nominal a la raíz. Sólo mostramos la regla que añade el acento:

```

Rule poner_acento : {
  Vocal<=>Vocal_acentuada
  surface_right_context = sX /* X = variable de caracter */
  lexical_right_context = s+ /* "+" = limite morfemático */
  morphemes_to_be_found = (nombre + sufijo_nominal)
}

```

Nótese que la regla "poner_acento" no necesita tener complejos contextos léxicos especificados, ni requiere una compleja GP.

2.1 Conclusiones

Las ventajas y consecuencias de nuestra propuesta son las siguientes:

1. La GP es un verdadero módulo independiente, puesto que las RDNs seleccionan los morfemas adecuados siempre que puede aparecer un conflicto. Así, el diseño de la GP puede atender sólo a consideraciones de formación de palabra.
2. SEGMORF permite al lingüista expresar en las RDNs el contexto morfológico adecuado, de manera que se desambigua lo más pronto posible y se mejora la eficiencia del sistema.
3. SEGMORF proporciona modos lingüísticamente motivados de expresar las relaciones morfografémicas y morfotácticas, evitando así algunos tratamientos no naturales.

3 SEGMORF. EL FORMALISMO

SEGMORF es una extensión al formalismo de segmentación morfografémico del sistema Alep ([CEC, 1994]); su principal característica es que las RDNs permiten especificar tanto el contexto morfografémico como el morfotáctico que restringen su aplicación.

3.1 El Contexto Morfografémico

Para empezar introducimos las siguientes abreviaturas con sus correspondientes significados:¹

- *SLC* = Contexto superficial izquierdo (Surface Left Context)
- *LHS* = Left Hand Side
- *SRC* = Contexto superficial derecho (Surface Right Context)
- *LLC* = Contexto léxico izquierdo (Lexical Left Context)
- *RHS* = Right Hand Side
- *LRC* = Contexto léxico derecho (Lexical Right Context)

¹Lo que se describe en este subapartado es una reformulación de [Alshawi *et al.*, 1991], pág 161.

donde SLC, LHS, SRC, LLC, RHS y LRC son una secuencia de cero o más caracteres, o variables cuyo valor puede ser un carácter. Las variables se interpretan via unificación. Es posible, además, restringir el rango de su valor; por ejemplo, el usuario puede especificar que el valor de una determinada variable debe ser una vocal.

A partir de las definiciones anteriores, el contexto morfografémico de las reglas TLM y su correspondiente interpretación pueden especificarse de las siguientes maneras²:

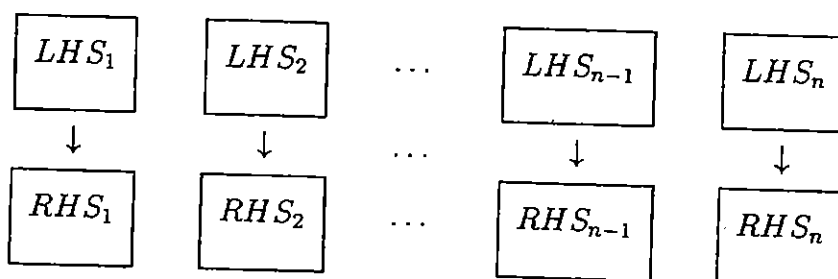
1. $SLC LHS SRC \implies LLC RHS LRC$
Si se satisfacen los contextos superficial y léxico, entonces si reconocemos la secuencia LHS a nivel superficial, a ésta le debe corresponder la secuencia RHS a nivel léxico.
2. $SLC LHS SRC \Leftarrow LLC RHS LRC$
Si se satisfacen los contextos superficial y léxico, entonces si reconocemos la secuencia RHS a nivel léxico, a esta le debe corresponder la secuencia LHS a nivel superficial.
3. $SLC LHS SRC \iff LLC RHS LRC$
Si se satisfacen los contextos superficial y léxico, entonces a la secuencia LHS le debe corresponder la secuencia RHS, y viceversa.
4. $SLC LHS SRC \text{ opt } LLC RHS LRC$
Si se satisfacen los contextos superficial y léxico, entonces a la secuencia LHS le puede corresponder la secuencia RHS, y viceversa.

Ya que los contextos, incluyendo el LHS y el RHS, pueden tener un número diferente de caracteres, las reglas TLM pueden aplicarse sobre más de un carácter al mismo tiempo. Algunos caracteres están reservados en SEGMORF:

- + marca explícitamente el final del morfema
- # marca explícitamente el final del string superficial

Naturalmente, ya que ciertas reglas son idiosincráticas de ciertos lexemas, se permite que éstos puedan bloquear la aplicación de determinadas reglas a través de la unificación de rasgos (ver por ejemplo [Bear, 1988] y [Trost, 1990]).

La aplicación de las reglas produce particiones de los strings léxico y superficial, donde el string léxico es una secuencia de morfemas y el string superficial una secuencia de caracteres. Una partición se interpreta como una división de los strings en secuencias de cero o más caracteres, de manera que haya idéntico número de secuencias en ambos strings. Es decir, a la *secuencia_i* a nivel superficial le corresponde la *secuencia_i* a nivel léxico:



²Esta es una de las diferencias principales de SEGMORF respecto al componente de dos niveles del sistema Alep. En Alep, únicamente los operadores 2 y 3 están disponibles; el primero se interpreta como regla opcional y el segundo como regla obligatoria.

```

Rule poner_acento : {
    Vocal<=>Vocal_acentuada
    surface_right_context = sX /* X = variable de caracter */
    lexical_right_context = s+ /* "+" = limite morfemático */
    morphemes_to_be_found = (nombre + sufijo_nominal)
}
    
```

Nótese que la regla "*poner_acento*" no necesita tener complejos contextos léxicos especificados, ni requiere una compleja GP.

2.1 Conclusiones

Las ventajas y consecuencias de nuestra propuesta son las siguientes:

1. La GP es un verdadero módulo independiente, puesto que las RDNs seleccionan los morfemas adecuados siempre que puede aparecer un conflicto. Así, el diseño de la GP puede atender sólo a consideraciones de formación de palabra.
2. SEGMORF permite al lingüista expresar en las RDNs el contexto morfológico adecuado, de manera que se desambigua lo más pronto posible y se mejora la eficiencia del sistema.
3. SEGMORF proporciona modos lingüísticamente motivados de expresar las relaciones morfografémicas y morfotácticas, evitando así algunos tratamientos no naturales.

3 SEGMORF. EL FORMALISMO

SEGMORF es una extensión al formalismo de segmentación morfografémico del sistema Alep ([CEC, 1994]); su principal característica es que las RDNs permiten especificar tanto el contexto morfografémico como el morfotáctico que restringen su aplicación.

3.1 El Contexto Morfografémico

Para empezar introducimos las siguientes abreviaturas con sus correspondientes significados:¹

- *SLC* = Contexto superficial izquierdo (Surface Left Context)
- *LHS* = Left Hand Side
- *SRC* = Contexto superficial derecho (Surface Right Context)
- *LLC* = Contexto léxico izquierdo (Lexical Left Context)
- *RHS* = Right Hand Side
- *LRC* = Contexto léxico derecho (Lexical Right Context)

¹Lo que se describe en este subapartado es una reformulación de [Alshawi *et al.*, 1991], pág 161.

donde SLC, LHS, SRC, LLC, RHS y LRC son una secuencia de cero o más caracteres, o variables cuyo valor puede ser un carácter. Las variables se interpretan via unificación. Es posible, además, restringir el rango de su valor; por ejemplo, el usuario puede especificar que el valor de una determinada variable debe ser una vocal.

A partir de las definiciones anteriores, el contexto morfografémico de las reglas TLM y su correspondiente interpretación pueden especificarse de las siguientes maneras²:

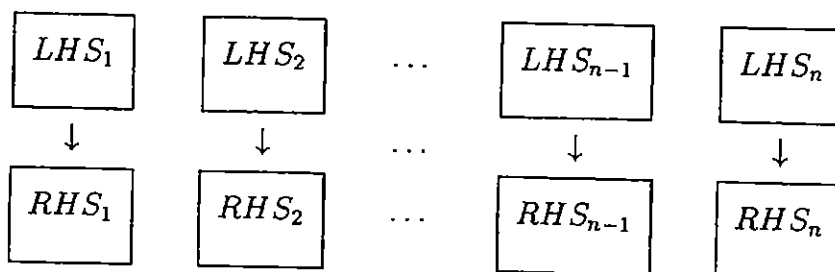
1. $SLC LHS SRC \implies LLC RHS LRC$
Si se satisfacen los contextos superficial y léxico, entonces si reconocemos la secuencia LHS a nivel superficial, a ésta le debe corresponder la secuencia RHS a nivel léxico.
2. $SLC LHS SRC \Leftarrow LLC RHS LRC$
Si se satisfacen los contextos superficial y léxico, entonces si reconocemos la secuencia RHS a nivel léxico, a esta le debe corresponder la secuencia LHS a nivel superficial.
3. $SLC LHS SRC \iff LLC RHS LRC$
Si se satisfacen los contextos superficial y léxico, entonces a la secuencia LHS le debe corresponder la secuencia RHS, y viceversa.
4. $SLC LHS SRC \text{ opt } LLC RHS LRC$
Si se satisfacen los contextos superficial y léxico, entonces a la secuencia LHS le puede corresponder la secuencia RHS, y viceversa.

Ya que los contextos, incluyendo el LHS y el RHS, pueden tener un número diferente de caracteres, las reglas TLM pueden aplicarse sobre más de un carácter al mismo tiempo. Algunos caracteres están reservados en SEGMORF:

- + marca explícitamente el final del morfema
- # marca explícitamente el final del string superficial

Naturalmente, ya que ciertas reglas son idiosincráticas de ciertos lexemas, se permite que éstos puedan bloquear la aplicación de determinadas reglas a través de la unificación de rasgos (ver por ejemplo [Bear, 1988] y [Trost, 1990]).

La aplicación de las reglas produce particiones de los strings léxico y superficial, donde el string léxico es una secuencia de morfemas y el string superficial una secuencia de caracteres. Una partición se interpreta como una división de los strings en secuencias de cero o más caracteres, de manera que haya idéntico número de secuencias en ambos strings. Es decir, a la *secuencia_i* a nivel superficial le corresponde la *secuencia_i* a nivel léxico:



²Esta es una de las diferencias principales de SEGMORF respecto al componente de dos niveles del sistema Alep. En Alep, únicamente los operadores 2 y 3 están disponibles; el primero se interpreta como regla opcional y el segundo como regla obligatoria.

3.2 El contexto morfotáctico

Dadas las siguientes abreviaturas:

- *LMC* = Contexto morfotáctico izquierdo (Left Morphotactical Context)
- *RMC* = Contexto morfotáctico derecho (Right Morphotactical Context)

definimos un contexto morfotáctico como un par $\langle LMC, RMC \rangle$, donde tanto LMC como RMC son una secuencia de cero o más descripciones morfotácticas. Una descripción morfotáctica incluye información morfosintáctica simple, como POS, concordancia, ...

Para que la aplicación de una regla sea válida los morfemas ya encontrados deben satisfacer el LMC de la regla, y el resto de morfemas que se obtendrán posteriormente deben satisfacer el RMC. Nótese que la posibilidad de asociar contextos morfotácticos a las reglas de dos niveles es una de las novedades principales de nuestro formalismo. Esta característica no está presente en el componente de dos niveles del sistema Alep.

3.3 Ejemplos de reglas

```
Rule identidad: {
  [] [X] [] opt [] [X] []
  {X not in [+,#]}      /* Se restringe el valor de X */
  [] []                 /* Sin contexto morfotactico */
}
```

```
Rule absorcion_a: {
  [] [a] [#] <= [] [e,+,a] []
  [] [adj + fem]
}
```

Fijémonos en que la regla de la identidad no especifica ningún contexto morfotáctico especial; por el contrario, la regla "absorcion_a" sí lo hace. Ésto permite que, por ejemplo, la forma incorrecta para el catalán "menyspra (menyspre + a (3ra sing, pres))" no sea ni reconocida ni generada. Puesto que las reglas de dos niveles del sistema Alep no permiten especificar contextos morfotácticos, lo máximo que se podría hacer es asociar un tipo a la regla "absorcion_a". Este tipo unificaría con el tipo de las entradas léxicas de adjetivos, pero no con el de las entradas verbales.

3.4 La Gramática de Palabra

La idea del formalismo SEGMORF es que éste permita mantener la GP independiente respecto al contexto morfografémico, es decir, que la GP no necesite información sobre qué reglas se han aplicado para tratar fenómenos como los que se describen en la sección primera. Por ejemplo, para tratar los casos "puedo-pudo" del apartado 2, en una arquitectura como la del sistema Alep³, y puesto que las reglas no permiten la especificación de contextos morfotácticos, no hay más remedio que transmitir información sobre qué regla se ha aplicado a la gramática de la palabra. Por razones de eficiencia, una GP parcial

³La arquitectura del sistema Alep está pensada para todos los estadios de procesamiento: desde el text handling hasta el análisis semántico de oraciones. Por el contrario, SEGMORF ha sido pensado para suministrar solamente información de tipo morfológica.

que elimine segmentaciones parciales erróneas resulta útil (sugerido ya, entre otros, por [Trost, 1991]).

3.5 Comentarios sobre la expresividad de SEGMORF

Aunque la eficiencia no ha sido un criterio decisivo para diseñar SEGMORF, ésta sí ha sido tenida en cuenta. Como resultado, algunas herramientas expresivas no han sido incorporadas cuando se ha creído que éstas podían perjudicar el rendimiento del sistema. Por ejemplo, para describir algunos fenómenos parecía necesario saber cuantas sílabas tenía el lexema que se estaba tratando; en vez de utilizar expresiones regulares para expresar este hecho, hemos preferido conseguir el mismo efecto marcando léxicamente el número de grupos vocálicos del lexema⁴. Por motivos similares, las descripciones morfotácticas no incorporan la utilización de tipos; puesto que se supone que éstas deben tener información morfosintáctica muy simple, nos ha parecido que la incorporación de tipos perjudicaría el rendimiento de SEGMORF.

4 APLICACIÓN DE SEGMORF AL CATALÁN

En este apartado mostramos los resultados obtenidos al aplicar SEGMORF al tratamiento de la morfología verbal catalana.

Se ha partido de la base que un verbo se compone de dos partes: la raíz, que es invariable para toda la flexión, y el sufijo, que varía según el modo, el tiempo, el número o la persona de cada forma verbal. Aunque hay algunas propuestas de descomposición de los morfemas flexivos en afijos separados ([Badia Margarit, 1994]), hemos optado por seguir la práctica habitual en los tratamientos computacionales de la morfología verbal catalana ([Martí, 1988]; [Hernández, 1993]) y hemos considerado un único sufijo para cada forma.

Tradicionalmente, los verbos catalanes se han clasificado en tres conjugaciones, según la terminación de la forma de infinitivo:

- la conjugación en *-ar*
- 2a conjugación en *-er* o *-re*
- 3a conjugación en *-ir*

Así, la gramática tradicional tiende a establecer un número muy pequeño de modelos de conjugación. Por ejemplo, en [Fabra, 1979] encontramos un modelo para la primera conjugación, dos para la segunda y dos más para la tercera. Este tratamiento es insuficiente para dar cuenta de la diversidad real de paradigmas verbales, si bien tiene la virtud de agrupar los más próximos entre sí. Esta deficiencia descriptiva se suple con una larga lista de verbos con la flexión completa.

El tratamiento que hasta ahora ha recibido la morfología verbal catalana desde un punto de vista computacional presenta, contrariamente a la gramática tradicional, un elevado número de modelos de conjugación (42 en [Martí, 1988], 94 en [Hernández, 1993]), de manera que cada verbo queda asignado a uno de estos modelos. El inconveniente de

⁴El número de grupos vocálicos de un morfema es predecible a partir de su ortografía

este análisis es el número de sufijos resultante, añadido al hecho de que muchos de estos sufijos son redundantes, porque pertenecen a varios modelos de conjugación a la vez.

Nuestro tratamiento de la morfología verbal aprovecha las posibilidades expresivas del formalismo de descripción de dos niveles, concretamente la capacidad de tratar los cambios grafémicos en los morfemas cuando se combinan con otros morfemas. De este modo se ha logrado reducir notablemente el número de entradas léxicas (raíces, sufijos y formas verbales lexicalizadas) y los modelos de conjugación.

Los resultados obtenidos mediante la aplicación de nuestro modelo son los siguientes:

1. En cuanto a la primera conjugación, muchas de las diferencias entre paradigmas verbales son predictibles por la forma de la raíz verbal y, por consiguiente, explicables por reglas. La descripción de esta conjugación se ha resuelto con un solo modelo de conjugación y 10 reglas.
2. En cuanto a la segunda y tercera conjugaciones, no existen paradigmas predictibles a partir de uno de más general. Como consecuencia, no se han podido establecer reglas; pero se ha logrado una descripción simple y no redundante de la totalidad de estos dos grupos de verbos mediante la agrupación bajo un mismo modelo de verbos que presentan sólo algunas diferencias respecto al modelo definido. Con ello quedan agrupados los verbos de características flexivas parecidas. Las diferencias restantes son lexicalizadas. Con esta estrategia, han bastado 19 modelos para la descripción completa de la segunda y tercera conjugaciones.

5 CONCLUSIONES

Hemos presentado un nuevo formalismo de TLM que permite expresar naturalmente los fenómenos morfológicos en que hay interacción entre las RDNs y la GP. Hemos mostrado también que nuestro formalismo permite que la GP sea bastante modular, simplificándose así su construcción.

References

- [Alshawi *et al.*, 1991] H. Alshawi, D.J. Arnold, R. Backofen, D.M. Carter, J. Lindop, K. Netter, S. Pulman, J. Tsuji and H. Uszkoreit, CEC. 1991. Eurotra ET6/1:Rule Formalism and Virtual Design Study (final Report).
- [Badia Margarit, 1994] A.M. Badia Margarit 1994. *Gramàtica Catalana*. Enciclopèdia Catalana (Biblioteca universitària, 22).
- [Bear, 1988] J. Bear. 1988. Morphology with Two level rules and Negative Rule Features. In *Proceedings of the 12th International Conference on Computational Linguistics (COLING-88)*, Budapest, Hungary
- [CEC, 1994] CEC. 1994. The Alep Linguistic System.
- [Fabra, 1979] P. Fabra. 1979. *Gramàtica Catalana*. Editorial Teide, Barcelona.
- [Hernández, 1993] L. Hernández. 1993. *Tots els verbs catalans*, Editorial Teide, Barcelona.

- [Karttunen *et al.*, 1992] L. Karttunen, R.M. Kaplan, A. Zaenen 1992. Two-Level Morphology with Composition In *Proceedings of the 14th International Conference on Computational Linguistics (COLING-92)*, Nantes, Francia
- [Koskenniemi, 1984] K. Koskenniemi. 1984. A General Computational Model for Word-form Recognition and Production. In *Proceedings of the 10th International Conference on Computational Linguistics (COLING-84)*, Stanford, U.S.A
- [Martí, 1988] M.A. Martí. 1988. Processament del llenguatge natural: un sistema d'anàlisi morfològica per ordinador. Ph.D Thesis, Departament de Filologia Romànica, Facultat de Filologia de la Universitat de Barcelona.
- [Trost, 1990] H. Trost. 1990. The Application of two-level morphology to non-concatenative German morphology. In *Proceedings of the 13th International Conference on Computational Linguistics (COLING-90)*, pages 371-376, Helsinki, Finland.
- [Trost, 1991] H. Trost. 1991. X2MORF: A Morphological Component Based on Two-level Morphology. Research Report DFKI-RR-91-04, Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI), Saarbrücken, Germany.