Loading MRD into LDB.
Characteristics of Vox Dictionary

Irene Castellón
María Antonia Martí
Germán Rigau
Horacio Rodríguez
Felisa Verdejo

Report LSI–91–24

LOADING MRD INTO LDB. CHARACTERISTICS OF VOX DICTIONARY

I. CASTELLóN (1)
G.RIGAU (1)
H.RODRíGUEZ (1)
M.A. MARTí (2)
M.F. VERDEJO (1)

(1) Universitat Politècnica de Catalunya
(2) Universitat de Barcelona

January 1991
LSI Departament
Universitat Politècnica de Catalunya
Barcelona
SPAIN

ESPRIT BRA-3030 ACQUILEX WP NO. 019

Table of contents

LOADING MRD INTO LDB
CHARACTERISTICS OF VOX DICTIONARY

Introduction

In this document we present an explanation of the loading of the MRD of the Spanish Vox dictionary into LDB structure. This work has been done inside the Esprit project Acquilex (BRA 3030). The aim of this project is the use of existing lexical resources in order to build lexicons for NLP systems. The loading of MRDs into a LDB is the first step in the process of information extraction from dictionaries.

The first section is devoted to the explanation of the characteristics of the dictionary Vox from a lexicographic point of view. The second deals with the conversion of the Vox MRD into the Lisp structure: the categorization of the text by means of a morphological analyser, the grammar of the entries and the conversion into the lisp structure. In section 3 we show the template of the Vox dictionary and the tags. Finally, in section 4 we present a first approach to the lexical information contained in our dictionary.

1. Source dictionary characteristics: Vox monolingual

1.1. Number and types of entries

1.1.1. Number of entries and senses

The Vox dictionary has a total of 89.793 entries and has an average of 1.6 senses - entry. The total number of senses is 143.700. The maximum number of senses -entry is 24. The minimum number of senses -entry is one.

1.1.2. Types of entries

The different types of entries in VOX are :

1.- Single words (compounds are considered as a single entry (1) or as a sense in an entry (2)):

'**pasacalles** (1)'
'**cobre** (2)..... 5 ~ *verde* , malaquita.'

2.- Proper nouns and geographic nouns:

'**Baal** m. '
'**Babia** Territorio de ....'
'**Cardona** n.pr.'

In the dictionary specifications there is a special code for proper nouns : n.pr., but, in the dictionary, proper nouns appear categorized also as 'm.', 'f.' as common nouns are (Baal), or without specifications (Babia). They have in common that begins with a capital letter.

3.- prefixes and suffixes

'a-,an-, '
'bl-, '
'- able,'

This kind of entry is characterized by hyphen, '-', after or before the entry.

4.- irregular word forms

4.1.- Apocopations

'mío, mía '
'mi, mis adj. Apóc de los adjetivos posesivos mío, mía....'

'san, apócope de santo.'
'tan , apócope de tanto.'

4.2 - irregular word form

'visto pp. de ver'
'obtuve pret. indefinido de obtener.'

5.- interjections

'¡bah! interj.'

6.- Abbreviations

6.1-Chemical names

'Ba, símbolo químico del Bario.'

6.2-Abbreviations of single words

'O, Abreviatura de Oeste.'

7.- Letters of the alphabet

'B, b'

8.- Contractions

'del, contracción de la prep. *de* y el artículo *el*.'

9.- Onomatopoeias

'¡paf! voz onomatopéyica...'

1.1.3. Storage of entries

Lexicographers who have worked on the Vox dictionary explain the criteria used for the storage of homonyms and senses as follows:

Homonyms

The storage of homonyms in the Vox monolingual dictionary is based on historical criterion. Usually when words with the same form differ in their etymology , different homonyms are created.

We have observed that the historical criterion does not apply in the entire dictionary. There are cases in which homonyms, with the same etymology, are distinguished according to their category. I. e.:

I) **nada** *pron. indef.* (V. *nada II*)

II) **nada** *f.* (l. *res nata,* cosa nacida)

Senses

The order of appearance of different senses, according to the explanations of VOX lexicographers, is based on the semantic proximity to the etymology : i.e.:

'banco (germ.bank) m. asiento largo y estrecho... 2 Parte inferior de un retablo...3 Mesa de trabajo...4 Mesa que usaban los cambistas. 5 Establecimiento público de crédito.'

When this criterion is not possible, the order is based on didactic criteria, i.e. grouping the senses which are analogous, i.e.:

'oso (l.ursu) m. Mamífero de la familia de ...: ~ blanco o marítimo,... 2 fig. fam. Hacer uno el ~,... 3 Nombre que p.ext. se da a ...~ hormiguero...; ~ panda,...'

Other criterion is grouping the senses with the same POS.

At the end of the entry appear the senses with a thematic subject code, followed by the senses with geographic code.

1.1.4. Structure of the entries on tape

1.1.4.1. Sub-divided entries.

In general, the content of the VOX dictionary entries is :

word form; etymology; POS; use labels; subject code; definition; scientific nouns; samples; variant form; cross references; figure references; lexical or derivative families; idioms; grammatical information and homophones

The Vox criterion for separate senses is the appearance of different POS, different labels of use, different geographical use, etc.

1.1.4.2. Non-subdivided entries. References

There are two types of non-subdivided entries :

1.- Entries which are compounded by only one sense.

2.- references: this type of entry is used to refer to back entries. This type of entries is characterized because its definition contains only one word which is another entry of the dictionary.

2.- Dictionary grammar and the generation of LISP structure

2.1. Morphological analyser. Categories, attributes and values.

The morphological analyser assigns categories to some markers in the MRD. These markers could be typographical codes, punctuation marks, labels and parts of the text.

The categories are:

a.- Categories of typographical codes

Every typographical code has been categorized. These categories correspond to terminal nodes in the MRD grammar, to identify the entries and determine theirs parts. These are:

<*CTTX1> begining of text
<*CTTX2> entry code
<*CTTX3> lexical relations
<*CTTX4> open capital versalitas
<*CTTX5> close capital versalitas
<*CTTX6> open versalitas
<*CTTX7> close versalitas
<*CTTX8> end of the text
<*CTTX9> end of the text
<*CTTL1> round letter
<*CTTL2> italics
<*CTTL3> bold letters

b.- Marks

Category	Values
<*COMA>	','
<*PUNG>	'-'
<*PUNPO>	'('
<*PUNPT>	')'
<*PUNTO>	'.'
<*DPUN>	':'
<*PCOMA>	';'
<*CONJ>	'y'

c.- Labels

Category	Values
<*CATG>	'm', 'f', 'adj', 'n', 'v',' intr'...
<*GEO>	'Alm.', 'Chile', 'Ar','Argen.'...
<*TEMA>	'MIN', 'MUS.', 'MED', 'ASTRON.'...
<*USO>	'ús', 'p.us', 'ant.',...
<*REG>	'fam.', 'fest.'...
<*SEM>	'fig.', 'eufem.',...
<*FORM>	'CONJUG.', 'HOMOF.'...
<*REL>	'REL.', 'SIN.', 'CONTR.',...

d.- Others

<*NH>	'I, II, III, IV, V, VI, VII, VIII, IX, X'
<*FIG>	'***'
<*NUM>	'1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,...'

e- Categories with a text values

<*TEXTO>
<*LEX>
<*SFX>
<*PFX>

2.2. The grammar of entries.

The grammar consists of a total of 39 rules.

Explanation of the Meta symbols:

[ ]	*optional*
< >	*non terminal node*
< >*	*terminal node*
<A>/<B>	*A or B.*
(<A>/<B>) <C>	*A C or B C*

The first node of the grammar is SENTENCE. Every SENTENCE is an entry in the MRD.

<SENTENCE> = <ENTRADA>[<ETIMOLOGIA>]<ACEPCIONS>
<FORMAS>[<NCOD>][<RELA>].

ENTRADA analyses a typographical code and the entry-word followed by another typographical code. An entry-word can be LEX , PREFIXE (prefix) or SUFIXE (suffix):

<ENTRADA > = <*CTTX2 > (<LEX>/<PREFIXE> /<SUFIXE>) <*CTTL1>.

LEX analyses the homonyms number <*NH>, the word-form of the entry <*ENTR> , and the inflection <FLEXS> which is a recursive rule.

<LEX> = [<*NH> <*PUNPT>]<*ENTR > <FLEXS>.
<FLEXS> = <*COMA> <*PUNG > <*FLEX> <FLEXS> /<*NULL>.

PREFIXE is a recursive rule and analyses the entry if the current word is a prefix:

<PREFIXE>= <PREF> <PREFIXE> /<*NULL>
<PREF> = <*PFX> <*PUNG> <*COMA>

SUFIXE is a recursive rule and analyses the entry if the current word is a suffix:

<SUFIXE> %%= <SUFIX> <SUFIXE>/<*NULL>.
<SUFIX> %%= <*PUNG> <*SFX> <*COMA>.

ETIMOLOGIA, analyses the etymologic information; that information appears parenthetically:

<ETIMOLOGIA> = <*PUNPO> <ETIMS> <*PUNPT> /<*NULL>.

ACEPCIONS takes the form of a recursive rule and analyses the sense part of the entry. A distinction is made between the first sense and the following ones.

<ACEPCIONS> = <ACEPCIO> (<ACEPCION2> /<*NULL>).

<ACEPCIO> = [<*CTTL2>] [<*CATG>] [<GEOS>] [<INFLEX>] <TEXT>.

In the first case, ACEPCIO groups the category and morphological information <*CATG>, the geographic information about the use of the entry <GEOS> , the scientific use of the entry <INFLEX>, and <TEXT> grouping the definition and the use code.

In the second case ACEPCION2 is also a recursive rule, and analyses the sense number <*NUM>, the category and morphological information <*CATG>, the geographic information about the use of the entry <GEOS>, the scientific use of the entry <INFLEX>, and <TEXT> grouping the definition and the use code.

<ACEPCION2> = <ACEPCIO2> (<ACEPCION2> /<*NULL>).

<ACEPCIO2> = [<*PUNG>] [<*CTTL2>] [<*NUM>] [<*CATG>] [<GEOS>]
[<INFLEX>] <TEXT>.

GEOS is also a recursive rule that analyses the geographic information about the use of the entry (i.e.: Alm., Chile,...):

<GEOS> = [(<*CTTL1><*CONJ> <*CTTL2> /<*COMA>)] <*GEO> <GEOS> /<*NULL>.

The following node INFLEX includes the node TEMA which gives information about the scientific use of the entry. (i.e.: MIN, BOT, DEP,..):

<INFLEX> = <*CTTX4> <*TEMA><*CTTX5>.

TEXT is the last node of the rule <ACEPCIO> , grouping the definition and the use code.

<TEXT> = <*CTTL1> [<ILS>] <TPS> [<*PUNTO>].

ILS groups different information about the word use:

<ILS> = (<*USO>/<*REG>/<*SEM>) <ILS>/<*NULL>.

USO analyses the use code (i.e.: p.us, ús,...)
REG analyses the register code (i.e: fest, fam, ...)
SEM analyses the semantic code (i.e.:fig, desp.,..)

TPS groups the definition text of the sense, the rule TPS is recursive and is composed of text and typographical codes, it is explained in the following paragraphs.

FORMAS includes different kinds of information attached to the entry, like homographs, the inflectional model, incorrect uses, superlatives, etc. All of them are introduced by a label. This rule is recursive.

<FORMAS> = <FORMA> <FORMAS>/<*NULL>.

<FORMA> = <*PAG> [<*CTTX4> <*FORM> <*CTTX5>] <TXFS>.

The node NCOD analyses differents kinds of information (colloquial uses, idioms, etc.) that are not introduced by any label but identifiable by typographical codes.

<NCOD> = <*CTTX6> <*CTTX7> <TXNS> .

The last node of SENTENCE is RELA.

<RELA> = <*CTTX3> <RELACIONS>.

The rule RELACIONS is recursive:

<RELACIONS> = <RELACIO> <RELACIONS> /<*NULL>.

<RELACIO> = <*CTTX6> <*REL> <*CTTX7> <TXRS>.

RELACIO groups information about the relation between the entry and anothers entries of dictionary.

There are some categories that are composed of text:

a)- <*ENTR> , <*SFX> and <*PRX> analyses the word entry.

b)- ETIM identifies the etymology marked by '(")':

<ETIMS> = <ETIM> <MESETIMS>.
<MESETIMS> = <ETIMS> /<*NULL> .
<ETIM> = <COSA> <CODIGOS1>.
<CODIGOS1> = <*CTTL1/<*CTTL2>/<*NULL>
<COSA> = <*TEXTO>/<*PUNTO>/<*COMA>/
<*PUNPT>/<*PUNPO>/<*DPUN>/
<*PCOMA>/<*CONJ>/<*PUNG>/
<SIGLO>/<*NULL>.

c)- TPS analyses the definition text, contains typographical codes and text. It's a recursive rule.

<TPS>= <TP> [<TPS>] .

<TP>= <COSA> <CODIGOS1>.

<CODIGOS1> = <*CTTL1/<*CTTL2>/<*NULL>

d)- TXFS, TXRS and TXNS are recursive rules that analyze the text and the typographical codes belonging to FORMA, RELACION and NCOD. These nodes are expanded in the following rules:

<TXRS>= <TXR> [<TXRS>] .
<TXFS>= <TXF> [<TXFS>] .
<TXNS>= <TXN> [<TXNS>] .
<TXR> = <COSA> <CODIGOS2>.
<TXF> = <COSA> <CODIGOS3>.
<TXN> = <COSA> <CODIGOS4>.

<CODIGOS2> = <*CTTL1/<*CTTL2>/<*CTTL3>/<*NULL>
<CODIGOS3> = <*CTTL1/<*CTTL2>/<*CTTL3>/<*NULL>
<CODIGOS4> = <*CTTL1/<*CTTL2>/<*CTTL3>/<*NULL>
<COSA> = <*TEXTO>/<*PUNTO>/<*COMA>/
<*PUNPT>/<*PUNPO>/<*DPUN>/
<*PCOMA>/<*CONJ>/<*PUNG>/
<SIGLO>/<*NULL>.

The remaining categories indicate the typographical codes of the dictionary. These are:

<*CTTX1> general text
<*CTTX2> entry text
<*CTTX3> synonymy text
<*CTTX4> small capitals
<*CTTX5> cancel small capitals
<*CTTX6> small capitals
<*CTTX7> cancel small capitals
<*CTTX8> cancel text
<*CTTX9> cancel text
<*CTTL1> round letter
<*CTTL2> italics
<*CTTL3> bold
<*CTTL4> bold and italics

This grammar analyses the 99% of the whole dictionary.

2.3.- Examples

See appendix 1.

2.4. Generating lisp structure from the grammar.

2.4.1. Lispification.

The lisp structure is the result of the grammar's analysis. Lisp structure is composed of the following fields:

```
((ENTRADA )
(FLEX:    )
(NH :     )
(ETIM:    )
(Sense:   )
(CATG:)
(GEO:     )
(TEMA:    )
(USO:     )
(REG:     )
(SEM:     )
(DEF:     )
(FORMA:   )
(TIPOF:   )
(TXF:     )
(RELA:    )
(TIPOR:   )
(TXR:     ))
```

2.4.2. From grammar to lisp structure.

The conversion from analyzed MRD to lisp structure is made taking into account the following correspondences:

grammar	lisp
<*ENTR >	ENTR
<*NH>.....	NH
<ETIM>.....	ETIM
<*FLEX>.....	FLEX
<*NUM>.....	NS
<*CATG>.....	CATG
<*TEMA>.....	TEMA
<*GEO>.....	GEO
<*USO>	USO
<*SEM>.....	SEM
<*REG>.....	REG
<TP>.....	DEF
<TXN>.....	TXR
<*FORM>.....	TIPOF
<TXF>.....	TXF
<*REL>.....	TIPOR
<TXR>.....	TXR

EXPLANATIONS:

ENTR	word entry
NH	homonym number
FLEX	inflection
ETIM	etymology
SENSE	sense number
CATG	category
TEMA	subject code
GEO	geographical code
USO	use code
SEM	semantic code
REG	register code
DEF	definiton text
FORMA	number form
TIPOF	label form
TXF	text form
RELA	number rela

TXF	text form
RELA	number rela
TIPOR	label rela
TXR	text rela

2.4.3. Some examples.

See appendix 2

3. - Template structure: vox monolingual

3.1. Template for lexical entry.

The template is a representation of the internal organization of the "maximal" entry of the dictionary, in this case of the Vox dictionary, and reflects its hierachical structure.

3.1.1. Abbreviated schema of the lexical entry.

Abbreviated Lexical Entry Template shows only the main Node_tags.

```
ENTRY
    HEADWORD_GROUP
        VARIANT_GROUP
        ETYMOLOGY_GROUP
            CROSS_REFERENCE_GROUP
        HOM_GROUP
            GRAM_INF_GROUP
                POS-GROUP
                MORPH_GROUP
            SENSE_GROUP
                CROSS_REFERENCE_GROUP
                DEFINITION_GROUP
                    SEMANTIC_LABEL_GROUP
                    DEF_GROUP
                    EXAMPLE_GROUP
                    COMPOUND_GROUP
                        DEF_GROUP...
```

SEMANTIC_LABEL_GROUP

SEMANTIC_RELATIONS_GROUP

SYNONIM_GROUP

ANTONYM_GROUP

"ALTERATE_GROUP

SEMANTIC_RELATIONS_GROUP

HOMOPH_GROUP

3.1.2. Tags and template.

The Tags are of two types: Node-tags that govern a group of constituent Tags, and Attribute-Tags that always take values.

- Tags at the Dictionary level are:

```
Dict_Source:
        LANGUAGE
                Lang:
        PHONETIC TRANSCRIPTION
                IPA: -
```

- Tags at entry level are:

```
ENTRY
        HEADWORD_GROUP
                Hdwd_type:
                Hdwd_form:
                (Hdwd_figure_ref):
                (Hdwd_text):
                (Hdwd_Homonym_No):

                (VARIANT_GROUP )
                        (Variant_label):
                        Variant_form:
                (ETYMOLOGY_GROUP):
                        Etymology_text:

                        (CROSS_REFERENCE_GROUP)
                                (Xref_text)
                                Xref_label
                                Xref_ewntry
                                (Xref_extens)


                +HOM_GROUP
                        Hom_No:
                        (Hom_form):
```

(Hom_Compact):

(GRAM_INF_GROUP)
Gram_Inf_text:

*(POS-GROUP):
(POS):
(subcat)
(subtype):
(gender):
(number):
(various):

*(MORPH_GROUP):
(Morph_text):
(Morph_label):
(Morph_form):

*(SENSE_GROUP)
Sense_No.

(CROSS_REFERENCE_GROUP)...

(DEFINITION_GROUP)

*(SEMANTIC_LABEL_GROUP)
(Semantic_label_text)
(Subject_label)
(Semantic_code)
(Register_code)
(Usage_code)
(Geographic_code)
(American_code)

(DEF_GROUP)
Def_text
*(Implicit_Xref)
(Figure_ref)

*(EXAMPLE_GROUP)
+Ex_text:
(Ex_label):
(Ex_explanation):

(COMPOUND_GROUP)
Cpd_label:
Cpd_form:
DEF_GROUP...
(SEMANTIC_LABEL_GROUP)...

(SEMANTIC_RELATIONS_GROUP)
(SYNONIM_GROUP)
Syn_label:
+Synonim:

+Synonim:

(ANTONYM_GROUP)
Ant_label:
+Antonym:

("ALTERATE_GROUP)
Alt_label
+"Alterate"

(SEMANTIC_RELATIONS_GROUP)...

(HOMOPH_GROUP)
Homoph_label:
Homoph_text:
Homoph_entry:

3.1.3. Some examples.

See appendix 3

3.2. Characteristics of tags and groups in Vox .

The tags and groups specific of Vox dictionary are:

Country_code: It's an important information for the Spanish language because the South American language differs from the peninsular language .

Ex_Label : It´s necessary in VOX

HOMOPH_GROUP: It´s an information produced in Vox .

3.3. Tags values.

See appendix 4

4.- A first approach to lexical information in VOX

4.1. Classification of information.

4.1.1.- Information on the form

The Vox dictionary contains information about the written form, spelling variants, stress information, variants forms and derivations; in the case of flection class, Vox has information of the verbs (tenses, past participle and auxiliary verb), nouns (plural and diminutive) and adjectives (superlative). Vox has not information about the stem and the phonetic representation.

The fields containing form information in the lisp structure are:

ENT, FLEX, FORM, and RELA.

4.1.2.- Grammatical information

The grammatical information in the Vox contains: the POS; verbs subcategorization: transitive or intransitive. There is also information about their arguments, but in an implicit way(i.e.: the verb *'macetear'* has the following definition: *Golpear [a alguien] con la maceta* , in this case *[a alguien]* indicates that this argument must be a person, must have a preposition *'a'* , and the category is a prepositional phrase); morphological characteristics of nouns, it's gender, gender but doesn't indicate the countability , semantic type nor valency; about the adjective doesn't appear grammatical informatior explicitly unless the POS.

The fields that contains grammatical information in the lisp structure are:

CATG , DEF, FORM and RELA

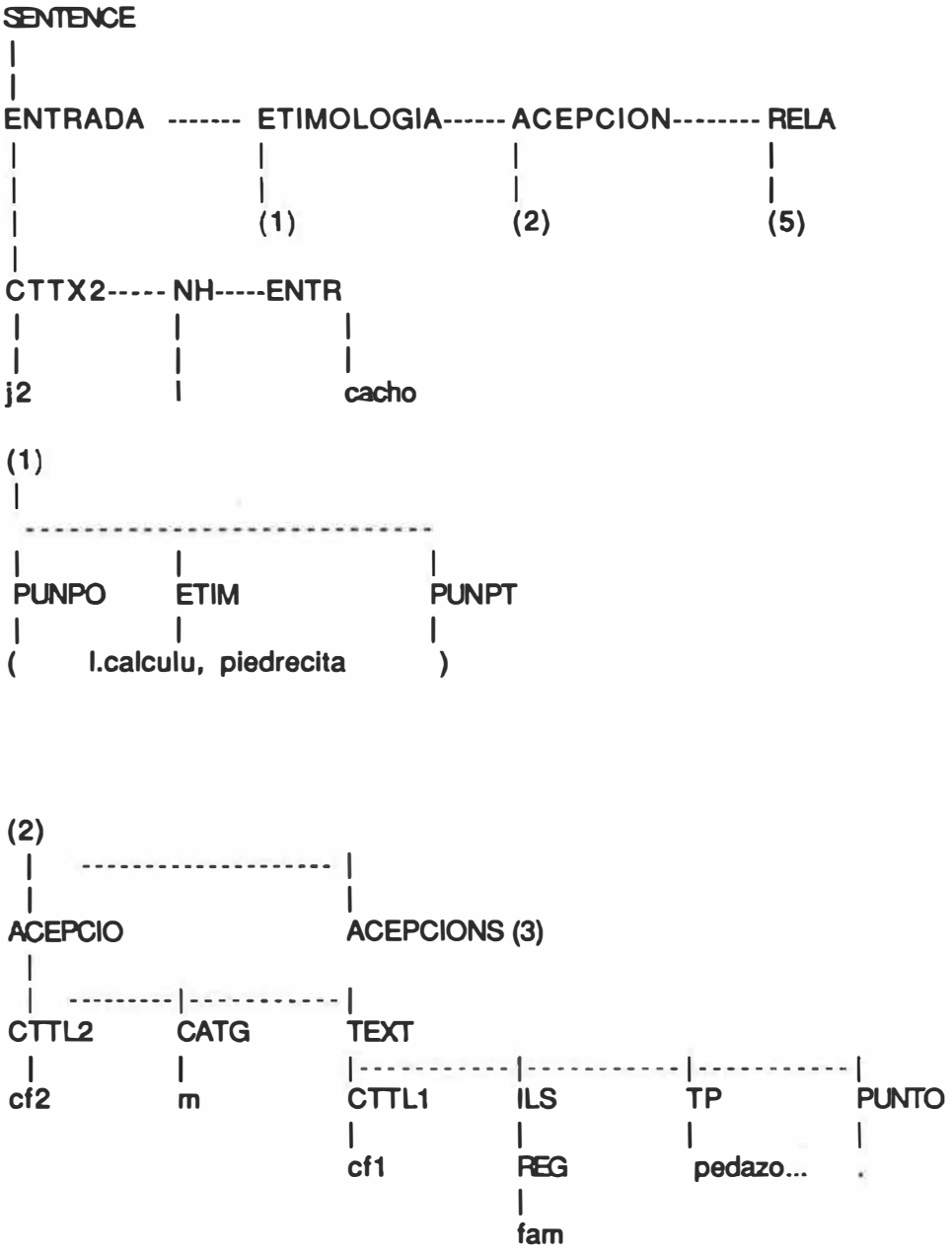
Appendix 1

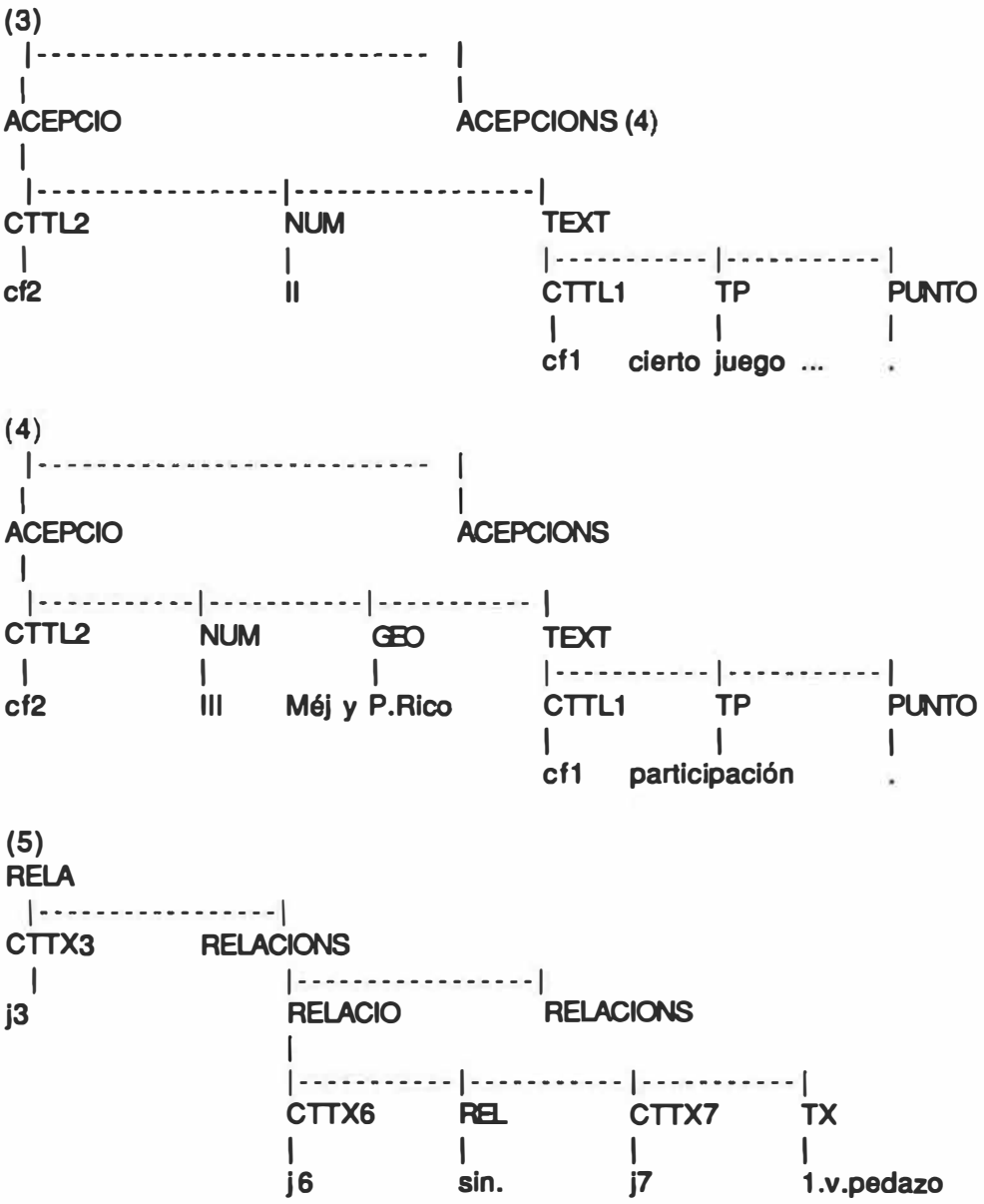
One example of grammar analysis.

Entry :

l) cacho (l.*calculu*, piedrecita) *m.* fam. Pedazo pequeño de alguna cosa. 2. Cierto juego de naipes.3. *Méj y P.Rico.* Participación pequeña en un número de la lotería.
SIN 1. V. **Pedazo.**

Analysis:




(3)
ACEPCIO
ACEPCIONS (4)
CTTL2
NUM
TEXT
cf2
II
CTTL1
TP
PUNTO
cf1
cierto
juego
...
.
(4)
ACEPCIO
ACEPCIONS
CTTL2
NUM
GEO
TEXT
cf2
III
Méj y P.Rico
CTTL1
TP
PUNTO
cf1
participación
.
(5)
RELA
CTTX3
RELACIONS
j3
RELACIO
RELACIONS
CTTX6
REL
CTTX7
TX
j6
sin.
j7
1.v.pedazo

Appendix 2

Some examples of lispification.

Entry :

I) cacho (l.*calculu,* piedrecita)*m.* fam. Pedazo pequeño de alguna cosa. 2. Cierto juego de naipes.3. *Méj y P.Rico.* Participación pequeña en un número de la lotería.
SIN 1. V. **Pedazo**

Lispification:

((ENT: cacho)
(NH : I)
(ETIM: l.*calculu*, piedrecita)
(SENSE 1)
(CATG: m)
(REG: fam)
(DEF: pedazo pequeño de alguna cosa)).
(SENSE 2)
(CATG: m)
(DEF: cierto juego de naipes)).
(SENSE 3)
(CATG: m)
(GEO: Méj ,P.Rico)
(DEF: participación pequeña en un número de la lotería)).
(RELA :1)
(TIPOR: SIN)
(TXR: 1.v. pedazo))).

Entry:

II) caho,- cha (l. *coactu;* pp. de *cogere*, coger, condensar) *adj.* Gacho.

Lispification:

((ENT: cacho)
(NH : II)
(FLEX: cha)
(ETIM: l.*coactu*; pp. de *cogere*, coger, condensar)
(SENSE 1)
(CATG:adj)
(DEF: Gacho)).

Appendix 3

An example of a template for lexical entry in the monolingual vox :

TAGS AT DICTIONARY LEVEL

Dict_Source: VOX

LANGUAGE
Lang: spanish

TAGS AT ENTRY LEVEL

ENTRY

* I) cacho (l. calculu, piedrecita) m. fam. pedazo pequeño de alguna cosa. 2 Cierto juego de naipes. 3 Mej. y P. Rico. Participación pequeña en un número de la loteria.
Sin.: 1 v. pedazo.

HEADWORD_GROUP
Hdwd_type: lemma
Hdwd_form: cacho
(Hdwd_figure_ref):
(Hdwh_text): "cacho"
(Hdwd_Homonym_No.): 1

(VARIANT_GROUP)
(Variant_label):
Variant_form:

(ETYMOLOGY_GROUP)
Etymology_text: l.calculu, piedrecita
(CROSS-REFERENCE_GROUP)
(Xref_text):
Xref_label:
+Xref_entry:
(Xref_extens):

+HOM_GROUP
Hom_No.: NIL
(Hom_form):
(Hom_Compact):

(GRAM_INF_GROUP)
Gram_Inf_text: m.

*(POS_GROUP)
(POS): n.
(subcat):
(subtype):
(gender): m.
(number):
(various):

```
*(MORPH_GROUP)
      (morph_text):
      (morph_label):
      +morph_form:

+(SENSE_GROUP)
      Sense.No.: 1
      (CROSS_REFERENCE_GROUP)...
      (DEFINITION_GROUP)
      *(SEMANTIC_LABEL_GROUP)
                                          (Semantic_label_text):fam.
                                          (Subject_code):
            (Semantic_code):
                                          (Register_code):fam.
                                          (Usage_code):
            (Geographic_code):
            (American_code):

      (DEF_GROUP)
                                          Def_text: pedazo pequeño
                                            de alguna cosa.
                                          *(Implicit_Xref):
            (Figure_ref):
            *(EXAMPLE_GROUP)
                                          +Ex_text:
                  (Ex_label):
                (Ex_explanation):

            (COMPOUND_GROUP)
            Cpd_label:
            Cpd_form:
            DEF_GROUP ...
            (SEMANTIC_LABEL_GROUP)...

(SEMANTIC_RELATIONS_GROUP)
      (SYNONYM_GROUP)
                                          Syn_label: SIN
                                          +Synonym: Pedazo
      (ANTONYM_GROUP)
                                          Ant_label:
            +Antonym:
      ("ALTERATE"_GROUP)
                                          Alt_label:
                                          +"Alterate":

+(SENSE_GROUP)
      Sense.No.: 2
      (CROSS_REFERENCE_GROUP)...
      (DEFINITION_GROUP)
      *(SEMANTIC_LABEL_GROUP)
                                          (Semantic_label_text):
                                          (Subject_code):
            (Semantic_code):
                                          (Register_code):
```

(Usage_code):
(Geographic_code):

(American_code):

(DEF_GROUP)

Def_text: cierto juego de naipes.
*(Implicit_Xref):

(Figure_ref):

*(EXAMPLE_GROUP)

+Ex_text:

(Ex_label):
(Ex_explanation):

(COMPOUND_GROUP)
Cpd_label:
Cpd_form:

DEF_GROUP ...
(SEMANTIC_LABEL_GROUP)...
(SEMANTIC_RELATIONS_GROUP)
(SYNONYM_GROUP)

Syn_label:
+Synonym:

(ANTONYM_GROUP)

Ant_label:

+Antonym:
("ALTERATE"_GROUP)

Alt_label:
+"Alterate":

+(SENSE_GROUP)
Sense.No.: 3
(CROSS_REFERENCE_GROUP)...
(DEFINITION_GROUP)
*(SEMANTIC_LABEL_GROUP)

(Semantic_label_text):
(Subject_code):

(Semantic_code):

(Register_code):
(Usage_code):
(Geographic_code):

(American_code): Mej. y P.Rico.

(DEF_GROUP)

Def_text: participaciÛn pequeÒa en un numero de la loteria.
*(Implicit_Xref):

(Figure_ref):

*(EXAMPLE_GROUP)

+Ex_text:
(Ex_label):
(Ex_explanation):

(COMPOUND_GROUP)
Cpd_label:
Cpd_form:

DEF_GROUP ...
(SEMANTIC_LABEL_GROUP)...
(SEMANTIC_RELATIONS_GROUP)
(SYNONYM_GROUP)
Syn_label:
+Synonym:
(ANTONYM_GROUP)
Ant_label:
+Antonym:
("ALTERATE"_GROUP)
Alt_label:
+"Alterate":

(SEMANTIC_RELATIONS_GROUP) ...
(HOMOF_GROUP)
homoph_label:
homoph_text:
homoph_entry:

ENTRY

* II) cacho,- cha (l.coactu;pp.de cogere,coger,condensar) adj. Gacho.

HEADWORD_GROUP
Hdwd_type: lemma
Hdwd_form: cacho
(Hdwd_figure_ref):
(Hdwh_text):"cacho, -cha".
(Hdwd_Homonym_No.): 2

(VARIANT_GROUP)
(Variant_label):
Variant_form:

(ETYMOLOGY_GROUP)
Etymology_text:l.coacto;pp.de cogere,coger,condensar.
(CROSS-REFERENCE_GROUP)
(Xref_text):coger, condensar
Xref_label:uso de redondilla
+Xref_entry: coger, condensar
(Xref_extens):

```
+HOM_GROUP
        Hom_No.: NIL
        (Hom_form):
        (Hom_Compact):

    (GRAM_INF_GROUP)
        Gram_Inf_text:  adj.

        *(POS_GROUP)
        (POS): adj.
        (subcat):
        (subtype):
        (gender):
        (number):
        (various):


        *(MORPH_GROUP)
            (morph_text):: ",  -cha".
            (morph_label): ",  -".
            +morph_form:  "cha".


        +(SENSE_GROUP)
            Sense.No.: 1
            (CROSS_REFERENCE_GROUP)...
            (DEFINITION_GROUP)
            *(SEMANTIC_LABEL_GROUP)
                                                (Semantic_label_text):
                                                (Subject_code):
                (Semantic_code):
                                                (Register_code):
                                                (Usage_code):
                                                (Geographic_code):
                (American_code):

            (DEF_GROUP)
                                                Def_text: Gacho
                                                *(Implicit_Xref):
                (Figure_ref):

                *(EXAMPLE_GROUP)
                                                +Ex_text:
                 (Ex_label):
                     (Ex_explanation):


                (COMPOUND_GROUP)
                Cpd_label:
                Cpd_form:
```

DEF_GROUP ...
(SEMANTIC_LABEL_GROUP)...
(SEMANTIC_RELATIONS_GROUP)
(SYNONYM_GROUP)
Syn_label:
+Synonym:
(ANTONYM_GROUP)
Ant_label:
+Antonym:
("ALTERATE"_GROUP)
Alt_label:
+"Alterate":

(SEMANTIC_RELATIONS_GROUP) ...
(HOMOPH_GROUP)
homoph_label:
homoph_text:
homoph_entry:

Appendix 4

Attribute values in the lexical entry template

Hdwd_type: lemma,
word_form,
abbreviation,
suffix,
preffix,
proper noun,
contracted form,
appocopation,
onomatopoeia.

Hdwd_Homonym_No.:NIL (if there is only one homonym)
1,2,3,4,...

Variant_label: VAR, var, También, Incorrecto, "-"

Hom_No.:NIL (if there is only one homograph)
1,2,3,n... (otherwise)

POS: adj = adjective
adv = adverb
conj= conjuction
s = noun
v, vb = verb
prep = preposition
pron = pronoun
interj = interjection
loc = locution
...

subcat: impers, intr, prnl, tr, tr.-prnl,

subtype:
(for adverb :) c (quantity)
m (manner),
l (place),
neg (negative),
o (order),
t (time)

(for noun:) pr (proper).

(for pronoun:) indef (indefinite)
relat (relative)

(for loc:) adj (adjectival)
conj (conjunctive)
adv (adverbial)
prep (prepositional)

gender: f (feminine),
m (masculine),
com (common),
amb (ambiguous).

number: sing, pl

morph_label: sing, pl, f, m, superl, aum, dim.

Sense.No.: NIL (if there is only one sense)
1,2,3,4,5, n... (otherwise)

Subject_code:
AERON., ARQUEOL., ASTRON., ALBAÑ.,
ANAT., ASTROL., ARTILL., BIB., BIOL., CARP.,
CETR., CINEM., CIR., COM., CONSTR., CRIST.,
DEP., DER., DIAL., ECON., ELECTR., EQUIT., ESC.,
ESGR., ETNOL., FIB.,FARM.,FILOL., FOS.,FISIOL.,
FON.,FORT., FOT., GEOD., GEOGR., GEOL., GEOM.,
GRAM., H.NAT.,IMPR., INFORM., LING., LIT.,
LITURG., L"G.,MAR., MAT., MEC., MED., METAL.,
METEOR., M...TR., MIL., MIN., MONT., MOR., M⁄S.,
N⁄M., NUMIS., PT., ORTOGR., PALEONT., PERS.,
PINT., POL., PSICO., QUOM., RET., TAUROM.,
TECNOL., TEOL., TOPOGR., TRIG., VETER.,
ZOOL., AGR., TECN.

Semantic_code: fig., burl., desp., despec., eufem.,
irón.,

Register_code: lit., rúst., fam., científ.,fest.,
pleb., poét.,vulg., neol.

Usage_code: ant., desus., inus., p.anal.,
p.ant., p.excel., p.ext., p.us., us.,

Geographic_code: dial., ¡l., Albac., Alic., Alm.,
And., Ar., Ast., Bad., Burg., Các., Cád., Can.,
Cord., C.Real.,Cuen.,Extr.,Gal., Gran., Guadal.,
Guip., Logr., Mál., Murc., Nav., Pal., Sal.,Sant.,
Seg., Sev., Sor., Tol., Val., Vallad., Zam., Zar., ...

American_code: Amér., Amér.Central., Amér.Merid.,
Ant.,Argent., Bol., Colomb.,
C.Rica., Ecuad., Filip., Guat., Hond.,Méj., Nicar.,
Pan., Parag.,P.Rico., R. de la Plata., Salv., S.Dom.,
Urug., Venez.,

Ex_label: FR, fr, frs, EXPR.

Without any label, an example can appear elsewhere in the definition. It is introduced by a colon and followed by the text of the example written in italics. The headword is sometimes referred to with an "~".

Syn_label: SIN

Ant_label: CONTR

Alt_label: aum., der., dim., superl.,SUPERL.,

Cpd_label: "~" text in italics

Xref_label: v., V., REL.,

homoph_label: HOMOF.

6. References

Alshawi H.- B. Boguraev - D. Carter (1989)
"Placing LDOCE on-line", in *Computational lexicography for Natural Language Processing,*
Longman, London, 41-64.

Calzolari, N.- C. Peters- A. Roventini (1990)
"Computational model of the dictionary entry", Acquilex, Esprit BRA 3030. April 1990.

Diccionario General Ilustrado de la Lengua Española VOX,(1987). Ed. Biblograf S.A., Barcelona.

Vossen, Piek (1989)
"Getting to grips with the structure of the VanDale Dictionary". Amsterdam University, December. ADDICT Group.

Acknowledgements

We would like to thank Bibliograf Publishers for allowing the Spanish group to use the MRD version of the Spanish Vox dictionary in/ this project.