Les premiers dictionnaires électroniques du roumain en format DELA[1]

Dana-Marina DUMITRIU

0. Introduction

Pour mieux répondre aux défis de notre époque, le monde scientifique a élaboré des outils de traitement automatique des textes bruts et des dictionnaires électroniques adaptés d'un côté aux exigences des automates, de l'autres, aux particularités de chaque langue. L'un des formats les plus répandus des dictionnaires électroniques est le format DELA élaboré par les chercheurs du LADL[2] de Paris VII et les laboratoires d'informatique linguistiques associés du monde entier qui forment le réseau RELEX. Les dictionnaires DELA ont été conçus d'abord pour le français, langue qui en bénéficie d'une variété très importante : dictionnaires morphologiques des formes simples (DELAS) et des formes composées (DELAC), dictionnaire des formes fléchies (DELAF), dictionnaire phonémique (DELAP), dictionnaire des radicaux de la langue (DELAR). Mais le format DELA a été exporté vite à d'autres langues.

Les premiers dictionnaires du roumain sont des dictionnaires des formes simples (le DELAS et le DELAF du roumain).

Les dictionnaires DELAS sont des dictionnaires morphologiques qui enregistrent seulement les formes canoniques qu'ils font suivre par un code précisant la classe de flexion auquel appartient la forme en question :

arbitru,N1
(DELAS)

Les entrées des dictionnaires DELAF sont les formes fléchies obtenues par l'application aux formes canoniques des DELAS des grammaires de flexion correspondant au code mentionné :

arbitru, arbitru. N1+Hum: iNms:iAms:iGms:iDms
arbitrul, arbitru. N1+Hum: :dNms:dAms
arbitrului, arbitru. N1+Hum: dGms:dDms
arbitri, arbitru. N1+Hum: iNmp:iAmp:iGmp:iDmp
arbitrii, arbitru. N1+Hum: dNmp:dAmp
arbitrilor, arbitru. N1+Hum: dGmp:dDmp
(DELAF)

Nous allons présenter dans ce qui suit quelques aspects de la flexion nominale du roumain par les automates, relevés dans la rédaction des premiers dictionnaires électroniques du roumain dans le format DELA. Pour cela nous commençons avec quelques considérations sur la flexion automatique et sur les particularités flexionnelles du roumain.

[1] Colloque international 'Coloquio de Lenguas Extranjeras', U.A.M. Azcapotzalco (Mexique), 9-10 sept 2005.
[2] Laboratoire d'Automatique Documentaire et Linguistique.

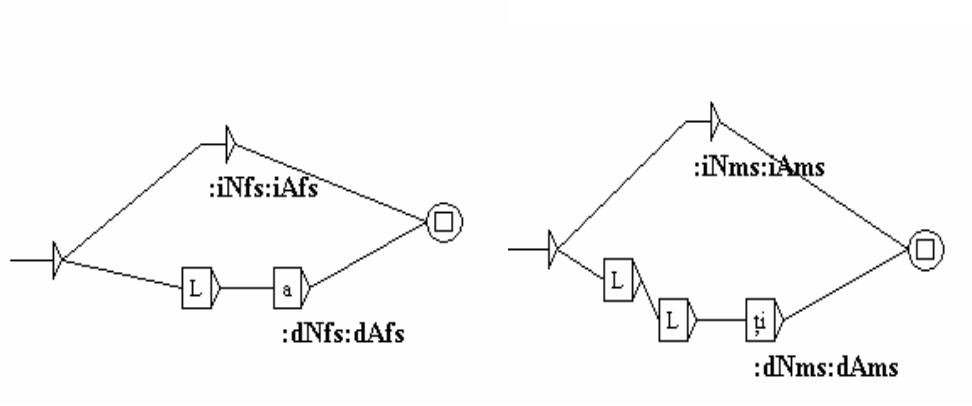
Philologica Jassyensia, An II, Nr. 1, 2006, p. 31-45

BDD-A764 © 2006 Institutul de Filologie Română „A. Philippide"
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-15 01:52:58 UTC)

1. Les opérateurs de la flexion automatique

La flexion automatique se réalise grâce à des opérateurs spécialisés qui agissent différemment sur la forme canonique. Les plus importants sont :

– *L* – qui déplace le curseur vers la gauche (*left*) de la forme canonique, ce qui a comme résultat l'enlèvement d'une lettre. Si la flexion suppose l'enlèvement de plusieurs lettres, à chacune on associera un opérateur *L* et on parlera d'enlèvement en cascade. Par exemple : pour aboutir à la forme définie de nominatif-accusatif du substantif *lumină*, il faut remplacer la lettre *ă* (désinence de féminin singulier nominatif – dans les grammaires traditionnelles) par la lettre *a* (article défini féminin singulier nominatif). Ce qui équivaut à l'application de l'opérateur *L* pour enlèvement de la lettre finale de la forme canonique et ajout sous une forme suffixale de la marque de la définitude. Pour décrire le chemin de flexion qui permet la réalisation de la forme de pluriel non défini du nom *munte*, il faut reculer jusqu'à la partie invariante du mot (*mun-*) par deux applications successives de l'opérateur *L* et ensuite ajouter le suffixe *-ți*.



(a) (b)

Fig 1. Applications simple (a) et en cascade (b) de l'opérateur *L*

– *R* – qui agit à la droite (*right*) d'un nom affecté déjà par l'opérateur *L* en rétablissant une lettre. Les noms neutres *abator*, *alofon* et *altoi* ont en commun l'alternance vocalique *o – oa* de l'avant dernière lettre, la dernière lettre étant différente. Au lieu de faire trois graphes à l'aide de l'opérateur en cascade *LL* auquel on ajouterait les suffixes *-are*, *-ane* et *-aie*, respectivement, on peut en envisager un seul à l'aide de l'opérateur *R* comme dans la grammaire N53. Leur unification sous une seule grammaire se base sur le fait que les suffixes en question sont formés des voyelles *a* et *e* entre lesquelles se trouve la dernière lettre de la forme canonique.

BDD-A764 © 2006 Institutul de Filologie Română „A. Philippide"
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-15 01:52:58 UTC)

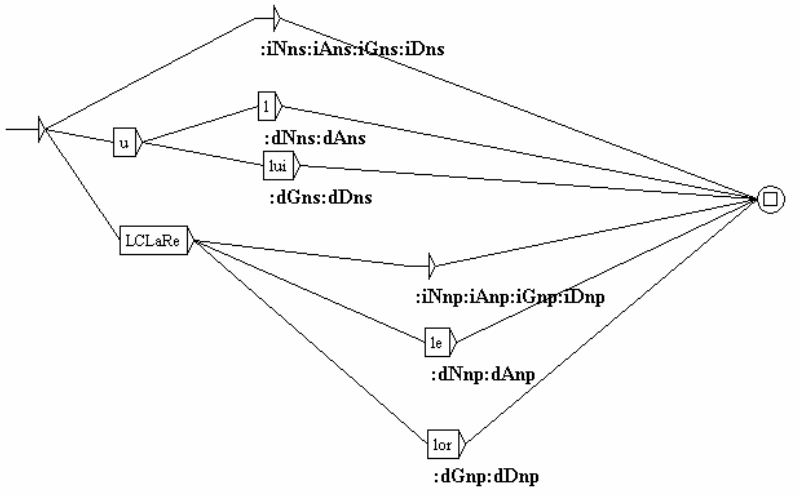


Fig 2. Exemple d'application de l'opérateur *R* (graphe N53)

– ***C*** – qui a pour effet de tout dupliquer à partir de la lettre qui se trouve à gauche de la position courante. Par exemple, si un graphe de flexion ne contient que l'opérateur C, cela aura pour effet de dupliquer la dernière lettre du lemme. Dans la flexion des mots roumain, il s'applique associé à d'autres opérateurs. La même grammaire N53 témoigne de l'application de cet opérateur.

2. Particularités du roumain

2.1. Particularités affectant le corpus

2.1.1. La composition

Les dictionnaires électroniques du roumain de la rédaction desquels nous nous sommes occupé étant des dictionnaires des formes simples, le problème de la distinction entre les formes simples et les formes composées devait être tranché dès le début et d'une façon claire pour permettre à l'ordinateur une analyse systématique et sans ambiguïté. Le problème n'est pas si simple qu'il semble au premier abord, car les grammaires traditionnelles sont unanimement d'accord qu'il y a trois types de composition : (a) avec trait d'union : *floarea-soarelui* 'tournesol', *rea-voință* 'mauvaise volonté', (b) sans trait d'union, mais avec des blancs : *punct de vedere* 'point de vue', (c) sans trait d'union, ni blancs : *bunăvoință* 'bonne volonté', *untdelemn* 'huile', *însumi* '(moi) même' . Ces derniers sont comparables aux mots simples par le fait d'être délimités par des blancs, mais, à la différence de ceux-ci, ils se laissent 'décomposer' (en *bună* + *voință* et, respectivement, *unt* + *de* + *lemn* , *însu(l)* + *mi*) et certains présentent même des variations flexionnelles au milieu du mot[3] (*însămi* – fém. sg. nominatif-accusatif/ *însămi* – fém. sg. génitif-datif).

La variation flexionnelle au milieu du mot s'avérant un critère non pertinent (homonymie totale des cas pour la forme masculine *însumi* et homonymie partielle – nominatif-accusatif /vs/ génitif-datif – de la forme féminine) et la décomposition un critère encombrant, nous avons décidé de facilité la capacité d'analyse de l'ordinateur

[3] Ces variations flexionnelles, acceptées par *rea-voință* (*relei-voințe* - GD) et refusées par *bunăvoință* (*bunăvoinței* - GD), sont la cause de l'orthographe différente des deux noms composés.

BDD-A764 © 2006 Institutul de Filologie Română „A. Philippide"
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-15 01:52:58 UTC)

par simplification et uniformisation considérant comme forme simple toute forme à l'intérieur de laquelle il n'y a pas de séparateurs (trait d'union, blanc). De ce point de vue (purement formel), *bunăvoință*, *untdelemn* et *însumi* sont des mots simples et lemmatisables, tout comme *măr* 'pomme' , *carte* 'livre', *frumos* 'beau'.

Fidèle à ce critère, nous traitons les dérivés des sigles comme des noms simples s'ils s'écrivent sans tiret : *pesedist* 'membre du parti PSD[4]» et *pedist* 'membre du parti PD[5]» seront répertoriés dans nos dictionnaires, mais pas PSD-ist et PD-ist.

L'application de ce critère comme critère primordial pour la réalisation du corpus a des conséquences importantes sur la représentativité de certaines parties du discours, qui ont uniquement des formes composées ou dont les formes composées sont en variation conditionnée avec des formes simples. Dans la première catégorie entre le pronom possessif qui a toujours des formes composées (*al meu* 'le mien', *a mea* 'la mienne', *al tău* 'le tien', *a ta* 'la tienne' etc.). Dans la deuxième, l'adjectif possessif, qui actualise une forme simple à droite du nom (*caietul meu* 'mon cahier') et une forme composée s'il est positionné à sa gauche (*al meu caiet* 'mon cahier').

Conséquences :

- les pronoms possessifs ne seront pas répertoriés ;
- seulement les formes simples de l'adjectif possessif seront enregistrées dans nos dictionnaires ;
- la reconnaissance de ces formes composées par l'ordinateur revient à un autre palier du traitement du texte, qui tient de la syntaxe[6].

2.1.2. L'autonomie

Généralement, les formes graphiquement autonomes sont lemmatisées. Deux cas mériteraient une discussion à ce propos.

Le premier concerne l'article défini. Celui-ci a des formes non autonomes qui se retrouvent dans les graphes des noms et des adjectifs et une forme autonome – *lui* – qui se rencontre devant les noms propres masculins et que nous posons comme forme canonique d'un Artdéf défectif au niveau du cas.

Les conséquences de cette approche sont les suivantes :

- représentation de l'article défini dans le corpus uniquement par la forme autonome, ce qui veut dire que le code Artdéf ne renvoie qu'à cette forme ;
- la conception de l'article défini forme autonome comme une partie du discours défective au niveau du cas.

Le deuxième cas intéressant d'autonomie concerne le pronom personnel, qui a des formes autonomes (*mie*, *mine*), non autonomes (*m-*) et graphiquement autonomes ou non autonomes (*te* ; *te-*, *-te*). Les premières correspondent à ce que la grammaire traditionnelle appelle formes 'accentuées' ou 'disjointes' ; les deuxièmes, aux formes 'non accentuées' ou 'conjointes' ; les dernières sont toujours des formes non accentuées, mais dont l'autonomie est conditionnée par la réalisation vocalique /vs/ consonantique du mot qui les précède ou leur suit :

[4] Parti Social-Démocrat.

[5] Parti Démocrat

[6] Ce n'est pas la même situation que pour *floarea-soarelui*, *rea-voință*, *punct de vedere* (qui imposent le recours à un DELAC). Il s'agit dans ce cas de la détection d'un modèle d'agencement de certaines parties du discours ayant comme résultat la création d'une autre partie : Artposs (*al*) + Pron (*meu*).

BDD-A764 © 2006 Institutul de Filologie Română „A. Philippide"
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-15 01:52:58 UTC)

***Te** văd.* 'je te vois'
***Te**-am văzut.* 'je t'ai vu'
*Văzându-**te**.* 'te voyant'

Nous avons codé les formes accentuées avec le code *a* et les formes non accentuées avec le code *b*.

Le pronom roumain connaît l'opposition *accentué* /vs/ *non accentué* uniquement au niveau des formes du Datif et de l'Accusatif. Comme nous avons décidé de faire un seul graphe pour les formes accentuées et les formes non accentuées et de poser comme lemme une forme *a*, nous avons dû prendre un décision concernant les formes de Nominatif et de Génitif : nous les avons considérées des formes *a*.

2.1.3. L'ambiguïté fonctionnelle (les homonymies des parties de discours)

L'ambiguïté fonctionnelle est résolue par dissociation et encodage individuel pour chaque valeur du lemme en question, aussi bien dans le DELAS que dans le DELAF:

acesta Adjdem +d
acesta PROdem

Si l'ambiguïté se produit entre une forme canonique d'une partie du discours (*lui*, article défini proclitique) et une forme non canonique d'une autre partie du discours (*lui*, pronom personnel de la troisième personne) ou entre des formes non canoniques, elle sera mise en évidence de la même façon, mais seulement dans le DELAF, le DELAS n'enregistrant pas les formes non fléchies :

lui,lui.Artdéf :Gms :Dms :Gfs :Dfs
lui,el.PRO3 :aGP3s

2.1.4. La position de l'adjectif pronominal par rapport au nom

La plupart des adjectifs pronominaux font une distinction, avec des conséquences sur la compositionnalité et la flexion, entre position à la gauche du nom (code : +*g*) et position à la droite du nom (code : +*d*) :

(*acest + al meu) caiet* /vs/ *caietul* (*acesta + meu)* '(ce + mon) cahier'
acest Adjdem +g
acesta Adjposs +d
meu Adjposs +d
(L'Adjposs +g (*al meu*) étant une forme composée n'entre pas dans notre discussion.)

Les DELAS et DELAF enregistreront séparément les formes positionnées à gauche du nom que les formes placées à sa droite. (Ce qui comme conséquence une augmentation des entrées.)

2.2. Particularités affectant la forme canonique

La forme canonique des mots qui connaissent la flexion nominale est celle de nominatif singulier (codée : Ns). Pourtant, le fait de dissocier l'article défini proclitique de l'article défini enclitique nous amène à une forme canonique de génitif-datif pour le premier. D'autre part, les noms *pluralia tantum* exigent une forme canonique de nominatif pluriel.

Dans le cas des parties de discours présentant des variations en fonction du genre, nous avons choisi comme forme canonique la forme de masculin singulier nominatif (msN). C'est ainsi que les formes féminines de l'article indéfini (*o*, *unei*) ne sont qu'une partie des sorties du graphe Artindéf posant comme forme canonique la forme *un*.

La complexité de la flexion du pronom personnel a exigé sa représentation par plusieurs sous graphes, chacun défini par une forme canonique. En échange, la réduction

BDD-A764 © 2006 Institutul de Filologie Română „A. Philippide"
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-15 01:52:58 UTC)

de la flexion du pronom de renforcement à une seule forme casuelle, a permis de réunir toutes les variations flexionnelles (en fonction du genre, du nombre, de la personne) dans un seul graphe ayant comme forme canonique la forme de la I^e personne masculine singulier.

2.3. Particularités affectant les graphes de flexion

2.3.1. Les alternances vocaliques et consonantiques

La flexion est accompagnée souvent des alternances consonantiques et/ou vocaliques qui affectent parfois non seulement la partie finale, mais aussi initiale ou médiane :

t/ț : *student/ studenți*; *elefant/ elefanți*
s/ş: *pas/paşi*; *serios/serioşi*; *ministru/miniştri*
d/z: *brad/brazi*; *rotund/rotunzi*
z/j: *obraz/obraji*
o/ oa: *om/oameni*; *popor/ popoare*; *gros/groasă*
ă/e: *văr/veri*; *măr/mere*
ea/e: *cafea/cafele*; *fereastră/ferestre*
oa/o: *moară/mori*; *floare/flori*

La conséquence la plus importante en est l'analyse du mot en *partie invariable + parties variables*. Cette analyse exigée par la flexion automatique ne se superpose pas dans tous les cas à la segmentation traditionnelle en *racine + désinence + article défini* (pour le nom et l'adjectif) et *racine + suffixe(s) + désinence(s)* (pour le verbe). C'est, par exemple, le cas du nom *om* qui, à cause de l'alternance vocalique *o – oa* sera réduit dans la flexion à sa partie invariante (*o*) à laquelle on ajoutera *-ameni* (pour iNmp:iAmp), *-amenii* (pour dNmp:dAmp), *-amenilor* (pour dGmp:dDmp).

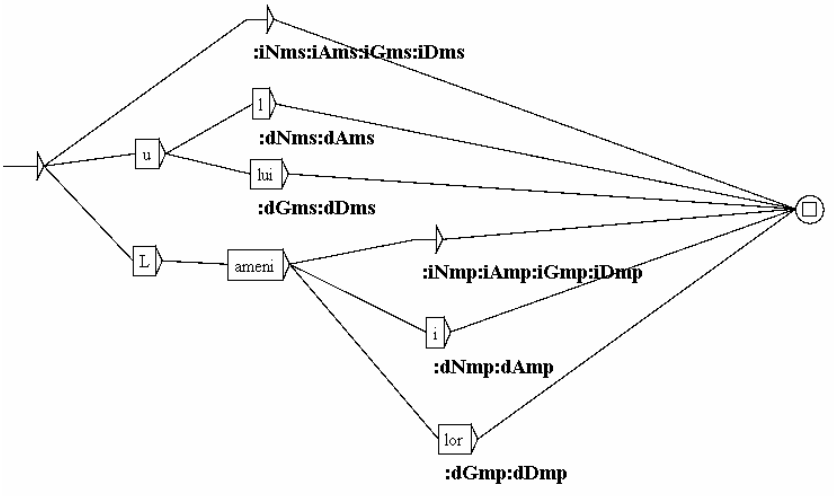


Fig 3. Graphe N3

C'est aussi le cas de l'article indéfini qui au pluriel nominatif-accusatif (*nişte*) ne garde rien de la forme canonique (*un*).

Une autre conséquence de cette différence au niveau du découpage du mot est l'augmentation du nombre des classes flexionnelles dans la flexion automatique par rapport aux classes établies par les grammaires traditionnelles.

BDD-A764 © 2006 Institutul de Filologie Română „A. Philippide"
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-15 01:52:58 UTC)

2.3.2. Les homonymies casuelles

Une caractéristique du roumain concernant les cas est l'homonymie entre le nominatif et l'accusatif et entre le génitif-datif[7]. Cette homonymie se rencontre aussi bien au singulier qu'au pluriel, quelle que soit le genre et affecte le nom, l'adjectif (qualificatif ou pronominal), l'article, les pronoms sans formes personnelles et une partie des pronoms avec formes personnelles. Ce qui nous détermine d'affirmer que l'homonymie N=A, G=D est une homonymie de base dans la langue. Nous considérons toujours comme homonymies de base, mais avec un degré plus petit de généralité (s'adressant soit aux noms masculins, soit aux noms féminins, soit aux neutres) les homonymies suivantes :

- Ns = As = Gs = Ds pour les noms masculins et neutres au singulier non défini[8]
- Np = Ap = Gp = Dp pour les noms masculins et neutres au pluriel non défini[9]
- Gs =Ds = Np =Ap = Gp = Dp pour les noms féminins sans article défini[10]

A part ces homonymies de base, il y a des homonymies qui caractérisent une certaine classe flexionnelle. Par exemple, il y a aussi bien parmi les noms masculins, que parmi les noms féminins et les neutres une classe qui présentent une homonymie à huit termes Ns = As = Gs =Ds = Np =Ap = Gp = Dp (il s'agit des classes N17, N39 et N55 de notre classification)[11].

Le nombre des homonymies a des conséquences sur le nombre des sorties de chaque graphe de flexion. Comparé au graphe N10, par exemple, qui a 6 sorties, le graphe N39 n'en a que 5 (v. Fig.9 et Fig.10).

2.3.3. Le défectivité

Le fait que certains noms sont des *singularia tantum* ou des *pluralia tantum* a des conséquences sur les graphes de flexion qui définissent leurs classes de flexion et qui constituent souvent une duplication de la partie du singulier ou, respectivement, du pluriel, des graphes de certaines classes de mots non défectifs. Ainsi, le graphe N54 (mot-type : *aboliționism*) reproduit-il la partie du singulier des graphes N51, N52 et N53 (mots-type : *abandon*, *abecedar*, *abator*). De même, le graphe N18 (mot-type : *lapți*) reproduit la partie du pluriel des graphes N7 et N9 (mots-type : *asistent*, *munte*).

2.3.4. La position de l'article défini

L'article défini roumain a un comportement de marqueur suffixal de la définitude ; il s'attache à la fin du nom ou de l'adjectif si celui-ci précède le nom :

copil 'garçon', *un copil frumos* 'un beau garçon'

*copilu**l*** [garçon-le], *copilu**l** frumos* [garçon-le beau] /vs/ *frumosu**l** copil* [beau-le garçon]

La non autonomie graphique de l'article défini empêche sa lemmatisation et impose la multiplication des parcours et des sorties de graphes des noms et des adjectifs,

[7] Ce qui fait certains linguistes parler de deux cas : le Nominatif-Accusatif et le Génitif-Datif.

[8] L'article défini réduit cette homonymie à la première homonymie de base (*omul* –NA, *omului* GD).

[9] L'article défini réduit cette homonymie à la première homonymie de base (*oamenii* –NA, *oamenilor* GD).

[10] Ex : *unei case*$_{Gs/Ds}$ 'de/à une maison' – *niște case*$_{Np/Ap}$ 'des maisons' - *unor case*$_{Gp/Dp}$ 'des/aux maison'.

[11] Les graphes de flexion illustrent très bien ces homonymies.

BDD-A764 © 2006 Institutul de Filologie Română „A. Philippide"
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-15 01:52:58 UTC)

par la prise en considération des formes avec article défini (codées : *d*) à côté des formes sans article défini[12] (codées : *i*).

3. Grammaires de la flexion nominale roumaine

Pour obtenir les formes flexionnelles enregistrées dans le DELAF, nous avons établi d'abord pour chaque terme du DELAS le type de flexion auquel nous avons associé ensuite une grammaire de flexion sous forme de graphe. Les grammaires de flexion (graphes) qui ont la forme d'un réseau avec une entrée (la forme canonique) et plusieurs chemins de parcours décrivant les opérations de flexion (ajout des suffixes, réduction de la forme canonique, rétablissement d'une ou plusieurs lettres enlevées, duplication d'un lettre en décalant tout ce qui se trouve à sa droite) subies par la forme canonique. Les sorties donnent les codes flexionnels qui se retrouveront dans les lignes du dictionnaire des formes fléchies après les codes grammaticaux et sémantiques.

3.1. Grammaires des articles

3.1.1. Grammaire de l'article indéfini (Artindéf)

La grammaire de l'article indéfini *un* (masculin singulier), *o* (féminin singulier), *nişte* (masculin et féminin pluriel) pose comme forme canonique la forme de masculin singulier. Ce qui implique l'application de l'opérateur LL pour les deux autres formes susmentionnées. Le graphe aura la forme suivante :

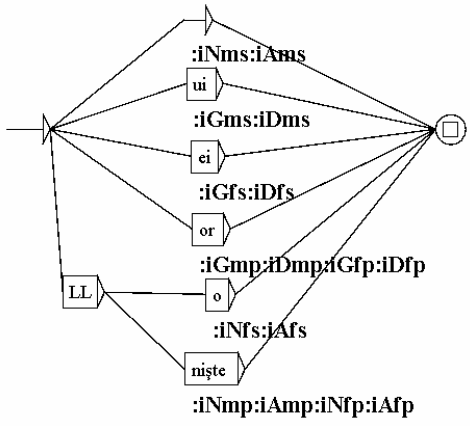


Fig 4. Graphe Artindef

Dans les dictionnaires DELA, l'article indéfini apparaîtra sous les formes :
un,Artindef
(DELAS)
unui,un.Artindef:iGms:iDms
o,un.Artindef:iNfs:iAfs
nişte,un.Artindef:iNmp:iAmp:iNfp:iAfp
unor,un.Artindef:iGmp:iDmp:iGfp:iDfp
unei,un.Artindef:iGfs:iDfs
un,un.Artindef:iNms:iAms
(DELAF)

[12] Sans aucun article ou avec un autre article qui a une forme autonome et qui n'affecte pas la forme du nom ou de l'adjectif.

BDD-A764 © 2006 Institutul de Filologie Română „A. Philippide"
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-15 01:52:58 UTC)

3.1.2. Grammaire de l'article défini (Artdef)

Les formes non autonomes de l'article défini étant prises en considération lors de la flexion nominale et adjectivale, seule la forme autonome bénéficiera d'une grammaire.

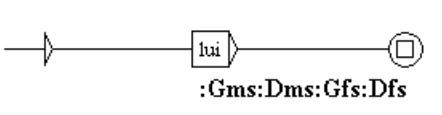


Fig 5. Graphe de flexion de la forme autonome de l'article défini

Nous ne codons pas la catégorie de la définitude dans ce cas – selon le modèle que nous avons établi pour l'article indéfini – parce que l'article défini proclitique n'apporte pas cette information au nom qu'il précède (toujours un nom propre) ; il n'est que morphème casuel dans la flexion des noms propres.

3.1.3. Grammaire de l'article possessif (Artposs) et de l'article démonstratif (Artdem)

Les formes de l'article possessif *al* (masculin singulier), *a* (féminin singulier), *ai* (masculin pluriel), *ale* (féminin pluriel) et celles de l'articles démonstratif *cel* (masculin singulier), *cea* (féminin singulier), *cei* (masculin pluriel), *cele* (féminin pluriel) présentent des similitudes de flexions qui seront illustrées grâce au sous graphe suivant :

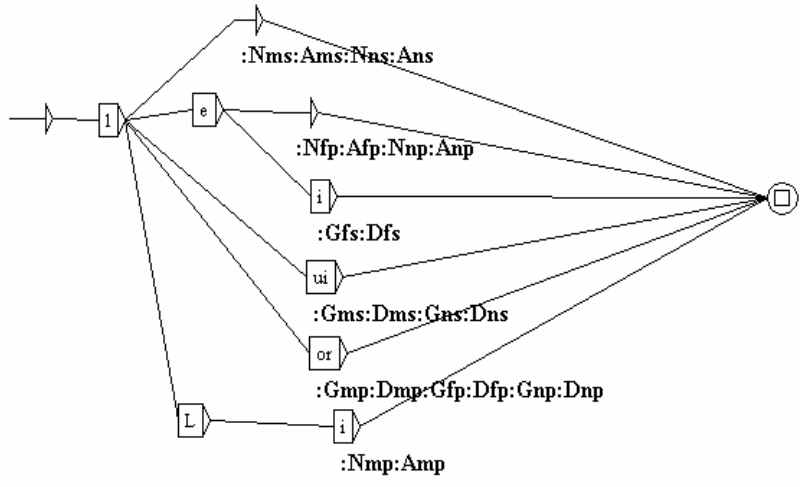


Fig 6. Sous graphe l.grf, auquel renvoient les graphes de l'article possessif et de l'article démonstratif

Les graphes de l'article possessif et de l'article démonstratif réalisés à l'aide de ce sous graphe seront:

BDD-A764 © 2006 Institutul de Filologie Română „A. Philippide"
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-15 01:52:58 UTC)

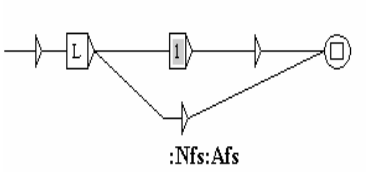


Fig 7. Graphe Artposs

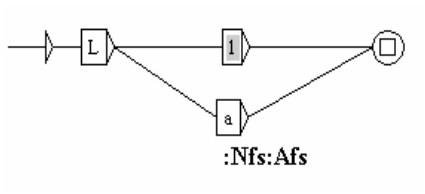


Fig 8. Graphe Artdem

Le même sous graphe apparaît dans la structure des graphes des pronoms et adjectifs démonstratifs.

3.2. Grammaires des noms

La flexion différente du nom en présence et en absence de l'article défini nous a déterminé à considérer les formes avec articles défini comme des formes flexionnelles du nom variant en fonction de la catégorie de la définitude et utiliser pour elles le code *d* s'opposant au code *i* des formes sans articles défini (à savoir: formes sans aucun article ou formes avec article indéfini[13]).

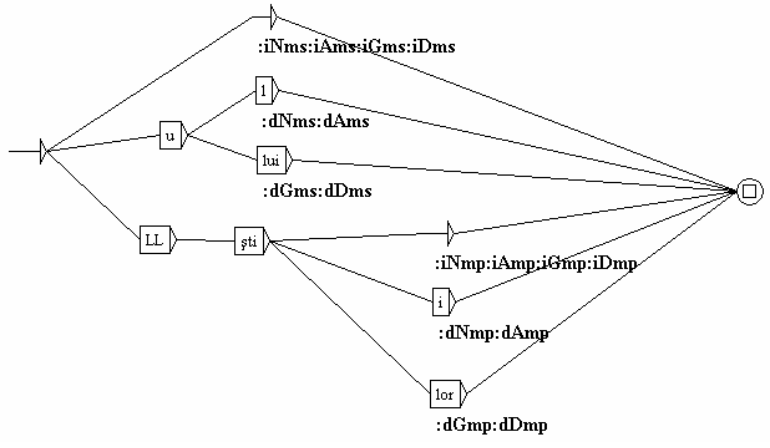


Fig. 9. Graphe N10

A côté des homonymies basiques (qui se rencontrent dans toutes les classes de la flexion nominales ; par exemple : nominatif = accusatif, génitif = datif), il y a des homonymies spécifiques à un nombre restreint de classes de flexion que l'on observe très bien en comparant les graphes. Par exemple, l'homonymie totale (à huit termes) des formes non définies est une caractéristique des noms des classes N17 (masculins), N39 (féminins) et N55 (neutres).

[13] Une forme casuelle sans article est identique à la même forme casuelle avec article indéfini : *un muncitor* 'un ouvrier' (masculin singulier nominatif-accusatif avec article indéfini) – *muncitor* 'ouvrier' (masculin singulier nominatif-accusatif sans article).

BDD-A764 © 2006 Institutul de Filologie Română „A. Philippide"
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-15 01:52:58 UTC)

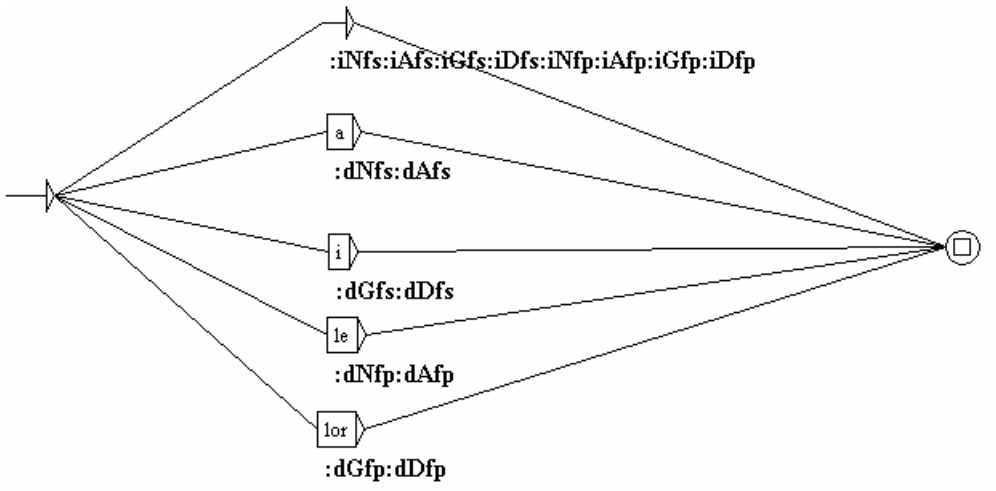


Fig. 10. Exemple d'homonymie à huit termes (graphe N39)

D'autres classes se caractérisent par blocage des formes de pluriel (les noms *singularia tantum*) ou de singulier (des *pluralia tantum*[14]).

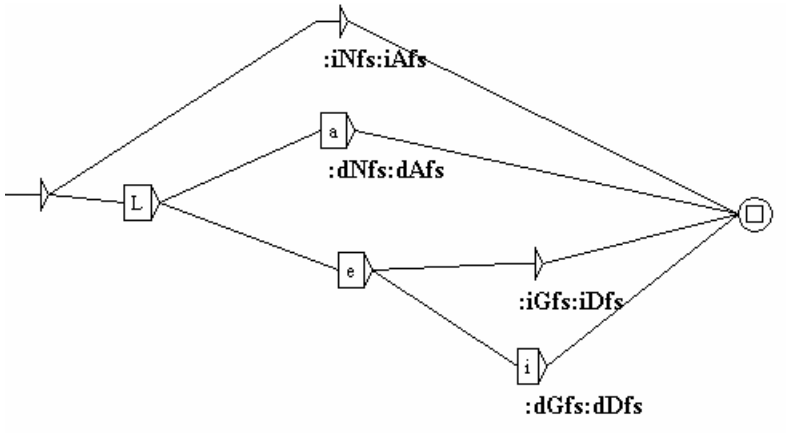


Fig. 11. Exemple de flexion des noms *singularia tantum* (graphe N34)

Les grammaires tiennent ensuite compte des opérateurs de flexion et des alternances vocaliques et consonantiques que nous avons résolues au niveau des suffixes, réduisant parfois le radical jusqu'à la lettre initiale de la forme canonique.

3.3. Grammaires des pronoms

Les graphes des pronoms se compliquent par rapport au graphes des noms à cause de la catégorie de la personne, qui rend parfois nécessaire le recours à des sous graphes.

Voilà, à titre d'exemple le graphe patron pour le pronom personnel proprement dit et le sous graphe de la 1^{e} personne :

[14] Nous n'avons pas fait pour l'instant une classification des noms *pluralia tantum*.

BDD-A764 © 2006 Institutul de Filologie Română „A. Philippide"
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-15 01:52:58 UTC)

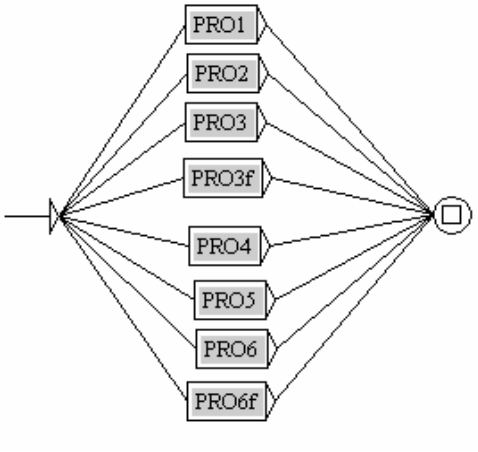


Fig. 12. Graphe PROpers

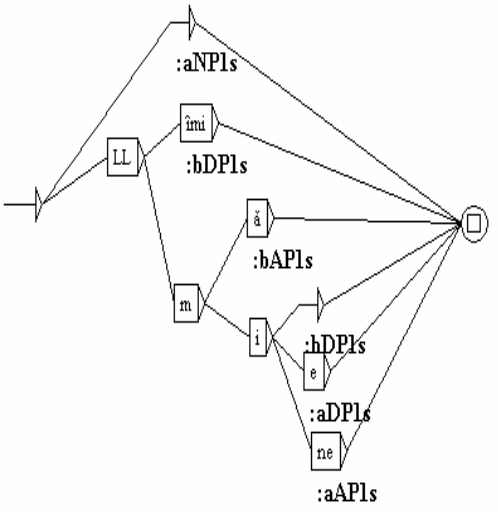


Fig. 13. Sous graphe PRO1

Les formes de datif et d'accusatif se divisent en formes clitiques et non clitiques. Nous avons utilisé les code *a* pour les dernières et *b* pour les premiers. Pour des raisons de notation unitaire, nous avons étendu le codage aussi aux formes de nominatif et de génitif, que nous considérons des formes de type *a*.

3.4. Grammaires des adjectifs

3.4.1. Grammaires de l'adjectif qualificatif

Les graphes de flexion tiennent compte des faits suivants :

- l'absence des formes spécifiques pour le neutre : celui-ci suit au singulier le modèle de flexion du masculin et au pluriel le modèle du féminin ;

- l'existence des formes avec article défini s'opposant aux formes sans article. Même si l'article n'est pas un attribut de l'adjectif, nous considérons nécessaire de donner aussi les formes avec article défini, vu le fait que celui-ci se joint non pas au nom, mais à l'adjectif quand l'adjectif précède le nom :

copilul frumos 'le beau garçon'
[garçon-le beau]
frumosul copil 'le beau garçon'
[beau-le garçon]

- les degrés de comparaison s'expriment analytiquement en roumain. Les rares formes synthétiques (ex : *rarisim*), ainsi que les formes sans degré de comparaison (*viu* 'vif', *mort* 'mort') seront traitées, du point de vue formel comme des formes de positif.

BDD-A764 © 2006 Institutul de Filologie Română „A. Philippide"
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-15 01:52:58 UTC)

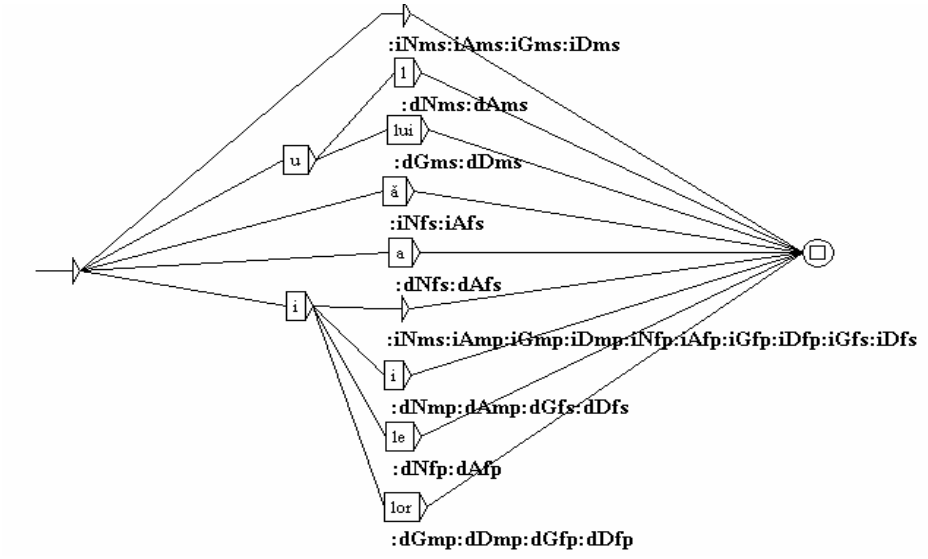


Fig. 14. Exemple de flexion des adjectifs qualificatifs (A9)

3.4.2. Grammaires des pronoms et adjectifs pronominaux

Entre les pronoms sans formes personnelles et les adjectifs correspondants il y a beaucoup de similitudes dans la flexion, allant parfois jusqu'à l'identité des graphes de flexion si les adjectifs pronominaux se placent à la droite du nom. Si, par contre, l'adjectif se place à la gauche, les graphes de flexion présentent seulement des similitudes.

Prenons comme exemple le pronom démonstratif *acesta* 'celui-ci' et les deux adjectifs qui lui correspondent *acesta* (+d) et *acest* (+g). Le pronom et l'adjectif homonyme se fléchiront d'après le modèle illustré dans Fig. 15, tandis que l'adjectif qui se place à la gauche du nom se fléchit comme dans Fig. 16.

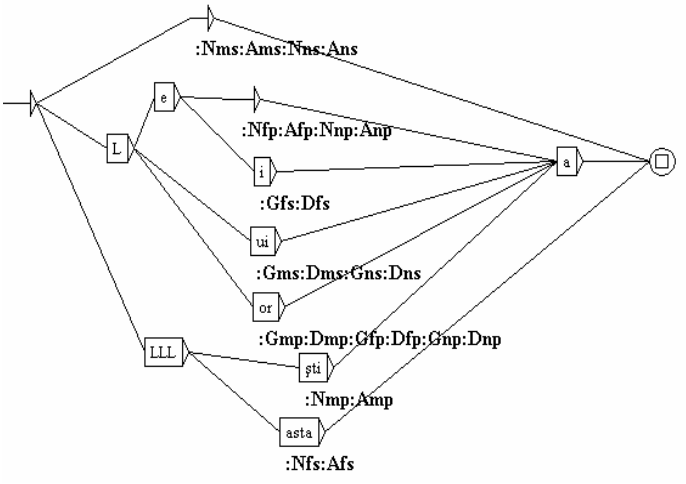


Fig. 15. Graphe de flexion du pronom démonstratif *acesta* et de l'adjectif homonyme

BDD-A764 © 2006 Institutul de Filologie Română „A. Philippide"
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-15 01:52:58 UTC)

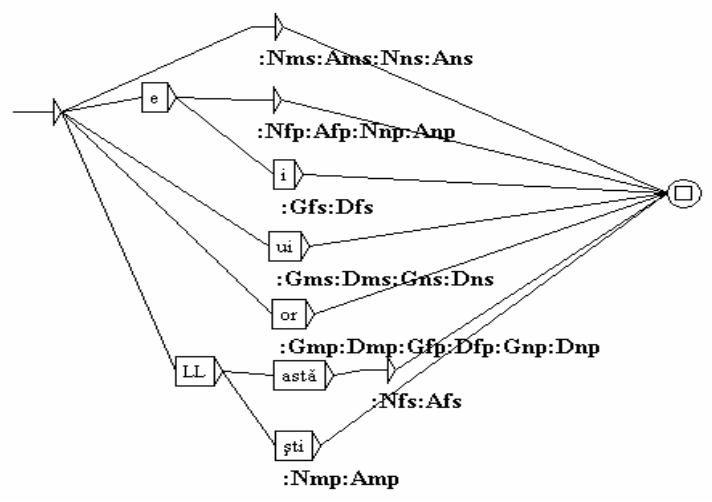


Fig. 16. Graphe de flexion de l'adjectif démonstratif *acest*

Si l'adjectif a la même forme qu'il se place à la gauche ou à la droite du nom (ex : *acelaşi* 'le même'), sa flexion est identique à celle du pronom :

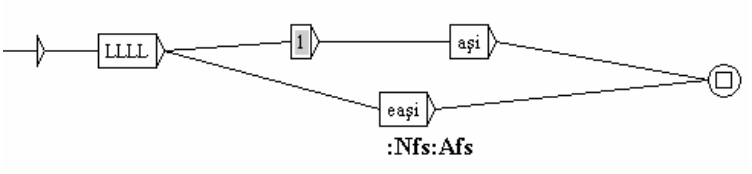


Fig. 17. Graphe de flexion pour le pronom démonstratif *acelaşi* et l'adjectif correspondant

4. Conclusion

La méthode de rédaction des premiers dictionnaires électroniques du roumain en format DELA et les critères de réalisation du corpus respectent le modèle offert par les dictionnaires élaborés sous la direction de Maurice Gross, mais les particularités du roumain (allant du comportement suffixal de l'article défini et jusqu'à la flexion différente de l'adjectif en fonction de sa position par rapport au nom) nous ont révélé des situations inouïes qui ont exigé des solutions qui vont quelque fois en contradiction avec les grammaires descriptives traditionnelles du roumain, mais qui se justifient du point de vue de l'analyse automatique[15].

[15] Nous remercions le directeur du Laboratoire d'Informatique Linguistique de l'Institut Gaspard Monge-CNRS, M. Eric Laporte, pour les conseils qu'il nous a donné lors de la rédaction de ces dictionnaires du roumain.

BDD-A764 © 2006 Institutul de Filologie Română „A. Philippide"
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-15 01:52:58 UTC)

Bibliographie

Courtois, Blandine , *Un système de dictionnaires électroniques pour les mots simples du français*, dans «Langue française» 87, 1990, pp. 11-22.

Courtois, Blandine , *DELAS : Dictionnaire électronique du LADL pour les mots simples du français, version V05*, rapport technique LADL n° 20 ; 1989.

Courtois, Blandine , *DELAS rangé alphabétiquement par parties du discours*, rapport technique LADL n° 19 ; 1988.

Dubois, Jean ; Françoise Dubois-Charlier, *Incompatibilité des dictionnaires*, «Langue française» 87, 1990, pp. 5-10.

Dugas, André, *La création lexicale et les dictionnaires électroniques*, «Langue française» 87, 1990, pp. 23-29.

Dugas, André, *La construction de dictionnaires électroniques*, rapport technique du Groupe de recherche en informatique linguistique, Université du Québec à Montréal ; 1989.

Dumitriu, Dana-Marina**,** *Grammaires de flexion du roumain en format DELA*, Rapport interne 2005-02 de l'Institut Gaspard Monge – CNRS, în http://univ-mlv.fr /LabInfo/rapportsInternes/2005/02.pdf.

Gross, Maurice, *La construction de dictionnaires électroniques* in Compte rendu du Programme de recherches coordonnées : informatique linguistique, LADL, Université Paris 7, 1987, pp. 3-28.

Guillet, Alain, *Reconnaissance des formes verbales avec un dictionnaire minimal*, «Langue française» 87, 1990, pp. 52-58.

Kyriacopoulou, Tita, *Les dictionnaires électroniques : la flexion verbale en grec moderne*, thèse de doctorat en linguistique, Université Paris7, LADL, 1990.

Laporte, Eric, *Le dictionnaire phonémique DELAP*, «Langue française» 87, 1990, pp. 59-70.

Leeman, Danielle ; Serge Meleuc, *Verbes en tables et adjectifs en -able*, «Langue française» 87, 1990, pp. 30-51.

Silberztein, Max, *Dictionnaires électroniques et reconnaissance lexicale automatique*, thèse de doctorat en informatique fondamentale, Université Paris7, LADL, 1989.

Los primeros diccionarios electrónicos del rumano en formato DELA

En este artículo presentamos algunos aspectos de la declinación rumana mediante autómatas, los cuales han sido encontrados durante la redacción de los primeros diccionarios electrónicos del rumano en formato DELA. Después de una introducción a los operadores automáticos abordamos el problema de las particularidades del rumano distinguiendo entre aquellas que afectan al corpus (la composición, la autonomía, la ambigüedad funcional y la posición dentro de un sintagma nominal), las que afectan a la forma canónica y las que afectan a los grafos de flexión (las alternancias vocálicas y consonánticas, la homonimia casual, la ausencia de forma y el comportamiento como sufijo de una cierta parte del discurso).

El artículo continua con consideraciones prácticas sobre las gramáticas de flexión con comentarios y grafos de nombres, artículos, pronombres, adjetivos y adjetivos pronominales.

L'Université Paris 3 Sorbonne Nouvelle/L'Institut Gaspard Monge, Marne-la-Vallée, Cedex 2, France
L'Université de Craiova
Roumanie

BDD-A764 © 2006 Institutul de Filologie Română „A. Philippide"
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-15 01:52:58 UTC)

BDD-A764 © 2006 Institutul de Filologie Română „A. Philippide”
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-15 01:52:58 UTC)