

ANALYSE AUTOMATIQUE: MOT A MOT OU PAR FORMES? ANALYSE  
MORPHOLOGIQUE, SYNTAXIQUE, SEMANTIQUE-AMPHIBOLOGIES

*Busa*

Dans l'analyse automatique des mots, est-ce-que vous aussi vous vous limitez au niveau morphologique ou non?

*Delatte*

Mais peut-être faudrait il dire un mot, que nous n'avons pas dit hier, à propos d'un fait qui marque une très grande différence par rapport à vous: c'est qu'il n'y a pas de mise en ordre alphabétique préalable chez nous dans l'analyse sur l'ordinateur. L'ordinateur analyse en continu, dans l'ordre du texte.

Le programme est ainsi constitué qu'il y a deux façons de procéder, dont la première est purement automatique.

On donne à l'ordinateur des cartes où les mots du texte sont repris à raison de trois mots par carte (avec ponctuation, parce que l'ordinateur référence en même temps qu'il analyse). Une légère prédiction résout le problème des enclitiques: nous les séparons préalablement avant de les mettre sur cartes. *-que*, par exemple, constitue chez nous un mot à part, *-ve* aussi. L'ordinateur sort les analyses à raison d'une carte par analyse possible, et on fait après-coup un listing sur une tabulatrice IBM 447. Le philologue prend ce listing et coche l'analyse correcte. La dactylo, ensuite, élimine les cartes des analyses incorrectes.

La seconde méthode d'analyse comporte un contrôle à la console. L'ordinateur commence d'abord par chercher toutes les analyses possibles, puis il écrit la forme du mot à la machine à écrire et il demande si la forme et la ponctuation sont correctes: en cas d'erreur, une correction à la console serait possible. Puis il tape un astérisque s'il a plusieurs analyses en mémoire. Ensuite, il donne l'analyse à la machine à écrire et il demande à l'opérateur si el-

le est correcte. En cas de réponse négative, il passe à la suivante. Dans ce mode de travail, il y a un enrichissement possible: quand l'ordinateur ne trouve pas le lemme voulu, l'opérateur le lui fournit et l'introduit en même temps dans les lexiques enregistrés dans les mémoires.

Cette manière de faire était séduisante, mais nous avons été obligés de l'abandonner pour des raisons psychologiques. C'est que le philologue qui est à la console fait des complexes par rapport à l'ordinateur. Il a l'impression que l'ordinateur attend et, pour ne pas retarder le travail, il lui arrive de répondre à la légère aux questions que lui pose la machine.

Nous travaillons donc en automatique; l'ordinateur peut sortir 15.000 cartes environ, par jour, alors que chaque philologue de la petite équipe que nous formons n'arrive à vérifier que 1.000 mots (c.a.d. environ 2.500 cartes) par jour.

L'analyse dans l'ordre du texte nous avait paru séduisante parce que nous pensions que notre programme actuel n'était qu'une première étape, la seconde étape étant une analyse du contexte destinée à lever les amphibologies. Jusqu'à présent, nous résolvons le problème par un contrôle philologique fait après coup. Mais si nous pouvons procéder de la sorte, c'est parce que nous ne traitons pas beaucoup de mots.

Si nous avons des textes aussi longs que les vôtres, ce serait à peu près impossible, de sorte que je comprend très bien votre position, qui est de faire du travail pratiquement automatique depuis le début jusqu'à la fin: c'est l'abondance de matière qui conditionne le choix de la méthode.

*Busa*

M. Delatte, quelles sont les raisons pour lesquelles vous n'ordonnez pas les mots en ordre alphabétique avant d'en faire l'analyse?

*Delatte*

C'est inutile: l'ordinateur est capable d'analyser en continu; pour lui il n'y a là aucune difficulté.

Mais quant à l'analyse syntaxique, elle pose un gros problème.

Elle se fait entièrement à la main, et nous ne voyons pas la possibilité de la faire automatiquement. Voulez-vous que nous abordions ce problème?

*Busa*

Supposons qu'une forme de mot soit employée 200 fois dans un texte. Il y a deux possibilités: examiner la concordance où les 200 contextes consé-

cutifs de la même forme sont recueillis et enregistrés l'un après l'autre ou bien sauter de l'un à l'autre dans l'ordre du **texte** c'est-à-dire les examiner de manière discontinue quand on les rencontre, éparpillés dans le texte. Si vous avez les contextes d'un même mot recueillis dans une concordance et si vous en avez tous les emplois avec le contexte de chacun, l'un sous l'autre, est-ce que ça n'aiderait pas aussi l'analyse syntaxique du mot?

*Delatte*

Nous ne croyons pas: qu'appellez-vous un contexte semblable? Est-ce exactement le même contexte, ou des contextes à peu près semblables?

*Busa*

Non, ce sont les différents contextes du même mot.

*Delatte*

Mais rien ne prouve jamais que ce soit la même forme.

*Busa*

La même forme graphique, oui.

*Delatte*

La même forme graphique, certes, mais qui n'a pas nécessairement la même analyse.

*Delatte*

Le programme d'analyse morphologique tel qu'il vous a été décrit est loin d'être parfait puisqu'il ne résout aucune amphibologie ainsi qu'en témoignent les listings que vous avez sous les yeux.

Ceci est évidemment une grande faiblesse et provoque un ralentissement considérable du travail car le philologue est obligé de vérifier artisanalement toutes les analyses données par la machine.

Conscients du retard apporté par la paradoxale nécessité de refaire le travail de l'ordinateur, nous nous sommes demandé s'il n'y avait pas moyen d'entrer plus profondément dans l'automatisation de l'analyse et d'obtenir par des critères purement formels que l'ordinateur lui-même, seul, résolve certains types d'ambiguïtés. Ainsi il nous paraissait évident que le problème de CUM, préposition ou conjonction devait pouvoir être réglé assez facilement. Nous pensions, en effet, que CUM préposition est suivi d'un ablatif

(à un, deux, trois, quatre mots après lui) et qu'on ne pouvait trouver un CUM conjonction suivi immédiatement d'un ablatif parce que, dans ce cas, l'amphibologie pouvait exister aussi pour un lecteur latin, du moins à certaines époques. Mais une phrase du type suivant a réduit nos illusions à néant: *in hoc omnes errore versamur ut non putemus ad mortem nisi senes inclinatoseque jam vergere, CUM ILLO infantia statim et iuventa omnisque aetas ferat* (Cons. ad Marc., XXI, 7).

Dans cette phrase, en effet, *cum* est conjonction et il est suivi d'une forme que l'on pourrait prendre pour un ablatif.

En réalité, tous les essais effectués montrent que le latin est une langue tellement dépourvue de structures fixes, tellement désarticulée (puisque l'ordre des mots, sans être indifférent, y est libre), qu'il nous a été impossible jusqu'à présent de découvrir des points de repère, des pivots, grâce auxquels on aurait pu apprendre à l'ordinateur quelques règles qui lui auraient permis de reconnaître la fonction de chaque mot dans la phrase.

M. Evrard vous montrera tout à l'heure ces impossibilités formelles par quelques exemples.

### *Evrard*

En désespoir de cause, nous nous sommes demandé si nous ne pourrions pas nous limiter à tenter de distinguer les verbes de propositions subordonnées des verbes de propositions principales et de perforer sur la carte des verbes subordonnés, les codes des modes de subordination. Nous sommes partis de l'idée que seuls les verbes à un mode personnel ne comportent aucune amphibologie et qu'à ce titre, ils pourraient servir de points de repère dans la phrase. Nous avons alors établi trois règles qui, dans notre esprit, pouvaient résoudre certains problèmes.

1 - Le premier verbe, après un point, est un verbe de proposition principale s'il n'y a pas de conjonction intercalée entre le point et le verbe.

2 - Le verbe qui suit immédiatement une conjonction est le verbe commandé par cette conjonction.

3 - Le verbe qui suit un verbe commandé par une conjonction est un verbe de proposition principale:

- a) si l'on trouve entre lui et le verbe de la subordonnée ni conjonction, ni relatif, ni aucun autre mode de subordination;
- b) si devant le verbe de la subordonnée ne se trouvent pas deux conjonctions qui se suivent sans verbe intercalaire à un mode personnel.

Ces conditions nous paraissaient assez strictes pour espérer aboutir à certains résultats et ne souffrir aucune exception. Hélas, l'examen de quelques phrases d'une Lettre à Lucilius nous a convaincus que ces règles ne servaient à rien, à l'exception de la première, mais qui couvre si peu de cas que cela ne nous a pas paru valoir la peine de faire un programme.

M. Bodson vous montrera, par quelques exemples, l'impossibilité d'aller très avant dans ce domaine.

### *Bodson*

Les trois textes que je vais utiliser ne sont pas particulièrement compliqués. Nous les avons choisis à peu près au hasard. Je voudrais d'abord vous montrer que, tant qu'on n'a pas levé les amphibologies posées par la plupart des mots, il est pratiquement impossible de déterminer la structure logique de la phrase. J'utiliserai à cette fin le texte de César que voici. (*B.G.*, II, 25)

Ubi . . . rem esse in angusto vidit neque ullum esse subsidium , quod submitti posset, scuto ab novissimis uni militi detracto, quod ipse eo sine scuto venerat, in primam aciem processit centurionibusque nominatim appellatis reliquos cohortatus milites signa inferre et manipulos laxare iussit, quo facilius gladiis uti possent.

La phrase est extraite de l'épisode de la bataille du fleuve Sabin. J'en ai supprimé une bonne partie: les points de suspension qui suivent *ubi* représentent, en fait, une dizaine de lignes. Voici d'abord une traduction:

"Dès que. . . César vit que la situation était critique, et qu'il n'y avait aucune troupe de secours qui pût être envoyée, ayant arraché son bouclier à un soldat du dernier rang, parce que lui-même était venu là sans bouclier, il s'avança jusqu'à la première ligne, et, ayant appelé les centurions par leur nom et ayant exhorté les autres soldats, il donna l'ordre de charger et d'espacer les manipules, pour qu'ils pussent plus facilement se servir de leurs glaives."

Je relève les mots qui seraient utiles pour déterminer la structure logique de cette phrase en supposant qu'aucune possibilité n'a encore été exclue. L'ordinateur devra tenir compte de *ubi*, qui est une conjonction possible. Il devrait retenir aussi *neque*, qui est une conjonction de coordination. Mais on ne peut rien fonder sur les conjonctions de coordination, car elles peuvent coordonner des compléments et rien ne permet de distinguer formellement les cas où elles coordonnent des verbes de propositions.

Après *ubi vidit*, qui est clair, nous trouvons *quod*, puis *posset*, qui en dépend. Je ne tiens pas compte de *submitti*, ni des participes, parce que ces formes impersonnelles n'influent pas sur la structure.

A la deuxième ligne, le subordonnant *quod* pourrait commander *eo* qui, pour l'ordinateur, est un verbe possible. *Sine*, lui aussi pourrait être un verbe (impératif de *sinere*). Ensuite, *venerat* et *processit* ne posent pas de problème, mais *appellatis* pourrait être un indicatif (vous appelez) et *cohortatus* pourrait être l'équivalent de *cohortatus est*. A la fin de la phrase, nous relevons *iussit, quo, uti* (qui peut être une conjonction) et *possent*. Il est évident, je crois, que l'ordinateur ne retrouvera jamais la structure de cette phrase.

Si nous reprenons les règles énoncées précédemment par M. Delatte, nous pouvons dire que *vidit* est introduit par *ubi* (1ère règle) et que *quod* introduit *posset* (2e règle). Mais, à partir du second *quod*, il n'y a plus aucune possibilité d'analyse, principalement en raison de l'amphibologie de *eo*, de *sine* et de *appellatis*.

On pourrait songer à faciliter la tâche de la machine, en levant, préalablement à l'analyse logique de la phrase, toutes les ambiguïtés; c'est-à-dire en précisant par exemple, que *uti* est ici un infinitif, que *sine* est une préposition, etc.

Mais une telle méthode ne résout pas non plus le problème. Je voudrais le montrer au moyen d'un texte de Sénèque (*Ep.*, 77, 1-2).

Omnis in pilis Puteolorum turba consistit et ex ipso genere velorum Alexandrinas quamvis in magna turba navium intellegit: solis enim licet siparum intendere, quod in alto omnes habent naves. Nulla enim res aeque adiuvat cursum quam summa pars veli; illinc maxime navis urgetur. Itaque quotiens ventus increbruit maiorque est quam expedit, antemna submittitur minus habet virium flatus ex humili. Cum intravere Capreas et promunturium ex quo "alta procelloso speculatur vertice Pallas", ceterae velo iubentur esse contentae; siparum Alexandrinarum insigne indicium est.

En voici la traduction.

"Toute la population de Pouzzoles s'est rassemblée sur le môle, et au type de voilure, reconnaît les navires venus d'Alexandrie, bien que ce soit dans une grande foule de bateaux. Eux seuls, en effet, peuvent hisser "le bourslet de lune", que tous les navires portent en haute mer. Rien en effet n'aide autant la marche du navire que la partie supérieure de la voilure. C'est par là que le navire est le plus poussé. C'est pourquoi chaque fois que le vent s'amplifie et devient plus fort qu'il ne convient, on abaisse la vergue. Le souffle qui vient d'en bas a moins de force. Lorsque les bateaux ont franchi la passe entre Capri et le promontoire du haut duquel Pallas regarde la mer, les autres navires doivent se contenter de la grande voile. Le bourslet de lune

est le signe particulier des navires d'Alexandrie". Imaginons qu'ici il n'y ait plus aucune ambiguïté concernant l'analyse des mots. Nous aurons bien séparé et bien isolé les verbes à un mode personnel et les conjonctions de subordination. Le premier verbe est "consistit". Comme aucune conjonction ne le précède, il est un verbe principal (1ère règle). Mais, comme *quamvis* est une conjonction de subordination et que le premier verbe qui la suit est *intellegit*, la 2e règle nous conduit à croire que *quamvis* introduit *intellegit*, ce qui est faux.

Dans la suite, il n'y a de problème ni pour *licet*, ni pour *quo habent*, ni pour *adiuuat*. Mais les difficultés recommencent à *cursum quam*: l'application de nos règles nous amènerait en effet à croire que le verbe introduit par *quam* est *urguetur*, ce qui est une deuxième erreur. Si nous avançons dans le texte, nous constaterons que la machine, à partir des règles énoncées plus haut par M. Delatte, reconnaît correctement les groupes *quotiens increbruit* et *quam expedit*, mais, qu'elle n'aurait aucune raison de rattacher *est* à *quotiens*. Et je pourrais ainsi multiplier les exemples.

On peut aller plus loin encore et désigner les verbes principaux.

Dans ce cas, la machine n'a plus qu'à déterminer les rapports entre les verbes subordonnés et les mots subordonnants. Là aussi on court également à un échec et un dernier texte va le montrer. Le voici.

Caesar necessariis rebus imperatis ad cohortandos milites, quam in partem fors obtulit, decucurrit et ad legionem decimam devenit. Milites non longiore oratione cohortatus, quam uti suae pristinae virtutis memoriam retinerent neu perturbarentur animo hostiumque impetum fortiter sustinerent, quod non longius hostes aberant quam quo telum adigi posset, proelii committendi signum dedit. (CESAR, *B.G.*, II, 9)

J'en donne rapidement la traduction.

"César, après avoir donné les ordres nécessaires descendit en courant, pour exhorter les soldats, dans le secteur où le hasard le porta, et il arriva auprès de la dixième légion. Après avoir exhorté les soldats, sans faire de longueurs excessives, à garder le souvenir de leur courage passé, à ne pas se laisser impressionner et à supporter avec courage l'assaut ennemi, comme les ennemis n'étaient plus qu'à une portée de javelot, il donna le signal d'engager le combat".

La première phrase n'offre pas de difficultés. Si l'ordinateur sait que *decucurrit* et *devenit* en sont les verbes principaux, il rattachera correctement *obtulit* à *quam*. Mais il n'en va pas de même pour la seconde phrase. Averti que *dedit* est le verbe principal, l'ordinateur, devant la série *quam uti*

*retinerent neu perturbarentur sustinerent*, va, s'il applique nos trois règles, rattacher correctement *retinerent* à *uti* et *perturbarentur* à *neu*, ce qui est correct, mais il continuera en rattachant *sustinerent* à *quam*, ce qui est une erreur. En continuant de la même manière, l'ordinateur verra bien que *aberant* dépend de *quod* et *posset* de *quo*. Mais, comme verbe introduit par *quam*, il ne trouvera que *dedit*, qui lui est pourtant désigné comme verbe principal.

Conclusion: même si l'on désigne à la machine les verbes principaux, c'est-à-dire en faisant pratiquement tout le travail, il est impossible d'automatiser la recherche des rapports de subordination. C'est assez désespérant.

#### *Delatte*

Il n'y a pas la moindre structure fixe à laquelle on puisse se raccrocher et ce qui est vrai dans le domaine de la syntaxe est aussi vrai dans le domaine de la morphologie, M. Evrard a préparé quelques exemples qu'il va vous donner à cet égard-là.

#### *Evrard*

On pourrait se demander si, au moment où l'ordinateur recherche ses analyses, il n'y aurait pas moyen d'éliminer certaines possibilités en mettant en programme des règles relatives aux petits groupes de mots, ou bien des règles relatives aux compatibilités des différentes analyses dans une même phrase ou dans un même groupe de mots.

Cela paraît assez séduisant, mais à l'expérience il faut bien se rendre compte que l'ordinateur ne pourrait pas non plus aller très loin dans ce domaine. M. Delatte nous parlait tantôt de *cum*.

Précisément à propos de cette préposition, il y a des expressions qui, à première vue, ne prêtent à aucune espèce de confusion. Ce sont les expressions du type *mecum*, *tecum*, etc., où il semble certain a priori que le *cum* est préposition et que la forme pronominale *me*, *te* etc. est un ablatif régi par cette préposition. Mais en fait, pour l'ordinateur, les deux mots de ces expressions sont analysés indépendamment l'un de l'autre, et, comme M. Delatte le rappelait tantôt, dans notre système, les deux mots, à la suite d'une légère prédiction, sont enregistrés séparément. D'ailleurs il faut admettre que la graphie continue qui est dans nos habitudes ne signifie guère plus que les ponctuations faibles, mises par les éditeurs pour faciliter la compréhension du texte ou pour marquer l'interprétation qu'ils en donnent. Dès lors le problème est de savoir si, dans tous les cas où *cum* suit une forme de la série *me*,



*te*, etc., il est préposition et régit cette forme pronominale. En fait, *cum* ainsi placé peut être ou une conjonction ou une préposition régissant un autre mot, et la forme pronominale peut être un accusatif tout autant qu'un ablatif. C'est d'ailleurs là ce qui rend possible, par exemple, l'interprétation que certains commentateurs, tels Heinsius, donnent de Virgile, *Buc.* 2, 12 où ils lisent non pas "*at mecum raucis tua dum vestigia lustris . . . Sole sub ardenti resonant arbusta cicadis*" mais bien *At me* (acc.) *cum raucis . . . resonant cicadis*.

Restons encore un moment dans le domaine des prépositions.

La position d'un verbe à un mode personnel placé entre une préposition et son régime, semble à première vue rare et exceptionnelle, au point que l'on serait tenté de la croire impossible. Par ailleurs, le régime de la préposition peut être sous-entendu et, dans ce cas, il est habituellement suggéré par une épithète ou par un génitif qui s'y rapporte. Dès lors, dans les cas où un verbe personnel survient avant que l'on ait détecté le régime d'une préposition, on pourrait voir dans cette circonstance un indice que le régime est sous-entendu, ou, s'il s'agissait d'un mot qui, tel *super*, peut être soit préposition, soit adverbe, on pourrait y voir l'indice qu'il est adverbe plutôt que préposition. Or on trouve par exemple dans Sénèque, *Cons. ad Marc.*, 26, 3, le bout de phrase suivant: "*In alieni percussoris venit arbitrium*". Dans ce passage, *arbitrium*, régime introduit par *in*, se trouve en quatrième place après la préposition et il en est séparé par un verbe à un mode personnel, dont, formellement, il pourrait être le sujet. La règle suggérée plus haut, conduirait donc ici à une erreur.

Je voudrais maintenant vous proposer une brève phrase, tirée elle aussi de la *Consolation à Marcia*. Sénèque, en un endroit, rappelle quelques femmes célèbres pour leur courage, lors de la mort de leurs enfants. Il l'annonce en ces termes:

"*Quod si tibi vis exempla referrī feminarum quae suos fortiter desideraverint*".

Après cette introduction, l'auditeur ou le lecteur s'attend à trouver une série d'exemples de femmes qui se sont montrées courageuses au moment où leurs enfants mouraient; et effectivement, le texte continue par l'exemple de Cornélie, la mère des Gracques, et par celui de la mère de Drusus.

A propos de la première, on lit la toute petite phrase que voici: "*Duodecim illa partus totidem funeribus recognovit*" (Sén., *Cons. ad Marc.*, 16, 3).

La situation, à première vue, est favorable, puisque, des six mots de cette phrase, trois ne permettent qu'une seule analyse: ce sont *duodecim, toti-*

*dem* et *recognovit*. De plus, parmi ces trois mots se trouve le verbe *recognovit*. Mais il convient de regarder les choses de plus près. Pour le lecteur humain, qui situe le passage grâce aux indications que je viens de rappeler, et qui par conséquent sait qu'il est question de femmes, d'enfants, de morts et de cérémonies funèbres, bref, d'un domaine de significations bien défini, le mot *partus* forme naturellement couple avec *funeribus*, tandis que *illa* rappelle évidemment Cornélie, dont il a été question une ou deux lignes plus haut.

Quand on choisit entre le datif et l'ablatif pour *funeribus*, c'est le sens, ici encore, qui nous permet de le faire.

Essayons maintenant de nous mettre dans la situation de l'ordinateur. Voici ce que ce dernier nous propose au moyen du programme d'analyse morphologique.

Pour *duodecim*, il n'a qu'une seule analyse, Pour *illa*, il va proposer une série d'analyses relatives au pronom *ille* (accusatif pluriel neutre, nominatif pluriel neutre, ablatif singulier féminin et nominatif féminin singulier), puis il ajoutera l'adverbe de lieu *illa*. Pour *partus* il nous offre sept analyses: d'abord les formes du substantif *partus*, *partus*, puis le participe passé de *pario*. Pour *totidem*, il donne une seule analyse, pour *funeribus*, deux analyses (datif ou ablatif), et pour *recognovit* une seule analyse.

Peut-on demander à l'ordinateur d'opérer un choix entre les suggestions qu'il nous offre? Il ne nous paraît guère possible, du moins à l'heure actuelle, de le faire travailler sur les significations. Ici, en effet, il ne s'agit pas seulement d'accoler un sens à un mot, mais de choisir parmi les diverses possibilités résultant à la fois de la polysémie et des homographies, et d'opérer ce choix en fonction du sens général du passage, tel qu'il apparaît dans ce qui précède la phrase qu'on est occupé à lire. On pourrait penser à des règles concernant les petits groupes de mots, ou le type de construction attendu après tel mot, etc. Mais tout cela est insuffisant, parce que l'application n'en est jamais régulière au point de ne pas supporter d'exceptions, et parce que cela laisse subsister bien des ambiguïtés. Les choses seraient différentes, si nous pouvions accepter un certain pourcentage d'erreurs, disons par exemple 5 ou 10% d'erreurs. Mais une telle latitude, admissible dans les travaux de caractère pratique, telle la traduction automatique, est absolument contraire à l'esprit de recherches désintéressées, comme celles de la philologie. Ceci dit, revenons à notre phrase.

*Illa*, dont nous avons vu tantôt qu'il reprenait Cornélie, pourrait être pris comme un accusatif pluriel neutre. Il deviendrait alors l'objet direct. Quant à *partus*, on ne voit pas quelle règle formelle interdirait de l'analyser

comme un participe apposé au sujet de *recognovit*. Par ailleurs, ce participe pourrait avoir pour complément d'origine l'ablatif féminin *illa*. C'est là une construction rare, mais qu'on ne peut pas rejeter en vertu d'une règle stricte.

C'est donc là une troisième analyse formellement possible pour *illa*.

Sans doute, dans ce cas, le verbe *recognovit* resterait-il sans objet direct. Mais je n'imagine pas très bien le moyen de déterminer de manière stricte les conditions dans lesquelles l'objet direct d'un verbe transitif doit être exprimé ou peut être sous-entendu. Même l'analyse de *partus* comme génitif singulier pourrait apparaître comme acceptable, ce mot étant alors complément d'un *illa*, qui pourrait être considéré comme un accusatif neutre. Quant à *funeribus*, l'embarras reste entier aussi longtemps qu'on ne fait pas intervenir le sens. On sait bien que le datif est rare, mais on ne voit pas ici non plus quelles conditions formelles autoriseraient à exclure un datif, celui-ci étant presque toujours possible, ne fût-ce que comme complément d'avantage, par exemple.

Sans doute estimerait-on qu'il serait plus facile d'écarter les analyses de nominatif et de vocatif pluriel, en raison du fait que le seul verbe est au singulier et à la troisième personne; mais, pour les écarter en toute sécurité, nous devrions être sûr qu'il n'y a aucune ellipse, aucun verbe sous-entendu.

Dans la phrase que je vous ai soumise, la chose paraît assurée; mais la présence d'un mot qui parmi ses analyses possibles, aurait celle des conjonctions de coordination ou de subordination (par exemple un *et*), suffirait à faire naître un doute. J'ajoute que l'on peut toujours se demander dans quelle mesure il ne pourrait pas y avoir une ponctuation, par exemple un point et virgule après *duodecim illa*. Ainsi donc, même si l'on s'efforce de faire intervenir des règles formelles de compatibilité ou de formation de petits groupes de mots, il ne semble guère possible d'arriver à des résultats intéressants à moins que de se référer à la signification globale de la phrase et même du passage. Et à cela, je ne crois pas qu'on y arrivera de sitôt avec l'ordinateur.

#### *Delatte*

Nous avons donné quelques cas qui nous paraissent particulièrement flagrants, où la machine ne pouvait absolument pas lever les amphibologies ni dans le domaine de la syntaxe, ni dans le domaine de la simple morphologie, étant entendu que l'ordre des mots est tellement libre que rien ne se raccroche à rien, sinon par le sens.

.....

#### *Busa*

Peut-être pourrait-on conclure sur ces coïncidences et différences de

*Busa*

Peut-être pourrait-on conclure sur ces coïncidences et différences de méthode entre les deux dictionnaires.

Le LEL de Liège est inclus dans un cycle de recherche qui aboutit à l'analyse syntaxique et sémantique complète, tandis que notre dictionnaire se borne à une première analyse morphologique, grâce à laquelle d'autres pourront approfondir l'analyse morphologique et aussi l'analyse syntaxique. Ça, c'est une première différence. Une deuxième différence c'est que pour vous la flexion automatique des mots est un moyen pour l'analyse: tandis que pour nous, du moins, elle a seulement pour but de corriger notre dictionnaire de machine.

Dans les prochains mois nous allons chercher les erreurs. Une des méthodes serait de flexionner entièrement à la machine tous les lemmes qui se trouvent dans notre dictionnaire, et après de comparer toutes ces formes pour ainsi dire artificielles avec les codes morphologiques et avec le développement de l'homographie que nous avons déjà assigné aux formes pour ainsi dire naturelles. Si on songeait à fondre les deux dictionnaires, on pourrait dire que les matériaux de notre dictionnaire sont prêts pour être introduits dans un dictionnaire qui doit servir à l'analyse automatique, n'est-ce-pas? Pour ça, on devrait examiner d'une part les différences dans les codes: pas exactement les différences de codage, qu'il est bien facile de transférer; mais les différences de catégorie de codes. Il faudra voir par ailleurs si le découpage entre constantes et désinences des lemmes suit des règles correspondantes ou non.