A FINITE-STATE MORPHOLOGICAL PROCESSOR FOR SPANISH

Evelyne Tzoukermann and Mark Y. Liberman
AT&T Bell Laboratories
600 Mountain Avenue
Murray Hill, NJ 07974

Abstract

A finite transducer that processes Spanish inflectional and derivational morphology is presented. The system handles both generation and analysis of tens of millions inflected forms. Lexical and surface (orthographic) representations of the words are linked by a program that interprets a finite directed graph whose arcs are labelled by n-tuples of strings. Each of about 55,000 base forms requires at least one arc in the graph. Representing the inflectional and derivational possibilities for these forms imposed an overhead of only about 3000 additional arcs, of which about 2500 represent (phonologically-predictable) stem allomorphy, so that we pay a storage price of about 5% for compiling these forms offline. A simple interpreter for the resulting automaton processes several hundred words per second on a Sun4.

1 Introduction

One useful way to look at computational morphology and phonology is in terms of **transductions**, that is, n-ary word relations definable by the element-wise concatenation of n-tuple labels along paths in a finite directed labeled graph. For instance, we can take one member of such a relation to be the spelling of an inflected form, another member to be the corresponding lemma, another to be a string representing its morphosyntactic features, another to represent its pronunciation, and so forth.

Inspired by the (unpublished) work of Kaplan and Kay (ongoing since the late 1970's), and that of Koskenniemi in [12], many researchers have used binary word relations to represent "underlying" and "surface" forms in the morphophonology of words. Much of the interest of this work has been focused on methods to combine multiple two-tape automata, which may be composed or run in parallel in order to compute the desired binary relation.

In this paper, we take a somewhat different approach to defining and computing word relations, and discuss its application in a morphological processor for Spanish orthographic words that covers more than forty millions forms generable from the approximately 55,000 basic words in the Collins Spanish Dictionary ([3]) [1]. The main advantage of this approach is the extreme simplicity both of its data structures and of their interpretation. As a result, an interpreter is easy to implement; time and/or space optimization issues in the implementation are straightforward to define; at the same time, it is extremely easy to compile traditional morphological information into the required form, at least for languages like Spanish that can be fairly well modeled in terms of the concatenation of stems and affixes. As is usually the case in automata-based approaches, the system treats analysis and generation symmetrically, and the same description can be run with equal facility in either direction.

Define[2] an *n-ary nondeterministic finite automaton* as a 5-tuple

$$\mathcal{A} = (Q, q_1, F, \Sigma, H)$$

where Q is a finite non-empty set of states, q_1 is a designated start state, F is a set of designated final states, Σ is a finite non-empty alphabet, and H is a finite subset of $\{Q \times (\Sigma^*)^n \times Q\}$, where $(\Sigma^*)^n$ is the set of n-tuples of (possibly empty) words over Σ. $\mathcal{A}$ can be thought of as a labeled directed graph, whose nodes are elements of Q, and whose edges are elements of H, each such edge being labeled with the appropriate n-tuple of words. The component-wise concatenation of labels along every path that begins in q_1 and ends in an element of F defines a set of n-tuples, $R \subseteq (\Sigma^*)^n$, which is the relation accepted by $\mathcal{A}$.

As a practical matter, we generally want to run a program that (explicitly or implicitly) searches this graph in order to find all the n-tuples in R with some interesting property, say those corresponding

[1] The present set can be increased almost exponentially by adding new derivational affixes.

[2] The name and the basic idea of these automata come from [5]. For simplicity of exposition we gloss over various authors' attempts to distinguish variously among *machines*, *automata* and *transducers*, as well as the profusion of precursors and descendants in ([15], [16], [2], [7], etc.). Our notation is eclectic.

to forms whose surface spelling is the string w, or those corresponding to the first person plural imperfect subjunctive of such-and-such a verb. Depending on the structure of H and the property selected, the search will be harder or easier. For the stem-and-affix kind of morphology exemplified by Spanish, the natural structure for H is quite easy to search. We do not have space to discuss search methods here, but will simply observe that a non-optimal method devised for convenience in another experiment ([6]) processes several hundred Spanish words per second on a Sun4.

For the application discussed in this paper, we want to relate inflected forms, lemmas, and morphosyntactic features, so that the elements of R should be 3-tuples like:

$(cambiaran, cambiar, 3rd\ plural\ perfect\ subjunctive)$.

Since most Spanish words consist of a stem, which mainly specifies the lemma, and a set of affixes that mainly specify the morphosyntactic features, it is appropriate to use 2-tuples made by concatenating the second and third elements.

The basis of our run-time system is the arc list H. For a large lexicon, it is inconvenient to write this list by hand, and so we compile it from a lexical table that reflects more directly the way that morphological information is represented in a standard dictionary, such as the Collins dictionary we began with. The program interprets recursively all the possible arcs of the lists. Therefore, more than one analyzed or generated form is given. For instance, the analysis for the input word "retirada" is of the form:

```
retirar past participle feminine singular
retirado adjective feminine singular
retirada noun feminine  singular
```

2 The arc-list compiler

The arc-list compiler starts with a list of lexical items with their morphological classes, applying morphophonological transformations to generate the arc list. For instance, each verb headword in the Collins dictionary is given an index that specifies one of 62 conjugation classes. Based on this information, the arc-list compiler calculates the set of stem allomorphs necessary for that verb's inflection, along with the set of endings that each stem allomorph selects. Spanish verbs have from one to five orthographic stem allomorphs. When the verb is regular there is only one stem, like "cambi-" in "cambiar" (to change). An irregular verb may have up to five stems, like "ten-", "teng-", "tien-", "tend-", "tuv-" for the verb "tener" (to have). This is common in Romance languages (see Tzoukermann 1986 for French). These different stems are the result of morphophonological changes occuring during the verbal flexion, usually related to the stress implications of the verbal ending or to the features of its initial vowel.

Depending on the conjugation class, the character string corresponding to the verb lemma is subjected to one or more rewriting rules. These rewriting rules are of different types:

- they can be the consequence of a stress change during the verbal flexion:

  (a) **e – ie** when the last syllable is not stressed like in **querer / quiero**.

- they can be a morphographic change that is general to Spanish orthography:

  (b) **c – qu** before "e" and "i" like in **sacar / saque**.

  or the reverse rule

  (c) **qu – c** before "a", "o", "u" like in **delinquir / delinco**.

Some verbs are subject to one type of rewriting rule such as in (a) - (c) above, and consequently produce one additional stem allomorph. The verb "sacar" (to take / pull out) will generate "sac-" and "saqu-", as well as "delinquir" (to offend) with "delinqu-" and "delinc-".

Some other verbs, less frequent in number but more frequent in actual use, are subject to two rewriting rules and need a more complex treatment. In "forzar" (to force), the morphophonological rule combines with the othographic one and produces a distribution of four stems, such as "forz-", "forc-", "fuerc-", "fuerz-". The same phenomenon occurs for "rogar" (to beg) with the stems "rog-", "rogu-", "rueg-", "ruegu-". For some verbs of the second group in "-er", the stem production is less predictable; for instance "tener" presents five stems "ten-", "teng-", "tien-", "tend-", "tuv-". Notice that some of them such as "teng-" do not follow the type of morphophonological rules mentioned above.

Because of Spanish orthographic conventions connected with the notation of stress, some nouns and adjectives also acquire more than one stem allomorph in a rule-governed way. In addition, of course, there must be a list of cases where the allomorphy is simply unique to the word in question.

3 The arc list

Using a state labeled 1 by convention as the start state, and a state labeled 0 by convention as the (unique) final state, we express all of the information needed to define our automaton $\mathcal{A}$ by enumerating the arcs in H, which now can be represented as lists of 4-tuples (q_i, q_j, u, v), where q_i and q_j are arbitrary identifiers for states, u is a substring of an inflected form, and v is a substring of the corresponding lemma + morphosyntactic category.

Used either for analysis or for generation, our program interprets this same arc list. The arc list can be conceptually divided in two parts: one contains the stems of the verbs, nouns and adjectives; the other contains a number of sub-lexicons that provide the endings for these lexical categories as well as the clitics.

Our Spanish system is defined by a set of about 58,000 such 4-tuples, (most of which are) generated by rule from head words and category information extracted from the typographer's tape for the Collins Spanish Dictionary. Affixes, assorted null-string transitions and clitics account for about 1000 elements of this set; the remainder are stems or stem allomorphs. Since we have about 55,000 lemmas, the overhead for compiling out predictable aspects of allomorphy is at worst the approximately 2,500 stem allomorphs and affix arcs, i.e. less than 5%. There are about 225 states in total.

3.1 Verbal stems

The verbal stem lexicon was obtained by extracting the verb headwords (about 6,800 Spanish verbs) from the Collins dictionary.

Once the grammar provides the stems, a state pair is associated to them. The first state is always the initial state "1", the second depends on the type of stem and its ending throughout the conjugation (digits or character strings can be used indifferently for labelling the states). For example, for the first verb conjugation, whose infinitives end in "-ar," the second states are spread out among 10 different states.

1	2	cambi	cambiar
1	6	cruc	cruzar
1	3	envi´	enviar
1	4	envi	enviar
1	3	situ´	situar
1	4	situ	situar
1	5	cruz	cruzar
1	6	cruc	cruzar
1	7	jug	jugar
1	8	jueg	jugar
1	9	juegu	jugar
1	10	jugu	jugar

Two verb stems x and y will share the same second state number if and only if:

- x has the same number of stems as y,
- x has the same ending distribution as y.

This permits a compression of the database since the set of stems are gathered under a common second state number. Other arguments in favor of this choice of representation are given in section 4.1.

For the 62 conjugation classes, grouped in three verb conjugations, the number of stems combined with the various ending distributions creates a number of verb-stem-final states close to 150.

Defective verbs, due to their idiosyncrasies, are listed separately.

3.2 The adjective stems

The adjective base forms (about 10,500) were derived from the masculine singular forms listed in the dictionary. The lexical representation of a regular adjective has an entry in the lexicon as follows:

```
1       300       buen       bueno
```

where "buen-" is the stem and "bueno" (good) the dictionary base form. Special attention needed to be paid to stressed adjectives like "musulmán" (Muslim) or "mandón" (bossy) where the inflected form does not keep the accent. Therefore, both forms (stressed and unstressed) needed to be stored.

3.3 The noun stems

About 30,700 nouns were extracted from the dictionary. These nouns are not inflected for gender, but are simply listed as masculine or feminine. Thus the arc label for a noun contains the complete form of the singular. Some examples of arcs for nouns are:

```
(a) 1     499     aerodromo
          aerodromo noun masculine
(b) 1     500     mariscos
          mariscos noun masculine plural
```

In the above examples, (a) can either generate a singular form or it can acquire the plural form in a further step, whereas (b), which occurs only in the plural, can have no further inflection added.

4 The affixes

Besides the stems, various sublexicons containing "intermediary states" and affixes of different types constitute the other part of the Spanish arc list.

4.1 Intermediary nodes or continuation classes

The regrouping of the verbal arc list by stem and person allows reduction of the number of states and therefore, of arcs. For instance, an intermediary state was added for the **tenses** only. The arc marked "#" shows a transition on an empty string.

```
2       150       #       #
```

This arc takes any verb stem of which the final state is 2 and links it to the indicative present node – labeled here 150 – of the "-ar" verbs. Consequently, there are as many nodes of that kind as tenses for each group and verb category.

4.2 Endings

A series of sublexicons lists the inflections for the verbs, nouns and adjectives. Verbal inflections are of the form:

```
150   500   o   1st singular present indicative
150   500   as  2nd singular present indicative
```

In the same way, the regular endings for the adjectives are of the form:

```
300   497   o   adjective
497   498   #   masculine
497   500   #   singular
498   500   s   plural
```

Each transition corresponds to the gender or number feature of the adjective.

4.3 Clitics

The eleven Spanish clitics can occur either alone or in combination ([1]). Over sixty-five combinations can be formed such as "seles", "noslas", etc. The infinitive, gerund and imperative are the only forms in which they can occur, for instance, "hacerlo" (to do it) or "diciéndooslo" (saying it to you). Nevertheless, they are sometimes subject to orthographic rules of the type: deletion of "s" for first person plural imperative verbs in front of the enclitic "nos", such as in "amamonos".

Consequently, about 300 arcs were listed to handle the general cases as well as the idiosyncrasies.

4.4 Reflexive verbs

In the case of reflexive verbs such as "afiliarse" (to afiliate, to join) or "abstenerse" (to abstain, to refrain), a special treatment is motivated. Such verbs have a paradigm like:

```
(a) me afilio, (I afiliate)
    te afilias, (you afiliate)
    me afiliaba,(I was afiliating)
    te afliabas, (you were afiliating)
(b) afiliandome (afiliating myself)
    afiliate! (afiliate!)
```

The reflexive pronouns generally precede the verb form, separated from it by white space as shown in (a), except for the infinitive, imperative and present participle (example (b) above) [3]. For the preceding reflexive pronouns, there is a dependency between the person-and-number of the pronoun and the person-and-number of the verbal ending, spanning the intervening verb stem. To capture such dependencies in a single automaton of the kind that we are using, we would have to use a separate path for each person-number combination, duplicating the verb stem (and its allomorphs, if any) six times. This seems like a bad idea. A better alternative, in such cases, is to set up the automaton to permit all reflexive pronouns to co-occur with all endings, and to filter the resulting set of tuples to remove the ones that do not match. This can be done, for example, by passing the output through a second automaton that does nothing but check person and number agreement in reflexive verbs.

We find it interesting that precisely those aspects of Spanish morphology that require such a treatment are those whose formatives are written as separate words.

[3] Note that some verbs (e.g. "afiliarse") occur only reflexively, while other (e.g. "lavar" (to wash, to clean)) may be used reflexively or non reflexively. Note also that object pronouns in general are cliticized, note only the reflexive ones.

4.5 Prefixes and suffixes

About 60 suffixes and 90 prefixes were added to the arc list for handling derivational morphology. Only the very productive ones were selected. The prefixes are of the form "aero-", "ante-", "auto-", "bio-" occurring with or without the dash; the suffixes are of the form "-ejo", "-eta", "-zuela", "-uelo", etc.

The resulting arc list, in addition to supporting an efficient computation of relations between surface and lexical forms, provides a good overview of the morphological structure of the Spanish verbal system, permitting easy access to the sets of verbs that behave in a similar way.

5 Conclusion

We have implemented a complete morphological processor for Spanish, one which generates and recognizes all (and only) well-formed inflected and derived forms. It covers about 95 % of Spanish text extracted from the EFE newswire text coming from Madrid. It has been linked to a browser for the Spanish newswire and to the Collins bilingual dictionary (see Appendix), is also being utilized in the construction of a Spanish parser (Donald Hindle at Bell Laboratories) and for further research in Spanish text analysis. We have found this model to be both simple and powerful. We plan to implement other Romance languages, and to experiment with German, where the treatment of compounds presents some special interest.

References

[1] Casajuana R. and C. Rodríguez 1985. *Clasificación de los verbos castellanos para un diccionario en ordenador.* I congreso de lenguajes naturales y lenguajes formales. Universidad de Barcelona. Facultad de Filología. Departamento de Lingüística General. Barcelona.

[2] Chomsky, N. 1962. *Context-free Grammars and Pushdown Storage*, M.I.T. Research Laboratory of Electronics Quarterly Progress Report #65, pp. 187-193.

[3] *Collins Spanish Dictionary: Spanish-English.* Collins Publishers, Glasgow, 1989.

[4] Corbin D. 1987. *Morphologie dérivationnelle et structuration du lexique.* Niemeyer Verlag: Tubingen.

[5] Elgot, C.C. and J.E. Mezei 1965. *On Relations Defined by Generalized Finite Automata*, IBM Journal Res. 9, pp. 47-68.

[6] Feigenbaum, J., M.Y. Liberman, R.N. Wright (forthcoming). Cryptographic Protection of Databases and Software. In *Proceedings of the DIMACS Workshop on Distributed Computing and Cryptography*, Feigenbaum and Merritt, Eds. AMS and ACM.

[7] Ginsburg, S. 1966. *The Mathematical Theory of Context-Free Languages*, McGraw Hill.

[8] Kay, M. 1982. *When Meta-rules are not Meta-rules.* In Spark-Jones & Wilks (eds.) *Automatic Natural Language Processing.* University of Essex, Cognitive Studies Center (CSM-10).

[9] Kartunnen, L. 1983. *KIMMO: A general morphological processor.* Texas Linguistic Forum, No. 22 pp 165-186.

[10] Kartunnen, L., K. Koskenniemi, R. Kaplan 1987. *A Compiler for Two-level Phonological Rules.* Ms. Xerox Palo Alto Research Center.

[11] Khan R. 1983. *A two-level morphological analysis of Roumanian.* Texas Linguistic Forum, No. 22 pp 153-170.

[12] Koskenniemi, K. 1983. *Two-level morphology: A General Computational Model for Word-Form Recognition and Production.* University of Helsinki, Dept. of General Linguistics, Publications, No. 11.

[13] Koskenniemi, K., K. W. Church 1988. *Complexity, Two-level morphology and Finnish.* Proceedings of the 12th International Conference on Computational Linguistics. Budapest, Hungary.

[14] Lun S. 1983. *A two-level morphological analysis of French.* Texas Linguistic Forum, No. 22 pp 271-277.

[15] Rabin, M.O. and D. Scott, 1959. *Finite Automata and their Decision Problems*, IBM J. Res. 3, pp. 114-125.

[16] Schützenberger, M.P. 1961. *A Remark on Finite Transducers*, Information and Control 4, pp. 185-196.

[17] Tzoukermann E., R. Byrd 1988. *The Application of a Morphological Analyzer to on-line French Dictionaries.* Proceedings of the International Conference on Lexicography, Euralex. Budapest, Hungary.

[18] Tzoukermann E. 1986. *Morphologie et génération des verbes français.* Unpublished PhD dissertation. Institut National des Langues Orientales, Sorbonne Nouvelle, Paris III, France.

(1)

```
XA3274 90/04/02 16:46 B ECO  CAFE COLOMBIANO__ AB sum exg icx
DP7665 90/04/02 16:49 U POX  WALESA: "NO TENGO DIFERENCIAS CON MAZOWIECKI"__ $
XJ9199 90/04/02 16:51 B ECO  COTIZACIONES ESPANOLAS EN WALL STREET__ AB exg i$
DP7670 90/04/02 16:59 U POX  CUMBRE CENTROAMERICA__PRESIDENTES CONFIADOS NUEV$
XA3275 90/04/02 17:04 B ECO  WALL STREET: CIERRE CON PEQUENAS PERDIDAS TRAS R$
DP7630 90/04/02 17:08 R SUC  MUNICIPALIDAD NUEVA YORK MAYOR ARRENDATARIO CLUB$
XA3276 90/04/02 17:09 B AAA    NO_BOD  ABesh wsd icx
XA3277 90/04/02 17:10 B AAA    NO_BOD  ABmus icx
FB7684 90/04/02 17:16 R CUL  MURIO ALDO FABRIZI, EL DON PIETRO DE "ROMA, CIUD$
XA3280 90/04/02 17:18 B ECO  CACAO: BAJA__ AB sum exg icx
XA3281 90/04/02 17:19 B ECO  CAFE: BAJA__ AB sum exg icx
XA3282 90/04/02 17:20 B ECO  SUBEN LOS PRECIOS DEL PETROLEO EN NUEVA YORK__ A$
XA3279 90/04/02 17:23 U POX  BID      __JAPON RECONOCE PERSISTENCIA GRAVEDAD D$
XA3283 90/04/02 17:27 U POX  BID__CANADA: ACUERDO CANJE DE DEUDA HONDURAS-MEX$
DP7632 90/04/02 17:36 R SUC  EEUU: CONDENADOS A MUERTE ACUARDAN DECISION TRIB$
XA3285 90/04/02 17:38 B ECO  PROMEDIOS DOW JONES: BAJA__ AB sum exg icx
EF7684 90/04/02 17:40 U POX  EEUU: APRUEBAN 38 MILLONES DOLARES PARA CARCEL E$
EF7685 90/04/02 17:42 U ECO  EEUU: "PAN AM" QUIERE REANUDAR VUELOS A MANAGUA_$
HH0105 90/04/02 17:49 R DEP  TENIS: IVAN LENDL, EN HORAS BAJAS__ AB dog sum d$
CB0589 90/04/02 17:51 B AAA    NO_BOD  ABpan aex lha wsh wsd std saj inf tfb
FB7650 90/04/02 17:54 R POX  1.- RDA: BOEHME DIMITE POR NO ACLARARSE ACUSACI$
FB7651 90/04/02 17:56 R POX  Y2.- RDA: BOEHME DIMITE POR NO ACLARARSE ACUSACI$
FB7680 90/04/02 18:05 U POX  RUMANIA: HERMANO CEAUCESCU RECHAZA ACUSACIONES A$
XA3288 90/04/02 18:07 B ECO  METALES PRECIOSOS: BAJA EL ORO, SUBE LA PLATA__ $
XA3289 90/04/02 18:08 B ECO  AZUCAR: BAJA__ AB sum exg icx
XA3290 90/04/02 18:10 B ECO  DOLAR SUBE EN NUEVA YORK ANTE MAYOR PARTE MONEDAS$
XA3291 90/04/02 18:12 U ECO  BID__PRESTAMO 18,8 MILLONES DOLARES PARA PARAGUA$
XH1252 90/04/02 18:13 B SVD    AB icx wsh tdu dxe
DP7672 90/04/02 18:18 U ECO  MEXICO: PAGOS DEUDA NACIONAL TRIPLICAN LOS DE DE$
DP7673 90/04/02 18:22 U ECO  BID__RESTAMO PARA CENTRAL HIDROELECTRICA ARGENTI$
XA3293 90/04/02 18:24 U ECO  VENEZUELA: EXPORTACIONES EEUU AUMENTARON 31 POR $
XH1253 90/04/02 18:27 U DEP  TENIS-RIO DE JANEIRO:__VICTORIAS DE ARGENTINO AR$
DP7674 90/04/02 18:34 U POX  EEUU: CONDICIONAN APROBACION AYUDA PANAMA Y NICA$
-XX-Emacs: efe-index          (EFE)----BOX----------
```

(2)

```
referente a la infraccion del reglamento que prohibe la posesion
de armas, pero se justifico diciendo que habia creido que, como
oficial, tenia derecho a llevar una consigo.
    El juez le recordo que solo durante la etapa de la revuelta
popular previa al derrocamiento de su hermano Nicolae se permitio a
los personas que portasen armas y aquellos que no las devolvieron
tras el triunfo de la revolucion fueron castigados con penas de
prision.
    Durante la primera sesion del juicio, Andruta Ceaucescu
cometio numerosas faltas gramaticales y de pronunciacion y
gesticulo de forma muy similar a como lo hizo su hermano Nicolae
en el proceso en que, al igual que su esposa Elena, fue condenado
a muerte y fusilado inmediatamente despues de emitirse la
sentencia.
    Andruta Ceaucescu, de 66 anos y detenido desde el pasado 27 de
diciembre, fue viceministro del Interior y jefe de campamento de
entrenamiento de Baneasa, el principal de la "Securitate", policia
politica secreta rumana.
    La acusacion por asesinato que pesa sobre Andruta Ceaucescu
senala que mato a tiros a seis personas en las manifestaciones
celebradas el pasado 21 de diciembre en Bucarest contra el regimen
comunista.
    Asimismo, es presunto culpable de instigacion al genocidio,
por creerse que dirigio una unidad de la "Securitate" en Bucarest
-----Emacs: RUMANIA: HERMANO      (Fundamental)----16%----------
```

(3)

```
(defun detener )
::pattern_1.1   Zk
::cat_1.1       vt
::def_1.1.a     (parar) [personal, [bal'on], [veh'iculo], [epidemia] (etc) to $
::phr_1.1.a.1   detener el progreso de
::def_1.1.a.1   to hold up the progress of
::phr_1.1.a.2   no quiero detenerte
::def_1.1.a.2   I don't want to delay you
::phr_1.1.a.3   se detuvo en la calle
::def_1.1.a.3   he stopped me in the street, he accosted me in the street.
::def_1.1.b     (retener) to keep, hold back, retain; [respiraci'on] to hold.
::def_1.1.c     (Jur) to arrest, detain.  >EM<
::form_1.2      detenerse
::fcat_1.2      vr
::def_1.2.a     to stop; to pause; to delay, linger
::phr_1.2.a.1   se detuvo a mirarlo
::def_1.2.a.1   he stopped to look at it
::phr_1.2.a.2   ^!no te detengas!
::def_1.2.a.2   don't hang about!, don't delay!
::phr_1.2.a.3   se detiene mucho en eso
::def_1.2.a.3   he's taking a long time over that.
(defun detenidamente )
::cat_1.1       adv
::def_1.1.a     carefully, thoroughly; at great length.
-XX-Emacs: D2.1                (Lisp)----BOX----------
```

(4)

```
(defun detenido )
::cat_1.1       adj
::def_1.1.a     (Jur) arrested, under arrest.
::def_1.1.b     [narraci'on etc] detailed; [examen] lengthy, thorough; careful$
::def_1.1.c     (fig: t'imido) timid.
::def_1.1.d     (fig: tacan~o) mean, niggardly.  >EM<
::cat_1.2       na
::alt_1.2       detenida
::altcat_1.2    nf
::def_1.2.a     person under arrest; detainee.
(defun detenimiento )
::cat_1.1       na
::def_1.1.a     care, thoroughness
::phr_1.1.a.1   con detenimiento
::def_1.1.a.1   carefully, thoroughly.
(defun detener )
::pattern_1.1   la
::cat_1.1       vt
-XX-Emacs: D2.1                (Lisp)----BOX----------
```

(5)

```
detenido

detener         past participle     masculine       singular
detenido        adjective           masculine       singular
detenido        noun                masculine       singular

-**-Emacs: *morph*             (Inferior Morph: run)----All----------
```

Appendix. Emacs windows showing the use of the Spanish browser.

(1) Spanish newswire giving indications of date, time, codes and title of the news.

(2) Sample of a Spanish text selected from (1).

(3) and (4) Collins Spanish-English dictionary translation of the entries "detener" (verb), "detenido" (adjective), "detenido" (noun).

(5) Morphological analysis of the word "detenido", appearing in (2).