

L'ANALYSE AUTOMATIQUE DE L'HEXAMETRE LATIN ET LES CONCORDANCES DE LA VULGATE

Ott

Je veux mentionner très brièvement deux projets que nous exécutons à Tübingen, au centre de calcul de l'Université. L'un de ces projets, l'analyse métrique automatique de l'hexamètre latin, est décrit dans deux articles de la "Revue" du L.A.S.L.A. L'analyse métrique a quelques rapports avec le problème de l'automatisation de l'analyse morphologique et aussi sémantique, comme a dit déjà M. le Prof. Delatte: quand l'ordinateur sait la quantité de certaines voyelles, par exemple du "a" final des mots de la première déclinaison, il peut réduire le nombre des analyses possibles d'une forme. Pensez aussi à certains verbes, dont les formes du parfait et du présent sont homographes, mais ont une différente quantité des voyelles.

L'autre projet qui touche un peu aux problèmes de cette séance, c'est une Concordance de la Vulgate que nous préparons à Tübingen. Père Bonifatius Fischer O.S.B., le directeur de l'Institut "Vetus Latina" de Beuron, a fait une nouvelle édition critique de la Vulgate. Il prépare en même temps une concordance à partir de cette édition. La part automatique de ce travail est faite par moi au centre de calcul de l'Université de Tübingen. En ce moment, la moitié du texte est déjà perforée sur bande perforée. La perforation est faite deux fois, pour automatiser la révision du texte, pour éliminer des erreurs de perforation. La division du texte en des petites phrases qui suivent chaque mot comme contexte est faite avant la perforation: on prend, plus ou moins, la division en versets qui se trouve dans l'édition.

Le gros travail qui reste encore est la lemmatisation des formes, parce que la concordance sera une concordance des mots, pas des formes. Pour l'automatisation de ce travail, je me sers du LEL Pisanum. J'en ai reçu une copie en bande magnétique de Père Busa. J'ai choisi cette version du dictionnaire et pas celle du L.A.S.L.A. de Liège par des raisons pratiques: je peux prendre les bandes magnétiques et les monter directement sur notre ordinateur à Tübingen. Il sera plus difficile de se servir du dictionnaire de Liège. D'autre part, il nous suffit pour la concordance d'avoir de chaque forme le

lemme, pas l'analyse morphologique complète que fournit le programme du L.A.S.L.A. Ca réduit aussi le nombre des analyses d'une forme qui sortent de l'ordinateur, et réduit en même temps le travail du philologue. Petit exemple: quand nous avons un mot comme "rosae", c'est égal pour la concordance qu'il soit le nominatif pluriel ou le génitif singulier ou le datif singulier ou le vocatif pluriel: pour nous, l'analyse est seulement "rosa". Seulement quand une forme peut être dérivée de différents lemmes, nous faisons imprimer les différents lemmes, avec le lemme le plus fréquent imprimé en première place; quand le philologue qui fait la révision de la liste des lemmes n'indique pas autre chose, l'ordinateur prend la première analyse. Ainsi, j'espère pouvoir lemmatiser environ 80% des mots de la Vulgate automatiquement, sans intervention humaine.