

## ORTOGRAFIA CATALANA I LINGÜÍSTICA COMPUTACIONAL

JAUME TIÓ i FELIP MANYÀ  
*Universitat de Lleida*

### 1. *Introducció*

Tant pel que fa a l'anàlisi com a la generació de llenguatge natural mitjançant tècniques computacionals l'ortografia pot representar un escull important a l'hora d'identificar o d'escollir adequadament les unitats lèxiques de què es compon o s'ha de compondre un text. En aquest camp, les llengües (o les regles ortogràfiques aplicables a les llengües) poden presentar problemàtiques molt diferents que cal estudiar separatament per tal de poder donar la solució que en cada cas sigui més eficient.

L'ortografia catalana presenta uns quants problemes que podríem resumir, sense ànim de ser exhaustius, en els punts següents:

- a) Elisions (p. ex., el pare -> l'home).
- b) Diacrítics (p. ex., coixins -> coixí).
- c) Duplicacions (p. ex., nas -> nassos).
- d) Alternances (p. ex., roca -> roques).

Pretenem resoldre aquests problemes utilitzant el llenguatge Prolog. El Prolog és un llenguatge de programació declarativa que permet especificar un problema amb un subconjunt del llenguatge de la lògica de primer ordre i, a partir d'aquesta especificació, resol el problema automàticament usant la potència deductiva de la lògica. Dit d'una altra manera, un cop expressat lògicament el coneixement necessari per a resoldre el problema no cal dir-li com l'ha de resoldre.

El problema que nosaltres volem tractar és saber si un enunciat compleix les regles gramaticals del català, tenint en compte les dificultats ortogràfiques que presenta. Per fer aquesta tasca usarem la notació DCG (definite clause grammar) (cf. Pereira i Warren 1980) que ja ve incorporada a la majoria de versions de Prolog. Aquesta notació, com tot seguit veurem, ens permet escriure les regles gramaticals de manera semblant a com es fa en lingüística. Per altra banda, un cop escrites les regles en notació DCG, el llenguatge Prolog permet comprovar, sense cap esforç addicional de programació, si una frase donada és ben formada, així com generar totes les frases ben formades a partir de les esmentades regles i un vocabulari.

Exemple: Aquest és un programa en notació DCG. Un cop donada aquesta especificació, el Prolog identifica que la frase "la nena menja" és ben formada, mentre que la frase "la nen plora" no és correcta. A més a més, també podríem generar totes les frases ben formades.

```
frase --> grup_nominal(Número,Gènere),verb(Número).
grup_nominal(Número,Gènere)-->
article(Número,Gènere),nom(Número,Gènere).
grup_nominal(Número,Gènere) -->
```

```
article(Número,Gènere),nom(Número,Gènere),adjectiu(Número,Gènere).
article(singular,masculí) --> [el].
article(singular,femení) --> [la].
nom(singular,masculí) --> [nen].
nom(singular,femení) --> [nena].
verb(singular) --> [menja].
verb(singular) --> [plora].
```

Observem que les regles estan formades per una part esquerra, que té un sol component que representa un símbol no terminal, i una part dreta, que té un o més components, separats per una coma, que tant poden ser símbols terminals com no terminals. El significat d'una regla és que la part dreta és una possible forma per a una expressió del símbol no terminal a l'esquerra de la regla. Així, donada la regla "grup\_nominal --> article,nom", es llegeix "un grup nominal està format per un article seguit d'un nom". De vegades aquest tipus de regles tenen un poder expressiu limitat i el podem augmentar introduint arguments als components bàsics de les regles. Aquests arguments són locals a la regla, o sigui que, per satisfer una regla en què un mateix argument apareix més d'un cop, les diferents ocurrencies de l'argument han de pendre el mateix valor. Per exemple, la regla

```
grup_nominal(Número,Gènere)-->
article(Número,Gènere),nom(Número,Gènere)
```

introdueix els arguments Número i Gènere. Aquesta regla sols es satisfà si els valors que prenen les variables Número i Gènere en el nom i en l'article coincideixen. Notem que els arguments que comencen per majúscules fan referència a variables, mentre que els que comencen per minúscules fan referència a valors concrets. Finalment, notem que les paraules entre claudàtors [ ] fan referència a símbols terminals (elements del vocabulari). El lector més interessat en la notació DCG pot consultar Bratko (1990[1986]) o Clocksin i Mellish (1981).

## 2. *Elisions*

Les elisions -és a dir: en ortografia, la pèrdua d'una lletra en una paraula qualsevol-les trobem bàsicament en el cas dels articles, de les preposicions i dels pronoms febles.

### 2.1. *Articles i preposicions*

L'article masculí singular *el* i la preposició *de* perden la *e* davant de vocal (o de *h+vocal*), llevat que la *i* sigui semiconsonant (p. ex., l'ocell, l'hivern; el iode, el hiatus; d'hivern, d'escola, de iode; ...), i l'article femení singular *la* perd la *a* també en els mateixos casos, llevat també que la *i* sigui semiconsonant i llevat també que la vocal sigui una *i* o una *u* àtones o que sigui una *a*-privativa o en companyia de

certs substantius (p. ex., l'escola, l'opinió; la hiena; la idea, la humitat; la anormalitat; la una, la ira; ...).

En els casos que acabem d'esmentar no deixa de tractar-se d'una alternança gràfica que es produeix en unes condicions ben definides. No obstant això, el fet que no sempre les mateixes condicions provoquin la mateixa alternança fa que informàticament hàgim de plantejar-nos quina podria ser la millar solució.

En Prolog, hauríem de definir les regles següents:

```
article(masculí,singular,consonant) --> [el].
article(femení,singular,consonant) --> [la].
article(_,singular,vocal) --> [l].
```

que podríem llegir "l'article masculí i singular que s'usa davant de consonant és *el*", "l'article femení i singular que s'usa davant de consonant és *la*" i "l'article de qualsevol gènere i singular que s'usa davant de vocal és *l*". El signe '\_' indica una variable qualsevol; dues ocurrences diferents de '\_' fan referència a diferents variables.

Això voldria dir que a l'hora de definir els noms hauríem d'incloure també aquesta noció ortogràfica. Noms com *idea*, *humitat*, *hiena*, *iode*, *anormalitat*, *ira*, etc., haurien de tenir, d'acord amb les regles ortogràfiques, l'argument *consonant*. El problema se'ns presenta quan volem introduir la preposició *de* (*d*). Si seguíssim el mateix model hauríem de proposar, tenint en compte que les preposicions no tenen arguments de gènere ni de nombre, el següent:

```
preposició(consonant) --> [de].
preposició(vocal) --> [d].
```

Però això provocaria la generació de casos erronis com 'de idea', 'de ira', 'de anormalitat', etc., al costat de casos correctes com 'de pare', 'de iode', etc.

Evidentment, hauríem d'introduir un ítem nou al costat dels de consonant/vocal, que podríem anomenar *àton* (com a etiqueta, qualsevol altre seria vàlid), i afegir a les llistes anteriors les regles següents:

```
article(femení,singular,àton) --> [la].
preposició(àton) --> [d].
```

Això ens obligaria a redefinir els noms on teníem *consonant* d'acord amb els usos de l'article i de la preposició, p. ex.:

```
nom(femení,singular,àton) --> {idea}.
nom(femení,singular,consonant) --> {hiena}.
nom(femení,singular,àton) --> {ira}.
```

perquè la generació fos correcta: la idea, d'idea, la hiena, de hiena, la ira, d'ira, etc.

Una altra problemàtica relacionada amb l'elisió en el cas dels articles i de les preposicions, que ens obligarà a definir més encara aquestes regles, és la que presenten els anomenats articles contractes.

Aquests articles contractes afecten només les preposicions *a*, *de* i *per* en les següents condicions:

a) quan l'article és masculí plural, i

b) quan, essent l'article masculí singular, el nom comença en consonant.

Sembla raonable definir les contraccions en regles independents, afegint als arguments de l'article l'argument semàntic aportat per la preposició:

contracte(masculí,singular,consonant,de) --> [del].  
 contracte(masculí,plural,\_,de) --> [dels].

on *de* seria l'argument semàntic de la preposició.

Aleshores seria funció d'una regla sintàctica l'elecció d'un contracte o d'una preposició + article. Ara bé, una regla com

... --> preposició(Inici),  
 article(Gènere,Nombre,Inicial),  
 nom(Gènere,Nombre,Inicial).

ens podria donar com a resultat formes incorrectes com p. ex. 'a el pare'. Per evitar-ho, ens caldria donar un altre argument, aquesta vegada per fer referència a l'actitud que tenen davant del fet de la contracció, tant a preposicions com a articles. Les regles quedarien definides de la manera següent:

article(masculí,singular,consonant,no\_contracte) --> [el].  
 article(masculí,plural,\_,no\_contracte) --> [els].  
 article(femení,singular,consonant,contracte) --> [la].  
 article(femení,singular,àton,contracte) --> [la].  
 article(\_,singular,vocal,contracte) --> [l'].  
 article(femení,plural,\_,contracte) --> [les].  
 preposició(consonant,contracte) --> [de].  
 preposició(vocal,fora\_contracte) --> [d'].  
 preposició(àton,fora\_contracte) --> [d'].  
 preposició(\_,contracte) --> [a].  
 preposició(\_,contracte) --> [per].  
 preposició(,\_ ) --> [amb].

Fem notar que, a més de *contracte* (etiqueta amb què nomès volem indicar que permet la successió preposició+article) i *no\_contracte* (etiqueta amb què nomès volem dir que la successió anterior no és possible si la preposició admet contracció), necessitem un altre ítem, que hem anomenat *fora\_contracte*, per a la preposició *d'*; és obvi que aquesta forma va sempre sense article i per tant no participa en aquest problema.

D'aquesta manera, una regla com

... --> preposició(Inici,Contracció),  
 article(Gènere,Nombre,Inicial,Contracció),  
 nom(Gènere,Nombre,Inicial).

ens generaria només les formes correctes 'a la mare', 'de l'home', etc. I una altra regla:

... --> contracte(Gènere,Nombre,Inicial,Preposició),  
 nom(Gènere,Nombre,Inicial).

ens generaria també només formes contractes correctes, p. ex.: 'del pare', 'als homes', 'pel iode'.

## 2.2. *Pronoms febles*

Els pronoms febles presenten una problemàtica especial que, per la seva complexitat, caldria estudiar en un article a part. Pensem només en els pronoms febles avantposats o posposats a les formes verbals, l'ús de l'apòstrof i del guionet, el fet d'haver-hi formes diferents en el cas que hi hagi només un pronom o una combinació de dos o tres pronoms, etc.

D'altra banda, les solucions giren a l'entorn de la mateixa problemàtica exposada en el punt anterior. Cal fer observar, però, un problema específic. Un programa de lectura que ens aïlli les paraules haurà de contenir indicacions sobre l'apòstrof: s'ha de llegir "l' " "home" o bé "l'" "home"? En el cas de l'article i de la preposició *de* és clar que la forma més eficient és la segona; d'altra manera hauríem d'indicar doblement (amb apòstrof i sense) tots els noms començats en vocal. Ara bé, en el cas de "me'l", "te'ls", etc., sembla que l'elisió que s'ha produït ens obligui a fer la lectura "me" "l'", "te" "ls", etc., la qual cosa ens obligarà a donar a l'ordinador instruccions perquè unes vegades ens llegeixi "l'" i unes altres "'l'". En definitiva només hi ha dues opcions: o donem aquestes instruccions o fem sempre la mateixa lectura, ajuntant l'apòstrof al final del primer element. Això últim ens faria tenir, a més de *me*, *te* i *se*, les formes *me'*, *te'* i *se'*, que potser no són des del punt de vista de l'ortografia plausibles, però que sí que poden ser des del punt de vista de la programació molt més eficients. Hi ha, però, una objecció important: no podem exposar-nos a haver de duplicar formes verbals quan tenim pronoms posposats mitjançant l'apòstrof, p. ex. *ajuda'm*, *dóna'ls*, etc.

## 3. *Diacrítics, duplicacions i alternances*

Aquests fenòmens presenten sempre, al capdavall, el mateix problema: el dels canvis ortogràfics que sofreixen els morfemes i que són fruit d'una morfotàctica determinada.

Pel que fa als diacrítics -l'accent gràfic i la dièresi- tenim alternances com és ara *camí-camins* (í-i), *aglà-aglans* (à-a), *llengua-llengües* (u-ü), *canti-canviï* (i-ï), etc.

Quant a les duplicacions -consonants dobles-, bàsicament ens referim a la *s* final que s'ha de mantenir sorda en els derivats, com passa a *nas-nassos*, *interès-interessos*, etc.; però també, p. ex., a la *l* en casos com *tranquil-tranquil·la* o *nul·nul·la*.

I pel que fa a les alternances en un sentit més estricte, és a dir: als canvis d'unes lletres per altres lletres o dígrafs per raons de conservació del mateix so en entorns vocàlics diferents, tenim com a exemples les parelles *roca-roques*, *platja-platges*, *plaça-places*, *amiga-amigues*, etc. En aquest apartat també hi hem d'incloure l'alternança sorda-sonora de les oclusives finals de lexema, com en *amic-amiga*, *nebot-neboda*, etc., i altres casos com *europèu-europea*, *meu-meva*, *mig-mitja*, etc.

### 3.1. *Alternances a la classe nominal*

Quan definim les regles referents als articles o a les preposicions podem permetre'ns el luxe d'introduir una regla per a cada cas diferent perquè es tracta de

classes tancades i que consten d'un número relativament reduït d'elements. Quan definim noms (o adjectius) o verbs no podem actuar de la mateixa manera. No podem definir:

```
nom(masculí,singular,consonant) --> [nen].
nom(femení,singular,consonant) --> [nena].
nom(masculí,plural,consonant) --> [nens].
nom(femení,plural,consonant) --> [nenes].
```

Això ens obligaria a allargar enormement i inútilment la llista de les regles. Hem de trobar una altra solució més econòmica. Evidentment, la primera que se'ns acut és la de definir el lexema com a un fet, p. ex.:

```
rel(consonant,nen).
```

i treure les altres formes mitjançant regles com és ara:

```
nom(masculí,singular,Inici) --> [Nom],
    {rel(Inici,Nom)}.
nom(masculí,plural,Inici) --> [Nom],
    {rel(Inici,Rel),concat(Rel,s,Nom)}.
nom(femení,singular,Inici) --> [Nom],
    {rel(Inici,Rel),concat(Rel,a,Nom)}.
nom(femení,plural,Inici) --> [Nom],
    {rel(Inici,Rel),concat(Rel,es,Nom)}.
```

Notem que hem ampliat la sintaxi de la regla permetent posar informació a la dreta enmig dels símbols {}. Aquesta informació expressa les condicions que s'han de complir per poder satisfer la regla. Però això ens depara alguns problemes importants. En primer lloc hauríem de disposar de diferents classes de noms: *nen* pertanyeria a la classe 1 perquè la mateixa rel generaria formes masculines, femenines, singulars i plurals diferents; *home* pertanyeria a la classe 2 perquè, encara que les variacions siguin les mateixes que en el cas anterior, només hauria de generar formes masculines; *padrí* pertanyeria a la classe 3, ja que els al·lomorfs de femení i de plural són diferents dels que teníem en el cas dels noms anteriors; etc. D'altra banda, una rel com *avi* seria només apta per generar masculins i hauríem d'introduir una altra rel (*àvi*) per generar els femenins; per no fer aquestes duplicacions hauríem d'introduir subclasses dintre de cada una de les classes anteriors segons els canvis ortogràfics que presentessin. Per exemple, hauríem de definir:

```
rel(consonant,classe1,subclasse1,nen).
rel(vocal,classe1,subclasse2,avi).
rel(vocal,classe2,subclasse1,home).
rel(consonant,classe3,subclasse1,padrí).
```

Les regles que hauríem d'introduir per generar tots els noms serien: en primer lloc, el masculí singular es podria generar directament, sense distinció de classes ni de subclasses:

```
nom(masculí,singular,Inici) --> [Nom],
    {rel(Inici,_,_,Nom)}.
```

En segon lloc, el masculí plural es podria generar encara sense gaires dificultats mitjançant dues regles: la primera exclouria les rels de *classe3* i la segona seria

específica per a les d'aquesta classe (però s'hauria d'afegir una regla per a cada subclasse; aquí, *substitueix* ens conduiria a un procediment que ens substituïria *i* (codi ASCII 146) per *i* (codi ASCII 105)):

```
nom(masculí,plural,Inici) --> [Nom],
    {rel(Inici,Classe,_,Rel),Classe\=classe3,concat(Rel,s,Nom)}.
nom(masculí,plural,Inici) --> [Nom],
    (rel(Inici,classe3,subclasse1,Rel),name(Rel,Y),
    substitueix(146,Y,105,Z),name(Rel2,Z),concat(Rel2,ns,Nom)).
```

En tercer lloc, la generació del femení singular hauria de disposar de tres regles (només per als exemples descrits): una per a la *classe1*, *subclasse1* (nen -> nena); una altra per a la *classe1*, *subclasse2* (avi -> àvia); i una altra per a la *classe3*, *subclasse1* (padrí -> padrina). Finalment, per a la generació del femení plural caldrien tres regles més com en el cas anterior.

Hem d'observar que si bé les classes poden ser poques (tinguem en compte la pobra variació al·lomòrfica del plural: -s, -os, -ns, -ø), no passa el mateix amb les subclasses. Uns quants exemples ens faran adonar de les múltiples variacions existents:

```
padrí-padrina-padrins-padrines
vef-veïna-veïns-veïnes
amic-amiga-amics-amigues
nebot-neboda-nebots-nebodes
nuvi-núvia-nuvis-núvies
promès-promesa-promesos-promeses
germà-germana-germans-germanes
mestre-mestra-mestres-mestres
europeu-europea-europeus-europees
soci-sòcia-socis-sòcies
boca-boques
esponja-esponges
desig-desitjos
llengua-llengües
interès-interessos
```

...

Si reflexionem una mica sobre la quantitat de regles necessàries per generar les formes nominals a partir de les seves rels, aviat veurem que aquest camí és impracticable.

Creiem que la solució més adequada a aquest problema és establir uns fets que incloguin tota la variació nominal, d'una manera ordenada, i aplicar a cada una de les variants, mitjançant regles, els arguments variables. D'aquesta manera tindriem, p. ex.:

```
rel(consonant,nen,nena,nens,nenes).
rel(consonant,padrí,padrina,padrins,padrines).
```

També es podrien definir com a fets simples paraules de rels diferents però semànticament complementàries, sempre que l'argument consonant/vocal/àton coincidís, com p. ex.:

rel(consonant,pare,mare,pares,mares).  
rel(consonant,gall,gallina,galls,gallines).

I fins i tot podríem definir de la mateixa manera noms que no presentessin variació de gènere, omplint els arguments buits amb un indicatiu d'absència de forma, que podria ser s:

rel(consonant,s,cosa,s,coses).  
rel(consonant,sostre,s,sostres,s).

Una dada interessant pot ser que en aquestes definicions tenim aproximadament els mateixos arguments que en el cas anterior, però és que a més a més anul·laríem totes les dades classificatòries -que no representen res més que treball intern- i podríem reduir encara més el nombre de fets a definir (com en el cas de pare-mare-pares-mares). Les regles que ens generarien tots els noms serien només quatre, amb la qual cosa l'economia, tant pel que fa a les regles com pel que fa als fets, és màxima:

nom(masculí,singular,Inici) --> [Nom],  
    {rel(Inici,Nom,\_,\_),Nom\s}.  
nom(femení,singular,Inici) --> [Nom],  
    {rel(Inici,\_,Nom,\_,\_),Nom\s}.  
nom(masculí,plural,Inici) --> [Nom],  
    {rel(Inici,\_,\_,Nom,\_,\_),Nom\s}.  
nom(femení,plural,Inici) --> [Nom],  
    {rel(Inici,\_,\_,\_,Nom),Nom\s}.

### 3.2. Alternances a la classe verbal

Si traspassem el problema del camp nominal al verbal, podem comprovar que la solució no pot ser pas la mateixa. Evidentment, les quatre formes de la variació nominal podem expressar-les com a diferents arguments d'un sol fet; però en el cas dels verbs tindríem una llista argumental exageradament llarga (pensem només en els set temps simples, exclosos l'imperatiu i les formes no finites, les tres persones i la diferència de nombre, la qual cosa faria que haguéssim de contemplar quaranta-dos arguments només de càire morfològic; a aquests caldria afegir-hi encara els corresponents a la sintaxi, semàntica, pragmàtica, etc.). La solució haurà de ser per tant, forçosament, mixta.

L'intent de sistematització ens hauria de dur a reduir com més millor les regles i a establir uns fets amb un nombre d'arguments que poguessin incidir en aquesta reducció, eliminant fins allà on fos possible arguments classificatoris.

Així, hauríem de tendir a poder donar només una regla per a la formació de cada una de les persones d'un temps verbal determinat; les irregularitats, morfològiques o ortogràfiques, i les variants segons les classes de conjugació haurien de quedar incloses ja en els arguments expressats en els fets. De fet, és al present d'indicatiu allà on els verbs catalans presenten la variació més gran:

començo	comences	...
fujo	fuges	...
pudo	puts	...



escric	escrius	...
sé	saps	...
cabo	caps	...
puc	pots	...
conec	coneixes	...
estic	estàs	...
sóc	ets	...

En aquest temps, l'única rel que es pot establir i que sigui vàlida per a tots els casos, sense cap excepció, és la que trobem a la primera i segona persones del plural, incloent en aquesta rel la vocal temàtica i deixant com a morfema gramatical simplement *-m* i *-u*, respectivament. Dit d'una altra manera, els fets haurien de contenir, només per al present, cinc arguments:

```
rel_verbal(començo,comences,comença,comence,comencen).
rel_verbal(fujo,fuges,fuig,fugi,fugen).
rel_verbal(pudo,puts,put,pudi,puden).
rel_verbal(escric,escrius,escriu,escrivi,escriuen).
```

Les regles aplicables serien<sup>1</sup>:

```
verb(present,indicatiu,primera,singular) --> [Verb],
  {rel_verbal(Verb,_,_,_),Verb\=s}.
verb(present,indicatiu,segona,singular) --> [Verb],
  {rel_verbal(Verb,_,_,_),Verb\=s}.
verb(present,indicatiu,tercera,singular) --> [Verb],
  {rel_verbal(Verb,_,_,_),Verb\=s}.
verb(present,indicatiu,primera,plural) --> [Verb],
  {rel_verbal(Verb,_,_,Rel),Rel\=s,concat(Rel,m,Verb)}.
verb(present,indicatiu,segona,plural) --> [Verb],
  {rel_verbal(Verb,_,_,Rel),Rel\=s,concat(Rel,u,Verb)}.
verb(present,indicatiu,tercera,plural) --> [Verb],
  {rel_verbal(Verb,_,_,Rel),Rel\=s,concat(Rel,u,Verb)}.
```

Per a l'imperfet d'indicatiu hauríem de disposar de dues rels: la primera per a tot el singular i la tercera persona del plural i la segona per a les dues persones del plural restants. Si hi parem una mica d'atenció, veurem que és l'única regularitat que pot englobar totes les irregularitats de tots els verbs. Les terminacions que hi deixem (*-a*, *-es*, *-a*, *-em*, *-eu*, *-en*) són també comunes a tots els verbs sense cap excepció<sup>2</sup>:

```
rel_verbal(...,començav,començàv).
rel_verbal(...,fugí,fugí).
rel_verbal(...,rei,rèi).
rel_verbal(...,cloï,cloï).
```

En canvi, per al futur i condicional, una sola rel ens podrà generar totes les formes. Les terminacions seran sempre *-é*, *-às*, *-à*, *-em*, *-eu*, *-an*, per al futur, i *-ia*, *-ies*, *-ia*, *-íem*, *-íeu*, *-ien*, per al condicional:

```
rel_verbal(...,començar).
rel_verbal(...,fugir).
```

<sup>1</sup> Preveiem també el cas de l'argument buit (*s*) per qüestions que quedaran explicades més avall.

<sup>2</sup> Per brevetat, els arguments que hem vist anteriorment els substituïm per punts suspensius.

rel\_verbal(...,haur).  
rel\_verbal(...,anir).

En el cas del perfet simple necessitarem tres rels: una per a la primera del singular (desí, comencí, canvií, esglaií, aní, apareguí, fui, viu, ...), a la qual no s'haurà d'afegir cap terminació; una altra per a la segona del singular i la tercera del plural (desa, comença, esglaiia, aparegue, reduí, fo, fe, ve, ...), a la qual afegirem mitjançant regles les terminacions -res i -ren; i la tercera per a les formes restants (desà, comença, esglaià, aparegué, reduí, fó, fé, vé, ...), a la qual afegirem -rem, -reu a les persones del plural. Només hi ha tres formes que s'han de tractar a part: les rels que hem posat com a exemple en aquest últim cas, corresponents als verbs ser, fer i veure, no són les formes corresponents a la tercera persona del singular, que són fou, féu i véu, respectivament. Aquestes formes les haurem d'incloure separatament al començament del llistat de les regles, com a excepcions:

verb(perfet\_simple,indicatiu,tercera,singular) --> [fou].  
verb(perfet\_simple,indicatiu,tercera,singular) --> [féu].  
verb(perfet\_simple,indicatiu,tercera,singular) --> [véu].

i eliminar-ne la possibilitat de generació mitjançant la regla general:

verb(perfet\_simple,indicatiu,tercera,singular) --> [Verb],  
{rel\_verbal(...,\_,\_,Verb),Verb\=fó,Verb\=fé,Verb\=vé}.

Per que fa al present de subjuntiu, ens caldran dues rels: una per a tot el singular i la tercera del plural (desi, pateixi, lloí, esglaií, redueixi, vagi, ...), a la qual s'haurà d'afegir la terminació -s a la segona del singular i la terminació -n a la tercera del plural; i l'altra per a la primera i segona del plural (dese, patí, lloe, esglaié, reduí, ane, ...), a la qual afegirem respectivament les terminacions -m i -u. Només haurem de fer excepció en el cas de la primera i de la tercera del singular dels verbs *cabere* i *saber*:

rel\_verbal(...,càpigüe,capigüe).  
rel\_verbal(...,sàpigüe,sapigüe).  
verb(present,subjuntiu,primera,singular) --> [càpiga].  
verb(present,subjuntiu,tercera,singular) --> [càpiga].  
verb(present,subjuntiu,primera,singular) --> [sàpiga].  
verb(present,subjuntiu,tercera,singular) --> [sàpiga].  
verb(present,subjuntiu,primera,singular) --> {Verb},  
{rel\_verbal(...,Verb,\_) ,Verb\=càpigüe,Verb\=sàpigüe}.  
verb(present,subjuntiu,tercera,singular) --> [Verb],  
{rel\_verbal(...,Verb,\_) ,Verb\=càpigüe,Verb\=sàpigüe}.

Finalment, en el cas de l'imperfet de subjuntiu, hem de disposar també de dues rels: una per a la segona del singular i la tercera del plural (cantes, reduís, ...), a la qual afegiríem -sis, -sin, respectivament; i una altra per a les formes restants (cantés, reduís, ...), a la qual afegiríem a les persones del plural les terminacions -sim i -siu. L'únic problema que se'ns presentaria el tindríem a la primera i tercera persones del singular de les formes monosil·làbiques (fes, fos, des -una de les formes del verb donar-), que a diferència de la primera i de la segona del plural respectives -bisil·làbiques- no duen accent gràfic; la solució hauria de ser la mateixa que hem vist anteriorment per al present de subjuntiu en el cas dels verbs *cabere* i *saber*.

Tot el que acabem d'exposar ens portaria a uns fets com és ara:

rel\_verbal(començo,comences,comença,comence,comencen,començav,  
començàv, començar,comenci,comença,començà,comenci,comence,  
comences,comencés).

rel\_verbal(fujo,fuges,fuig,fugi,fugen,fugi,fugi,fugir,fugí,fugi,fugí,fugi,fugi,  
fugis,fugís).

rel\_verbal(escric,escrius,escriu,escrivi,escriuen,escrivi,escriví,escriur,  
escriví,escrivi,escriví,escrigue,escrivis,escrivís).

No ens ha de sorprendre veure arguments repetits; aquestes repeticions ens calen per poder generar les formes de tots els verbs d'una manera uniforme, és a dir: amb les mateixes regles. Així podem veure que en el cas de començar, els arguments 2 i 14 són idèntics (el 2 correspon a la segona del singular del present d'indicatiu i el 14 a les formes comencassis i comencassin), però hem de veure que aquests mateixos arguments en el cas de fugir i escriure són diferents.

Pel que fa a les regles, n'haurem d'escriure només una per a cada persona de cada temps perquè el programa ens generi totes les formes possibles de tots els verbs (a part, naturalment, de les excepcions que hem anat apuntant -fou, féu, véu, càpiga (2), sàpiga (2), fes (2), fos (2), des (2)-, que sumen un total de tretze). També haurem de tenir en compte que necessiten un tractament especial els verbs defectius com ploure, caldre, etc., que haurem de considerar com a excepcions.

Per acabar, i fent referència a una nota a peu de pàgina que hem fet anteriorment, hem de pensar que alguns verbs els haurem d'introduir en dos fets diferents. Per exemple, en casos de formes dobles com nasqués-naixés, cresqués-creixés, etc., o en casos en què unes rels supletives comencin en vocal i altres en consonant, com el cas de vagi-anem, en què l'argument Inicial (vocal/consonant, que necessitem per a la sintaxi amb pronoms febles: el vagi portar, l'anem a veure, ...) haurà de ser diferent. En aquests casos, els arguments que no representen complementarietat han de ser arguments buits (que escrivim s).

#### 4. *Conclusió*

A tall de conclusió volem dir només que el tractament computacional del llenguatge escrit presenta uns problemes afegits a la descripció lingüística que, si seguim els mateixos paràmetres i les mateixes tècniques de descripció morfològica que són aplicables a aquesta última, ens abocaríen a una enorme complicació que aniria en detriment de l'eficàcia del programa. El nostre intent ha estat el de resoldre aquesta problemàtica amb la màxima economia descriptiva i l'eliminació de totes les dades de treball morfològic que no tenen incidència en la sintaxi.

Jaume Tió  
Dpt. de Filologia I - Secció de Lingüística General  
Universitat de Lleida  
Apartat de Correus 471  
E-25080 Lleida

Felip Manyà  
 Dpt. d'Informàtica i Enginyeria Industrial  
 Universitat de Lleida  
 Apartat de Correus 471  
 E-25080 Lleida

## BIBLIOGRAFIA

- BADIA MARGARIT, A. M. (1962): *Gramàtica catalana*, 2 vols., Madrid:Gredos.  
 BRATKO, I. (1990[1986]): *Prolog. Programming for Artificial Intelligence*, Wokingham, England:Addison-Wesley, 21990.  
 CLOCKSIN, W.F.; MELLISH, C.S. (1981): *Programming in Prolog*, Berlin:Springer. (Traducció castellana: *Programación en Prolog*, Barcelona:Gustavo Gili, 1987).  
 FABRA, P. (1937): *La conjugació dels verbs en català*, Barcelona:Barcino.  
 PEREIRA, F.; WARREN, D.H.D. (1980): *Definite Clause Grammars for Language Analysis: A Survey of the Formalism and Comparison with Augmented Transition Networks*, *Artificial Intelligence* 13(1980), pp. 231-278.  
 STERLING, L.; SHAPIRO, E. (1986): *The Art of Prolog*, The MIT Press.

## RESUM

Els autors sostenen que el tractament computacional del llenguatge escrit presenta uns problemes afegits a la descripció lingüística que no poden atènyer-se als mateixos paràmetres i les mateixes tècniques de descripció morfològica que són aplicables a aquesta última. Així, el que ens proposen és resoldre aquesta problemàtica, utilitzant el llenguatge Prolog, amb la màxima economia descriptiva i l'eliminació de totes les dades de treball morfològic que no tenen incidència en la sintaxi.

## SUMMARY

### Catalan Orthography and Computational Linguistics

The authors of this paper argue that computational processing of written language poses specific problems regarding their linguistic description which cannot be dealt with by using parameters and techniques for morphological description that are otherwise applicable in linguistics. Their line of argument is that these problems can be solved by maximum descriptive economy when using the programming language Prolog and by eliminating all of the morphological data that do not affect the syntactic structures.