

POUR UNE RECONSTRUCTION ALGORITHMIQUE DES GRAMMAIRES : RÉCIT D'UNE EXPÉRIENCE SUR L'ALLEMAND

Raymond SAVARY

*"Les bouliers se perfectionnent
Les formulations succèdent aux formulations,
Les algorithmes demeurent".*

Propos apocryphes attribués à Al-Khawarizmi
(éponyme de l'algorithme).

L'application de l'informatique à la linguistique peut se proposer plusieurs démarches :

- 1) la reconnaissance des formes : c'est un processus d'*analyse* ;
- 2) la génération des formes : c'est le processus inverse, la *synthèse* ;
- 3) la correction des erreurs dans un texte inconnu au départ ;
- 4) la traduction assistée par ordinateur ;
- 5) la réponse à des questions.

Les trois dernières démarches sont des mises en oeuvre complexes des deux premières.

Dans l'expérience inédite que nous avons menée avec les étudiants d'allemand de Reims (non formés à l'informatique), nous avons choisi la démarche qui nous paraissait la plus facile et ce, dans un secteur de la grammaire que les étudiants de ce niveau considèrent comme archi-connu et qui est pourtant rarement entièrement maîtrisé : la *déclinaison du groupe nominal*. Cette démarche a été dans un premier temps restreinte à un problème très particulier : la *génération des terminaisons de l'adjectif épithète* à partir de la forme de *l'article défini*. La discussion a montré que l'ordinateur avait rapidement besoin, dans beaucoup de cas, d'informations plus complètes pour aboutir à une génération correcte. Cette découverte a été très féconde car elle faisait prendre

LE BULLETIN DE L'EPI N° 49 RECONSTRUCTION ALGORITHMIQUE DES GRAMMAIRES

conscience aux étudiants des mécanismes complexes qui régissent les formes qu'ils considéraient jusqu'à présent comme existant en soi.

	Masculin	Neutre	Féminin	Pluriel des 3 genres
Nominatif	der - e	das - e	die - e	die - en
Accusatif	den - en	das - e	die - e	die - en
Datif	dem - en	dem - en	der - en	den - en
Génitif	des - en	des - en	der - en	der - en

Le problème posé a paru tout d'abord presque enfantin puisque, dans bon nombre d'occurrences, la simple considération de la forme de l'article entraîne automatiquement le choix correct de la désinence de l'adjectif : les étudiants ont tout de suite découvert que c'était le cas pour dem, den, des entraînant toujours -en¹, ainsi que pour das, toujours suivis de -e. Cette première constatation prouvait que, dans ces contextes, l'ordinateur pouvait à coup sûr et sans aucune autre information, prévoir et générer la désinence de l'épithète et ceci par une méthode non-intelligente, fort différente de celle qu'ils avaient apprise dans les classes d'allemand.

Les autres formes de l'article défini (der et die) plaçaient l'ordinateur devant un choix fort restreint puisqu'il ne lui était offert que l'alternance entre -e et -en. Ayant réfléchi préalablement à ces problèmes, je proposai de commencer par die pour ne pas décourager mes néophytes en informatique. Le choix offert par die se résout en effet par un simple problème de nombre : si le substantif est au singulier, l'adjectif se termine par -e ; dans le cas où le substantif est au pluriel, la terminaison est -en.

L'affaire paraissait vite classée : il suffisait (!) de fournir à l'ordinateur la liste des formes des substantifs singuliers et/ou celle des substantifs pluriels. Mais quelques étudiants se rappelèrent qu'un certain nombre de substantifs avaient au pluriel une forme identique à celle du singulier. L'ordinateur devait-il alors déclarer forfait ? L'instant de découragement passé, quelques-uns se servirent d'un raisonnement logique très simple pour sortir de l'impasse : il y a en effet bon nombre de substantifs allemands invariables au pluriel mais ces substantifs sont du genre *masculin* ou *neutre*. Par chance, *aucun* féminin ne conserve la même forme au pluriel. Or die ne convient aux trois genres que s'il représente l'article pluriel. Si, en revanche, die est devant un nom au

singulier il ne peut s'agir que d'un substantif *féminin*. Ce simple raisonnement dictait la démarche à indiquer à l'ordinateur : comparer la forme du substantif du texte avec celles trouvées dans la liste des *féminins singuliers* : si la forme se trouvait dans la liste, l'adjectif devait prendre la terminaison -e, dans tous les autres cas, il ajoutait -en. Ouf !

Restait le dernier carré de résistance : celui constitué par l'article der. Cette fois encore, il s'agissait de choisir entre -e et -en, mais la démarche paraissait beaucoup plus complexe dans la mesure où elle semblait faire intervenir à la fois trois paramètres : le genre, le nombre et le cas puisque la désinence -e caractérise le nominatif masculin singulier, la terminaison -en apparaissant dans toutes les autres occurrences. Les étudiants se rendirent vite compte que, contrairement à un élève, l'ordinateur aurait bien de la peine à déduire du texte que le substantif étudié se trouvait au *nominatif*. Par contre, en se servant de la même démarche que pour die, on pouvait demander à la machine de rechercher s'il était dans la liste des *masculins singuliers*. Mais alors, l'obstacle qu'un petit raisonnement logique avait pu écarter dans l'étude de die revint à l'esprit : un certain nombre de masculins ont la même forme au singulier et au pluriel.

Plusieurs modes d'emploi furent proposés alors, mais je rappelai qu'il fallait retenir le procédé le plus *économique* en temps-machine : demander à l'ordinateur si le substantif considéré se trouve dans la "liste WAGEN", liste qui regroupe les masculins invariables au pluriel. (En fait, aucune grammaire de l'allemand (aussi bien en allemand qu'en français) ne donne cette liste (seuls, Helbig et Buscha fournissent une cinquantaine de termes). On y trouve seulement des indications complexes (et négatives) sur les masculins terminés en -el, en, -er qui n'en font *pas* partie : nouvelle preuve qu'il faudra réécrire les grammaires à l'usage des ordinateurs). Si le substantif étudié ne figure pas dans cette liste (ce qui est le cas le plus probable)², il suffit de consulter la liste des masculins singuliers variables. Si le substantif s'y trouve, la désinence est -e, sinon, c'est -en puisqu'il s'agira d'un masculin pluriel ou d'un pluriel d'un autre genre.

Le seul problème épineux est donc constitué par le petit groupe de substantifs masculins qui ont la même forme au nominatif singulier et au pluriel (liste WAGEN). Il semble donc que l'on soit obligé de doter la machine de capacités bien plus subtiles comme la délimitation des "propositions"³. En supposant ce problème résolu, on dispose d'un indice formel trahissant l'existence d'un groupe nominal au singulier, c'est la

forme du verbe conjugué qui, à la troisième personne des verbes allemands, est toujours différente au singulier et au pluriel. Si c'est une forme du singulier, il y a fort à parier pour que le substantif lui-même soit au singulier (la désinence sera alors -e, sinon -en) mais cela n'est vrai à coup sûr que dans le cas où ce substantif constitue le seul groupe nominal de la proposition. (Il faut donc que la machine soit capable d'isoler et de compter les groupes nominaux d'une "proposition"). Dans les propositions contenant plusieurs groupes nominaux, l'ordinateur est bien incapable d'identifier si l'article der introduit un nominatif mais certains indices formels décelables dans le contexte peuvent le conduire plus facilement à conclure qu'il s'agit d'un *génitif*. Les étudiants proposent la présence d'un autre groupe nominal, ou d'une préposition du groupe während ou d'un verbe du type bedürfen, en général à gauche du substantif considéré. Si l'un de ces critères est présent, la désinence sera très probablement -en, sinon -e⁴. Nous insistons sur le "très probablement". En effet, on peut imaginer des propositions hors-contexte pour lesquelles l'ordinateur pourra donner une désinence fautive en appliquant les critères énoncés. Ce serait le cas pour un exemple, assez peu probable il est vrai, comme :

Der ernst- Techniker bedarf der alt- Kellner.

Cette phrase pourrait signifier : "Le technicien sérieux a besoin des vieux garçons de café" et, exceptionnellement dans un contexte approprié, "Le vieux garçon de café a besoin des techniciens sérieux". Mais on peut remarquer que, mis en présence de cet exemple isolé, l'être humain demeurerait aussi perplexe que l'ordinateur.

Cette incertitude fournit l'occasion de montrer concrètement aux étudiants l'opposition fondamentale de deux démarches informatiques. Autant pour l'article die, l'ordinateur pouvait suivre une démarche *algorithmique*, c'est-à-dire prédire à coup sûr, à partir des informations dont il disposait, la désinence de l'épithète, autant, pour l'article der, les critères deviennent plus lâches et font appel, comme la plupart des comportements humains, à des probabilités statistiques plus ou moins intuitives, donc à une démarche *heuristique*, c'est-à-dire qui implique des tâtonnements et des possibilités d'erreurs.

Le récit simplifié de cette expérience montre comment la démarche informatique peut féconder et renouveler la présentation des grammaires. Nul besoin pour les apprenants de connaître le fonctionnement pratique des ordinateurs ni tel ou tel "langage de programmation" (Basic, Pascal, LSE, Lisp...) qu'il faut souvent

réapprendre dès qu'on change de machine. Sans quitter leur spécialité, les apprenants sont introduits d'emblée au coeur des mécanismes logiques et formels qui constituent l'essence de l'informatique.

Nous avons choisi un exemple tiré de la grammaire de l'allemand mais, pour nous, il va de soi que beaucoup de chapitres de la grammaire de *chaque* langue peuvent se prêter à une reconstitution analogue à celle que nous avons tentée. Pour en rester au groupe nominal, nous songeons à l'emploi de l' en français, de l'article an en anglais⁵.

Nous n'avons pas encore tenté de traduire nos algorithmes en "programmes" tournant sur les ordinateurs dont nous disposons à l'U.F.R. de Lettres et Sciences Humaines. Ce serait naturellement le moyen idéal de tester leur adéquation et d'aviver la curiosité des étudiants. Les instructions seraient assez simples (sauf pour le découpage en "propositions" et en groupes nominaux) mais exigeraient la constitution de listes impressionnantes de substantifs et de formes verbales. On pourrait se limiter, dans un premier temps, à un échantillon formé d'un petit groupe de substantifs et de verbes appartenant à divers types.

CONCLUSIONS

Cette expérience s'est révélée très enrichissante pour l'enseignant et pour certains étudiants, mais non pour tous. En effet, tandis que les esprits plus logiques se piquaient au jeu de la découverte progressive de règles fondées sur des critères formels, d'autres, plus attachés à la présentation traditionnelle de la grammaire, ont objecté que, pour des "littéraires", c'était une gymnastique torturante et peu utile de "se mettre à la place de l'ordinateur".

Pour tous, cette démarche inhabituelle a été l'occasion de voir sous un jour nouveau des faits notoirement connus comme les formes de pluriel des substantifs. Les étudiants ont pris conscience progressivement des difficultés qui peuvent se présenter dans l'application d'une règle en se rendant compte du nombre toujours plus élevé d'informations qu'il faut fournir à l'ordinateur. Qui plus est, ils ont découvert que ces informations ne figuraient pas toujours dans les grammaires actuelles, mêmes les plus complètes, et qu'une "linguistique informatique" était encore à construire.

Il va de soi qu'une telle *présentation* ne s'adresse pas aux débutants dans une langue car elle nécessite une connaissance synthétique des mécanismes de la langue considérée. Cependant, on peut se demander si les *résultats* de cette démarche ne pourraient pas remettre en question certains aspects de la *pédagogie* traditionnelle. Par exemple, pour en rester au domaine que nous avons abordé, on pourrait proposer, dans l'apprentissage des cas de l'allemand et de la déclinaison des groupes nominaux, un appel plus systématique à des critères formels comme ceux que nous avons établis et qui sont sans doute plus faciles à faire passer au statut d'habitude, passage nécessaire à une expression rapide dans la langue considérée.

Raymond SAVARY
Université de Reims

Ouvrages cités

Der Duden in 10 Bänden.

Band 4 :

DIE GRAMMATIK, Bibliographisches Institut, Mannheim, 1984.

EGGERS Hans (et collaborateurs) :

ELEKTRONISCHE SYNTAXANALYSE DER DEUTSCHEN GEGENWARTSSPRACHE, Max Niemeyer Verlag, Tübingen, 1969.

HELBIG Gerhard, BUSCHA Joachim :

DEUTSCHE GRAMMATIK, UEB, Verlag Enzyklopädie, Leipzig, 1972.

Notes

1 Nous avons laissé de côté - pour mettre une autre investigation - le problème du choix entre les désinences *-en* et *-n*.

2 A plus de 95%, si l'on se fie aux indications statistiques de la grammaire DUDEN (édition de 1984, p.245) et si l'on suppose que les neutres invariables sont nettement plus nombreux que les masculins.

3 Eggers, dès 1969, montre la possibilité de résoudre ce problème.

4 Faute de place, nous avons renoncé à faire figurer l'organigramme qui représente l'imbrication des divers choix, mais il est très facile à construire à partir de nos indications.

5 Pour une approche plus globale de la déclinaison du groupe nominal en allemand, voir : JANITZA J. (et collaborateurs) : ENSEIGNEMENT ASSISTÉ PAR ORDINATEUR DES LANGUES ÉTRANGÈRES. Hatier, 1985.