

FLANOM: Flexionador y lematizador automático de formas nominales*

Autores: Santana, O.; Pérez, J.; Carreras, F.; Duque, J.D.; Hernández, Z.; Rodríguez, G.
Departamento de Informática y Sistemas
Universidad de las Palmas de Gran Canaria

RESUMEN

El propósito del presente trabajo consiste en presentar una aplicación informática desarrollada por el *Grupo de Estructuras de Datos del Departamento de Informática y Sistemas de la Universidad de Las Palmas de Gran Canaria* (<http://protos.dis.ulpgc.es>) útil para: a) lematizar las diferentes formas nominales (identifica la forma canónica, su categoría gramatical y la flexión o derivación que la produce); y b) generar una forma nominal a partir de su forma canónica y flexión o derivación —en ambos procesos, se consideran las modificaciones necesarias debidas a la composición por prefijación. El reconocimiento y la generación de las formas se realizan de manera bidireccional por medio de una misma estructura de datos —recorrerla en sentidos contrarios implica que se opere con ella como herramienta lematizadora o flexionadora. Considera: género y número en los sustantivos, adjetivos, pronombres y artículos; heteronimia por cambio de sexo en los sustantivos; grado superlativo en los adjetivos y adverbios; adverbialización y adverbialización del superlativo en los adjetivos; derivación apreciativa en los sustantivos, adjetivos y adverbios; formas canónicas múltiples en todas las categorías gramaticales; formas invariantes tales como preposiciones, conjunciones, exclamaciones, palabras de otros idiomas y locuciones o frases. El universo está compuesto por 109 194 formas canónicas e incluye todas las entradas no verbales del *Diccionario de la Lengua Española* de la Real Academia Española (70 056), del *Diccionario General de la Lengua Española* Vox (83 709), del *Diccionario de Uso del Español* de María Moliner (66 099), del *Gran Diccionario de la Lengua Española* de Larousse Planeta (58 605), del *Diccionario de voces de uso actual* dirigido por Manuel Alvar Ezquerro (4 644), del *Gran Diccionario de Sinónimos y Antónimos* de Espasa-Calpe (31 011) y del *Diccionario Ideológico* de Julio Casares (56 533).

0.— INTRODUCCIÓN.

A pesar de la popularización de los ordenadores no se ha resuelto la comunicación entre las personas y las máquinas en lenguaje natural. La ingeniería lingüística constituye un campo de investigación cada vez más estratégico; con el reciente desarrollo de las

* Este artículo complementa al “FLAVER: Flexionador y lematizador automático de formas verbales que trata conjugación y pronombres enclíticos” publicado en *Lingüística Española Actual*, XIX-2, 1997, págs. 229/282.

industrias de la lengua, los usuarios de la tecnología lingüística disponen de recursos que tienden hacia la sociedad de la información multilingüe. El tratamiento automatizado de la morfología del español es la piedra angular sobre la que construir cualquier procesador de lenguaje natural que habrá de considerar ulteriormente la sintaxis y la semántica. La disponibilidad de un procesador morfológico facilita además adecuada solución a una amplia gama de problemas actuales, tales como consultas en bases de datos documentales, corrección ortográfica, lematización, etiquetado, análisis y generación de textos escritos, tratamiento de corpus, etcétera.

El presente trabajo se enmarca dentro de un conjunto de herramientas orientadas a servir de ayuda en la elaboración de documentos escritos —dedicadas a analizar el texto que se produce y a ofrecer facilidades que enriquezcan la expresión— tales como frecuencia de aparición de palabras, empleo de formas verbales y no verbales, corrección ortográfica, búsquedas en texto libre, detección de vicios, depuración de estilos, análisis sintáctico, diccionarios de sinónimos o ideológicos, etc.

Se ha desarrollado un instrumento con la capacidad de:

1. Lematizar una forma no verbal que obtiene junto a la forma canónica, su categoría gramatical y la correspondiente relación de flexión o derivación.
2. Generar una determinada forma flexionada o derivada a partir de la forma canónica.

Considera las siguientes flexiones y derivaciones:

- Género y número en los sustantivos, adjetivos, pronombres y artículos.
- Heteronimia por cambio de sexo en los sustantivos.
- Grado superlativo en los adjetivos y adverbios.
- Adverbialización y adverbialización del superlativo en los adjetivos
- Derivación apreciativa en los sustantivos, adjetivos y adverbios.
- Formas canónicas múltiples en todas las categorías gramaticales.
- Formas invariantes tales como preposiciones, conjunciones, exclamaciones, palabras de otros idiomas y locuciones o frases.

Trata los prefijos que sean necesarios en el análisis y los incorpora en la generación de la forma solicitada.

1.– MORFOLOGÍA NO VERBAL.

Para el estudio de la morfología no verbal del español conviene tener en cuenta las diferentes categorías gramaticales y los accidentes o flexiones que pueden sufrir cada una de ellas. Se consideran las siguientes categorías gramaticales: sustantivos, adjetivos, pronombres, artículos, adverbios y otras formas que carecen de flexión —preposiciones, conjunciones, exclamaciones, palabras de otros idiomas y locuciones o frases.

Los sustantivos poseen las desinencias flexivas de género y número y, mediante los correspondientes sufijos apreciativos, pueden derivar en sus aumentativos, diminutivos o peyorativos. En general, y con respecto al género, se clasifican en masculinos o femeninos; algunos cambian de género, con o sin repercusiones semánticas, y varían su morfología

bien en la desinencia, bien en toda su forma —heteronimia—, o conservan la misma forma —comunes—; los menos tienen ambigüedad respecto al género.

Los adjetivos tienen desinencias flexivas de género, número y grado superlativo —ciertas irregularidades afectan al lexema. Admiten derivación apreciativa mediante sufijos aumentativos, diminutivos o peyorativos; en general pueden ser adverbializados añadiendo a la forma femenina —de su forma original o de la superlativizada— la terminación *-mente*. Hay adjetivos de dos terminaciones —una para cada género—, y de una terminación —no cambian su morfología con el género.

En los pronombres y artículos se consideran las desinencias flexivas de género (masculino, femenino y neutro) y de número (singular y plural).

Los adverbios que mayor interés suponen son los caracterizados por su terminación en *-mente*: provienen de un adjetivo, originariamente son de modo y no tienen flexión con respecto al género ni al número.

Las preposiciones, conjunciones, exclamaciones, palabras de otros idiomas y locuciones o frases se tratan como invariantes, no admiten flexión ni derivación.

Aunque sean cuestiones bien sabidas, no está de más recordar cuáles son las reglas morfológicas que rigen estos procesos, pues de esta manera se entenderá mejor la formalización a la que se quiere llegar.

1.1.— LA FORMACIÓN DEL FEMENINO EN SUSTANTIVOS Y ADJETIVOS.

Son de uso genérico las siguientes reglas básicas:

1. Para los sustantivos:

1.1. Los terminados en *o* cambian la *o* por *a*.

niño ⇒ *niña*

1.2. Los terminados en consonante añaden una *a*.

horticultor ⇒ *horticultora*

1.3. Los terminados en *e* permanecen invariables,

el conferenciante ⇒ *la conferenciante*

aunque en ocasiones cambian la *e* por *a*.

franchute ⇒ *franchuta* ...

1.4. Los terminados en *a* permanecen invariables,

el malabarista ⇒ *la malabarista*

salvo excepciones.

poeta ⇒ *poetisa* ...

2. Para los adjetivos:

2.1. Los gentilicios que acaban en consonante añaden una *a*.

andaluz ⇒ *andaluza*

2.2. Los que acaban en *o*, *ote* o *ete* cambian la vocal final por *a*.

malo ⇒ *mala*

grandote ⇒ *grandota*

vejete ⇒ *vejeta*

2.3. Los que acaban en *an*, *on* u *or* añaden una *a*,

haragán ⇒ *haragana*

bribón ⇒ *bribona*

agresor ⇒ *agresora*

salvo los comparativos latinos: *exterior*, *inferior*, *menor*, *peor*,...

2.4. El resto permanecen invariantes,

amable, *azul*, *decente*, *verde*,...

salvo excepciones.

el lebrél ⇒ *la lebrela*

el cliente ⇒ *la clienta* ...

Además de las implícitas en las reglas anteriores existen otras excepcionalidades que se pueden concretar en:

3. Excepciones a la formación del femenino:

3.1. Muchos sustantivos son sólo masculinos.

libro, *objeto*,...

3.2. Muchos sustantivos son sólo femeninos.

casa, *legaña*,...

3.3. Hay sustantivos femeninos que acaban en *o*.

la dinamo, *la bonoloto*,...

3.4. Hay sustantivos masculinos que acaban en *a*.

el califa, *el fotograma*,...

3.5. Hay sustantivos de género común y adjetivos de una terminación, cuya morfología no cambia al usarlos en masculino o en femenino.

el cónyuge ⇒ *la cónyuge*

el mártir ⇒ *la mártir*

el testigo ⇒ *la testigo*

el tratable ⇒ *la tratable* ...

3.6. Hay sustantivos de género ambiguo —aunque el uso va reduciéndolos o dándoles una nueva distribución, se crean otros casos por analogía, ignorancia, etc.—, normalmente pueden ser usados en ambos géneros.

el linde / *la linde*

el mar / *la mar*

el pus / *la pus* ...

3.7. En casos especiales pueden aparecer otras terminaciones como *esa*, *isa*, *ina*, *iza* o *triz*.

conde ⇒ *condesa*

sacerdote ⇒ *sacerdotisa*

héroe ⇒ *heroína*

róbalo ⇒ *robaliza*

emperador ⇒ *emperatriz*

bisector ⇒ *bisectriz* ...

3.8. Algunos sustantivos tienen dos formas para el femenino.

actor ⇒ *actora* / *actriz*

diablo ⇒ *diabla* / *diabla*

generador ⇒ *generadora* / *generatriz*

juez ⇒ *juez* / *jueza*

tigre ⇒ *tigra* / *tigresa* ...

3.9. Algunos sustantivos poseen heteronimia por cambio de sexo, cambian la forma de la palabra y no sólo su desinencia.

toro ⇒ *vaca*

carnero ⇒ *oveja* ...

3.10. Se ha de resaltar aquí el problema que surge de la incorporación de la mujer a trabajos, oficios, etc., tradicionalmente de hombres —unas veces la Academia admite el femenino y otras no, unas veces los admite la sociedad y otras no—; en un esfuerzo por lograr la máxima generalidad en este trabajo se recogen tanto los aceptados por la Academia como los admitidos por el uso (ya incorporados en otros diccionarios).

3.11. También se han tenido en cuenta, aún siendo más raros, los masculinos regresivos.

modista ⇒ *modisto* (antietimológico)

comadrona ⇒ *comadrón* ...

4. Conviene tener en cuenta que existen voces que cambian su significado con el género,

el cólera / *la cólera*

el corte / *la corte*

el frente / *la frente*

el guardia / *la guardia* ...

o según la forma del femenino;

la actriz / *la actora*

la directora / *la directriz* ...

y que el género *no es igual* al sexo.

la foca macho / *el cocodrilo hembra*

1.2.– LA FORMACIÓN DEL PLURAL EN SUSTANTIVOS Y ADJETIVOS.

Son de uso genérico para los sustantivos y los adjetivos las siguientes reglas básicas:

1. Para las palabras terminadas en consonante:

1.1. Distinta de *z, c, x, n, s* añaden *es*,

árbol ⇒ *árboles*

ley ⇒ *leyes*

con excepciones que añaden *s*

réquiem ⇒ *réquiems* ...

y, en este caso, si terminan en *y* se hace la corrección ortográfica.

guirigay ⇒ *guirigáis* ...

1.2. Las palabras terminadas en *z*, precisan corrección ortográfica, la sustituyen por *ces*.

matriz ⇒ *matrices*

1.3. Las palabras terminadas en *c*, precisan corrección ortográfica, la sustituyen por *ques*,

ruc ⇒ *ruques*

con excepciones que añaden *s*.

coñac ⇒ *coñacs* ...

1.4. Las palabras terminadas en *x* permanecen invariables,

el clímax ⇒ *los clímax*

con excepciones que añaden *es* y precisan modificación ortográfica.

carcax ⇒ *carcajes*

fénix ⇒ *fénices* ...

1.5. Las palabras terminadas en *n* o *s* que sean agudas añaden *es*,

clan ⇒ *clanes*

gas ⇒ *gases*

si llevan tilde la pierden,

camión ⇒ *camiones*

japonés ⇒ *japoneses*

salvo que sea para formar hiato

mohín ⇒ *mohínes*

o diacrítica.

quién ⇒ *quiénes*

1.6. Las palabras terminadas en *n* que no sean agudas añaden *es*

oxímoron ⇒ *oxímorones*

y las llanas ganan una tilde al pasar a esdrújulas.

velamen ⇒ *velámenes*

1.7. Las palabras terminadas en *s* que no sean agudas permanecen invariables.

el guardagujas ⇒ *los guardagujas*

2. Para las palabras terminadas en vocal:

2.1. No acentuada añaden *s*,

copa ⇒ *copas*

excepto los monosílabos que añaden *es*,

el yo ⇒ *los yoes*

la a ⇒ *las aes*

la i ⇒ *las íes*

salvo las notas musicales que añaden *s*.

do, re, mi, fa, la, si ⇒ *dos, res, mis, fas, las, sis*

2.2. *á* u *ó*, añaden *es* y pierden la tilde,

abacá ⇒ *abacaes*

paletó ⇒ *paletoes*

con excepciones que añaden *s*.

dominó ⇒ *dominós ...*

2.3. *é*, añaden *s*.

café ⇒ *cafés*

2.4. *í* o *ú*, añaden *es*,

baladí ⇒ *baladíes*

bambú ⇒ *bambúes*

con excepciones que añaden *s*.

cañí ⇒ *cañís*

canesú ⇒ *canesús ...*

Además, hay un número importante de excepciones a la formación del plural que pueden concretarse en:

3. Excepciones a la formación del plural:

3.1. Palabras que sólo se usan en singular.

cariz, cenit, salud, ...

3.2. Palabras que sólo se usan en plural,

albricias, ambages, anales, ...

aunque los hablantes ponen algunas de ella bajo la forma singular.

braga, gafa, ...

3.3. Palabras invariantes —igual forma en singular y plural.

el caos ⇒ *los caos*

el quórum ⇒ *los quórum*

el superávit ⇒ *los superávit ...*

3.4. La desinencia del plural no aparece al final de la palabra.

hijodalgo ⇒ *hijosdalgo*

medianoche ⇒ *mediasnoches* ...

3.5. Cambian la sílaba tónica al formar el plural.

régimen ⇒ *regímenes*

espécimen ⇒ *especímenes*

carácter ⇒ *caracteres* ...

3.6. Otras irregularidades.

desiderátum ⇒ *desiderata*

hipérbaton ⇒ *hipérbatos*

vermut ⇒ *vermús*

zinc ⇒ *zines* ...

3.7. Palabras que tienen más de una forma para el plural.

champú ⇒ *champús* / *champúes*

accésit ⇒ *accésit* / *accésits* / *accessis* ...

3.8. Palabras que cambian el género al formar el plural.

la orina ⇒ *los orines* ...

3.9. Palabras de género ambiguo que forman el plural con un solo género.

el mar / la mar ⇒ *los mares* ...

4. Conviene tener en cuenta que hay palabras que pueden cambiar su significado con el número.

anal / *anales*

celo / *celos*

prez / *preces*

seso / *sesos* ...

1.3.– EL GÉNERO Y EL NÚMERO EN LOS PRONOMBRES.

Se resume en los consabidos esquemas:

PRONOMBRES PERSONALES DE PRIMERA PERSONA					
	Masculino singular	Femenino singular	Neutro singular	Masculino plural	Femenino plural
<i>Sujeto</i>	Yo	yo	—	nosotros	nosotras
<i>Objeto átono</i>	Me	me	—	nos	nos
<i>Objeto tónico</i>	mí, conmigo	mí, conmigo	—		

PRONOMBRES PERSONALES DE SEGUNDA PERSONA					
	Masculino singular	Femenino singular	Neutro singular	Masculino plural	Femenino plural
<i>Sujeto</i>	tú, usted	tú, usted	—	vosotros, ustedes	vosotras, ustedes
<i>Objeto átono</i>	te	te	—	os	os
<i>Objeto tónico</i>	ti, contigo	ti, contigo	—		
	vos	vos	—	vos	vos

PRONOMBRES PERSONALES DE TERCERA PERSONA (*)					
	Masculino singular	Femenino singular	Neutro singular	Masculino plural	Femenino plural
<i>Sujeto</i>	él	ella	ello	ellos	ellas
<i>Objeto directo</i>	lo	la	lo	los	las
<i>Objeto indirecto</i>	le	le		les	les
<i>Reflexivo y objeto átono con otro pron.</i>	se	se		se	se
<i>Reflexivo y objeto tónico</i>	sí, consigo	sí, consigo		sí, consigo	sí, consigo

(*) Salvo en el sujeto la tercera persona se usa también como segunda persona de cortesía.

Los pronombres demostrativos¹ se distribuyen de la siguiente manera:

PRONOMBRES DEMOSTRATIVOS				
Masculino singular	Femenino singular	Neutro singular	Masculino plural	Femenino plural
éste	ésta	esto	éstos	éestas
ése	ésa	eso	ésos	ésas
aquél	aquélla	aquello	aquéllos	aquéllas

¹ Según la *Ortografía* académica, pueden escribirse sin tilde cuando no resulte ambigüedad.

Los pronombres indefinidos —generalmente tienen función adjetiva— no poseen siempre variación de género y número; son invariantes: *algo, nada, alguien, nadie, más, menos, cada, demás,...*; con frecuencia tienen como características de género y número las terminaciones *o/a/os/as*, aunque hay excepciones. La siguiente tabla muestra algunos ejemplos:

PRONOMBRES INDEFINIDOS				
Masculino singular	Femenino singular	Neutro singular	Masculino plural	Femenino plural
alguno	alguna	alguno	algunos	algunas
ninguno	ninguna	ninguno		
otro	otra	otro	otros	otras
todo	toda	todo	todos	todas
mucho	mucha	mucho	muchos	muchas
poco	poca	poco	pocos	pocas
demasiado	Demasiada	demasiado	demasiados	demasiadas
bastante	bastante	bastante	bastantes	bastantes
cualquiera	Cualquiera		cualesquiera	cualesquiera
quienquiera	Quienquiera		quienesquiera	quienesquiera
uno	una	uno	unos	unas

Los pronombres relativos se limitan a ser representación de un sustantivo enunciado antes de la proposición; son átonos y se pueden concretar en: *que, quien, cual, cuanto* y *cuyo*. Los pronombres interrogativos se utilizan para eliminar la imprecisión de identidad o cantidad y por tanto aparecen en oraciones interrogativas; se corresponden con los relativos (excepto *cuyo* que resulta inusitado) y se diferencian de éstos en que son tónicos y llevan acento ortográfico. La siguiente tabla expresa sus formas en función del género y del número con el que se usen:

PRONOMBRES RELATIVOS				
Masculino singular	Femenino singular	Neutro singular	Masculino plural	Femenino plural
que	que	que	que	que
quien	quien		quienes	quienes
cual	cual	cual	cuales	cuales
cuanto	cuanta	cuanto	cuantos	cuantas
cuyo	cuya		cuyos	cuyas
PRONOMBRES INTERROGATIVOS				
qué	qué	qué	qué	qué
quién	quién		quiénes	quiénes
cuál	cuál	cuál	cuáles	cuáles
cuánto	cuánta	cuánto	cuántos	cuántas

Los pronombres posesivos —en general se consideran adjetivos posesivos— se pueden clasificar en formas monosilábicas y formas bisilábicas. Las primeras son átonas, invariantes respecto al género y forman el plural añadiendo *-s*; las bisilábicas tienen como desinencia para el masculino singular una *-o* que también se usa para el neutro singular, forman el femenino cambiando la desinencia por *-a* y los plurales del masculino y del femenino añadiendo una *-s*.

PRONOMBRES POSESIVOS				
Masculino singular	Femenino singular	Neutro singular	Masculino plural	Femenino plural
mi	mi		mis	mis
tu	tu		tus	tus
su	su		sus	sus
mío	mía	mío	míos	mías
tuyo	tuya	tuyo	tuyos	tuyas
suyo	suya	suyo	suyos	suyas
cuyo	cuya		cuyos	cuyas
nuestro	nuestra	nuestro	nuestros	nuestras
vuestro	vuestra	vuestro	vuestros	vuestras

1.4.– EL GÉNERO Y EL NÚMERO EN LOS ARTÍCULOS.

El artículo tiene las formas sabidas *el, la, los, las* y el neutro singular *lo*.

Para las formas *un / una* véase lo que se explica en la parte correspondiente de los numerales.

1.5.– EL GÉNERO Y EL NÚMERO EN LOS NUMERALES: CARDINALES, ORDINALES, FRACCIONARIOS Y PROPORCIONALES.

Los numerales son adjetivos, nombres o pronombres, y, en ocasiones, adverbios según la función gramatical que desempeñen. Se clasifican en cardinales, ordinales, partitivos o fraccionales y múltiplos o proporcionales.

Los cardinales sirven para designar o cuantificar mediante un número exacto: *cero, un/uno, dos, tres, cuatro, cinco, seis, siete, ocho, nueve, diez, once, doce, trece, catorce, quince, dieciséis, diecisiete, dieciocho, diecinueve*, —entre *dieciséis* y *diecinueve* la grafía con menor frecuencia permite escribir *diez y seis,...*, *diez y nueve*—, *veinte, veintiuno, veintidós, veintitrés, veinticuatro, venticinco, veintiséis, veintisiete, veintiocho, veintinueve, treinta, cuarenta, cincuenta, sesenta, setenta, ochenta, noventa, cien/ciento, doscientos, trescientos, cuatrocientos, quinientos, seiscientos, setecientos, ochocientos, novecientos, mil, millón, billón, trillón*, etcétera. Para la expresión de los restantes números se recurre a la combinación de dos o más palabras de la lista precedente intercalando la conjunción *y* entre decenas y unidades. No tienen variación de número: *uno* es singular y el resto son plurales; salvo *millón/millones, billón/billones*, etcétera. Son invariantes con el género salvo *uno/una, veintiuno/veintiuna* y los acabados en *-ientos* que para el femenino usan *-ientas*.

Los ordinales sirven para designar el orden en una secuencia: *primero, segundo, tercero, cuarto, quinto, sexto, séptimo, octavo, noveno/nono, décimo, undécimo, duodécimo, decimotercero, decimocuarto, decimoquinto, decimosexto, decimoséptimo, decimoctavo, decimonoveno, vigésimo, trigésimo, cuadragésimo, quincuagésimo, sexagésimo, septuagésimo, octogésimo, nonagésimo, centésimo, ducentésimo, tricentésimo, cuadringentésimo, quingentésimo, sexcentésimo, septingentésimo, octingentésimo, noningentésimo, milésimo, millonésimo, billonésimo*. Para la expresión

de los restantes números se recurre a la combinación de dos o más palabras de la lista precedente. Admiten variación de género y número con las desinencias *o/a/os/as*.

Los fraccionarios o partitivos sirven para designar las partes iguales de un todo: *medio, tercio/tercera*², *cuarto, quinto, sexto, séptimo, octavo, noveno, décimo*. A partir de once se forman con los cardinales y el sufijo tónico *-avo*: *onceavo*³, *doceavo, treceavo, catorceavo, quinceavo, dieciseisavo*,...; en algunos casos, se admite con números menores, *seisavo*. Además se usan como fraccionarios los ordinales en *-ésimo*. Todos ellos admiten variación de género y número con las desinencias *o/a/os/as*.

Los proporcionales o múltiplos sirven para expresar cantidades que contienen algo varias veces exactamente: *doble* o *duplo, triple, cuádruple* o *cuádruplo, quíntuple* o *quíntuplo, séxtuplo, séptuplo, óctuple* u *óctuplo, nóñuplo, décuplo, undécuplo, duodécuplo, terciodécuplo, céntuplo*. La variación de género sólo la tienen los terminados en *-o* que forman el femenino con *-a*, todos forman el plural añadiendo *-s*.

1.6.– EL GRADO SUPERLATIVO.

En los adjetivos se puede considerar el grado superlativo como una flexión con los morfemas flexivos: *-ísimo* para el masculino singular, *-ísima* para el femenino singular, *-ísimos* para el masculino plural e *-ísimas* para el femenino plural.

grande ⇒ *grandísimo / grandísima / grandísimos / grandísimas*

— Hay que tener en cuenta que los acabados en *-ble* lo forman con *-bilísimo/a/os/as*.

amable ⇒ *amabilísimo / amabilísima / amabilísimos / amabilísimas*

— Se precisan las correcciones ortográficas pertinentes:

— Si acaba en *ca, co* o *cu* se convierte la *c* en *qu*.

se-c-a ⇒ *se-qu-ísima*

ri-c-o ⇒ *ri-qu-ísimo*

francu-c-u ⇒ *francu-qu-ísimo*

— Si acaba en *z, za, zo* o *zu* se convierte la *z* en *c*.

efica-z ⇒ *efica-c-ísimo*

carro-z-a ⇒ *carro-c-ísima*

cazu-z-o ⇒ *cazu-c-ísimo*

En *zu* no se ha encontrado ningún caso.

— Si acaba en *ga, go* o *gu* se convierte la *g* en *gu*.

va-g-a ⇒ *va-gu-ísima*

cuellilar-g-o ⇒ *cuellilar-gu-ísimo*

En *gu* no se ha encontrado ningún caso.

² *Tercio* se emplea como principal y *tercera* como adjunto (en femenino).

³ En general, las formas de base acabadas en vocal permiten reducción: *onzavo, dozavo, dieciochavo*,...

— Existen adjetivos con irregularidades en la formación del superlativo.

iniquo ⇒ *iniquísimo*

sabio ⇒ *sapientísimo* ...

— Otros que además de la forma regular admiten una o varias formas irregulares.

pobre ⇒ *pobrísimo* / *paupérrimo*

bueno ⇒ *buenísimo* / *bonísimo* / *óptimo*

enemigo ⇒ *enemiguísimo* / *enemicísimo* / *inimicísimo* ...

— No todos los adjetivos admiten el grado superlativo en una formación regular:

— Por su significado: *absoluto*, *omnipotente*, *infinito*,...

— Por tener una marca superlativizadora: *buenísimo*, *mínimo*, *óptimo*,...

— Por ser gentilicios: *asturiano*, *canario*,...

— Por otros motivos: *exiguo*, *político*, *público*,...

— Aunque el grado superlativo es una característica propia de los adjetivos, existen adverbios que lo admiten.

cerca ⇒ *cerquísima*

lejos ⇒ *lejísimos*

tarde ⇒ *tardísimo* ...

1.7.— LA ADVERBIALIZACIÓN.

Los adjetivos permiten la formación de adverbios de modo mediante la concatenación de su forma femenina con la terminación *-mente*;

irónico ⇒ *irónicamente*

si son de una terminación, se añade directamente,

afable ⇒ *afablemente*

sin embargo, hay adjetivos que no admiten esta formación.

mucho, *ninguno*,...

También es posible adverbializar el superlativo de los adjetivos, añadiendo la terminación *-mente* a la forma femenina del superlativo de los adjetivos.

claro ⇒ *clarísimo* ⇒ *clarísimamente*

1.8.— SUFIJOS: CONSIDERACIONES GENERALES Y REGLAS DE DERIVACIÓN.

Una manera habitual de formar palabras nuevas es añadiendo sufijos en uso a palabras ya existentes. La palabra original se denomina primitiva y la compuesta derivada; cuando a un vocablo ya derivado se le añade otro sufijo resulta un derivado secundario. Por muy conocido que sea, a continuación se hará una detallada recapitulación de la derivación, con el objeto de ejemplificar bien todo lo que el sistema de lematización y flexión que se ha desarrollado tiene en cuenta.

— Si la primitiva termina en vocal: pierde las letras finales *a*, *e*, *o*,

cervez-a ⇒ *cervec-ero*

verd-e ⇒ *verd-or*

blanc-o ⇒ *blanc-ura*

y en algunas ocasiones pierde todo el diptongo final

palac-io ⇒ *palac-ete*

y puede alterar la forma del lexema;

rab-ia ⇒ *ráb-ico*

cuando termina en *u*, *i* normalmente el sufijo se añade sin sufrir modificaciones

tribu ⇒ *tribu-al*

cursi ⇒ *cursi-lería*

y es frecuente la reducción del diptongo resultante

tribu ⇒ *tribual* ⇒ *tribal*

o la vocal repetida.

dandi ⇒ *dandi-ismo* ⇒ *dandismo*

— Si termina en consonante, normalmente se añade el sufijo sin más,

verbal ⇒ *verbal-ismo*

las terminadas en *-dad* pierden el *-ad* final,

humed-ad ⇒ *humed-ecer*

sin embargo, existen excepciones.

virus ⇒ *viro-logía* ...

— Al unir el sufijo, si la palabra —o lo que quede de ella— acaba en *z*, ante *e* o *i*, se convierte en *c*.

rapa-z ⇒ *rapa-c-ejo*

tena-z ⇒ *tena-c-idad*

cabe-z-a ⇒ *cabe-c-ear*

ra-z-a ⇒ *ra-c-ial*

— Si acaba en *c* con sonido /z/ (*ce*, *ci*), ante *a*, *o* o *u* se convierte en *z*.

índi-c-e ⇒ *índi-z-ar*

avan-c-e ⇒ *avan-z-ón*

dul-c-e ⇒ *dul-z-ura*

Las palabras acabadas en *ci* no pierden la *i*.

— Si acaba en **c** con sonido /k/ (*c, ca, co, cu*), ante **e** o **i** se convierte en **qu**.

ticta-c ⇒ *ticta-qu-ear*

Isaa-c ⇒ *isaa-qu-ista*

barra-c-a ⇒ *barra-qu-ero*

Lor-c-a ⇒ *lor-qu-iano*

taba-c-o ⇒ *taba-qu-ería*

sa-c-o ⇒ *sa-qu-ilada*

Las palabras acabadas en *cu* no pierden la **u**.

rutucu ⇒ *rutucu-ejo*

cucurucu ⇒ *cucurucu-ito*

— Si acaba en **qu** ante **a**, **o** o **u** se convierte en **c**.

che-qu-e ⇒ *che-c-azo*

miriña-qu-e ⇒ *miriña-c-ote*

reben-qu-e ⇒ *reben-c-udo*

Las palabras acabadas en *qui* no pierden la **i**,

saltinbanqui ⇒ *saltinbanqui-ucho*

salvo excepciones.

güis-qu-i ⇒ *güis-c-azo ...*

— Si acaba en **g** con sonido /x/ (*ge, gi*), ante **a**, **o** o **u** se convierte en **j**.

tin-g-e ⇒ *tin-j-azo*

larin-g-e ⇒ *larin-j-ota*

esfin-g-e ⇒ *esfin-j-ucha*

Las palabras acabadas en *gi* no pierden la **i**.

— Si acaba en **g** con sonido /g/ (*ga, go, gu*), ante **e** o **i** se añade una **u**

dro-g-a ⇒ *dro-gu-ería*

bode-g-a ⇒ *bode-gu-illa*

tra-g-o ⇒ *tra-gu-ear*

pa-g-o ⇒ *pa-gu-ito*

y si el sufijo comienza por **u** seguida de vocal débil se debe añadir la diéresis para conservar el diptongo.

lla-g-a ⇒ *lla-g-üela*

Las palabras acabadas en *gu* no pierden la **u**.

— Si acaba en **gu**:

- a) Tras eliminar **a** u **o**, ante **e** o **i** se pone diéresis para mantener el sonido /gu/.

len-gu-a ⇒ *len-gü-eta*

a-gu-a ⇒ *a-gü-ita*

ambi-gu-o ⇒ *ambi-gü-edad*

exi-gu-o ⇒ *exi-gü-idad*

- b) Tras eliminar **e**, ante **a**, **o** o **u** se elimina la **u** para mantener el sonido /g/

meren-gu-e ⇒ *meren-g-ada*

plie-gu-e ⇒ *plie-g-ote*

potin-gu-e ⇒ *potin-g-ucho*

- y si el sufijo comienza por **u** seguida de vocal débil se debe reducir la doble vocal y añadir la diéresis para conservar el diptongo.

pira-gu-a ⇒ *pira-g-üela*

anti-gu-o ⇒ *anti-g-üelo*

li-gu-e ⇒ *li-g-üelo*

Una clase importante de sufijos la constituyen los apreciativos: se caracterizan por imprimir un matiz semántico subjetivo sin alterar normalmente la categoría gramatical y son tónicos. Según el tipo de eufemismo que producen se clasifican en: aumentativos, diminutivos o peyorativos. Además de los sufijos apreciativos discretivos en cada caso, se tiene en cuenta una importante cantidad de sufijos adicionales extraídos de las fuentes estudiadas; se recogen formas derivadas con más de un matiz semántico —*abogadillo*, diminutivo despectivo de *abogado*—; en este trabajo no se han tratado los meliorativos ya que en ninguna palabra de los diccionarios consultados figura tal característica. No suelen llevar apreciativos los sustantivos abstractos —la mayoría de los terminados en *ad* lo son, *libertad*, *igualdad*, *fraternidad*,... en cambio *ciudad* no lo es.

1.8.1.— LA FORMACIÓN DE AUMENTATIVOS.

Morfema normalmente sufijo, que añade al significado de la base léxica a la que se une la noción de magnitud o agrandamiento. Puede aportar, a la vez, otros valores, especialmente el de desprecio. Forman aumentativos los sustantivos, los adjetivos y algunos adverbios.

Los sufijos principalmente utilizados para la apreciación aumentativa son **-ón** y **-azo** que poseen sus formas femeninas **-ona** y **-aza** y los plurales correspondientes **-ones**, **-onas** y **-azos**, **-azas**; también se utilizan **-ote** y **-acho** con sus femeninos y plurales.

— Las palabras que terminan en vocal tónica, añaden el interfijo **-z-** para la formación de sus aumentativos (mantienen la vocal, aunque pierden la tilde).

papá ⇒ *papa-z-ote*

chalé ⇒ *chale-z-azo*

capó ⇒ *capo-z-acho*

— Además de éstos existe un importante número de sufijos que se utilizan con menor frecuencia para la formación de aumentativos, algunos son combinaciones de otros: *-achón, -achote, -ajón, -allón, -anchón, -ancón, -ardo, -arrón, -azón, -ejón, -ellón, -erón, -etón, -onazo, -orrón, -otón, -ucón, -udo, -ujón, -ullón, -urrón,...*

hues-o ⇒ hues-arrón

pícar-o ⇒ picar-onazo

grand-e ⇒ grand-ullón ...

— Existen algunas voces que reducen un diptongo en el lexema.

valient-e ⇒ valent-ón

fuert-e ⇒ fort-achón

sinvergüenz-a ⇒ sinvergonz-ón ...

— En ocasiones, sustantivos femeninos forman aumentativos en masculino

la vel-a ⇒ el vel-ón

la brum-a ⇒ el brum-azón ...

y puede darse además la reducción de diptongo en el lexema.

cazuel-a ⇒ cazol-ón ...

— A veces la formación de aumentativos es muy irregular.

bobo ⇒ bobalicón

nariz ⇒ narigón ...

— Aunque no es frecuente, en algunos adverbios se forman aumentativos.

antañ-o ⇒ antañ-azo

lej-os ⇒ lej-otes ...

1.8.2.— LA FORMACIÓN DE DIMINUTIVOS.

Morfema normalmente sufijo, que añade al significado de la base léxica a la que se une la noción de pequeñez en cantidad o tamaño. Puede aportar, a la vez, valores apreciativos, especialmente de afecto, aunque también irónicos y de desprecio. Los sustantivos, los adjetivos y algunos adverbios pueden tener diminutivos.

Los sufijos principalmente utilizados para la apreciación diminutiva son *-ito* e *-illo* con sus formas femeninas *-ita* e *-illa* y los plurales correspondientes *-itos*, *-itas* e *-illos*, *-illas*. También se utilizan *-ico* (aunque más bien es regional: Aragón, Navarra, Murcia y algunas zonas de Andalucía y Sudamérica), *-ín* (frecuente en Asturias) y *-uelo* (de aplicación más restringida y en ocasiones con carácter peyorativo) —admiten femeninos y plurales: *-ica/-icos/-icas*, *-ina/-ines/-inas*, *-uela/-uelos/-uelas*. Los mencionados sufijos se emplean con carácter general para la formación de diminutivos, aunque existen casos particulares en los que se añade algún interfijo.

— Añaden el interfijo **-c-** las palabras de dos o más sílabas:

— Agudas terminadas en **n** o en **r**.

camión ⇒ *camion-c-ito*

sabor ⇒ *sabor-c-illo*

— Acabadas en vocal tónica (mantienen la vocal, aunque pierden la tilde).

mamá ⇒ *mama-c-ita*

rondó ⇒ *rondo-c-illo*

bebé ⇒ *bebe-c-ico*

— Llanas acabadas en **n**.

dictamen ⇒ *dictamen-c-illo*

Se convierte en **-z-** si el sufijo no empieza por **e** ni por **i**, para mantener el sonido.

ladrón ⇒ *ladron-z-uelo*

— Añaden el interfijo **-ec-**:

— Los monosílabos acabados en consonante

son ⇒ *son-ec-illo*

y si acaban en **z** se convierte en **c**.

pe-z ⇒ *pe-c-ec-ito*

— Los bisílabos terminados en **e**.

cort-e ⇒ *cort-ec-ito*

— Los bisílabos con la primera sílaba en **ue**, **eu**, **ie**, **ei**

cuent-o ⇒ *cuent-ec-ito*

deud-a ⇒ *deud-ec-illa*

biel-a ⇒ *biel-ec-ita*

pein-e ⇒ *pein-ec-illo*

y si acaban en **za** o **zo** se cambia la **z** por una **c**,

fuer-z-a ⇒ *fuer-c-ec-illa*

lien-z-o ⇒ *lien-c-ec-ico*

si acaban en **ca** o **co** se cambia la **c** por una **qu**,

tuer-c-a ⇒ *tuer-qu-ec-ita*

peu-c-o ⇒ *peu-qu-ec-ín*

si acaban en **ga** o **go** se cambia la **g** por una **gu**.

mei-g-a ⇒ *mei-gu-ec-illa*

fue-g-o ⇒ *fue-gu-ec-ito*

— Los bisílabos con la última sílaba en *ia*, *io* o *ua*

savi-a ⇒ *savi-ec-illa*

orjol ⇒ *orjol-ec-ito*

ajuar ⇒ *ajuar-ec-illo*

y si acaban en *z* se cambia la *z* por una *c*,

arria-z ⇒ *arria-c-ec-ito*

si acaban en *gua* se añade la diéresis sobre la *u*,

le-gu-a ⇒ *le-gü-ec-illa*

aunque en algunos casos no se añade el interfijo.

a-gu-a ⇒ *a-gü-ita*

len-gu-a ⇒ *len-gü-eta*

— Las voces con dos o más sílabas terminadas en *io*.

radi-o ⇒ *radi-ec-ito*

cenobi-o ⇒ *cenobi-ec-ito*

Se convierte en *-ez-* si el sufijo no empieza por *e* ni por *i*, para mantener el sonido

lot-e ⇒ *lot-ez-uelo*

y en ocasiones hay reducción de diptongo.

pie-z-a ⇒ *pe-c-ez-uela*

— Añaden el interfijo *-ecec-* los monosílabos acabados en vocal.

té ⇒ *t-ecec-ito*

Se convierte en *-ecez-* si el sufijo no empieza por *e* ni por *i*, para mantener el sonido.

pie ⇒ *pi-ecez-uelo*

— Además de los mencionados, existe un importante número de sufijos que se utilizan con menor frecuencia para la formación de diminutivos, algunos son combinaciones de otros: *-aco*, *-ajo*, *-ancho*, *-anchuelo*, *-arajo*, *-ato*, *-ejo*, *-ellín*, *-encho*, *-ete*, *-etín*, *-eto*, *-ezno*, *-ijo*, *-ino*, *-iño*, *-irrinchín*, *-irritín*, *-itín*, *-oncho*, *-orritín*, *-orrotín*, *-ote*, *-uco*, *-ucho*, *-ugo*, *-ujo*, *-ullo*,...

alegr-e ⇒ *alegr-ete*

cuerp-o ⇒ *corp-iño*

caf-é ⇒ *caf-etín*

bob-o ⇒ *bob-irrinchín*

chic-o ⇒ *chiqu-itín* / *chic-orrotín* / *chiqu-irritín* ...

— Hay algunas voces que pierden el diptongo del lexema al formar diminutivos.

cuern-o ⇒ *corn-ecito*

vej-o ⇒ *vej-ecillo* ...

— En ocasiones, sustantivos femeninos forman diminutivos en masculino.

la faj-a ⇒ *el faj-ín*

la fald-a ⇒ *el fald-ellín* ...

— A veces la formación de diminutivos es muy irregular.

azúcar ⇒ *azuquítar*

capazo ⇒ *caparete* ...

— Aunque no es frecuente, con algunos adverbios se forman diminutivos

apen-as ⇒ *apen-itas*

ahor-a ⇒ *ahor-ita* ...

y en ocasiones de manera irregular.

ahor-a ⇒ *hor-ítica* ...

1.8.3.– LA FORMACIÓN DE PEYORATIVOS.

Morfema normalmente sufijo, que añade al significado de la base léxica a la que se une el valor de desprecio. Se unen a sustantivos, adjetivos y muy raramente a adverbios.

Los sufijos más utilizados para la formación de peyorativos son: *-ejo* y *-ucho*, con sus femeninos *-eja* y *-ucha*, y sus plurales *-ejos*, *-ejas* y *-uchos*, *-uchas*.

— Las palabras que terminan en vocal tónica, añaden el interfijo *-c-* para la formación de sus peyorativos (mantienen la vocal, aunque pierden la tilde).

puré ⇒ *pure-c-ejo*

Se convierte en *-z-* si el sufijo no empieza por *e* ni por *i*, para mantener el sonido.

sofá ⇒ *sofa-z-ucho*

— También se utilizan con frecuencia: *-acho*, *-aco*, *-ajo*, *-anco*, *-aracho*, *-ardo*, *-arraco*, *-aruco*, *-astre*, *-astro*, *-azo*, *-ejo*, *-ete*, *-orrio*, *-ote*, *-ucho*, *-uco*, *-ujo*, *-ujón*, *-usco*, *-uzco*, *-uzo*, ... Con poca frecuencia: *-achón*, *-achuelo*, *-aino*, *-ajón*, *-alla*, *-allón*, *-ancho*, *-anchón*, *-ancón*, *-arajo*, *-arrajo*, *-arrón*, *-azón*, *-ejón*, *-ellón*, *-encho*, *-erón*, *-etón*, *-istrajo*, *-onazo*, *-oncho*, *-orrón*, *-otazo*, *-ucón*, *-ullo*, *-ullón*, *-urrón*, ... Y ocasionalmente: *-ango*, *-arasco*, *-arazco*, *-aresco*, *-arezco*, *-ario*, *-arisco*, *-arizco*, *-arracho*, *-arro*, *-arusco*, *-aruzco*, *-enco*, *-iche*, *-il*, *-illo*, *-ina*, *-oche*, *-ón*, *-orio*, *-orro*, ...

bich-o ⇒ *bich-arraco*

cald-o ⇒ *cald-ibache*

cur-a ⇒ *cur-ángano* ...

— Existen algunas voces que pierden un diptongo intermedio al formar peyorativos.

ardient-e ⇒ *ardent-ejo*

cuerv-o ⇒ *corv-ucho* ...

— En ocasiones, sustantivos femeninos forman peyorativos en masculino.

la alde-a ⇒ *el alde-orro*

la cam-a ⇒ *el cam-astro* ...

— A veces la formación de peyorativos es muy irregular,

mezcla ⇒ *mezcolanza*

francés ⇒ *franchute*

llegando a la heteronimia

animal ⇒ *alimaña*

e incluso a la composición.

cojo ⇒ *cojitranco* ...

— Son infrecuentes los adverbios que forman peyorativos.

arrib-a ⇒ *arrib-ota* ...

1.9.— FORMAS INVARIANTES.

El resto de categorías gramaticales se consideran invariantes —no poseen flexión ni derivación—; entre ellas se cuentan las preposiciones, las conjunciones, las exclamaciones, las locuciones o frases y las palabras de otros idiomas.

1.10.— FORMAS CANÓNICAS MÚLTIPLES.

Algunas palabras vacilan entre varias posibles grafías —*cardiaco/cardíaco*, *gambuj/gambujo/gambux*,...—; para reflejar este heteromorfismo se establece el concepto de forma canónica múltiple que relaciona estas formas entre sí.

1.11.— LA PREFIJACIÓN.

La prefijación es una operación de derivación o de composición —según la teoría— que normalmente matiza, corrige o modifica el significado de la palabra, con independencia de la flexión, sin cambiar su categoría gramatical. No son aplicables en las formas consideradas invariantes, en los pronombres ni en los artículos.

Los prefijos considerados se pueden agrupar según su significación:

1. A ambos lados o alrededor de: *anfi-*, *circun-*

anfi- + *teatro* ⇒ *anfiteatro*

circun- + *yacente* ⇒ *circunyacente*

— El prefijo *anfi-* se convierte en *anfi-* ante palabras bisilábicas llanas.

anfi- + *podo* ⇒ *anfípodo*

— El prefijo *circun-* toma la forma *circum-* ante *b*, *p* o *m*

circun- + *beaterio* ⇒ *circumbeaterio*

circun- + *polar* ⇒ *circumpolar*

circun- + *molar* ⇒ *circummolar*

y en ocasiones ante *n*.

circun- + *nutación* ⇒ *circumnutación*

circun- + *navegación* ⇒ *circunnavegación*

— El prefijo *circun-* en ocasiones elimina la *e* ante palabras que comienzan por esta vocal.

circun- + *escrito* ⇒ *circunscrito*

2. A distancia o lejos: *tele-*

tele- + *guías* ⇒ *teleguías*

— El prefijo *tele-* puede convertirse en *telé-* ante palabras bisilábicas llanas.

tele- + *fono* ⇒ *teléfono*

tele- + *grama* ⇒ *telegrama*

— Además reduce la *e* ante palabras que comienzan por esta vocal.

tele- + *espectador* ⇒ *telespectador*

3. A través de, cambio o al otro lado: *tras-*, *trans-*, *ultra-*, *meta-*

tras- + *patio* ⇒ *traspatio*

trans- + *alpino* ⇒ *transalpino*

ultra- + *mar* ⇒ *ultramar*

meta- + *lenguaje* ⇒ *metalenguaje*

— Los prefijos *tras-* y *trans-* reducen la *s* ante palabras que comienzan por esta letra

tras- + *señalador* ⇒ *traseñalador*

trans- + *sexual* ⇒ *transexual*

y pueden hacer desaparecer la inicial en palabras que comienzan por *e*;

trans- + *emisión* ⇒ *transmisión*

trans- + *embrionario* ⇒ *transebrionario*

la reducción de la *s* puede ocurrir tras la desaparición de la *e*.

tras- + *esquilador* ⇒ *trasquilador*

trans- + *escribir* ⇒ *transcribir*

— Los prefijos *tras-* y *trans-* pueden convertirse en *trás-* y *tráns-* en ocasiones.

tras- + *fuga* ⇒ *trásfuga*

trans- + *fuga* ⇒ *tránsfuga*

4. Acción secundaria o atenuación del significado: *so-*, *sub-*, *za-*, *zam-*, *sus-*, *entre-*

so- + *frito* ⇒ *sofrito*

sub- + *afluente* ⇒ *subafluente*

za- + *heridor* ⇒ *zaheridor*

zam- + *bombazo* ⇒ *zambombazo*

entre- + *abierto* ⇒ *entreabierto*

— El prefijo *sus-* a veces hace perder la *e* a palabras que comienzan por esta vocal.

sus- + *estrato* ⇒ *sustrato*

5. Arriba, en alto o sobre: *ana-*, *epi-*, *supra-*

ana- + *tema* ⇒ *anatema*

epi- + *centro* ⇒ *epicentro*

supra- + *clavicular* ⇒ *supraclavicular*

6. Alejamiento, separación y privación: *ab-*, *abs-*, *dis-*, *dia-*, *di-*

ab- + *oral* ⇒ *aboral*

abs- + *tenido* ⇒ *abstenido*

dis- + *culpa* ⇒ *disculpa*

dia- + *crítico* ⇒ *diacrítico*

di- + *fluencia* ⇒ *difluencia*

7. Aumento, encarecimiento o refuerzo del significado: *re-*, *rete-*, *requete-*, *archi-*, *super-*, *sobre-*, *hiper-*, *ultra-*, *extra-*

re- + *agudo* ⇒ *reagudo*

rete- + *bueno* ⇒ *retebueno*

requete- + *guapo* ⇒ *requeteguapo*

archi- + *diócesis* ⇒ *archidiócesis*

super- + *población* ⇒ *superpoblación*

sobre- + *esfuerzo* ⇒ *sobreesfuerzo*

hiper- + *mercado* ⇒ *hipermercado*

ultra- + *feminista* ⇒ *ultrafeminista*

extra- + *plano* ⇒ *extraplano*

— El prefijo *sobre-* puede admitir además la forma con reducción de *e* ante palabras que comienzan por esta letra.

sobre- + *esdrújulo* ⇒ *sobreesdrújulo* / *sobresdrújulo*

8. Bajo o debajo: *so-*, *sub-*, *za-*, *zam-*, *sus-*, *sota-*, *soto-*

so- + *portal* ⇒ *soportal*

sub- + *normal* ⇒ *subnormal*

za- + *borda* ⇒ *zaborda*

zam- + *bullidor* ⇒ *zambullidor*

sus- + *pendido* ⇒ *suspendido*

sota- + *banco* ⇒ *sotabanco*

soto- + *bosque* ⇒ *sotobosque*

9. Compañía: *co-*, *con-*

co- + *existencia* ⇒ *coexistencia*

con- + *catredalidad* ⇒ *concatredalidad*

— El prefijo *con-* cambia la *n* por *m* ante *b* o *p*.

con- + *beneficiado* ⇒ *combeneficiado*

con- + *paternidad* ⇒ *compaternidad*

10. Conforme a: *ana-*

ana- + *bautismo* ⇒ *anabautismo*

— El prefijo *ana-* no se usa ante palabras que comienzan por *a* y se apocopa ante vocal.

ana- + *ion* ⇒ *anión*

11. Contra, contrariedad, contrario, oposición, opuesto o rechazo: *contra-*, *anti-*, *di-*, *dis-*, *para-*, *re-*

contra- + *veneno* ⇒ *contraveneno*

anti- + *biótico* ⇒ *antibiótico*

di- + *sentimiento* ⇒ *disentimiento*

dis- + *conforme* ⇒ *disconforme*

para- + *cronismo* ⇒ *paracronismo*

re- + *acción* ⇒ *reacción*

— El prefijo *anti-* puede convertirse en *anti-* sin regla conocida.

anti- + *tesis* ⇒ *antítesis*

12. Delante, anterioridad en tiempo y espacio o prioritario: *pro-*, *pre-*, *ante-*, *proto-*

pro- + *genitura* ⇒ *progenitura*

pre- + *clásico* ⇒ *preclásico*

ante- + *brazo* ⇒ *antebrazo*

proto- + *mártir* ⇒ *protomártir*

13. Dentro: *intro-*, *endo-*, *en-*

intro- + *vertido* ⇒ *introvertido*

endo- + *esqueleto* ⇒ *endoesqueleto*

en- + *cefálico* ⇒ *encefálico*

— El prefijo *endo-* en ocasiones hace perder la *e* a palabras que comienzan por esta vocal.

endo- + *esfera* ⇒ *endosfera*

— El prefijo *en-* cambia la *n* por *m* ante *b* o *p*.

en- + *bolso* ⇒ *embolso*

en- + *pasto* ⇒ *empasto*

14. Después de en el sentido de detrás: *pos-*, *post-*, *meta-*

pos- + *guerra* ⇒ *posguerra*

post- + *dorsal* ⇒ *postdorsal*

meta- + *carpo* ⇒ *metacarpo*

15. Doble o dos: *bi-*, *bis-*, *anfi-*, *di-*

bi- + *mensual* ⇒ *bimensual*

bis- + *nieto* ⇒ *bisnieto*

anfi- + *bio* ⇒ *anfíbio*

di- + *sílaba* ⇒ *disílaba*

16. En vez de o por substitución: *pro-*

pro- + *creación* ⇒ *procreación*

17. Entre o en medio: *inter-*, *dia-*

inter- + *nacional* ⇒ *internacional*

dia- + *tónico* ⇒ *diatónico*

18. Extensión o dilatación: *di-*

di- + *solución* ⇒ *disolución*

19. Exterior, junto o próximo: *para-*, *epi-*, *yuxta-*, *ad-*

para- + *militar* ⇒ *paramilitar*

epi- + *cicloide* ⇒ *epicicloide*

yuxta- + *posición* ⇒ *yuxtaposición*

ad- + *junto* ⇒ *adjunto*

— El prefijo *para-* puede convertirse en *pará-* sin regla conocida.

para- + *metro* ⇒ *parámetro*

para- + *poco* ⇒ *parapoco*

20. Fuera, más allá o externo: *extra-*, *ex-*, *es-*, *des-*, *ecto-*, *meta-*, *supra-*, *ultra-*, *exo-*

extra- + *muros* ⇒ *extramuros*

ex- + *céntrico* ⇒ *excéntrico*

es- + *tiramiento* ⇒ *estiramiento*

des- + *hora* ⇒ *deshora*

ecto- + *plasma* ⇒ *ectoplasma*

meta- + *física* ⇒ *metafísica*

supra- + *sensible* ⇒ *suprasensible*

ultra- + *puertos* ⇒ *ultrapuertos*

exo- + *térmico* ⇒ *exotérmico*

— El prefijo *exo-* reduce la *o* ante palabras que comienzan por esta vocal;

exo- + *oftálmico* ⇒ *exoftálmico*

si aparece tildada, puede mantenerse o no la tilde

exo- + *ósmosis* ⇒ *exósmosis* / *exosmosis*

y ante bisílabas llanas en ocasiones se convierte en *exó-*,

exo- + *gamo* ⇒ *exógamo*

pero en otras no.

exo- + *dermis* ⇒ *exodermis*

— El prefijo *exo-* a veces hace perder la *e* a las palabras que comienzan por esta vocal.

exo- + *espora* ⇒ *exospora*

21. Hacia atrás, de nuevo, tiempo anterior o inversión de la acción: *retro-*, *des-*, *re-*, *ana-*

retro- + *activo* ⇒ *retroactivo*

des- + *hecho* ⇒ *deshecho*

re- + *flujo* ⇒ *reflujo*

ana- + *pesto* ⇒ *anapesto*

22. Igual: *equi-*, *iso-*

equi- + *distante* ⇒ *equidistante*

iso- + *fónico* ⇒ *isofónico*

23. Inferioridad: **hipo-**

hipo- + **función** ⇒ **hipofunción**

— A veces se convierte en **hipó-**

hipo- + **tesis** ⇒ **hipótesis**

y en ocasiones se pierde la *e* en palabras que comienzan por esta vocal.

hipo- + **estasis** ⇒ **hipostasis**

24. Lejos de o separado de: **apo-**

apo- + **morfina** ⇒ **apomorfina**

25. Medio, casi o mitad: **semi-**, **hemi-**

semi- + **círculo** ⇒ **semicírculo**

hemi- + **ciclo** ⇒ **hemiciclo**

— El prefijo **hemi-** a veces hace perder la *e* a las palabras que comienzan por esta vocal.

hemi- + **esfera** ⇒ **hemisferio**

26. Origen o procedencia: **di-**, **ab-**

di- + **manación** ⇒ **dimanación**

ab- + **origen** ⇒ **aborigen**

27. Parte de acá: **cis-**, **citra-**

cis- + **andino** ⇒ **cisandino**

citra- + **montano** ⇒ **citramontano**

28. Parte de atrás: **opisto-**

opisto- + **gástrico** ⇒ **opistogástrico**

29. Preeminencia, primacía, superioridad o prioridad: **archi-**, **proto-**

archi- + **duque** ⇒ **archiduque**

proto- + **médico** ⇒ **protomédico**

30. Privación, negación o ausencia: **re-**, **a-**, **des-**, **de-**, **dis-**, **ex-**, **ana-**, **in-**

re- + **probado** ⇒ **reprobado**

a- + **cromático** ⇒ **acromático**

des- + **confianza** ⇒ **desconfianza**

de- + **foliación** ⇒ **defoliación**

dis- + **continuo** ⇒ **discontinuo**

ex- + **animación** ⇒ **exanimación**

ana- + **crónico** ⇒ **anacrónico**

in- + **correcto** ⇒ **incorrecto**

— El prefijo *a-* se convierte en *an-* ante vocal.

a- + *estesia* ⇒ *anestesia*

— El prefijo *in-* se convierte en *im-* ante *b* o *p*,

in- + *borrable* ⇒ *imborrable*

in- + *procedente* ⇒ *improcedente*

en *i-* ante *l* o *r*

in- + *legítimo* ⇒ *ilegítimo*

pero, en el caso de *r* ésta se dobla para mantener el mismo sonido.

in- + *realidad* ⇒ *irrealidad*

31. Progreso, continuidad de acción o hacia adelante: *pro-*, *para-*

pro- + *motor* ⇒ *promotor*

para- + *fraseador* ⇒ *parafraseador*

32. Que suple, hace las veces de, ocupa el segundo lugar en categoría o subalterno: *vice-*,

viz-, *vi-*, *sota-*, *soto-*, *sub-*

vice- + *presidente* ⇒ *vicepresidente*

viz- + *conde* ⇒ *vizconde*

vi- + *rey* ⇒ *virrey*

sota + *cura* ⇒ *sotacura*

soto- + *ministro* ⇒ *sotoministro*

sub- + *director* ⇒ *subdirector*

33. Situación o calidad intermedia: *entre-*, *entro-*

entre- + *fino* ⇒ *entrefino*

entro- + *metido* ⇒ *entrometido*

34. Unión: *co-*, *con-*, *sin-*

co- + *acusado* ⇒ *coacusado*

con- + *socio* ⇒ *consocio*

sin- + *genésicos* ⇒ *singenésicos*

— Los prefijos *con-* y *sin-* cambian la *n* por *m* ante *b* o *p*.

con- + *balanceo* ⇒ *combalanceo*

con- + *patriota* ⇒ *compatriota*

sin- + *biótico* ⇒ *simbiótico*

sin- + *pétalo* ⇒ *simpétalo*

— El prefijo *sin-* en ocasiones se convierte en *sín-*.

sin- + *tesis* ⇒ *síntesis*

Además, el proceso de unión entre prefijos y formas debe tener en cuenta las siguientes reglas:

1. Cuando se añade un prefijo a un monosílabo sin tilde que acaba en vocal, *n* o *s* debe tildarse.

requete- + *bien* ⇒ *requetebién*

2. Cuando se añade un prefijo terminado en vocal a una forma que comienza por *r* debe duplicarse la *r* para mantener el sonido fuerte.

contra- + *reloj* ⇒ *contrarreloj*

3. Cuando se añade un prefijo terminado en vocal fuerte a una forma que comienza por vocal débil tónica no tildada, ésta debe tildarse porque se forma un hiato,

entre- + *ido* ⇒ *entreído*

aunque vaya precedida de *h*.

re- + *hilo* ⇒ *rehílo*

2.– PRODUCCIÓN DE LAS FORMAS NO VERBALES FLEXIONADAS Y DERIVADAS.

En este trabajo, se parte de una base inicial con un total de 109 194 formas canónicas. Se han incluido todas las entradas no verbales del *Diccionario de la Lengua Española* de la Real Academia Española (70 056), del *Diccionario General de la Lengua Española* Vox (83 709), del *Diccionario de Uso del Español* de María Moliner (66 099), del *Gran Diccionario de la Lengua Española* de Larousse Planeta (58 605), del *Diccionario de voces de uso actual* dirigido por Manuel Alvar Ezquerro (4 644), del *Gran Diccionario de Sinónimos y Antónimos* de Espasa-Calpe (31 011) y del *Diccionario Ideológico* de Julio Casares (56 533).

Cada registro de la base inicial contiene: a) la forma canónica representante del registro, b) su categoría gramatical, c) la parte invariante, d) las terminaciones que permiten los cambios de género y de número —producto de la aplicación de las reglas descritas en los apartados que van del 1.1 al 1.5 considerando todas las excepciones e irregularidades—, e) la información sobre irregularidades y excepciones en la formación de apreciativos y en la del superlativo, f) las excepciones a la adverbialización y g) las formas canónicas relacionadas por multiplicidad o heteronimia.

Se construye el léxico mediante un proceso de generación que opera sobre los registros de la base inicial y produce las formas flexionadas y derivadas asociadas a cada forma canónica. Tal expansión se lleva a cabo gracias a la información contenida en el registro y a la aplicación de las reglas estudiadas en los apartados 1.8 y 1.9 —cada elemento del léxico conserva su relación de flexión o derivación con la forma canónica de la que procede.

El léxico obtenido se compone de más de tres millones y cuarto de formas (3 328 283). No se ha considerado la exagerada ampliación que produciría la prefijación (se cuenta con más de ochenta prefijos y con que una misma forma puede admitir varios); se deja a la discrecionalidad del usuario en la aplicación final ya que las reglas del apartado 1.11 se ejecutan con eficacia y no se justifica el desproporcionado aumento del volumen de información.

3.– ESTRUCTURACIÓN DE LOS DATOS.

La solución aportada se orienta a datos más que a reglas, con el fin de obtener unos mejores resultados. Dado el considerable volumen de datos, se ha diseñado una estructura

para su almacenamiento en memoria secundaria que consigue un adecuado equilibrio entre ocupación y velocidad de recuperación.

Debido al carácter flexivo de la lengua española, se opta por un conjunto de terminaciones que permita la generación de todas las formas flexionadas y derivadas por simple concatenación a partir de la raíz de la forma canónica —las irregularidades y los cambios ortográficos se manifiestan con la aparición de un cambio en la raíz. Se usa un criterio de corte que genera un número mínimo de raíces, aunque no siempre dé lugar a la raíz lingüística; tal conjunto de raíces tiene una cardinalidad bastante menor que el de formas —180 775 frente a 3 328 283—; por tanto, resulta más favorable afrontar la solución haciendo la partición de las formas en raíces y terminaciones. Se organiza la estructura por raíces y cada registro contiene la información de las terminaciones de flexión o derivación que esa raíz acepta y la referencia al registro donde se encuentra la raíz de su forma canónica —útil para el reconocimiento en caso de raíces múltiples. Los registros correspondientes a formas canónicas incluyen además la categoría gramatical y la referencia a los registros de sus raíces alternativas —útil para la generación.

3.1.— MÓDULO DE ÍNDICES.

El *módulo de índices*, figura 1, es el encargado de obtener la dirección correspondiente en la base de raíces mediante una función de dispersión —*dispersión nuclear*.

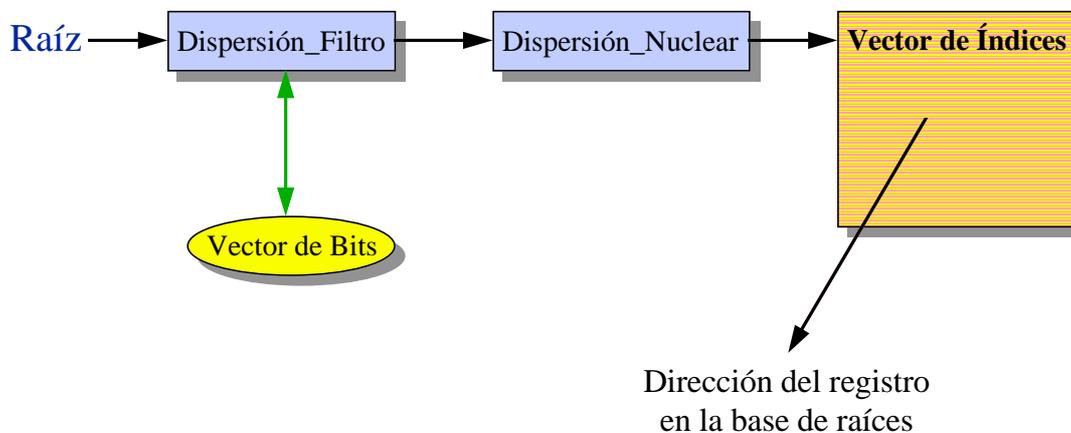


Figura 1: Módulo de índices

A fin de obtener una función que produzca pocas colisiones se transforma la cadena de caracteres en un valor numérico mediante un tratamiento adecuado de la representación binaria de sus caracteres. Con el propósito de reducir accesos infructuosos a dispositivos externos, se sitúa en la memoria principal un vector de bits que gestiona otra función de dispersión —*dispersión filtro*— y ya que las colisiones disminuyen al crecer el tamaño de la tabla, conviene su módulo muy grande —las posiciones libres aseguran que tales raíces no existen y hacen innecesaria la *dispersión nuclear*. A partir de la dirección que proporciona la *dispersión nuclear*, un *vector de índices* determina la dirección del registro en la base de raíces; este mecanismo trabaja con una base compacta aun cuando el

módulo de la función de dispersión sea relativamente grande, porque aprovecha todos los registros sin tener que dejar ninguno intermedio vacío —la reestructuración de los registros proporciona una recuperación óptima de las colisiones.

3.2.– MÓDULO DE ACCESOS EXTERNOS.

Este módulo, figura 2, gobierna las bases que guardan la información referente a raíces, terminaciones y las flexiones o derivaciones que representan sus uniones.

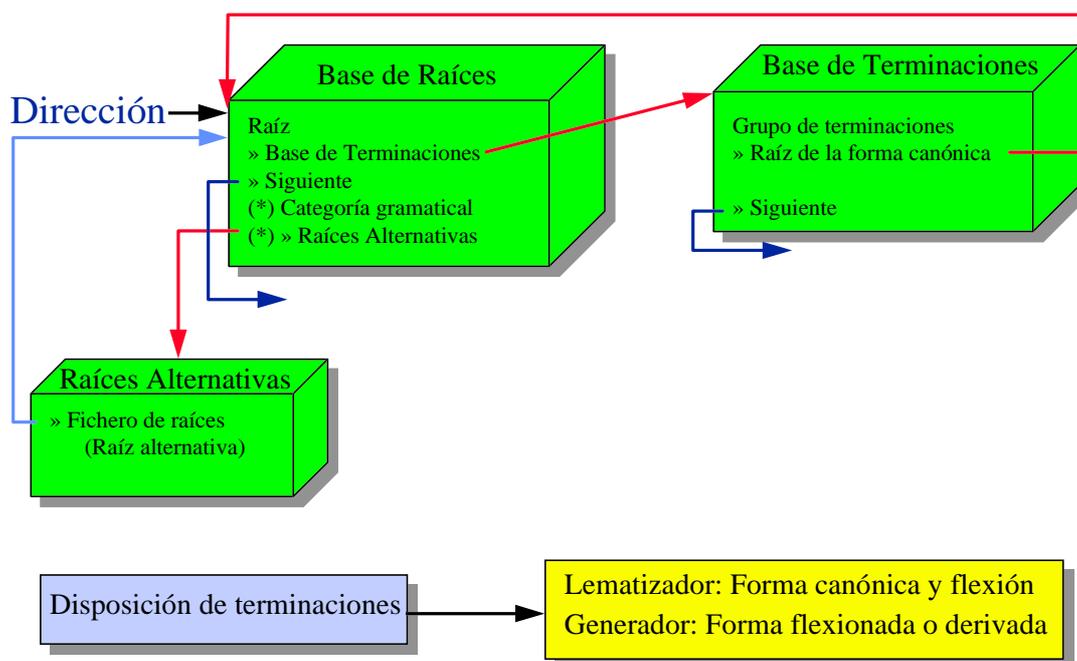


Figura 2: Módulo de accesos externos

A partir de la dirección de una raíz, accede a la base de raíces de la que obtiene su posición en la base de terminaciones. Junto a la raíz de una forma canónica, aparecen su categoría gramatical —útil para el proceso de lematización— y las raíces alternativas —útil para la generación de formas flexionadas o derivadas que no posean la misma raíz que la forma canónica. La base de terminaciones contiene la información acerca de las terminaciones que admite una raíz y de la formación de su forma canónica. El grupo de terminaciones permite averiguar qué terminaciones puede llevar una raíz y qué flexión representa para esa raíz cada una de ellas.

La existencia de raíces alternativas y de los grupos de terminaciones que admiten obedece al conjunto de terminaciones que se considera en el proceso de construcción. Las posibilidades varían desde no considerar ninguna terminación, con lo que cada forma sería una raíz, hasta considerar un conjunto de terminaciones lo suficientemente amplio como para poder generar cualquier forma derivada a partir de una única raíz por simple concatenación. Dadas las características morfológicas del español como lengua flexiva, se opta por una solución que proporciona un número de terminaciones suficiente para que: a) en los casos regulares se puedan obtener todas las formas a partir de una raíz única,

porque las operaciones de flexión y derivación no afectan a la raíz gramatical de la forma y b) cuando la raíz gramatical se ve afectada aparezcan raíces alternativas. Se consigue así un equilibrio entre abundancia de raíces alternativas y número de grupos de terminaciones.

4.- LEMATIZACIÓN.

El proceso de identificación actúa sobre una palabra de entrada por medio de un *segmentador* que la descompone en: a) los posibles pares raíz-terminación y b) los prefijos que pudieran poseer, figura 3. La raíz pasa al *módulo de índices* que determina su localización. El *módulo de accesos externos*: a) comprueba si la raíz admite la terminación, b) determina a qué flexión o derivación corresponde, c) deduce la forma canónica de la que proviene y d) proporciona su categoría gramatical.

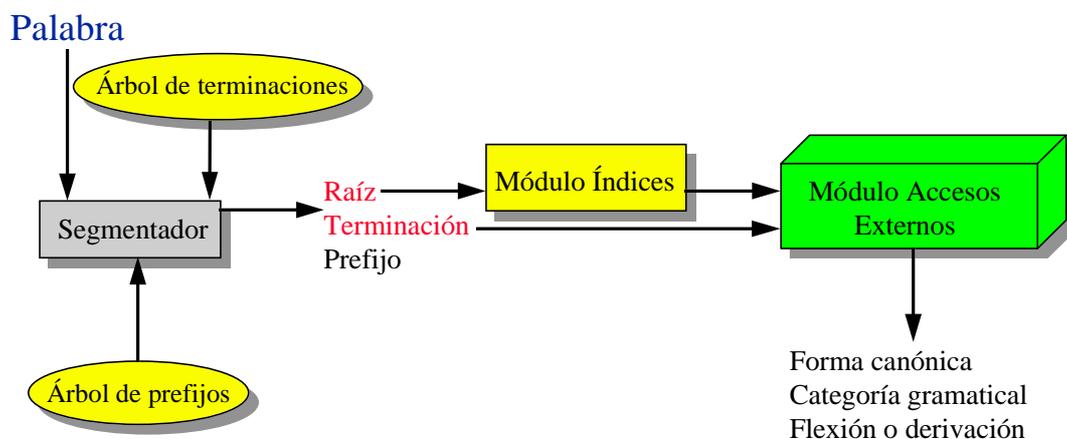


Figura 3: Esquema del lematizador

El *segmentador* se apoya en dos estructuras arbóreas: una para las terminaciones y otra que soporta los prefijos. Los posibles pares raíz-terminación se obtienen tras confirmar la existencia de una terminación en el árbol de terminaciones.

crédulamente ⇒ *crédulament-e*, *crédula-mente*, *crédul-amente*

En el *módulo de accesos* externos se rechazan los pares no localizados con el *módulo de índices*, pero en los casos exitosos se pasa a lematizar.

crédulament-e ⇒ forma canónica del adverbio de modo *crédulamente*

crédul-amente ⇒ adverbialización del adjetivo *crédulo*

La figura 4 ofrece una representación del subárbol de terminaciones que acaban en **e** —se organiza desde el último carácter hacia el primero, porque el sentido lo impone la segmentación al evolucionar en tal dirección. El punto negro (•) señala el final de la cadena de caracteres correspondiente a una terminación. Para el ejemplo propuesto se resalta el recorrido con línea más gruesa, se distinguen con sombreado los nodos donde hay coincidencias y con distintos grosores las terminaciones encontradas que además se señalan bajo el nodo terminal.

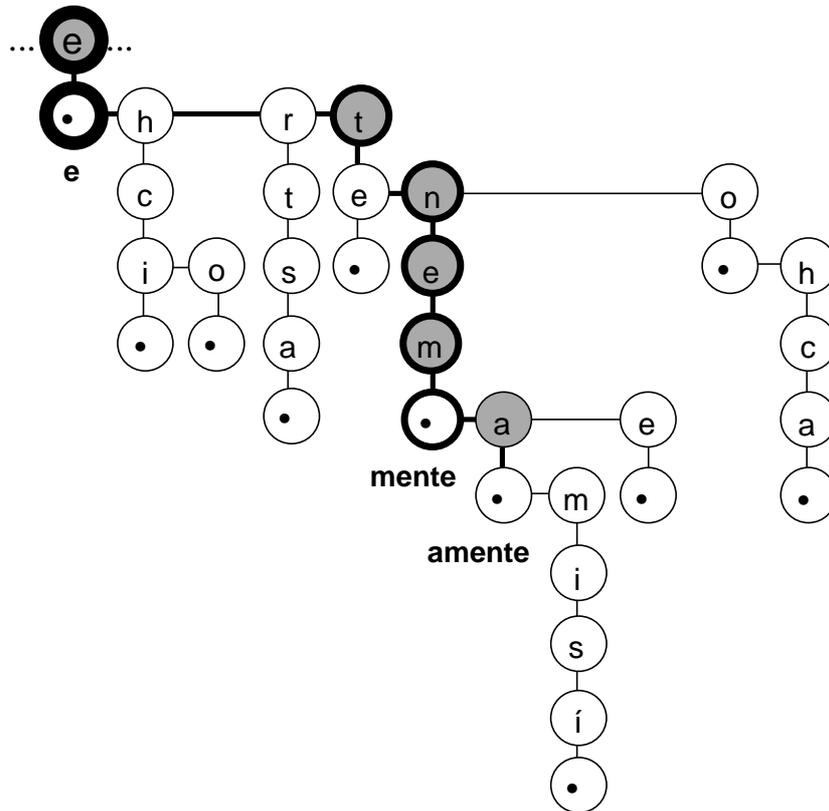


Figura 4: Árbol de terminaciones

Para tratar los prefijos se aíslan situando sus cadenas de caracteres en el índice correspondiente y se procede a segmentar en pares raíz-terminación la palabra descargada de prefijos.

archimaleado ⇒ *a+rchimaleado* ○ *arc+himaleado* ○ *archi+maleado*

La figura 5 presenta una vista del árbol de prefijos que comienza por **a**. El punto negro (•) señala el final de la cadena de caracteres correspondiente a un prefijo. Para el ejemplo propuesto se resalta el recorrido con línea más gruesa, se distinguen con sombreado los nodos donde hay coincidencias y con distintos grosores los prefijos encontrados que aparecen bajo el nodo terminal.

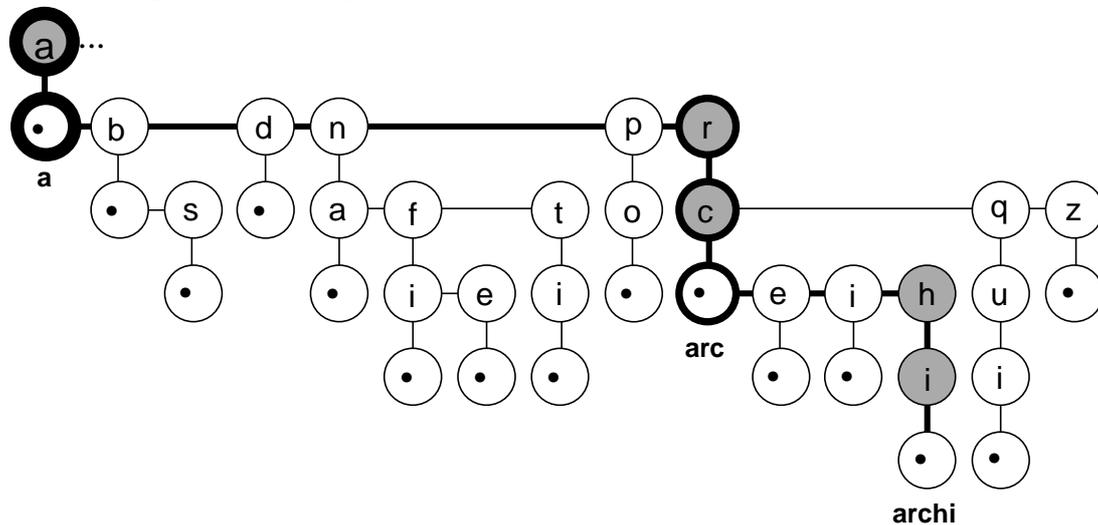


Figura 5: Árbol de prefijos

El árbol para los prefijos se construye tomando los caracteres desde el principio hacia el final de la palabra —el sentido lo impone la segmentación por evolucionar en tal dirección. Se estructuran como prefijos los enumerados en la sección 1.11 dedicada a la prefijación; se aplican las reglas de corte correspondientes, gracias a un procedimiento que se encarga de la tilde. Se emplea un separador silábico y un comprobador de las reglas ortográficas de acentuación que han sido desarrollados al efecto; permite realizar la separación en sílabas de cualquier palabra, determinar cuál es su sílaba tónica, así como cambiar de posición, situar o hacer desaparecer la tilde dada la sílaba tónica.

5.- GENERACIÓN.

Al disponer de una estructura de datos que permite conocer para cada palabra cuál es su forma canónica, qué raíces tiene, qué terminaciones admite cada raíz y qué flexión o derivación presenta cada una de ellas, es posible generar con poco esfuerzo una forma a partir de la canónica y de la flexión o derivación propuesta. Basta con disponer de la capacidad de acceder a las distintas raíces que aparecen al flexionar o derivar una determinada forma canónica y conocer qué terminaciones corresponden a la flexión o derivación dada, figura 2.

La entrada al generador está constituida por a) una forma canónica, b) la flexión o derivación correspondiente y c) los prefijos, figura 6.

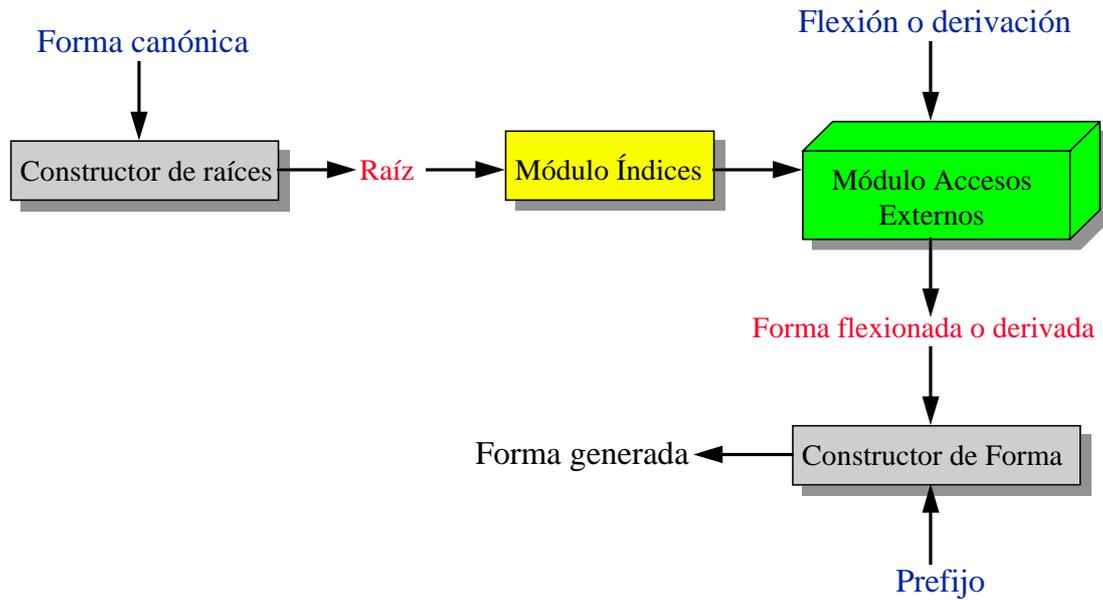


Figura 6: Esquema del generador

El *constructor de raíces* descarta la terminación y obtiene la raíz de la forma canónica, el *módulo de índices* localiza esa raíz en la base de raíces y gracias al *módulo de accesos externos* se llega a la forma flexionada o derivada. Para determinar la *forma generada*, el *constructor de forma* añade el prefijo o conjunto de prefijos aplicando las reglas de unión detalladas en la sección 1.11 dedicada a la prefijación.

La figura 7 muestra las posibilidades que ofrece la generación.

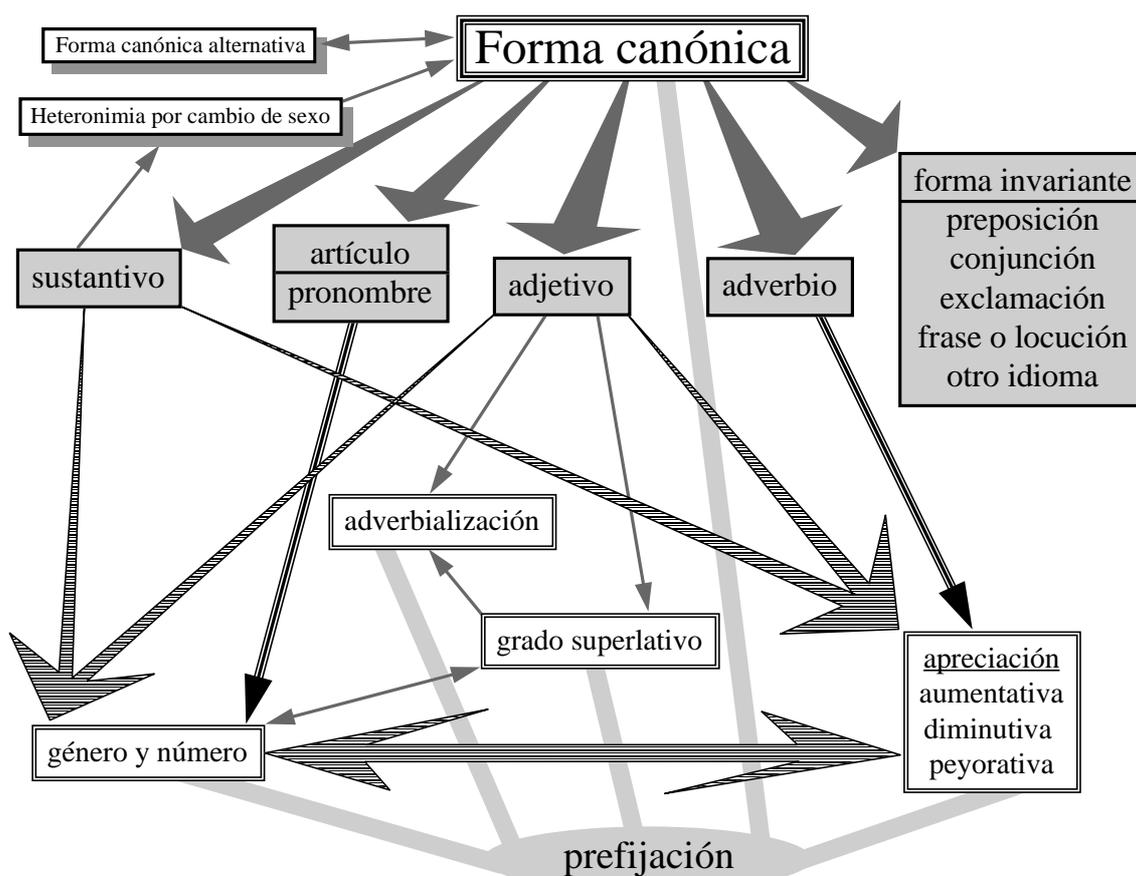
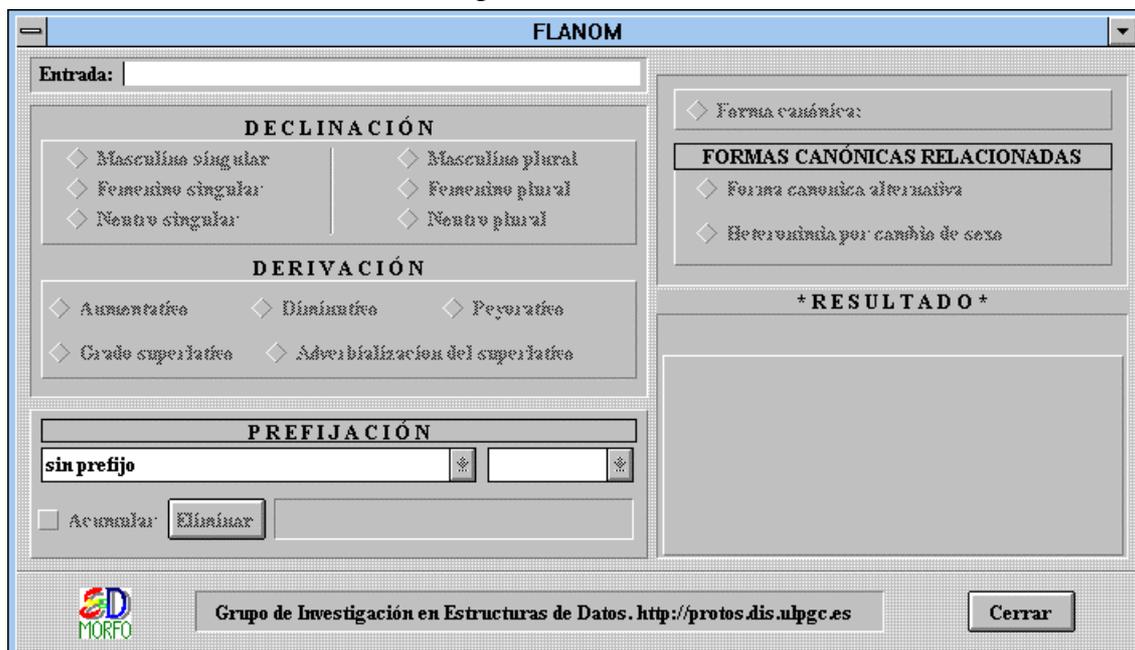


Figura 7: Esquema conceptual del generador

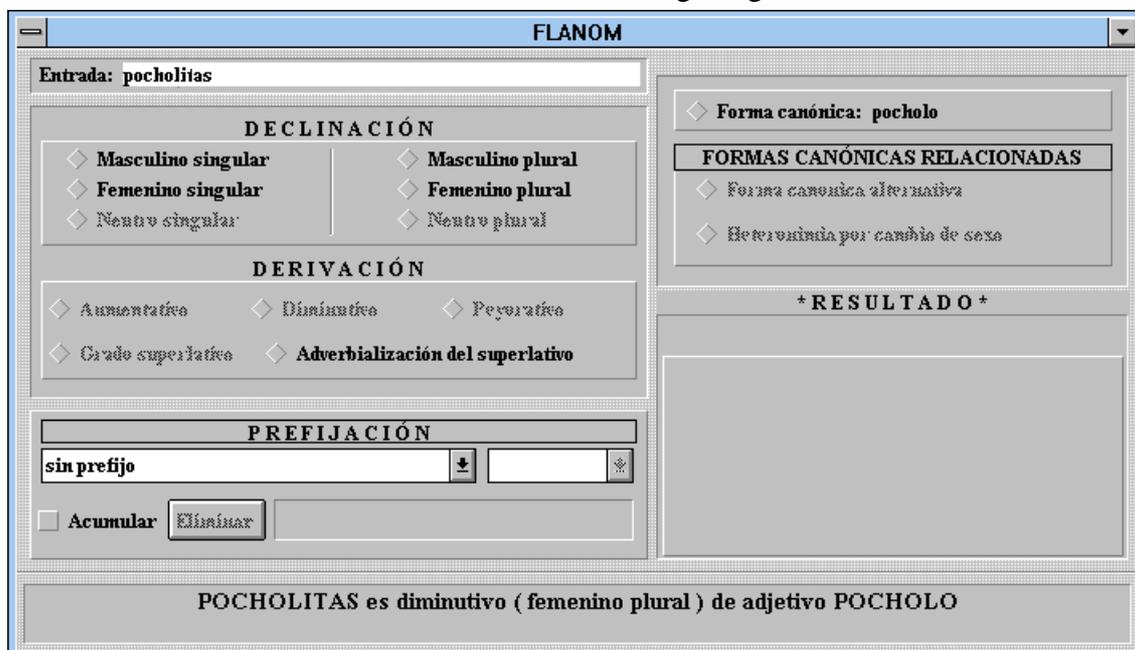
Dada una forma canónica existe un haz de categorías gramaticales admisibles (sustantivo, artículo, pronombre, adjetivo, adverbio y forma invariante). Los sustantivos y adjetivos admiten género y número y sufijación apreciativa, el género y el número determinan la apreciación y la apreciación puede modificar las posibilidades en cuanto a género y número. Los adverbios no admiten género ni número pero pueden apreciarse mientras que los artículos y pronombres sólo admiten género y número —y variación de caso. De los adjetivos se obtiene además el grado superlativo con sus diferentes géneros y números y la adverbialización que incluye la del superlativo. La heteronimia por cambio de sexo proporciona formas canónicas a partir de los sustantivos. Cualquier forma generable de sustantivo, adjetivo o adverbio puede incorporar prefijos. Las formas canónicas múltiples tienen alternativas.

6.- INTERFAZ DE USUARIO.

Con el objeto de lograr una utilización amigable y eficaz para una amplia gama de usuarios, se ha diseñado una interfaz gráfica en entorno Windows.



El *texto de la entrada* es de libre edición y permite introducir la forma no verbal que se analiza automáticamente; cuando la lematización produce una sola forma canónica la coloca junto a *Forma canónica* y muestra en el *área de mensajes* tanto la relación de la forma de entrada con la forma canónica como su categoría gramatical.



Si el análisis arroja múltiples posibilidades despliega una *lista de formas canónicas* bajo el *texto de la entrada* y solicita una elección —al recorrer las respuestas, muestra en el *área de mensajes* la relación entre la forma de entrada y cada forma canónica. El *área de mensajes*: a) usa la zona muerta en la parte inferior del diálogo, b) sólo está presente cuando es imprescindible, c) llama la atención de manera discreta,

d) no oculta ninguna información relevante y e) permite continuar sin ningún requerimiento extra.

The screenshot shows the FLANOM interface with the following elements:

- Entrada:** chiquitines
- Formas sugeridas:** chiquitín, chiquito, chico, chica.
- DERIVACIÓN:** Aumentativo, Diminutivo, Peyorativo, Grado superlativo, Adverbialización del superlativo.
- PREFIJACIÓN:** sin prefijo.
- FORMAS CANÓNICAS RELACIONADAS:** Forma canónica, Forma canónica alternativa, Heteronimia por cambio de sexo.
- * RESULTADO *:** (Empty area)
- Nota:** CHIQUITINES es masculino plural de adjetivo usado también como sustantivo CHIQUITÍN.

Si la entrada es reconocida como forma no verbal —afectada o no por prefijos— habilita: *Forma canónica*, las variaciones de *GÉNERO Y NÚMERO* que posea, la *PREFIJACIÓN* (sólo para sustantivos, adjetivos y adverbios de modo), la *Heteronimia por cambio de sexo* y la *Forma canónica alternativa* cuando existan, la *Adverbialización del superlativo* en los adjetivos y las formas derivativas (*Forma superlativa*, *Forma aumentativa*, *Forma diminutiva* y *Forma peyorativa*) en los adverbios que las tengan — aparecerán inhabilitadas las opciones que no se permitan.

The screenshot shows the FLANOM interface with the following elements:

- Entrada:** arriba
- DECLINACIÓN:** Masculino singular, Masculino plural, Femenino singular, Femenino plural, Neutro singular, Neutro plural.
- DERIVACIÓN:** Forma aumentativa, Forma diminutiva, Forma peyorativa, Forma superlativa, Adverbialización del superlativo.
- PREFIJACIÓN:** sin prefijo.
- FORMAS CANÓNICAS RELACIONADAS:** Forma canónica: arriba, Forma canónica alternativa, Heteronimia por cambio de sexo.
- * RESULTADO *:** Forma canónica: arriba.
- Nota:** ARRIBA es adverbio de lugar e exclamación.

Actuar sobre *Forma canónica* permite: a) obtener su categoría gramatical en la *área de mensajes*, b) mostrarla en la *ventana de resultado* y c) incorporarle los prefijos seleccionados. Al pulsar en *Heteronimia por cambio de sexo* o en *Forma canónica*

alternativa se obtienen otras formas canónicas que, si se seleccionan en **RESULTADO**, permiten reiniciar el proceso de generación.

The screenshot shows the FLANOM software interface. At the top, the title bar reads 'FLANOM'. Below it, the 'Entrada' field contains 'tertuliano'. The interface is divided into several sections:

- DECLINACIÓN:** A grid of radio buttons for 'Masculino singular', 'Femenino singular', 'Neutro singular', 'Masculino plural', 'Femenino plural', and 'Neutro plural'. None are selected.
- DERIVACIÓN:** A grid of radio buttons for 'Aumentativo', 'Diminutivo', 'Peyorativo', 'Grado superlativo', and 'Adverbialización del superlativo'. None are selected.
- PREFIJACIÓN:** A dropdown menu set to 'compañía', a 'co-' prefix field, and an 'Eliminar' button.
- Forma canónica:** A field containing 'tertuliano'.
- FORMAS CANÓNICAS RELACIONADAS:** A section with a selected 'Forma canónica alternativa' and an option 'Determinada por cambio de sexo'.
- * RESULTADO *:** A section titled 'Forma canónica alternativa' containing a list: 'cotertuliente' and 'cotertulio'.

At the bottom, there is a logo for 'SD MORFO', the text 'Grupo de Investigación en Estructuras de Datos. http://protos.dis.ulpgc.es', and a 'Cerrar' button.

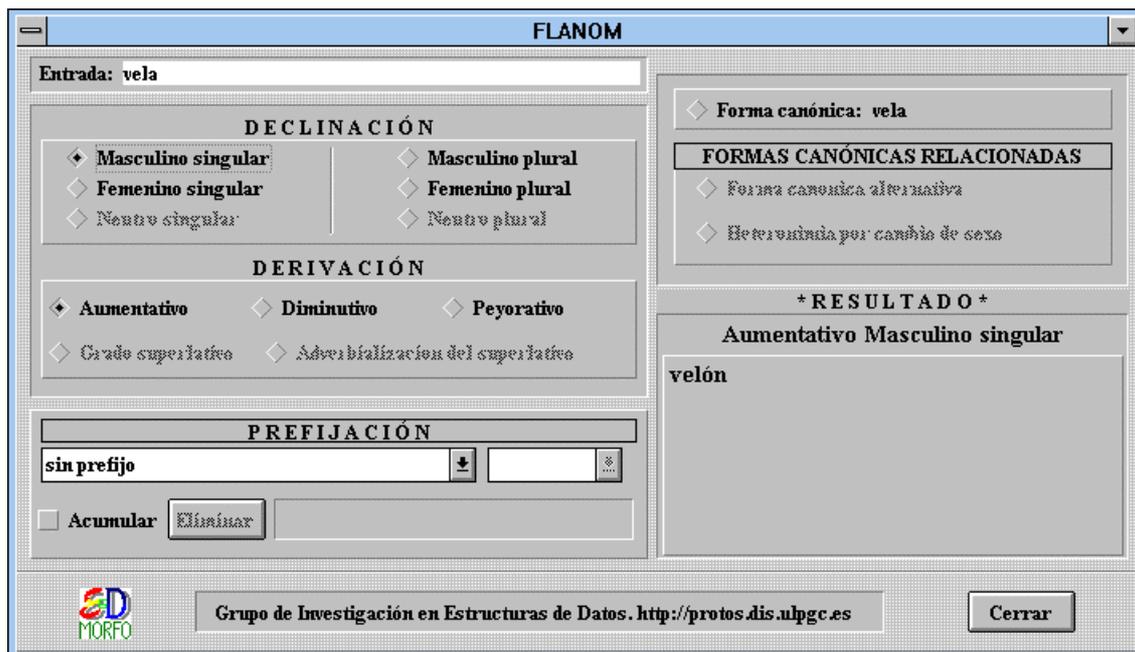
Para los sustantivos y los adjetivos la elección de un **GÉNERO Y NÚMERO** activa la **DERIVACIÓN** y en los adjetivos además el **Grado superlativo**. En cuanto se solicita una flexión o derivación se presenta en la *ventana de resultado* (encabezada por la opción elegida) una lista con las formas resultantes —cuando no existe ninguna, se notifica en el *área de mensajes* y si es necesario activa las barras de desplazamiento.

The screenshot shows the FLANOM software interface. At the top, the title bar reads 'FLANOM'. Below it, the 'Entrada' field contains 'viejo'. The interface is divided into several sections:

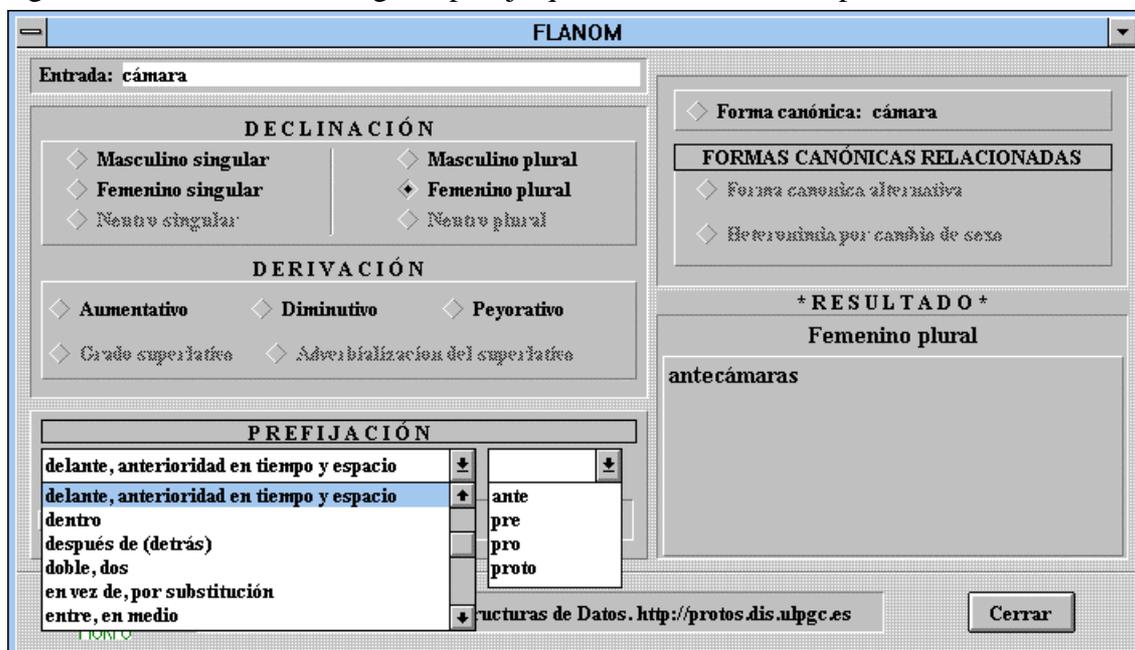
- DECLINACIÓN:** A grid of radio buttons for 'Masculino singular', 'Femenino singular', 'Neutro singular', 'Masculino plural', 'Femenino plural', and 'Neutro plural'. 'Masculino plural' is selected.
- DERIVACIÓN:** A grid of radio buttons for 'Aumentativo', 'Diminutivo', 'Peyorativo', 'Grado superlativo', and 'Adverbialización del superlativo'. 'Peyorativo' is selected.
- PREFIJACIÓN:** A dropdown menu set to 'sin prefijo', an empty prefix field, and an 'Eliminar' button.
- Forma canónica:** A field containing 'viejo'.
- FORMAS CANÓNICAS RELACIONADAS:** A section with a selected 'Forma canónica alternativa' and an option 'Determinada por cambio de sexo'.
- * RESULTADO *:** A section titled 'Peyorativo Masculino plural' containing a list: 'vejuchos', 'vejotes', 'vejetes', 'vejarrones', 'vejarrones', 'vejarrones', 'vejarrones', and 'vejestorios'.

At the bottom, there is a logo for 'SD MORFO', the text 'Grupo de Investigación en Estructuras de Datos. http://protos.dis.ulpgc.es', and a 'Cerrar' button.

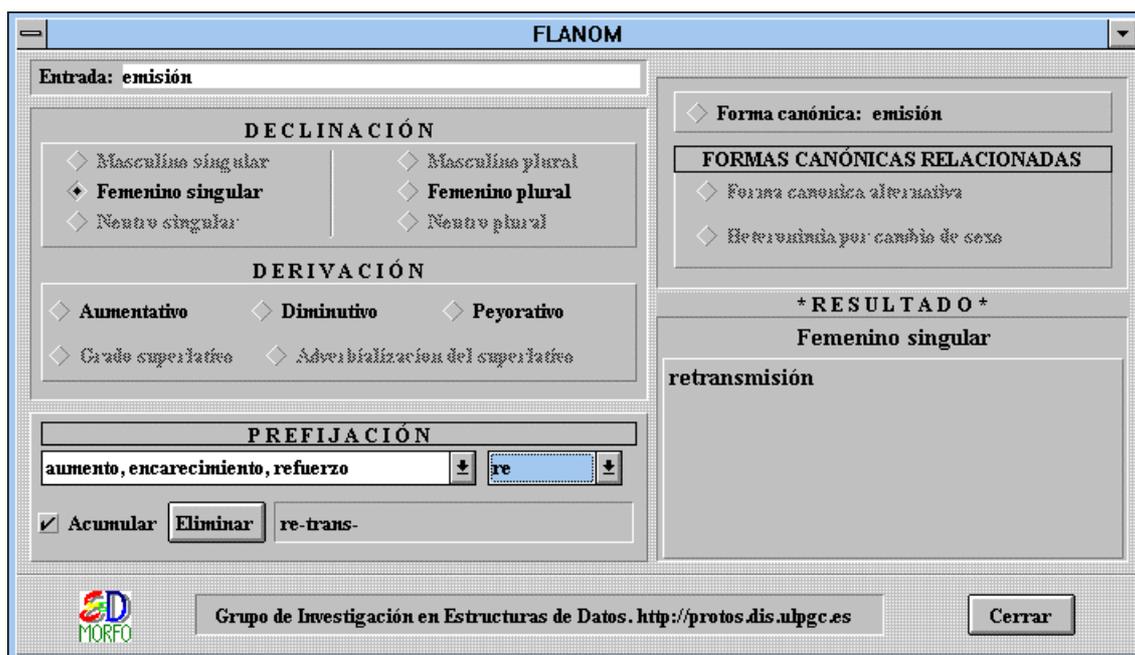
La **DERIVACIÓN** afecta las capacidades de **GÉNERO Y NÚMERO** ya que es posible obtener apreciativos con género diferente al de la forma canónica.



Se dispone de la facultad de incorporar prefijos que se seleccionan en las listas clasificadas según los distintos significados que poseen. La abundancia de significados sugiere su ubicación en una ventana desplegable donde aparecen los nombres de los grupos ordenados alfabéticamente: al fijar un conjunto, se abre la lista correspondiente. Incluye la opción de *sin prefijo* al objeto de eliminarlos con comodidad y la de *todos los prefijos* para escoger alfabéticamente uno cualquiera —el área de mensajes informa de sus significados. En cuanto se elige un prefijo queda modificada la respuesta en **RESULTADO**.



El desconocimiento de las reglas que rigen la secuencialización de prefijos aconseja la opción *Acumular* que añade por la izquierda cualquier prefijo a los ya existentes; cuando está desactivada un nuevo prefijo sustituye al anterior. La opción de *sin prefijo* elimina todos los prefijos de una vez; como resulta interesante poder descartarlos de uno en uno, el botón *Eliminar* lo hace de izquierda a derecha.



7.- RESULTADOS EXPERIMENTALES.

La aplicación se ha desarrollado en C++. Los requerimientos para su funcionamiento son los que necesita Windows y el espacio mínimo en memoria externa para datos y código no supera los 20 Mb. Las estructuras de datos utilizadas por el programa ocupan un total de 19,1 Mb en disco; contienen compactadamente los 53,4 Mb que ocupan las 3 328 283 formas flexionadas y derivadas, además de la flexión o derivación que representan, la reseña sintáctica sobre categorías gramaticales y toda la información referente a prefijación.

El universo de formas tanto reconocibles como generables se compone de: 3 328 283 formas. Gran parte de este universo puede ser multiplicado por un factor cercano a 80, tantas veces como prefijos se combinen, por ejemplo combinando con un prefijo el universo supera los 200 millones de palabras —este corpus ocuparía cerca de tres gigabytes.

Sobre un procesador Pentium II a 300 Mhz con 128 Mb de memoria RAM, se generan las formas flexionadas o derivadas de las formas canónicas a una velocidad de 1010 formas por segundo y se identifican a razón de 480 formas por segundo. Si se incorporan prefijos a las formas se generan 670 formas por segundo. Todos los casos, tanto de generación como de reconocimiento, se han presentado de forma aleatoria para evitar los efectos, favorables o desfavorables, de seguir un orden. Sobre un texto literario constituido por 111 690 palabras, el reconocimiento —sin considerar prefijos— se efectúa a razón de 590 formas por segundo; si no se consideran las formas verbales —constituyen el 10% del texto— la velocidad de lematización es de 570 formas por segundo, ya que se detectan las formas no reconocibles a razón de 800 palabras por segundo; lematizar con tratamiento de prefijos supone 450 formas por segundo y se reconoce un 1,3% más del

texto; el 0,7% del texto lo constituyen 527 palabras que no son reconocidas por tratarse de nombres propios, toponimia, jerga o numeración.

8.– CONCLUSIONES.

Se ha logrado una herramienta que permite identificar a partir de una forma no verbal:

- a) la forma o formas canónicas de las que proviene,
- b) su categoría gramatical,
- c) la flexión o derivación que incluye:
 - c.1) el género y el número en los sustantivos, adjetivos, pronombres y artículos,
 - c.2) la derivación apreciativa —aumentativo, diminutivo y peyorativo— en los sustantivos, adjetivos y adverbios,
 - c.3) el grado superlativo en los adjetivos y adverbios,
 - c.4) la adverbialización y la adverbialización del superlativo en los adjetivos,
 - c.5) la heteronimia por cambio de sexo,
 - c.6) las formas canónicas múltiples,
 - c.7) y se tienen en cuenta formas invariantes tales como preposiciones, conjunciones, exclamaciones, locuciones o frases y palabras de otros idiomas,
- d) los prefijos que pueda poseer.

Además, es capaz de generar formas derivadas o flexionadas a partir de una forma canónica, una flexión o derivación y los prefijos que se deseen incorporar.

Tanto el lematizador como el generador operan sobre una única estructura de datos de manera bidireccional —recorrerla en sentidos contrarios implica pasar de la operación lematizadora a la flexionadora—; se aporta una solución equilibrada entre grado de operatividad, tiempo de respuesta y cantidad de almacenamiento.

La aplicación es de fácil manejo y tiene una gran versatilidad para ser usada por los profesionales que tratan con documentos. Admite como entrada una palabra aislada tanto para su lematización como para generar a partir de ella formas flexionadas o derivadas, y, lo que es más importante desde nuestro punto de vista, también admite la lematización automática de un corpus, cualquiera que sea su extensión.

Este módulo no verbal (*FLANOM*) funcionará conjuntamente con la aplicación verbal ya desarrollada (*FLAVER*) para garantizar el análisis completo de textos —la abundancia de flexiones y derivaciones consideradas posibilita una gran profundidad de detalle. Próximamente, se incluirán relaciones entre formas canónicas que permitan derivaciones con cambio de categoría gramatical (sustantivación, adjetivación, adverbialización, sustantivos y adjetivos deverbales,...) tanto en la identificación como en la generación. El procesador morfológico resultante será de gran interés para la puesta a punto de motores de búsquedas complejas en bases documentales ya que abriría las

puertas a la localización con independencia de la flexión o derivación que afecte a la forma considerada.

Ya se está trabajando en el enriquecimiento del léxico con palabras cultas con elementos compositivos muy variados como las que aparecen en el *Diccionario etimológico de helenismos españoles* de Eseberri Hualde o en el *Diccionario de raíces griegas léxico castellano científico y médico* de Quintana Cabanas.

AGRADECIMIENTOS.

Queremos agradecer al profesor Dr. Manuel Alvar Ezquerra de la Universidad Complutense de Madrid y a la profesora Dra. María Auxiliadora Castillo Carballo de la Universidad de Sevilla su colaboración en cuantas consultas le hemos formulado a lo largo del desarrollo del presente trabajo.

REFERENCIAS:

- [Als90] Alsina, R.: *Todos los Verbos Castellanos Conjugados*. 17ª Edición. Teide. Barcelona, 1990.
- [Alv93] Alvar Ezquerra, M.: “La formación de palabras en español”. Arco/Libros. Madrid, 1993.
- [Alv94] Alvar Ezquerra, M.: *Diccionario de voces de uso actual*. Arco/Libros. Madrid, 1994.
- [Cas90] Casares, J.: *Diccionario Ideológico de la Lengua Española*. 2ª Edición. Ed. Gustavo Gili, s.a. Barcelona, 1990.
- [DGLE97] *Diccionario General de la Lengua Española VOX en CD-ROM*. Biblograf, s.a. Barcelona, 1997.
- [DLE95] *Diccionario de la Lengua Española*. Edición electrónica. Versión 21.1.0. Real Academia Española y Espasa-Calpe. Madrid, 1995.
- [Ese79] Eseberri Hualde, C.: *Diccionario etimológico de helenismos españoles*. Ediciones Aldecoa. Burgos, 1979.
- [Fer87] Fernández Ramírez, S.: *Gramática Española*. Arco/Libros, S.A. Madrid, 1987.
- [DUE96] *Diccionario de Uso del Español* de María Moliner. Edición en CD-ROM. Gredos. Madrid, 1996.
- [Gar97] García Platero, J. M.: “Sufijación apreciativa y prefijación intensiva en español actual”. *Lingüística Española Actual*, XIX/1, 1997, págs. 51-61.
- [GDL96] *Gran Diccionario de la Lengua Española*. Larousse Planeta, s.a. Barcelona, 1996.
- [GDS91] *Gran Diccionario de Sinónimos y Antónimos*. 4ª edic. Espasa-Calpe. Madrid, 1991.
- [Gil85] Gili Gaya, S.: *Curso superior de sintaxis española Vox*. Biblograf, s.a. Barcelona, 1985.
- [Góm91] Gómez Torrego, L.: *Manual de Español Correcto*. Arco/Libros, s.a. Madrid, 1991.

- [Góm92] Gómez Torrego, L.: *El buen uso de las palabras*. Arco/Libros, s.a. Madrid, 1992.
- [Lyo75] Lyons, J.: *Nuevos horizontes de la lingüística*. Alianza Editorial. Madrid, 1975.
- [Mar78] Martinet, A.: *Elementos de lingüística general*. Gredos. Madrid, 1978.
- [Per96] Pérez Aguiar, J. R.: “Reconocimiento y generación integrada de la morfología del español: Una aplicación a la gestión de un diccionario de sinónimos y antónimos”. Tesis Doctoral bajo la dirección del Dr. O. Santana Suárez. Universidad de Las Palmas de Gran Canaria, 1996.
- [Qui97] Quintana Cabanas, J. M.: *Diccionario de raíces griegas léxico castellano científico y médico*. 2ª edic. Editorial Dykinson. Madrid, 1997.
- [RAE89] Real Academia Española: *Esbozo de una nueva gramática de la lengua española*. 1ª edic. Espasa-Calpe. Madrid, 1989.
- [RHS93] Rodríguez, A.; Hernández, Z.; Santana, O.: “Agrupaciones de Tiempos Verbales en un Texto”. *Anales de las II Jornadas de Ingeniería de Sistemas Informáticos y de Computación*, Quito (Ecuador). Abril, 1993, págs. 132-137.
- [Sec91a] Seco, M.: *Diccionario de dudas y dificultades de la lengua española*. 9ª Edición. Espasa-Calpe. Madrid, 1991.
- [Sec91b] Seco, M.: *Gramática esencial del español: Introducción al estudio de la lengua*. 2ª edición, revisada y aumentada. Espasa-Calpe. Madrid, 1991.
- [SHR93] Santana, O.; Hernández, Z. J.; Rodríguez, G.: “Conjugaciones Verbales”. *Boletín de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN)*. Nº 13. Febrero, 1993, págs. 443-450.
- [SHRPCB94] Santana, O.; Hernández, Z.; Rodríguez, G.; Pérez, J.; Carreras, F.; Bogliani, S.: “Reconocedor automático de formas verbales que trata conjugación y pronombres enclíticos”. *Lingüística Española Actual*, XVI-1, 1994, págs. 125-133.
- [SPCSRH97] Santana, O.; Pérez, J.; Carreras, F.; Santos, S.; Rodríguez, G. Hernández, Z.: “GEISA: Un diccionario de sinónimos en formato electrónico”. *Revista de Lexicografía*. La Coruña 1997.
- [SPHCR97] Santana, O.; Pérez, J.; Hernández, Z.; Carreras, F.; Rodríguez, G.: “FLAVER: Flexionador y lematizador automático de formas verbales”. *Lingüística Española Actual*, XIX-2, 1997, págs. 229/282.
- [SRG93] Santana, O.; Rodríguez, J.C.; González, J.D.: “FRECTEXT: Una Aplicación de Ayuda a la Elaboración de Documentos”. *Boletín de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN)*, Nº 13, Febrero 1993, págs. 451-462.
- [Ver95] *VerbiCard: Todos los verbos castellanos conjugados*. Castellnou Editorial. Barcelona, 1995.