

# La plasticité fondamentale de l'écriture et le parcours du Laboratoire (A.R.I.T.) d'Analyse Relationnelle Informatique des Textes

Véronique HUYNH-ARMANET

## Abstract.

Studies in our laboratory originated from our own basic and methodological research. It is centred on the perceivable variability—in the linear order in homogeneous, non-specific texts in contemporary Spanish—of the relations between morphological classes and syntactic functions. After testing our hypotheses and hand procedures between 1972 and 1974 with the computational help of multidimensional analysis, we launched into the complete study of structurally complex texts, and we gradually allowed the computer to have an increasingly greater role (though with constant checking). The micro revolution enabled us, especially from 1988 onwards to speed up the creation of procedures better suited to our purpose, viz. to describe gradually and inductively the inner dynamics of textual writing, and to introduce young students to our own working methods. At the moment, as indicated in Javier Sanchez's paper published in this volume, we are able to create a "Data Basis from Concordance" enabling a study of complex texts and even hypertexts, in intra- and intertextuality, in structurally homogeneous chunks. Furthermore we are now using new complementary procedures meant, as in the first phase, to investigate the "Utterance contexts" in structurally complex texts with the permanent and controlled aid of the computer.

**Keywords:** computational relational analysis of texts, computational linguistic.      **Mots-clés :** analyse relationnelle informatique des textes, linguistique computationnelle.

Lorsqu'en 1975 ont été organisés pour la première fois, dans nos universités françaises, des enseignements collectifs pour les étudiants qui entraient en 3<sup>e</sup> Cycle, nous avons proposé, en tant que l'un des Directeurs de Recherche de la Formation Inter-Universitaire d'études ibériques des Universités de Paris III (Sorbonne Nouvelle) et Paris VIII, d'intégrer au D.E.A. (Diplôme d'Études Approfondies) que nous venions de créer, une structure pouvant aider ceux de nos futurs chercheurs qui le souhaiteraient à prendre contact avec l'ordinateur. Nous avons pris une telle initiative parce qu'à la suite de plusieurs années

---

✉ 1 rue des Martyrs 39-45; F-77920 Samois S/Seine (France).  
Fax : 33 1 64 24 64 41

---

de recherches personnelles centrées sur l'approche énonciative de la syntaxe textuelle hispanique, nous avons jugé indispensable de mettre à l'épreuve, à travers les données que nous avons recueillies, nos propositions théoriques et méthodologiques, sur l'approche des relations entre classes morphologiques et fonctions syntaxiques. Cette expérimentation multidimensionnelle conduite selon les procédures de l'*analyse des correspondances* nous avait confortée sur nos options de base mais nous savions dès cette époque que la saisie informatique des données textuelles en perspective énonciative, était méthodologiquement exigeante dans la mesure où elle ne pouvait s'accommoder des procédures réductrices les plus généralement pratiquées alors. D'où la nécessité, selon nous, de poursuivre les essais de logiciels plurilinguistiques de concordances et la mise au point, pour les textes à structure complexe, de systèmes de codage permettant d'extraire automatiquement à la demande, les blocs homogènes repérés, pour ensuite les comparer<sup>1</sup>.

La conduite de ce programme, qui, pour l'essentiel, a été définie avant 1977 a été possible grâce au concours technologique du Laboratoire d'Analyse des Données de l'Université de Paris VI, et du Centre de Calcul de la Maison des Sciences de l'Homme de Paris qui ont en outre aidé, dans toute la mesure du possible, à la poursuite d'applications de niveau Maîtrise. Les problèmes d'adaptation généralement rencontrés par nos étudiants de 3<sup>e</sup> Cycle de la décennie 1975/1984, obligatoirement confrontés jusqu'à l'apparition d'une micro-informatique performante avec des procédures relativement pesantes nous avait préoccupée au point de nous faire souhaiter pour nos chercheurs, une formation aussi précoce que possible.

Aussi, dès que fut annoncée, pour la rentrée de 1985, la mise en place au plan national d'une initiation généraliste de 1<sup>er</sup> Cycle au maniement du micro-ordinateur, nous avons proposé d'explorer expérimentalement dès la rentrée 1984, la création pour l'espagnol, d'une filière d'Enseignement/Recherche ciblée sur la *Linguistique Textuelle Informatique* destinée à donner, petit à petit, une meilleure assise à nos enseignements de 3<sup>e</sup> Cycle qui avaient déjà une dizaine d'années d'existence. Il s'agissait pour nous, tout en assurant un enseignement défini comme un enseignement de masse par le plan dit d'*Informatique Pour Tous*, de préparer pour l'avenir quelques étudiants de 2<sup>e</sup> et de 3<sup>e</sup> Cycle particulièrement intéressés par notre discipline.

Ce programme d'initiation étroitement lié aux recherches de notre laboratoire a comporté deux phases essentielles : une phase de rodage (1984–1988) où

---

<sup>1</sup> Cf. HUYNH-ARMANET (Véronique) : 1977, *Des structures syntaxiques de l'espagnol à l'analyse relationnelle des textes* (Paris : Honoré Champion), chapitre IX.

l'accent a été mis sur les techniques de saisie-clavier en perspective plurilinguistique puisque nos étudiants des années 1984/1988 étaient tout à fait novices et qu'en revanche, dans ce domaine, nos recherches en micro-informatique avaient sensiblement progressé. Ainsi, nous avons fait réaliser entre 1984 et 1986, pour le laboratoire, donc aussi pour notre enseignement, un logiciel expérimental couplé avec le logiciel de traitement de textes *Wordstar* seul disponible alors dans notre université. Certes, nous percevions les limites, au plan de la recherche, de ce type d'approche encore trop statique. Restant donc très attentive aux nouveautés de 1988 dans le domaine des logiciels, nous avons très vite souhaité la création de séquences instrumentales plus sophistiquées pour l'étude dynamique des textes. C'est pourquoi nous avons lancé, dès le début des vacances d'été de 1988, dans le cadre d'un DEA puis d'une thèse<sup>2</sup> qui a été soutenue au début de 1992, un deuxième programme instrumental, cette fois, à partir d'un logiciel intégré facilement accessible *Works 1* de Microsoft et d'un logiciel de concordances également nouveau mais dont nous connaissions expérimentalement certains antécédents : le logiciel *Micro-OCP* d'Oxford University Press. C'est alors que le Laboratoire d'Analyse Relationnelle des Textes nous a semblé pouvoir prendre le sigle A.R.I.T., qui mettait en évidence son option informatique<sup>3</sup>.

Parallèlement, et pour rester dans la logique de notre projet d'Enseignement/Recherche, nous avons initié, dès la rentrée 1988, ceux de nos étudiants des 2<sup>e</sup> et 3<sup>e</sup> Cycles qui travaillaient uniquement sur des applications, au maniement de *Works* et aux principes du codage *Cocoa* d'OCP, avec lesquels, par ailleurs, nos publications aussi bien antérieures que plus récentes avaient pu les familiariser<sup>4</sup>.

Pour des raisons conjoncturelles, cette expérience est actuellement interrompue. Dans les contextes où elle a été conduite, elle a eu le temps de prouver son efficacité. Tout en assurant aux étudiants de 1<sup>er</sup> Cycle qui le souhaitaient un apprentissage minimal de la *Linguistique Relationnelle Informatique*, elle nous a permis en particulier de former une nouvelle génération de chercheurs parmi lesquels quelques-uns se sont révélés, moyennant bien sûr un complément d'études en informatique, capables de travailler efficacement sur notre projet instrumental d'exploration dynamique des textes. Mais les publics changent depuis peu

---

<sup>2</sup> Cette thèse intitulée, « Recherches méthodologiques sur les conduites opérationnelles alternées en Analyse Relationnelle des Textes » vient pour l'essentiel d'être publiée sous un titre différent : SANCHEZ PEREZ (Francisco Javier) : 1992, *Méthodologie et outils de l'analyse relationnelle informatique des textes* (Saint-Denis, coll. « Analyse textuelle et nouvelles technologies », n° 1).

<sup>3</sup> On se reportera à l'article de Francisco Javier SANCHEZ PEREZ, « Qu'est-ce que l'analyse relationnelle informatique des textes ? » dans ce même volume. C'est dans cet article que sont publiées les références bibliographiques des publications du laboratoire A.R.I.T., auxquelles nous renvoyons dans notre texte.

<sup>4</sup> Cf. Bibliogr. 1.1, 2.8, 3.1, 3.3, 3.4, 4.3, et 4.5.

relativement vite et de plus, la stratégie instrumentale du Laboratoire a considérablement évolué depuis notre profonde mutation de 1988. Il faudrait par conséquent prévoir, si un programme d'Enseignement/Recherche était à nouveau mis en place, de sérieuses modifications dans l'apprentissage des techniques d'exploration des textes.

En effet, au plan de la Recherche notre programme instrumental s'est, en quatre ans, sensiblement enrichi : il associe par exemple au noyau des deux logiciels sélectionnés en 1988, un logiciel de saisie automatique récent et surtout depuis l'été 1991, un logiciel d'analyse des correspondances infiniment plus convivial que les précédents. De plus et surtout il permet maintenant, toujours à partir d'index de concordances globaux ou sélectifs, l'étude morpho-fonctionnelle de gros corpus ainsi que le démontrent les applications en cours. Ces applications en intra- et inter-textualité, nous paraissent désormais tout à fait prometteuses aussi bien au plan de la saisie automatique avec constitution ultérieure de blocs que de l'exploration *a posteriori* des modèles qui sous-tendent les textes. Cette exploration des modèles que nous avons dû laisser en veilleuse tant que nous ne disposions pas d'un logiciel d'analyse des correspondances à la fois convivial et performant, est de mieux en mieux intégrée à nos stratégies descriptives par paliers. Elle est également en passe de l'être dans des recherches moins avancées mais qui devraient nous faire progresser sensiblement dans les méthodologies d'approche de l'ordre linéaire et de l'insertion dans des textes autres que techniques, de terminologies d'élaboration rapide. Ces études ont été engagées au laboratoire dans le cadre d'une autre thèse<sup>5</sup>, cette fois, à partir d'une description sur tableur d'un texte espagnol du début du XVI<sup>e</sup> siècle qui pose des problèmes à la fois linguistiques et culturels intéressants. C'est en effet, l'un de ceux qui ont le plus contribué à la connaissance de quelques termes amérindiens dont certains sont tout à fait bien intégrés aux langues européennes<sup>6</sup>.

À ce stade d'une recherche qui informatise de plus en plus ses procédures de description du texte en perspective multidimensionnelle, on peut se demander, comme nous l'avions fait en 1972, si la problématique définie alors pour l'analyse de la plasticité syntaxique de l'énoncé, demeure adéquate. Il est certain que grâce à nos expérimentations, nous restons fondamentalement attachés à la reconnaissance de la variabilité des équilibres entre composantes et refusons par conséquent de les hiérarchiser *a priori*.

---

<sup>5</sup> L'étude expérimentale en cours porte sur deux termes devenus courants pour nous : *Ouragan* et *Savane* et un terme qui n'est pas entré dans notre langue : *Jagüey* extraits avec leurs énoncés de la *Historia general y natural de las Indias* de Fernandez de Oviedo.

<sup>6</sup> Cf. Bibliogr. 2.7 et 2.9.

De ce fait, nous reconnaissons un rôle potentiellement prépondérant à l'auteur d'un texte et laissons dans l'ombre, sans problème méthodologique, le code oral que ne calque d'ailleurs jamais le code écrit, même si nous savons que ce code oral peut à chaque instant être activé par la lecture à *voix* haute. Ce respect constant de l'équilibre particulier des composantes pour un texte donné, nous permet parallèlement de moduler le rôle de l'auteur selon par exemple, qu'on a affaire à une création individuelle ou collective. Mais dans tous les cas, nous respectons la structure, homogène ou non, du texte à décrire. D'où l'importance dans nos stratégies, du codage structurel bien sûr complété — mais il est bien plus facile à définir — du repérage éditorial. Sur ces points, nos procédures de codage, tout en restant fondamentalement proches de nos options initiales, se sont affinées surtout dans la dernière période<sup>7</sup>.

Parallèlement, tout en conservant notre définition générale (LXG) pour toutes les unités de vocabulaire avec également reconnaissance d'équilibres variables entre le ou les éléments lexémiques et le ou les éléments grammémiques, nous avons pu opérer à partir d'une structure du lexique qui ne pose pas de problèmes majeurs même quand on doit confronter deux langues. Nous distinguons d'une part, les formes appartenant à des inventaires dits fermés parce que statistiquement les plus lourds en règle générale et de toutes façons relativement stables diachroniquement et les inventaires ouverts, plus labiles. Nos études parce qu'elles ne prévoient aucune prédiction commencent toujours par un examen des inventaires fermés et la désambiguïsation des homographes dont nous préparons par l'étude des systèmes en situation une automatisation plus poussée mais contrôlable. Pour toutes ces procédures de repérage, nous avons conservé, en l'enrichissant sur certains points, notre code<sup>8</sup> d'avant 1972 qui reste tout à fait opérationnel.

Il apparaît à travers nos expériences qu'une description par paliers, mais non réductrice de la plasticité potentielle de l'écriture, s'accorde très bien avec les outils informatiques d'aujourd'hui à condition d'exploiter leur multidimensionalité essentielle. Certes dans notre stratégie, l'envol modélisateur peut sembler à certains trop timide, à d'autres pas assez prudent. Là aussi, les progrès technologiques rapprochent les échéances dans la mesure où ils élargissent le champ des comparaisons et de ce fait peuvent contribuer à enrichir la connaissance des dynamiques de l'écriture. Il est certain que l'ordinateur, au fur et à mesure qu'il devient pour nous plus performant, nous aide à progresser dans la découverte de nos paroles et à travers elles, de notre Langue. À condition peut-être cependant,

---

<sup>7</sup> Cf. Huynh-Armanet : 1977, ch. IX et Bibliogr. 3.3, 3.4, 3.6, 4.3 et 4.5.

<sup>8</sup> Cf. Huynh-Armanet : 1977, et Bibliogr. 4.5.

de développer nos capacités d'échanges et tout en restant cohérents bien sûr, de ne pas trop redouter ou paraître redouter « les matériaux impurs »...<sup>9</sup>.

---

<sup>9</sup> BRUNET (Étienne) : 1991, « Tous les chemins mènent-ils au CD-ROM ? », *RISSH* 27, p. 70.