

DIVISION LINGUISTIQUE DU CNUCE. ACTIVITÉ DE 1974 À 1976

ANTONIO ZAMPOLLI

1. De 1974 à 1976 la Division Linguistique du CNUCE (que nous indiquerons par l'abréviation DL) a continué à développer tous les secteurs d'activité mis en place au cours des années précédentes et dont la description se trouve dans Zampolli 1973 (a), (b).

De nouveaux projets ont également été pris non seulement pour essayer d'approprier l'ensemble de nos activités à notre conception théorique du secteur du 'linguistic data processing' mais également sous l'impulsion de quelques facteurs nouveaux, à savoir d'importantes subventions allouées par des Organismes publics pour des projets en commun, l'intensification de la collaboration avec des Universités et des Instituts étrangers, des responsabilités prises officiellement vis-à-vis des organisations scientifiques et politico-culturelles du secteur.

Les limites d'espace auxquelles nous sommes bornés nous dissuadent de faire une liste détaillée de tous les projets en cours; aussi préférons-nous exposer les principaux filons de notre activité en espérant toutefois mettre en évidence l'unité profonde et insécable de ces projets ainsi que l'étroite connexion avec les activités des Instituts italiens et étrangers qui collaborent avec nous.

En outre, étant donné que j'ai déjà eu l'occasion d'illustrer en d'autres lieux l'activité du secteur du 'text processing', je m'étendrai ici davantage sur les projets de linguistique quantitative et sur

ceux de l'analyse automatique aux différents niveaux du système linguistique.

2. *'Text-Processing'*.

Comme on le sait la DL collabore pratiquement à tous les projets en cours en Italie portant sur le 'linguistic and literary data processing' ¹. La DL se charge de la partie informatique de tous ces projets dans leurs diverses phases: organisation méthodologique, planification et analyses, programmation, exécution des élaborations.

Comme on l'a déjà observé plusieurs fois, en particulier au sein des principales organisations scientifiques du secteur, cette collaboration entre Instituts assure quelques avantages fondamentaux pour les recherches.

Les acquisitions méthodologiques, les expériences, les procédures et les programmes mis au point pour un projet sont immédiatement disponibles pour tous les projets analogues.

Tous les perfectionnements ou améliorations apportés aux procédures et aux méthodes grâce aux exigences d'un Institut sont immédiatement acquis pour les autres Instituts.

Tous les projets élaborent électroniquement les données selon les standards communs proposés par la DL. Aussi les textes élaborés par un Institut ou par un chercheur sont-ils immédiatement accessibles (après autorisation préalable) aux autres Instituts ou chercheurs.

¹ Ces projets se situent dans les domaines suivants: lexicologie et lexicographie, philologie classique et moderne, éditions critiques et critique textuelle, linguistique historique, dialectologie et géographie linguistique, langue et littérature (italienne et étrangère), psycholinguistique et sociolinguistique, linguistique générale, enseignement des langues, terminologie, démolologie, documentation automatique sur base textuelle, statistique linguistique et ses applications audiologiques et phoniatiques, etc...

Un texte (ou un autre matériel) reçu par un Centre étranger et inséré dans la Banque Italienne des Données Linguistiques de la DL reste à la disposition de tous les chercheurs italiens.

Cette Banque compte aujourd'hui plus de 5000 textes en plus de 20 langues, enregistrés selon les mêmes standards et opérables avec les mêmes programmes.

L'ensemble des faits que l'on vient de citer implique que le coût du dépouillement d'un texte est aujourd'hui réduit de moitié par rapport à ce qu'il serait si chaque Institut s'occupait aussi de l'élaboration électronique.

Les projets à long terme, en cours depuis plusieurs années, sont poursuivis. (Certains de ces projets sont mentionnés dans les communications de Avalle, Bartoletti-Colombo, Duro, Gregory pour ce même 'Colloque'). Grâce à eux la Banque Italienne des Données Linguistiques continue toujours à s'enrichir avec de nouveaux textes élaborés par les nouveaux projets entrepris au cours des trois dernières années. Comme il est impossible de tous les énumérer je cite seulement quelques-uns de ces textes en différentes langues: sanscrit: *Rgveda* (et lemmatisation semi-automatique, Oct. 1977); grec: *Scholia Pyndari, Legum Iustiniani Imperatoris Novellae, Pars Graeca*; latin: les grammairiens latins, Hippocrate; français: corpus à cheval entre le XVI et le XVIIème siècle; allemand et anglais: oeuvres complètes de Freud; espagnol: *El conde lucanor*; italien: oeuvres complètes de Montale et de Ungaretti, échantillons d'italien parlé (Radio, télévision, conversations familiales, conversations dans les magasins, Marioni 1974) et de l'italien des journaux (Crocetti 1975) etc...

Du point de vue méthodologique il faut souligner l'accroissement de la procédure standard de dépouillement avec des programmes demandés au fur et à mesure des exigences particulières des nouveaux projets. A titre d'exemple je cite le système mis au

point pour le dépouillement des textes pour lesquels il existe plusieurs versions (rédactions successives, traductions etc...) et qui a été expérimenté sur l'*Itinerarium Antonini Placentini* et sur l'*Orlando Furioso* (Segre-Zampolli 1971).

Il faut signaler également le début de l'activité dans le secteur historique qui se subdivise en trois secteurs principaux:

- élaboration quantitative de données historiques (verbales ou non);
- dépouillement de sources documentaires textuelles.

La préparation de procédures généralisées est en cours en particulier en collaboration avec l'Université de Pise.

3. *Banque Internationale de Données Linguistiques.*

Les buts de ce projet ont déjà été illustrés dans la première édition de ce Colloque (Zampolli 1976). Le projet qui a reçu le patronage d'importantes organisations politico-culturelles internationales opère essentiellement dans trois directions:

Textes: il s'agit de recherches qui ont pour but de comparer entre eux les divers systèmes d'enregistrement des textes en 'machine redeable form' utilisés dans les principaux centres européens et extraeuropéens et de la mise en place du software nécessaire pour leur échange entre ces centres. Les rapports déjà établis avec 13 Centres de 8 pays (Belgique, France, Angleterre, Allemagne, Espagne, Mexique, USA, Canada) permettent d'ores et déjà aux chercheurs italiens d'accéder, sans difficulté d'ordre technique ou d'organisation, aux archives de textes élaborées en diverses langues dans ces Centres, à travers les canaux technico-organisatifs de la DL, avec une importante économie de temps et d'argent. Plusieurs Instituts italiens qui travaillent avec nous, et avant tout, le Lessico qui est notre hôte dans ce 'Colloque', ont déjà profité de ce 'network' projeté et réalisé par la DL.

Analyse linguistique: le but est celui de rendre comparables et échangeables non seulement les textes mais également les analyses que les divers Centres opèrent sur les textes eux-mêmes, en introduisant à côté des unités du texte (mots, syntagmes, phrases) leur description linguistique, ou bien des renvois qui les relient à un système linguistique mémorisé une fois pour toute (lexique automatique, grammaire formalisée). Nous estimons qu'il est très utile et même indispensable, pour cette tâche, de recourir aux lexiques automatiques dotés de structures flexibles et adaptables à chaque fois, dans une certaine mesure, aux exigences d'un projet particulier. Pour atteindre cet objectif il est nécessaire de construire avant tout une description formelle et explicite des critères adoptés par les différents Centres dans l'analyse linguistique, et en particulier dans la lemmatisation. L'examen des différences et des analogies doit permettre d'établir à quel degré et à quel niveau d'analyse il est possible de rendre comparable les résultats des lemmatisations respectives et des autres classifications de données (voir Zampolli 1976, pp. 182-188).

Software: Des échanges de procédures et de programmes de divers types sont effectués avec les centres étrangers qui participent au projet en question.

Technologie: La DL étudie la possibilité d'appliquer dans le 'linguistic data processing' de nouvelles découvertes technologiques et l'opportunité d'en coordonner l'utilisation (mass-storage, computer produced microfiches, machine d'imprimerie laser, etc...).

4. *Dictionnaires de Machine.*

Les dictionnaires de machine sont en général considérés comme des instruments qui allègent radicalement le travail et le coût de la phase de lemmatisation et qui favorisent la cohérence des

critères entre des chercheurs différents (Duro-Zampolli 1968; Zampolli 1975): en effet ces dictionnaires proposent automatiquement pour chaque occurrence du texte toutes les analyses lexicales et morphosyntaxiques possibles selon le système linguistique dont le texte est une réalisation. C'est ainsi que la DL a entrepris la préparation de dictionnaires de machine pour l'espagnol (Zampolli 1977) et pour le latin; (il est à signaler que l'algorithme pour la génération automatique — flexion et segmentation — des lemmes de Forcellini, a considérablement progressé dans le cadre du projet discuté ici même au Lessico au mois de novembre dernier, de même que le système pour faire converger en une structure unique les divers lexiques actuellement existants). La DL a en outre acquis des dictionnaires préparés dans d'autres pays (par exemple le dictionnaire grec de Packard 1976 et celui de Liège). Le projet du Dictionnaire Italien de Machine (DMI) se propose également d'autres buts. Chacun de ses lemmes (soit près de 100.000 lemmes tirés des dictionnaires normaux de consultation et en particulier du Zingarelli (Xème éd.) et un million environ de formes correspondantes) est accompagné de quelques informations linguistiques de différent niveau (par exemple: catégories grammaticales, genre, nombre, temps etc...; type de construction syntaxique pour les verbes; limites d'usage: archaïque, dialectale etc...; codes de domaine pour les termes techniques); d'autres informations seront également ajoutées (codification des affixes, de l'étymologie, etc...).

Ces informations sont utilisées pour les recherches qui ont pour but de recueillir le plus grand nombre possible de données aux différents niveaux du système linguistique italien et de les structurer au moyen de modèles construits en élaborant heuristiquement les données elles-mêmes (cf. Zampolli 1968, Zampolli et alii 1977).

4.1. *Morphophonologie*

On a construit un algorithme pour la flexion automatique qui représente une description explicite complète et formalisée du système morphophonologique de l'italien, entièrement vérifiée au moyen de l'application aux 106.000 lemmes du DMI. Cette description va bientôt être complétée avec un modèle de la dérivation altérative et de celle des affixes en italien (cf. les contributions de E. Picchi, N. Ruiny, G. Turrini et de I. Prodanof in Zampolli et alii 1976). Ce modèle est constitué par des règles de dérivation exprimées sous forme de traits sélectifs classificateurs, associés à chaque suffixe du système italien et à chaque base particulière du DMI. Des procédures heuristiques appliquées au DMI devraient faciliter l'identification de ces traits, probablement de nature phonétique, syntaxique et sémantique (Halle 1973 et Jackendoff 1974 traitent ce problème dans une perspective générativo-transformationnelle). L'algorithme pour la transcription phonématique automatique des formes du DMI présente l'intéressante caractéristique de pouvoir exécuter sur demande la transcription selon n'importe lequel des divers inventaires phonématiques de l'italien proposés jusqu'à présent. En ce sens l'algorithme contient une description formalisée et opérativement applicable des équivalences et des divergences entre ces inventaires (cf. Zampolli 1973B).

4.2. *Syntaxe*

Les modèles relativement efficaces et vraiment connus pour le traitement automatique de la syntaxe (parser et génération de phrases) ne manquent pas: comme pour d'autres langues, pour l'italien également les efforts doivent converger surtout vers la recherche de faits et de types syntaxiques de l'italien dont il man-

que aujourd'hui encore des inventaires satisfaisants, et vers l'interaction entre composante syntaxique et composante sémantique. Une tentative de formalisation de la syntaxe relative à un sous-ensemble assez étendu de l'italien écrit contemporain a été entreprise, de façon à pouvoir décrire cette syntaxe avec le formalisme adopté par le GETA de Grenoble et à l'appliquer à un parser automatique inspiré de l'Augmente Transition de Woods (1976). Un but non secondaire de ces études est celui de contribuer à un système pour automatiser au moins en partie la distinction des homographes (Zampolli 1972). On se demande cependant si ce but n'est pas plutôt réalisable avec des parser limités qui ne se proposent pas de reconstruire la structure entière d'un énoncé, mais seulement d'identifier 'localement' la structure des syntagmes qui la composent.

4.3. *Sémantique.*

Il n'est pas nécessaire d'insister, je crois, sur la totale impossibilité d'appliquer 'in extenso' au lexique d'une langue aucune des théories embryonnaires formulées au cours de la «tragic history of semantics» (Katz 1972 p. 2). Pour répondre aux exigences de certains utilisateurs qui demandent au DMI l'indication non seulement des homographes mais également la signalisation des polysémies, de façon à pouvoir les distinguer dans la lemmatisation, on a ajouté au DMI la définition de toutes les acceptions indiquées, pour chaque lemme, pour les dictionnaires de référence. Pour chaque acception on a noté, avec l'indication des éventuelles limites stylistiques ou des limites relatives aux langages spéciaux, la définition lexicale fournie par le dictionnaire ou, le plus souvent, une version 'semi-formalisée'. Le projet de N. Calzolari, L. Moretti, L. Pecchia (Calzolari-Moretti 1976) se propose de tirer profit de ce corpus de définitions presque unique pour ces dimen-

sions, avec le double but de récupérer grâce à la gestion électronique, la somme de connaissances, de données et d'avènements théoriques et méthodologiques plus ou moins conscients produits au cours des ans par la tradition lexicographique et implicites dans les dictionnaires et de donner une première structuration explicite au matériel ainsi recueilli. Cette structuration devrait être réalisée au moyen d'un network ou graphe orienté, dans lequel les noeuds sont occupés par des markers sémantiques et les flèches par des relations sémantiques de diverse nature, interprétable au moins en partie dans la logique du calcul des prédicats. Cette organisation devrait permettre une évaluation également quantitative de l'indice de «circularité» du lexique (Calzolari 1977).

Le projet synonymie (Cappelli - Picchi 1976) qui opère sur le corpus des Synonymes indiqués comme tels dans le DMI, se propose en premier lieu de déterminer et d'explicitier les processus de classification qui opèrent dans la compétence «sémantique» du lexicographe.

5. *Linguistique quantitative.*

Il semble opportun de maintenir la distinction entre statistique sur les textes et statistiques sur le système, distinction déjà introduite et justifiée théoriquement par l'école de Prague (Troubetzkoi 1937 chap. 7).

5.1. *Statistique linguistique sur les textes.*

Comme on le sait, les dépouillements de textes qui se sont multipliés grâce à la diffusion des ordinateurs ont démontré la nécessité de modifier les modèles qui dans les années 50, avec la célèbre proportion langue — parole / univers statistique — échantillon, avaient essayé d'insérer les études sur les aspects quantitatifs

du langage directement au centre de la théorie linguistique structuraliste de type européen. Il n'existe pas encore de modèles substitutifs qui puissent donner à la statistique linguistique une base théorique adéquate. Des chercheurs ont apporté récemment quelques contributions relatives aux caractéristiques que ces modèles devraient avoir, aux exigences auxquelles ils devraient répondre, aux faits qu'ils devraient expliquer.

En d'autres termes ces chercheurs proposent de mesurer et de décrire avec soin les fréquences et les variations de fréquence du plus grand nombre possible de «caractéristiques» linguistiques (unité, structure, construction aux divers niveaux linguistiques) et dans le plus grand nombre de textes de type différent (auteur, genre, époque, sujet etc...). Des méthodes statistiques appropriées devraient permettre de subdiviser les textes en classes homogènes par rapport à une certaine «caractéristique» (ou faisceau de «caractéristique») c'est-à-dire de regrouper ensemble des textes dans lesquels on peut considérer que cette 'caractéristique' (ou faisceau) présente le même comportement du point de vue quantitatif. On devrait donc essayer de déterminer si les textes d'un groupe ont en commun des facteurs pragmatiques de production et lesquels: facteurs liés à l'interlocuteur et à l'effet que l'on veut produire, au thème et au contenu, au genre littéraire etc... Il en résulterait un processus heuristique récursif visant à déterminer la «nature statistique spécifique» des diverses 'caractéristiques' et à préciser quelles sont celles qui, parmi elles, ne peuvent pas être décrites statistiquement, et celles qui montrent une certaine corrélation avec certains facteurs de production.

Certains chercheurs italiens ont été parmi les premiers à donner un essor concret aux procédures de ce type et il semble que cette orientation commence à prévaloir également dans les autres pays (cf. par ex. Muller). De nombreux travaux de statistique lin-

guistique exécutés sur la Banque des Données de la DL peuvent être ramenés, qu'ils le déclarent ou non explicitement, à une problématique de ce type.

Par exemple, au niveau phonématique, l'analyse statistique du corpus de l'italien «parlé» exécutée par B. Marioni (1974) a démontré que même les fréquences des phonèmes varient selon le contexte de production (type de conversation: familière, cultivée; monologue, dialogue etc...).

Au niveau syntaxique l'étude de De Mauro-Policarpi a fourni les matériaux pour mettre en parallèle l'inventaire des types syntaxiques et les fréquences correspondantes dans l'italien littéraire et dans l'italien populaire (De Mauro 1974, Policarpi 1974) et elle a été complétée par une étude sur les types de proposition et de période dans la versification italienne entre le XIXème et le XXème siècle (Rombi 1974).

Au niveau lexical les matériaux fournis par le Frequency Dictionary of Italian Words (FDIW) de A. Juilland et V. Traversa (1973), par le Lessico di Frequenza della Lingua Italiana Contemporanea (LIF) de U. Bortolini, C. Tagliavini, A. Zampolli (1971) et par le corpus des journaux dépouillés par M. Crocetti (1975) ont été élaborés. Cette mise en parallèle a montré quelques extraordinaires régularités de variation des fréquences de divers paramètres (lexicaux, grammaticaux etc...) parmi les sous-ensembles qui constituent les 3 corpora cités (Zampolli 1975). Des variations analogues ont été rencontrées parmi des échantillons de langage de diverses catégories de malades mentaux (Castrogiovanni 1973).

Les résultats obtenus doivent cependant être complétés par d'ultérieures recherches conduites sur des corpora plus vastes. La Banque de mots en 'machine redeable forme' existante à la DL constitue un matériel précieux sur lequel il est possible de travail-

ler par induction. Il est cependant indispensable que la définition des unités linguistiques à «compter» soit cohérente dans la banque toute entière: il est nécessaire en particulier que la lemmatisation des textes d'une même langue soit exécutée selon des critères uniformes ou tout ou moins compatibles entre eux.

5.2. Statistiques sur le système linguistique.

Bien qu'elles aient été suggérées et recommandées par d'illustres linguistes de diverses écoles (Troubetzkoï 1939, Martinet 1960, Tagliavini 1968) les statistiques sur le système sont bien moins connues et diffuses que celles sur les textes. Même si les difficultés d'ordre théorique ne manquent pas, les difficultés qui fondamentalement ont empêché la réalisation des programmes souvent tracés semblent essentiellement opératives. En premier lieu, alors que les textes sont, pour ainsi dire, des 'objets' donnés naturellement, les inventaires d'unités, de fonctions, de règles doivent être construits par le linguiste et ce travail est souvent long et pénible.

La disponibilité du DMI répond enfin, tout au moins en partie, à ces exigences. Comme on l'a dit, le DMI a été constitué principalement par deux inventaires distincts: 1) un ensemble de lemmes, 2) un ensemble de formes correspondant au premier. En outre chacune des unités de ces inventaires est composée de deux parties: a) l'unité lexicale (respectivement lemme et forme) en graphie normale, en transcription phonématique et décomposée en base, suffixe, désinence; b) une description en code de ses propriétés et caractéristiques. Toute une série de recherches devient ainsi possible, qu'il s'agisse de celles déjà conclues ou en voie de conclusion, au niveau phonématique (fréquence des phonèmes, des syllabes, de divers types de structures phoniques, des oppositions phonologiques etc...), (cf. Muljacic 1969, p. 307) ou de celles qui

sont à peine commencées ou en cours sur le rendement des divers mécanismes de production lexicale par dérivation au moyen d'af-fixes, ou encore de l'étude du rendement des divers patterns synta-xiques des verbes et de la composition étymologique, etc... (cf. Zampolli 1973).

Il reste encore beaucoup à faire, mais la présence du DMI permettra d'accomplir pour l'italien des recherches souvent pro-grammées mais jamais menées à terme pour d'autres langues.

6. *Activité didactique.*

En plus des cours d'information et de mise à jour organisés pour les chercheurs qui opèrent dans le secteur du 'linguistic data processing' en Italie (la plupart travaillent à des projets financés par le CNR) Zampolli a continué la série des 'International Summer Schools' qui ont lieu tous les deux ans à partir de 1970. La troisième édition de l'Ecole a eu lieu en 1974 et a eu pour objet les méthodes pour la description formelle de la syntaxe et de la sémantique des langages naturels. Les cours ont été tenu par E. Bach (Amherst, Mass.), Ch. Fillmore (Berkeley), M. Gross (Pa-ris), B. Hall-Partie (Amherst), M. Kay (Palo Alto, Calif.), J. Lyons (Edimburg), S. Petrick (New York), E. Sanderwall (Up-psala), T. Winograd (MIT), W. Woods (Harvard), A. Zampolli (Pise). La 4ème édition de l'Ecole aura lieu du 8 au 26 août 1977 sur le thème «Natural language understanding by compu-ter». Voici les principaux cours qui sont prévus:

J. Bresnan (MIT, Cambridge, Mass.), A realistic transforma-tional grammar.

R. Dilligan (University of Southern California à Los Ange-les), Introduction to computers and programming fundamentals.

M. Halliday (University of Sidney), A functional approach to grammar.

R. Kaplan (Xeros Palo Alto Research Center), Psycholinguistics and procedural models for syntactic analysis.

L. Karttunen (University of Texas à Austin), Introduction to logics.

L. Karttunen (University of Texas à Austin), Model-theoretic semantics for natural languages.

M. Kay (Xeros Palo Alto Research Center), Computational linguistics and artificial intelligence.

F. Keenan (University of California à Los Angeles), Grammatical relations and universal grammar.

G. Lakoff (University of Bielefeld), Structures and functions of a text-theory.

I. Melčuk (University of Montreal), Linguistic Models of the Meaning.

Y. Wilks, Language, linguistics and processing.

A. Zampolli (Université de Pise, Italie), Introduction to computational linguistics.

Et en outre les workshop suivant:

E. Hajcova (Université de Prague), Functionalism and topic focus in generative description.

M. Kay (Xeros Palo Alto), Advanced methods in automatic syntactic analysis.

C. Ramassa (Université de Gênes), Introduction to computational linguistics.

7. Relations et collaborations internationales.

La DL ne travaille pas seulement au projet déjà cité «Banque Internationale des Données Linguistiques», elle opère également dans le secteur de la collaboration et des relations internationales avec l'organisation déjà citée de l'Ecole d'été, avec l'organisation périodique de Congrès et de Séminaires; pour la période qui nous

intéresse ont eu lieu à Pise les réunions du CIRPHO 1974, celle du groupe pour la «machine assisted translation» en rapport avec la Commission des Communautés Européennes (1975), celle du Comité Exécutif de l'AILA (1975), celle du groupe d'Etude pour le lexique automatique du latin (1975). Des cours et des conférences ont été tenus par G. Waite (Darmouth), D. Packard (UCLA), R. Dilligan (SCU), P. Bratley (Montréal), G. Lakoff (Berkeley), M. Gross (Paris), Kittredge (Montréal), M. Saltarelli (Urbino), M. Kay (Palo Alto), E. Keenan (UCLA), J. Smith (Manchester), C. Petracco (Gênes), J. Desclès (Paris), B. Hall-Partie et E. Bach (Amherst, Mass.), B. Quemada (Paris), J. Petöfi et H. Rauschner (Bielefeld), D. Olney (Los Angeles), L. Delatte, S. Govaerts, J. Denooz (Liège).

La DL, en la personne de son directeur, représente l'Italie aux activités de toutes les principales organisations scientifiques internationales du secteur.

8. *Conclusions.*

Le développement du secteur 'élaboration de textes' semble satisfaisant, surtout pour ce qui est des projets et des réalisations relatives à la Banque Nationale et à la Banque Internationale de données Linguistiques. Une attention toute particulière devra être attribuée semble-t-il au cours des prochaines années aux directions de travail suivantes:

- compléter les actuelles archives de textes italiens avec des corpora provenant du XIX^{ème} et du XX^{ème} siècle opportunément choisis et stratifiés.

- généraliser l'emploi du DMI comme instrument de lemmatisation des textes de façon à poursuivre une comparabilité des résultats des dépouillements qui suive la standardisation déjà obtenue dans la phase d'enregistrement.

— mettre au point des méthodologies et des instruments informatico-linguistiques qui permettent de rendre interactive et semi-automatique la phase de consultation et d'utilisation des matériaux linguistiques contenus dans la langue des textes et des données linguistiques, d'en étendre l'utilisation à de nouvelles classes d'utilisateurs (terminologues, sociolinguistes etc...) et qui permettent en outre une diffusion plus rationnelle et économique des données elles-mêmes (microfiches etc...).

Le DMI se présente comme un instrument des plus utiles pour la phase de regroupement et d'organisation d'inventaires, même exhaustifs, de connaissances sur le système linguistique italien et sur son fonctionnement. Il semble donc que le moment est venu de promulguer et de développer les recherches dans le secteur des modèles pour la représentation de structures et de connaissances pragmatiques, sémantiques, syntaxiques ainsi que pour la représentation de stratégies et d'algorithmes nécessaires pour leur «calcul» à partir d'énoncés italiens et vice versa. A ce sujet il semble qu'une étroite coopération entre les linguistes computationnels proprement dits, les linguistes en général et les chercheurs en linguistique italienne soit indispensable.

En ce qui concerne la statistique linguistique la crise théorique dans laquelle cette discipline se trouve à cause de l'absence de modèles adéquats rapproche la situation italienne de celle des autres pays.

Toutefois la présence de la Banque des Données et celle du DMI fournissent des instruments assez valides et stimulants pour l'élaboration de données sur lesquelles il sera possible de tenter la construction d'un nouveau modèle d'explication.

BIBLIOGRAFIA

- D'ARCO SILVIO AVALLE (1976), *La formalisation des graphies dans le domaine de la rime*, in Zampolli-Calzolari (1977).
- U. BORTOLINI, C. TAGLIAVINI, A. ZAMPOLLI (1971), *Lessico di frequenza della lingua italiana contemporanea*, IBM Italia.
- G. BUCCELLATI, F. SAPORETTI (1974), *The Old Babylonian Linguistic Analysis Project: Goals, Procedures und First Results*, in Zampolli-Calzolari (1977).
- N. CALZOLARI, L. MORETTI (1976), *A Method for a Normalization and a possible Algorithmic Treatment of Definitions in the Italian Dictionary*, in *Preprints COLING 76*.
- E. CAMPANILE, A. ZAMPOLLI (1976), *Some Experiments in Historical Computational Linguistics*, in Zampolli-Calzolari (1977).
- A. CAPPELLI, E. PICCHI (1976), *Synonymie lexicale: une tentative d'analyse*, in *Preprints COLING 76*.
- P. CASTROGIOVANNI, A. TELARA (1973), *Primi risultati di un'analisi statistica morfologica e lessicale delle risposte al test di Rorschach nella prospettiva di uno studio dei rapporti tra psicopatologia e linguaggio*, in A. Zampolli (1973) pp. 307-323.
- C. CROCETTI (1976), *Criteri di analisi del testo per una statistica del lessico dei quotidiani*, Tesi di Laurea dell'Università di Pisa.
- T. DE MAURO (1974), *Premessa a una raccolta di tipi sintattici*, in SLI (1974), pp. 551-574.
- M. DURANTE (1975), *La linguistica sincronica*, Torino.
- A. DURO (1973), *Elaborations électroniques de textes effectuées par l'Accademia della Crusca, pour la préparation du Dictionnaire Historique de la langue italienne*, in Zampolli (1973A), pp. 53-75.
- A. DURO (1975), *L'elaborazione elettronica dei dati linguistici: problemi filologici*, in Accademia Nazionale dei Lincei (1975), pp. 11-21.
- A. DURO, A. ZAMPOLLI (1968), *Analisi lessicali mediante elaboratori elettronici*, in Accademia Nazionale dei Lincei (1968).

- M. FATTORI-M. BIANCHI (a c. di) (1976), *Atti del I Colloquio Internazionale del Lessico Intellettuale Europeo*, Roma.
- G. FERRARI, I. PRODANOF (1976), *Machine Dictionary and Lexicon*, in *Preprints COLING 76*.
- CH. FILLMORE (1976), *Scene-and-Frames Semantics*, in Zampolli (1977B).
- A. JUILLAND, V. TRAVERSA (1973), *Frequency Dictionary of Italian Words*, The Hague.
- J.J. KATZ, (1972), *Semantic Theory*, Harper and Row, Publishers. *L'uomo e la Macchina, Atti del XXI Congresso Nazionale di Filosofia (Pisa, 22-25 Aprile 1967)* (1967) Torino.
- M. ROMBI, (1974), *Tipi di proposizione e periodo nella versificazione tra Otto e Novecento*, in *Fenomeni morfologici e sintattici nell'italiano contemporaneo. Atti SLI 7*, Roma, pp. 575-650.
- B. M. MARIONI, (1974), *Studi di statistica fonematica su un corpus di italiano parlato*, Tesi di Laurea dell'Università di Pisa.
- Z. MULJACIC, (1969), *Fonologia generale e fonologia della lingua italiana*, Bologna.
- G. POLICARPI, (1974), *Tipi di proposizione e periodo nell'italiano popolare contemporaneo e in Croce*, in *Fenomeni morfologici e sintattici nell'italiano contemporaneo. Atti SLI 7*, Roma, pp. 575-650.
- C. SEGRE, A. ZAMPOLLI, (1976), *Le Concorde diacroniche dell'Orlando Furioso*, in *Ludovico Ariosto: lingua, stile, tradizione*, Milano, (in stampa).
- W. WOODS, (1976), *Lunar Rocks in Natural English: Explorations in Natural Language Question Answering*, in Zampolli (1976B).
- A. ZAMPOLLI (a c. di) (1973A), *Linguistica Matematica e Calcolatori. Atti del Convegno e della prima Scuola Internazionale*, Firenze.
- A. ZAMPOLLI (1973B), *La Section Linguistique du CNUCE*, in Zampolli (1973A), pp. 133-199.
- A. ZAMPOLLI (1973C), *L'automatisation de la recherche lexicologique: état actuel et tendances nouvelles*, in «META», XVIII, 1-2, pp. 101-136.

- A. ZAMPOLLI (1973E), *Humanities Computing in Italy*, in «Computers and the Humanities», 7, 6, pp. 343-360.
- A. ZAMPOLLI (1975), *L'elaborazione elettronica dei dati linguistici: stato delle ricerche e prospettive*, in Accademia Nazionale dei Lincei (1975), pp. 23-107.
- A. ZAMPOLLI-N. CALZOLARI (eds.) (1977A), *Computational and Mathematical linguistics*, Olschki, Firenze.
- A. ZAMPOLLI (1977B), (ed.), *Linguistic Structures Processing*, Amsterdam, Noth-Holland, (in stampa).
- A. ZAMPOLLI (1976), *Les dépouillements électroniques: quelques problèmes de méthode et d'organisation*, in Fattori-Bianchi (1976), pp. 173-197.
- A. ZAMPOLLI, N. CALZOLARI, A. CAPPELLI, G. FERRARI, L. MORETTI, L. PECCHIA, E. PICCHI, I. PRODANOF, N. RUINY, G. TURRINI, *Il Dizionario di Macchina dell'Italiano*, in Atti del Convegno: «Logiche, Calcoli, Formalizzazioni e Lingue Storico-naturali», Catania 17-19 Settembre 1976, (in stampa).