

1988

ID

7

Bd

621 - VILLEURBANNE

DIPLOME D'ETUDES SUPERIEURES SPECIALISEES

INFORMATIQUE DOCUMENTAIRE

NOTE DE SYNTHESE

**LES MODELES LINGUISTIQUES DE L'ESPAGNOL
DANS LE TRAITEMENT DE L'INFORMATION**

**PRESENTE PAR :
MARIA-ANGELES RODRIGUEZ**

**SOUS LA DIRECTION DE :
MONSIEUR M. LE GUERN**

ANNEE 1987-1988

1988
ID
4

UNIVERSITE CLAUDE BERNARD
LYON I

43, Bd du 11 Novembre 1918
69621 VILLEURBANNE

DIPLOME D'ETUDES SUPERIEURES SPECIALISEES
INFORMATIQUE DOCUMENTAIRE

NOTE DE SYNTHESE

LES MODELES LINGUISTIQUES DE L'ESPAGNOL
DANS LE TRAITEMENT AUTOMATIQUE DE L'INFORMATION



PRESENTE PAR :

Maria Angeles RODRIGUEZ

SOUS LA DIRECTION DE :

Monsieur M. LE GUERN

ANNEE 1987-1988

1988
ID
7

SOMMAIRE

INTRODUCTION.....	2
I.PRESENTATION DU SUJET.....	4
II.METHODOLOGIE DE LA RECHERCHE DOCUMENTAIRE.....	6
1.Recherche manuelle.....	7
1.1.Demarche.....	7
1.2.Résultats.....	7
2.Contacts personnels.....	8
3.Recherche automatisée.....	8
3.1.Définition du sujet et choix des mots-clés.....	8
3.2.Choix des bases de données.....	8
3.2.1.Présentation de la base de données FRANCIS.....	9
3.2.2.Présentation de la base de données PASCAL.....	9
3.2.3.Autres bases consultées.....	10
3.3.Strategie de recherche.....	11
3.4.Résultats.....	13
4.Répérage des documents primaires.....	13
5.Considerations générales.....	14
III.THEORIE DE CAS DE BERNARD POTTIER	
A.L'ANALYSE AUTOMATIQUE DES LANGUES NATURELLES.	16
1.L'Analyse morphologique.....	16
1.1.La morphologie.....	16
1.2.Les analyseurs morphologiques....	16
2.L'analyse syntaxique.....	18
1.1.La syntaxe.....	18
1.2.Les analyseurs syntaxiques.....	18
3.Interdependance :	
l'analyse morphosyntaxique.....	19
B.THEORIE DE CAS DE BERNARD POTTIER.....	21
1.Morphologie.....	21
1.1.Les unités de base.....	21
1.2.La classification des unités.....	25
1.2.1.Substantifs (Sb).....	25
1.2.2.Verbes (Vb).....	26
1.2.3.Adjectifs (Adj).....	26
1.2.4.Indicateurs (I).....	27
1.2.5.Déterminants (Det).....	27
1.2.6.Quantificatifs (Quant)....	27
1.2.7.Auxiliaires (W).....	27
1.2.8.Relateurs (Rel).....	28
1.2.9.Substituts (Z).....	28
1.2.10.Afixes.....	30

2.Syntaxe.....	31
2.1.L'analyse syntaxique.....	32
2.2.Les niveaux d'analyse syntaxique.	32
2.2.1.Le fonctème.....	32
2.2.1.1.Fonctème nominal...	32
2.2.1.2.Fonctème adjectival	34
2.2.1.3.Fonctème verbal....	34
2.2.2.Les syntactèmes.....	35
2.2.2.1.FN <--- FN'.....	35
2.2.2.2.FN <--- FA.....	35
2.2.2.3.FN <--- FV1.....	35
2.2.2.4.FN ---> FV2.....	36
C.CONCLUSION.....	37
ANNEXE.....	39
BIBLIOGRAPHIE.....	41

INTRODUCTION

INTRODUCTION

Le sujet m'a été proposé par Monsieur Michel LE GUERN, professeur de linguistique à l'Université Lumière de Lyon II.

L'objectif du travail proposé est l'élaboration et analyse d'une bibliographie sur les modèles linguistiques de l'espagnol, en vue de son utilisation dans la conception d'un système de traitement automatique de l'information.

Dans la première partie apparaissent la présentation et la situation du sujet ; par la suite j'expose la méthodologie suivie dans la recherche documentaire.

En vue des résultats obtenus et compte-tenu de l'objectif de cette étude, on a décidé de réaliser une analyse aboutissant au classement des informations obtenues et à l'étude approfondie du traité de linguistique générale de Bernard Pottier.

Le classement des informations s'est avéré nécessaire pour faire le point sur l'état de recherches dans le domaine défini.

L'analyse du traité de Pottier nous a paru fondamentale du fait qu'on y retrouve toute une théorie des modèles linguistiques, qui n'a pas été réperée au cours de la recherche documentaire. Cette analyse sera utilisée pour dégager les modèles qu'il propose et leur adéquation au traitement automatique de l'espagnol.

Au début de cette analyse on rappelle quelques éléments de morphologie et syntaxe. Ensuite on expose la théorie de cas de Bernard Pottier, quant aux catégories morphologiques et niveaux syntaxiques.

La dernière partie est consacrée à la présentation des références bibliographiques obtenues. Cette présentation suit un classement par matières et les notices sont rédigées d'après les normes AFNOR.

PREMIERE PARTIE

PRESENTATION DU SUJET

I. PRESENTATION DU SUJET

Depuis la naissance des ordinateurs, le problème de la communication avec eux en langue naturelle, a été un souci fondamental, non seulement par la facilité d'utilisation que cela représente, mais aussi par les multiples applications qui en découlent : systèmes experts, E.A.O. (enseignement assisté par ordinateur), reconnaissance et génération de la langue parlée, interrogation de bases de données, traduction et indexation automatique, etc.

Parmi celles-ci, la traduction automatique fut un des premiers centres d'intérêt et très tôt des réalisations furent envisagées en s'appuyant sur la capacité de mémoire et la rapidité des machines.

On essaya de faire une traduction plus ou moins mot à mot, ce qui donna des résultats catastrophiques. Bientôt les chercheurs comprirent qu'il fallait attaquer le problème autrement, c'était plus une question de conception que de puissance des moyens.

En même temps les linguistes apportaient des nouvelles idées et la reformulation de la syntaxe par la grammaire transformationnelle revint au premier plan, avec des concepts linguistiques comme la structure profonde et la structure superficielle, dues à Noam Chomsky.

Cela orienta aussi des recherches en informatique et des nouveaux travaux sur la langue naturelle apparurent : leur idée de base était de faire à fond une analyse syntaxique, pour aborder plus tard le problème sémantique.

L'introduction de l'ordinateur dans la langue ne pouvait que rendre encore plus évidente la nécessité d'une analyse linguistique sérieuse.

Dans tout système de communication, la communication homme-machine n'y échappe pas, il y a un processus de compréhension où l'on analyse le message, et un processus de génération où l'on produit le message. Transféré à l'informatique, la première tâche est du ressort de l'analyse automatique et la deuxième de la génération automatique.

D'où les deux types de systèmes :

---> systèmes de reconnaissance ou analyse automatique des langages naturels, qui produisent une représentation formelle reflétant l'analyse d'un texte

---> systèmes de génération des langages naturels, qui produisent un texte à partir d'une représentation formelle

Compte-tenu du sujet, ce sont les systèmes de reconnaissance qui précisent les modèles linguistiques, orientés vers la mise au point d'un analyseur morpho-syntaxique.

Les modèles conçus pour l'analyse automatique sont très diverses, mais leurs composants de base sont, jusqu'aujourd'hui, communs :

- > un lexique qui décrit le vocabulaire à utiliser
- > une syntaxe qui décrit les règles grammaticales

Ces données sont nécessaires à l'application d'un système de reconnaissance. Par contre, des questions comme leurs sources et fondements, leur représentation, leur structuration, etc, sont des questions ouvertes auxquelles on apporte diverses solutions.

De toutes façons, de nos jours, il y a des tendances définies, comme l'utilisation de PROLOG comme langage de programmation, du à la concision et à la clarté d'écriture, et à la facilité pour écrire une grammaire.

L'étude du sujet nous montre que l'analyse automatique des langues naturelles, apparaît aujourd'hui comme une direction fondamentale de développement de l'informatique, qui tente de plus en plus des concepteurs de systèmes informaticiens.

Les implémentations réalisées ne prétendent pas avoir résolu le problème de la compréhension. D'abord elles ne s'intéressent pas à toutes les langues, et leurs modèles ne comprennent qu'un sous-ensemble de phrases syntaxiquement correctes d'une langue, et ceci dans un domaine restreint.

Cette limitation provient de la difficulté de décrire une grammaire générale et automatisable d'une langue.

D'autre part, on néglige souvent l'étude des mécanismes linguistiques qui sont tout à fait indispensables pour l'analyse automatique d'une langue naturelle.

Le traitement automatisé des langues naturelles fait appel à une compétence à la fois linguistique et informatique. Une étroite collaboration entre linguistes et informaticiens est, sans doute, nécessaire si on veut arriver à la conception d'un système intégré, capable d'enlever l'ambiguïté des expressions et de couvrir l'ensemble d'une langue.

DEUXIEME PARTIE

METHODOLOGIE

DE LA

RECHERCHE DOCUMENTAIRE

II. METHODOLOGIE DE LA RECHERCHE DOCUMENTAIRE *****

1. RECHERCHE MANUELLE

1.1. DEMARCHE

Etant donné la spécificité du sujet, la recherche dans les bibliographies courantes - commerciales ou non - s'est révélée tout à fait infructueuse.

Il a donc fallu débiter la recherche à partir des bibliographies spécialisées dans le domaine de la linguistique.

Compte-tenu de l'actualité du sujet, le dépouillement a été fait de 1976 à nos jours.

Dans un premier temps, la recherche a été faite dans la "Bibliographie linguistique", publiée par le Comité international permanente des linguistes, bibliographie courante qui dépouille environ 800 publications périodiques dans le domaine, outre les congrès, thèses, rapports et mélanges.

Un deuxième temps m'a conduit au dépouillement des documents primaires du laboratoire d'Informatique Documentaire de Lyon I, dépouillement infructueux compte-tenu de la limitation du sujet à la langue espagnole.

Trois publications ont été dépouillées :

---> Information processing & management : an international Journal Libraries and Information retrieval systems and Communication. Pergamon Press.

---> T.A. informations : revue internationale du traitement automatique du langage. CNRS.

---> Interface : Automatique. Bureautique. Gestion. AFCET (Association Française pour la cybernétique économique et technique).

1.2. RESULTATS

Cette recherche manuelle m'a permis de repérer 7 références concernant le traitement automatique de l'espagnol. Mais si on se limite au sujet, aux applications informatiques des modèles linguistiques théoriques, on plonge dans le néant : une seule référence s'est révélée pertinente.

2. CONTACTS PERSONNELS

Ils ont constitué une source d'information capitale pour la poursuite de mon travail, en se révélant comme le seul moyen d'être au courant des derniers développements dans le domaine considéré.

A travers de l'ISOC (Instituto de Informacion y Documentacion en Ciencias Sociales y Humanidades), j'ai pu contacter deux centres de recherche en Espagne, qui font des recherches sur ce sujet :

---> Facultad de Informatica de l'Université de Catalogne, à Barcelone

---> Centro de Investigacion UAM-IBM, à Madrid

Ces contacts m'ont permis de recueillir une documentation sur le sujet qu'autrement je n'aurais pas pu obtenir.

3. RECHERCHE AUTOMATISEE

La manque de références dans la recherche manuelle ainsi que la nécessité d'être au jour, inhérente à un sujet d'actualité, (le délais de publication de la "Bibliographie linguistique" étant de deux ans), m'ont conduit à la consultation de bases de données.

3.1. DEFINITION DU SUJET ET CHOIX DES MOTS-CLES

Toute interrogation de bases de données suppose, au préalable, une préparation intellectuelle pour bien cerner le sujet et aboutir au choix des mots-clés.

A partir des termes employés dans les références obtenues au cours de la recherche manuelle et de la consultation d'études sur le sujet, on a retenu des descripteurs susceptibles de l'indexer.

Ces mots-clés sont à réajuster en les étudiant de près pour dégager les termes généraux et termes spécifiques, et en vérifiant leur existence dans le lexique des bases de données qui seront interrogées.

3.2. CHOIX DES BASES DE DONNEES

Le sujet choisi faisant partie du domaine de la linguistique appliquée, en relation avec les sciences de l'information, et en utilisant le Répertoire des banques de données en conversationnel de l'Association Nationale de la Recherche Technique (ANRT), deux bases de données ont été interrogées : PASCAL et FRANCIS/SCIENCES DU LANGAGE.

3.2.1. Présentation de la base de données FRANCIS

La base de données FRANCIS/SCIENCES DU LANGAGE est une base de données bibliographiques, produite par le Centre de Documentation Sciences Humaines (CDSH) du Centre National de la Recherche Scientifique (CNRS).

Elle couvre les principaux domaines de la linguistique :

- linguistique théorique
- linguistique descriptive
- linguistique appliquée
- linguistique historique
- sémiotique
- sociolinguistique
- psycholinguistique
- pathologie du langage
- études sur la communication et l'information

FRANCIS/SCIENCES DU LANGAGE est constituée essentiellement d'articles de périodiques (environ 700), et recense aussi des ouvrages, rapports, comptes-rendus de congrès et thèses universitaires.

Les documents signalés proviennent du monde entier sans restriction d'origine ou langue.

Les références sont tirées des quelque 8.000 périodiques du réseau CDSH, parmi lesquels un millier sont consacrés, au moins en partie, à la linguistique.

Elle est constituée d'environ 55.000 références de documents publiés à partir de 1972, et s'accroît chaque année de 3.200 références, avec une mise au jour trimestrielle.

FRANCIS/SCIENCES DU LANGAGE est accessible sur Telesystèmes en langage d'interrogation QUESTEL PLUS.

3.2.2. Présentation de la base de données PASCAL

La base de données PASCAL est une base de données bibliographique et analytique des sciences et techniques, produite par le Centre de Documentation Sciences et Techniques (CDST) du Centre National de la Recherche Scientifique (CNRS).

Elle couvre les principaux domaines des sciences et techniques :

- sciences appliquées
- sciences de l'ingénierie
- chimie
- physique
- sciences de la vie
- sciences de l'information
- technologie
- sciences exactes

PASCAL analyse tous les articles de périodiques majeurs français et étrangers, ainsi que les rapports scientifiques, thèses et comptes-rendus de congrès.

Elle est accessible sur Telesystèmes en langage d'interrogation QUESTEL PLUS.

3.2.3. Autres bases consultées

Il faut mentionner ici la consultation d'autres bases de données, bien que les résultats ne le méritent pas.

Bases de données espagnoles

Grâce au service P.I.C. (Puntos de Informacion Cultural) du Ministère de la Culture en Espagne, j'ai eu accès à différentes bases de données, dont le propre Ministère est le serveur. Mais ne dépouillant pas les périodiques, je n'ai pas pu repérer qu'une monographie, malgré l'exhaustivité de l'équation de recherche.

Téléthèses

Téléthèses est une base de données sur les thèses de doctorat soutenues devant les Universités françaises.

Elle recense des thèses de droit, lettres, sciences humaines et sociales, sciences et économie de gestion, depuis 1972 ; et celles de médecine, pharmacie et odontostomatologie depuis 1983.

Téléthèses est réalisée par :

- > le fichier central des thèses de Nanterre, pour les thèses littéraires
- > le Centre de Documentation Scientifique et Technique du CNRS, pour les thèses scientifiques
- > le fond national de thèses de la bibliothèque interuniversitaire de Clermont-Ferrand, pour les thèses médicales.

Elle est accessible sur Minitel et interrogeable sans formation préalable.

Etant donnée la localisation d'une thèse française à partir de la recherche manuelle, l'interrogation de Téléthèses s'est avérée pertinente.

Pour la localisation de cette thèse dans le fonds de la base, des nombreux essais ont été effectués. Le nom de l'auteur n'y donnant pas accès, on a dû interroger par d'autres critères de sélection tels que la discipline, l'année de soutenance, les mots-clés et les mots du titre.

D'autre part l'inexistence de l'opérateur booléen "OU" fait impossible la recherche par union des différents mots-clés, ce qui oblige à balayer la base mot à mot. Par conséquence l'accès est lourde et le rappel douteux.

Des nombreux passages ont été donc réalisés pour arriver à la visualisation du document. Par la suite on a retenu les descripteurs, à partir desquels on a continué l'interrogation.

Le résultat s'est révélé tout à fait infructueux, car aucune référence n'a été localisée, malgré l'exploration du fonds à partir de tous les critères de sélection possibles.

3.3 STRATEGIE DE RECHERCHE

A l'aide des lexiques publiés pour les bases de données choisies, on a retenu les descripteurs susceptibles d'indexer le sujet dans chacune d'elles.

A partir d'ici la stratégie a été construite tout au long de la recherche, en fonction de la fréquence d'indexation de chacun des termes.

Dans une première approche, l'interrogation a été faite à partir du terme le plus spécifique.

On a interrogé par mots-simples, avec l'utilisation des opérateurs de troncature et de proximité, de manière à explorer les champs titre, résumé, descripteurs et terme générique.

L'équation a pris la forme suivante :

1 MODELE? AV LINGUISTIQUE
2 1 ET ESPAGNOL

Une telle spécificité a donné un résultat de "0" références, à cause de la forte sélection effectuée.

Une deuxième approche a été faite à partir des mots composés des lexiques.

Il a donc fallu être exhaustif pour augmenter le rappel, malgré le risque de bruit que ça entraîne.

Sur FRANCIS, étant donné la possibilité de limiter la recherche à un descripteur donné, en premier temps on s'est limité à l'espagnol et, par la suite, on a introduit tous les descripteurs:

1	..LIM ESPAGNOL
2	RECONNAISSANCE AUTOMATIQUE
3	ANALYSE AUTOMATIQUE
4	SYNTHESE AUTOMATIQUE
5	TRADUCTION AUTOMATIQUE
6	INDEXATION AUTOMATIQUE
7	3 OU 5

Remarque : les question 2,4 et 6 ont donné un résultat de "0" reponses.

Sur PASCAL, la stratégie a pris la même forme, mais avec les descripteurs de son lexique et en limitant la langue à la fin :

1	TRAITEMENT AUTOMATIQUE OU ANALYSES SYNTAXIQUE
2	TRADUCTION AUTOMATIQUE OU INDEXATION AUTOMATIQUE
3	RECONNAISSANCE FORME OU RECONNAISSANCE AUTOMATIQUE
4	1 OU 2 OU 3
5	4 ET ESPAGNOL

L'utilisation de l'opérateur booléen "OU" a permis la récupération des notices indexées sur n'importe quel descripteur.

3.4. RESULTATS

La recherche automatisée a donné, en tout, 27 références qui ont été visualisées en ligne.

Un premier examen a permis de conclure que si l'ensemble des références se rapportait bien aux différentes mots-clés choisis, pourtant moins de la moitié était en corrélation plus ou moins étroite avec le sujet qui avait été défini, et seulement deux références étaient tout à fait pertinentes (dont une avait été repérée manuellement).

Des 17 références visualisées en PASCAL, 7 ont été retenues. FRANCIS a donné 10 références dont 2 retenues. Ce qui fait un taux de pertinence de 35% pour PASCAL et de 20% pour FRANCIS.

Ces problèmes de manque de pertinence tiennent, en grande partie, au fait qu'on a utilisé des descripteurs généraux.

En outre, la recherche manuelle a donnée des résultats voisins, ce qui m'amène à conclure que le silence du système documentaire est dû à l'absence de références dans les fonds interrogés. Le vide doit être attribué à la spécificité du sujet et à la rareté des travaux qui lui ont été consacrés.

4. REPERAGE DES DOCUMENTS PRIMAIRES

L'ensemble recherche manuelle et automatisée a fourni au total 17 références, dont 1 doublon. Si on ajoute les 16 références obtenues par les contacts personnels, on a un total de 32 documents primaires à repérer.

Dans cette partie du travail, les services de l'Ecole Nationale Supérieure de Bibliothécaires ont été d'une très grande efficacité, d'une part pour la consultation de thèses et ouvrages par le prêt-inter-bibliothèques, et d'autre part pour la localisation des périodiques grâce au Catalogue Collectif National.

De plus, des contacts en Espagne m'ont permis l'accès aux documents.

5. CONSIDERATIONS GENERALES

Quant à la méthodologie :

Ce travail m'a permis d'appréhender de façon concrète les problèmes qui se posent lors d'une recherche bibliographique tant manuelle qu'automatisée.

Les résultats obtenus marquent l'inadéquation de la méthodologie classique de recherche au sujet traité. Le contact avec les chercheurs qui travaillent sur ce domaine, s'est avéré comme le moyen qui contribue le plus à fournir "l'état des arts".

Il faut souligner ici que, par l'intermédiaire de Sylvie Lainé du laboratoire d'Informatique Documentaire, j'ai eu connaissance qu'il existe en France des groupes qui font des recherches dans ce domaine. Pourtant aucune trace n'apparaît dans les systèmes documentaires. Ce qui m'incite à poser la question : Où sont-elles recensées?

Quant à la recherche :

Une stricte limitation au sujet a fourni si peu de références, qu'on a du retenir des travaux théoriques de modélisation et d'autres documents sur le traitement automatique, pour analyser leurs composants.

Quant au sujet :

Tout au long de la recherche, j'ai pu constater l'orientation différente qui a marquée historiquement la linguistique espagnole.

On s'est intéressé à l'informatique comme point de départ pour faire des analyses stylistiques : pour cerner le vocabulaire d'une époque ou d'un auteur, ou son changement au cours de diverses époques. C'est l'informatique au service de la philologie, on y exploite les possibilités les plus immédiates de l'ordinateur, c'est à dire sa grande capacité de calcul. Mais on n'a pas mis la linguistique au service de l'informatique, comme base théorique pour une application.

TROISIEME PARTIE

**THEORIE DE CAS
DE
BERNARD POTTIER**

A.L'ANALYSE AUTOMATIQUE DES LANGUES NATURELLES

L'analyse automatique d'une langue precise de la conception d'un modèle ou simulation abstraite.

Ce modèle est un ensemble de termes primitifs et un ensemble d'opérations qui, exécutées sur ces termes primitifs, permettent de construire la simulation de l'original.

La confection d'un modèle grammatical a comme base l'élaboration d'une théorie, dont la tâche principale sera de définir par abstraction, à partir du réel observable, un objet stable, homogène et clos.

Ce n'est que si certaines correspondances entre l'original et le modèle sont respectées, que l'on peut parler du caractère démonstratif du modèle.

Dans la conception d'un modèle il y a deux niveaux

- > analyse des constituants (mots) pour dégager les unités linguistiques : modèle morphologique
- > établissement des rapports entre les unités dégagées, c'est à dire leur regroupement en syntagmes : modèle syntaxique

1. L'ANALYSE MORPHOLOGIQUE

1.1. LA MORPHOLOGIE

La morphologie est la partie de la science linguistique qui étudie les règles gouvernant la structure interne des mots, leurs formes grammaticales.

1.2. LES ANALYSEURS MORPHOLOGIQUES

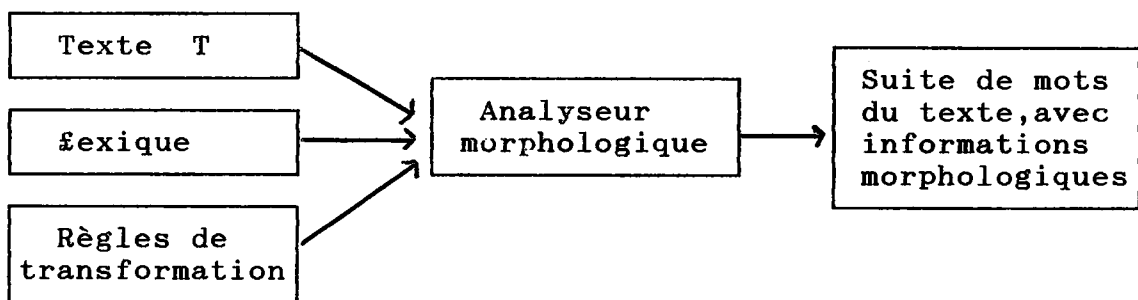
"Un analyseur morphologique est un programme d'opérations automatisé s'appliquant à un texte orthographique T et une grammaire de référence G, pour obtenir une représentation métalinguistique conventionnelle R" (RAMB)

Un analyseur morphologique sert à décrire les éléments de base constitutifs d'une phrase que sont les mots. Il étudie alors leur structure interne et leurs différentes formes.

Diverses méthodologies ont été utilisées, qui en général donnent sur ceux-ci des informations plus ou moins détaillées :

- la catégorie morphologique ou lexicale dans laquelle on les range
- ainsi que des variables supplémentaires que l'on utilise pour lever les ambiguïtés

Exemple d'analyseur morphologique (RAMB) :



Le produit de sortie d'un analyseur morphologique comporte donc, deux sortes d'information :

- > reconnaissance des mots, par isolation de chaque unité minimale du texte
- > catégorisation, par affectation des traits linguistiques

L'élaboration d'un analyseur morphologique efficace est avant tout un problème d'ordre linguistique.

Les problèmes rencontrés dans la conception de tels analyseurs concernent principalement :

- > le choix arbitraire des règles morphologiques
- > le choix arbitraire des catégories morphologiques, au niveau du découpage des mots et au niveau du choix précis de la catégorie pour un mot donné
- > la génération des ambiguïtés d'ordre technique (dues aux mécanismes d'analyse ou à la pauvreté de la représentation typographique) et des ambiguïtés inhérentes à une langue

Le résultat d'une analyse morphologique ne vaut que par son utilité au regard de ce qu'on envisage d'en faire par la suite. Il se justifie dans la mesure où il peut servir de base à d'autres formes de traitement automatique. Sa raison d'être ce sont les composants ultérieures: analyse syntaxique, module d'indexation automatique, etc (BERR)

2. L'ANALYSE SYNTAXIQUE

2.1. LA SYNTAXE

La syntaxe est la partie de la science linguistique qui étudie l'organisation des mots dans le discours, les combinaisons appropriées de mots en unités plus vastes.

On peut étudier la syntaxe à deux niveaux :

- de la phrase entière (on aura besoin d'une méthode linguistique très complexe)
- de segments de phrase (ou syntagmes)

2.2. LES ANALYSEURS SYNTAXIQUES

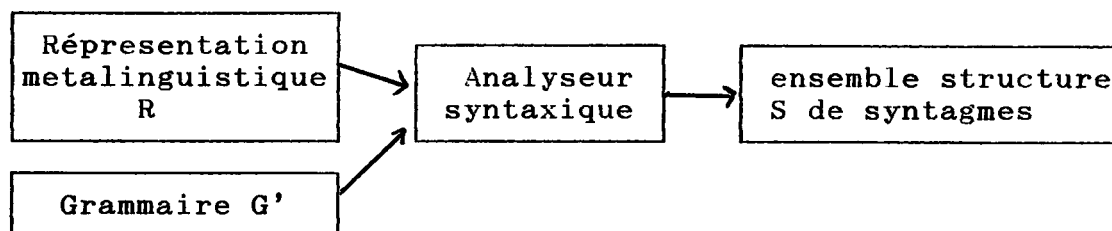
"Un analyseur syntaxique est un ensemble d'opérations automatisé s'appliquant à la représentation métalinguistique conventionnelle R, résultant d'une analyse morphologique d'un texte T, et à une grammaire G' de règles de regroupement et d'accord de mots pour obtenir un ensemble structuré S de syntagmes" (RAMB)

Les analyseurs syntaxiques rendent compte de la validité d'une phrase par rapport à une grammaire donnée, c'est à dire l'exactitude de la combinaison des mots.

Les premiers analyseurs ont été des analyseurs syntaxiques généraux qui utilisent des grammaires non contextuelles quelconques. Après ont été développés des analyseurs particuliers qui ne s'appliquent qu'à une certaine classe de grammaires non contextuelles.

Les méthodes utilisant les analyseurs généraux posant des problèmes, d'autres modèles linguistiques ont été étudiés. C'est le cas des grammaires de dépendance, qui définissent des relations de dépendance entre catégories de mots.

Exemple d'analyseur syntaxique (RAMB) :



Le produit de sortie d'un analyseur syntaxique comporte deux sortes d'informations :

- > identification des syntagmes
- > mise en évidence de l'organisation de chaque syntagme

Les problèmes rencontrés dans la conception de tels analyseurs concernent principalement les ambiguïtés syntaxiques :

- > ambiguïtés artificielles, liées à la structure de la grammaire et de l'analyseur utilisés
- > ambiguïtés structurelles, liées à la langue

3. INTERDEPENDANCE : L'ANALYSE MORPHOSYNTAXIQUE

La séparation des formes morphologiques et fonctions syntaxiques ne convient pas comme méthodologie, c'est pour ça qu'on parle de morphosyntaxe.

L'analyse morphologique et l'analyse syntaxique doivent s'entredeterminer et ne former qu'une seule entité. (RAMB).

L'analyse morphologique jusqu'à présent, se contentait de reconnaître et étiqueter les formes de surface d'un texte, indépendamment de l'analyse syntaxique. Mais les catégories morphologiques doivent être choisies sur la base de véritables propriétés syntaxiques et non plus sur les seuls critères morphologiques.

Pour pouvoir servir à la reconnaissance des structures syntaxiques, il convient que les catégories fonctionnent comme termes pertinents dans ces structures.

De même, la levée des ambiguïtés morphologiques suppose souvent que l'on ait déjà des informations partielles sur la structure syntaxique dans laquelle se trouve engagée l'unité analysée.

L'utilisation des techniques de traitement automatique des langues, nécessite donc l'emploi d'analyseurs morphosyntaxiques.

Le premier objectif d'un tel analyseur sera d'arriver à un faible taux des phrases non analysables.

B.THEORIE DE CAS DE BERNARD POTTIER

Cette étude a comme but d'analyser le traité de linguistique générale de Bernard Pottier (POTT-1), pour essayer de dégager les modèles nécessaires à une application automatique de la langue espagnole et cela sans altérer la réalité linguistique.

"Tout élément grammatical a un signifiant (son corps phonique) et un signifié. Le signifié est de deux natures : fonctionnel (un élément fait partie d'une catégorie grammaticale qui joue un rôle déterminé dans la constitution de l'énoncé); et sémantique (il signifie quelque chose)" (POTT-2)

substance signifié	forme signifié
signifiant	

Pottier a une subdivision des mots au niveau de la forme, de la fonction et de la signification qui révèle des modèles théoriques susceptibles d'être utilisés pour une analyse automatique.

Chez Pottier on trouve une classification des unités de base et une théorie des niveaux d'analyse syntaxique.

Cette étude comportera donc deux parties, qui correspondent à la morphologie (formes ou catégories grammaticales) et à la syntaxe (fonction des catégories dans l'énoncé). Morphologie et syntaxe sont toujours en relation avec le niveau sémantique.

1. MORPHOLOGIE

1.1. LES UNITES DE BASE

Pour Pottier l'unité initiale, de base est le morphème ou unité minimale de signification.

Au niveau de la forme on a des morphèmes lexicaux (lexèmes) et des morphèmes grammaticaux (grammèmes).

Le **lexème** de Pottier correspond à ce que la grammaire appelle "radical". Il est l'élément du mot qui porte la substance predicative. Cet élément, qui déterminera entre mille objets et notions du réel, la substance unique actualisée par le mot.

Le lexème correspond au niveau sémantique à ce que Pottier appelle "designation".

Le **gramème** correspond à la forme grammaticale d'une lexie (mot simple ou groupe de mots résultant d'une intégration sémantique ou d'une lexicalisation d'une séquence ou énoncé).

Le gramème correspond au niveau sémantique à ce que Pottier appelle relations et formulations. À partir des relations et formulations on a, au niveau de la forme, toute une variété de gramèmes qui permettent d'identifier les catégories grammaticales.

Les lexies prennent le rang de catégorie grammaticale grâce aux gramèmes qui correspondent au type de catégorie.

La lexie a donc un aspect significatif comme langue et un aspect fonctionnel comme catégorie.

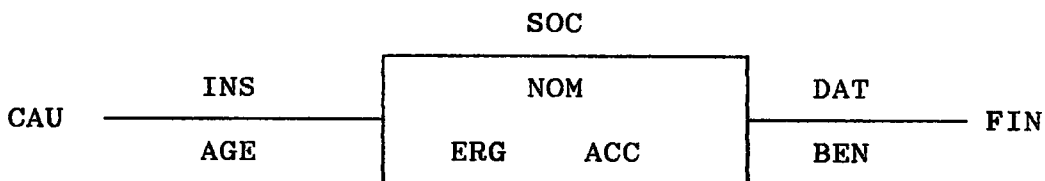
On va faire le point sur ce que Pottier appelle les relations et les formulations, puisque tout au long de cette étude on va se rapporter.

Les **RELATIONS** sont les liens établis entre les constituants d'un énoncé.

Il y a des marques formelles qui déterminent les niveaux et les types de relations.

Pottier distingue 4 types de relations :

- a) La voix, qui peut être :
 - attributive : *ella duerme*
 - active : *ella corta...*
- b) Le système casuel, ayant 11 zones :



CAU : causatif	INS : instrumental	AGE : agentif
SOC : sociatif	NOM : nominatif	ERG : ergatif
AC : acusatif	LOC : locatif	DAT : datif
BEN : bénéfactif	FIN : final	

- c) La vision ou point de vue du locuteur, ayant trois niveaux :
- orientation
 - directe (*el Sr. Garcia ha comprado la casa*)
 - inverse (*la casa ha sido comprada por*)
 - la visée ou point de départ (*la casa* ou *la compra*)
 - sélection ou nombre des éléments retenus
- d) Les intégrations ou passage d'une syntaxe libre à une syntaxe en voie de fixation. On a :
- [adjectivation : *la politica economica italiana*
 - [synthese : *lava-vajilla*
 - [aspectivation :
 - [d'un substantif : *el productor*
 - [d'un adjectif : *refrescante*
 - [des formes nominales du verbe
 - [relativation :
 - [parasyntese : *embarcar*
 - [relativation simple : *interurbano*

Les **FORMULATIONS** sont déterminées par l'intention du locuteur.

Pottier distingue deux types de formulation :

- > Liées au procès de communication (modales et locutives)
- > Descriptives (indications qualitatives ou quantitatives)

a) La formulation modale :

- Modalité ou critique subjective du propos :
Yo creo que Pedro es un insoportable
- Assertion (asertion;interrogation;negation ou emphase)
Yo no creo que Pedro sea insoportable
- Déroulement :
Yo empiezo a creer que Pedro es insoportable
- Détermination :
 - Article
 - Présentation
 - Substituts
 - Application

b) La formulation locutive :

- Personne
- Locution (relation entre personnes)
- Deixis :
 - spatiale : *aqui, alla, aca, alli, ahi*
 - temporelle : *ayer, hoy, manana*
 - notionnelle : *asi, de otro modo*

c) La formulation qualitative. On reproduit ici le
schema classificatoire de Pottier

NATURE

PROPIETES

Structure :

- matériel
- continuun

Dinamiques :

- transitif
- transformatif

Vitalité :

- animé
- sexué
- vivant
- puissance

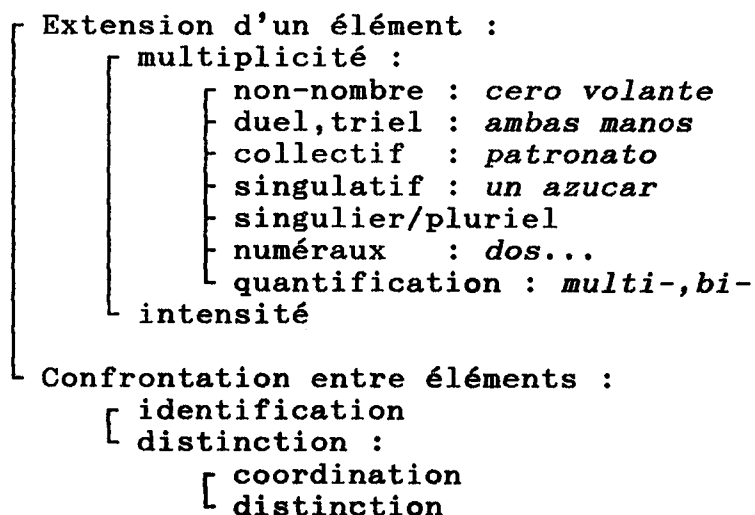
Extension :

- perfectif
- duratif

Divers :

- spécificité
- aliénabilité
- classificateurs variés

d) La formulation quantitative peut s'appliquer à
l'extension d'un élément ou à la confrontation entre
éléments.



1.2. LA CLASSIFICATION DES UNITES DE BASE

A partir de la distinction lexeme/grameme, Pottier établi les types de catégories suivantes :

---> Lexies composées de lexème et grammeme :

- Substantif
- Verbe
- Adjectif

---> Lexies grammaticales (seulement gramemes)

- Particules :
 - Indicateurs
 - Determinants
 - Quantificatifs
 - Auxiliaires
 - Relateurs
- Substituts
- Afixes

Ensuite on va analyser chaque catégorie

1.2.1. SUBSTANTIFS (Sb)

Du point de vue de la forme un substantif est du type :

| L, G1 |

G1 = +Q, C

Où L c'est le lexème et G1 les gramemes ou formants. Ces formants sont de deux classes :

- Constitutifs : qui peuvent être du type quantitatif (nombre) et qualitatif (sexe et genre)
- Facultatifs : parmi lesquels on distingue les infixes, prefixes et suffixes

re	-	distribu-	-	cion	-	es
-	-	-	-	-	-	-
prefixe		L		suffixe		Q

1.2.2. VERBES (Vb)

Le verbe est de la forme :

| L, G2 |

G2 = +F, +Q

Où G2 sont les formants constitutifs, inhérentes à la catégorie verbale :

- F = formulation modale (ind. et subj.) et locutive (deixis temporelle ou époque)
- Q = accord de personnes

La base du système verbal est constituée par le mode et l'époque

habl - ar - a

 époque mode

1.2.3. ADJECTIFS (Adj)

On retrouve dans l'adjectif les éléments formateurs qui composent le substantif.

L'adjectif est de la forme :

| L, G3 |

G3 = C, Q

Où G3 ce sont les morphèmes d'accord :

- C = sexe et genre
- Q = nombre, ou *-mente* dans le cas d'incidence verbal

Peut aussi comporter des formants facultatifs, comme le substantif.

archi - millon - ario - s

 prefixe L C Q

L'adjectif et le substantif, quoique possédant les mêmes éléments formateurs, se différencient d'abord par la fonction (l'adjectif dépend du substantif) et par la nature du gramème genre : le substantif apporte avec lui un genre, l'adjectif non.

1.2.4. INDICATEURS (I)

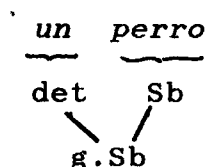
Les indicateurs, liés à la formulation modale et dont l'incidence est variable :

no, aun, sobre todo

1.2.5. DETERMINANTS (Det)

Le déterminant est l'indice habituel de l'intégration du substantif dans le discours.

Du point de vue de la forme il est toujours intégré au groupe substantival.



Le déterminant est en relation avec la formulation modale.

1.2.6. QUANTIFICATIFS (Quant)

Explicitent la pluralité de la catégorie du nom. Ils ont des formes très diverses :

nada, algo, cualquier cosa, cada cosa, todo
en ninguna parte, en alguna parte, nunca, etc

Ils sont liés à la formulation quantitative.

1.2.7. AUXILIAIRES (W)

Ce sont des éléments de la classe verbale qui correspondent à la formulation modal :

- > déterminés par la modalité (Wm)
querer, hacer, decir
- > déterminés par le déroulement (Wd)
ser, haber, ir, estar

On retrouve dans les auxiliaires les éléments formateurs des verbes.

1.2.8. RELATEURS (Rel)

Les relateurs, ou éléments de relation sont invariables quant à la forme (ni genre, ni nombre), puisqu'ils n'expriment qu'une relation entre deux syntagmes. On trouve trois types de relateurs :

- liés à la classe de relations : ce sont des grammemes qui explicitent le système casuel, la fonction syntaxique de l'élément qui suit
a, de, con, a causa de
- liés à la classe des formulations modales. Ce sont les liens modaux, qui expriment le degré de relation entre ce qui lui précède et ce qui lui suit
si, que
- liés à la classe de formulations quantitatives :
 - coordinatifs :
 - d'égalité (*o*)
 - d'addition (*y*)
 - de subtraction (*pero*)
 - comparatifs :
 - d'égalité (*tanto*)
 - de superiorité : (*mas*)
 - d'inferiorité (*menos*)

1.2.9. SUBSTITUTS (Z)

L'économie du discours a entraînée la création de trois sortes de substituts :

- Les substituts fonctionnels, qui évitent la répétition d'un même mot
- Les substituts lexicaux qui condensent, sous une forme simple, une tournure analytique lorsque certaines conditions se présentent
- Les substituts génériques

1.2.9.1. Les substituts fonctionnels

* Des substantifs :

- les pronoms personnels (*lo, le, la, ella...*)
- les adjectifs indéfinis (*ciertos...*)
- les adjectifs numéraux (*tres...*)
- les possessifs (*mi...*)
- les articles (*el, la...*)
- les démonstratifs (*ese, aquel...*)
- les interrogatifs ,dont la réponse se fait aussi avec des substituts (*quien, que, nada, nadie...*)

* Des adjectifs :

- les articles

1.2.9.2. Les substituts lexicaux

* Des adjectifs possessifs :

- de + personne = *mi, tu...*

* Des démonstratifs :

- *el libro que esta aqui = este libro*

* Des locatives, au niveau de :

- deixis spatiale : *aqui, alla...*
- deixis temporelle : *hoy, ayer...*
- deixis notionnelle : *asi*

* Des relateurs :

- domaine temporel :
 - a + que = cuando*
 - en + que = cuando*
- domaine spatial : *donde*

* Des quantitatifs :

- domaine temporel : *como*

* Substituts d'assertion, au niveau de l'énoncé : *si, no*

* Plus les interjections et les "muletillas" :

ah, y luego

1.2.9.3. Les substituts génériques

La syllabe générique "que", dont l'espagnol se sert pour une séquence non identifiée :

paralele "que"?

La lexie générique, c'est à dire les nouveaux lexèmes qui ont les propriétés formelles de la catégorie désirée et est en attente de signification :

se puede "tipotar" en Espana

1.2.10. AFIXES

Les afixes sont des particules liées morphologiquement à un lexème. Trois positions sont possibles :

- préfixe : *sobre-, ex-, des-, in-, ni-, per-, re-, bi-*
- infixe : *-o-, -i-, -it-*
- suffixe : *-oso, -cion, -orum, -iz, -isimo*

2. SYNTAXE

La syntaxe étudie l'emploi fonctionnel et la combinaison distributionnelle des unités de base, des signes linguistiques.

Chez Pottier on trouve une théorie des niveaux d'analyse syntaxique, dont on va essayer de faire le point.

2.1. L'ANALYSE SYNTAXIQUE

Un message doit satisfaire un minimum formel au niveau syntaxique, que Pottier appelle énoncé.

L'analyse syntaxique a donc comme base l'énoncé. Cet énoncé est composé d'un nucléus et d'éléments facultatifs marginaux :

EN = +NU †MA

Pedro vendra si tu quieres
+NU †MA

Le nucléus se caractérise en espagnol par la combinaison d'un élément de fonction nominal (Base) déterminé par le choix sémantique, et d'un Predicat qui peut être un groupe verbal, adjectival ou substantival. D'ici les 4 types de fonctèmes, ou unités fonctionnelles constitutives d'énoncé :

FN : fonctème nominal
FA : fonctème adjectival
FV : fonctème verbal
FN' : fonctème nominal de la Base

Au niveau syntaxique, l'énoncé est appelé syntactème, dont Pottier distingue 4 types :

FN <-- FN'
FN <-- FA
FN <-- FV1
FN --> FV2

On va analyser toutes ces unités.

2.2.LES NIVEAUX D'ANALYSE SYNTAXIQUE

2.2.1.LES FONCTEMES

Pottier fait la distinction entre F par nature (classés comme F par une langue) et F par transfert, créés par une équivalence fonctionnelle entre deux formes et qui viennent déterminés par le choix sémantique

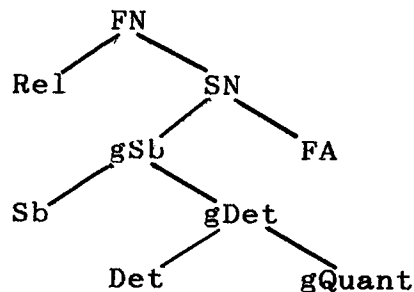
2.2.1.1. LE FONCTEME NOMINAL (FN)

Les FN par nature

=====

Comme on a déjà dit, en espagnol la base est un FN qui assure la fonction nominale.

On reprend ici le schéma de Pottier pour définir la composition d'un FN



Les éléments facultatifs peuvent se répéter dans un même FN :

- subordonnés : *la politica francesa del acero*
(adj1) (adj2)
- coordonnés : *la dificil pero necesaria iniciacion*
(adj1) (adj2)

Les FN varient selon le système casuel déterminé par les relations entre les constituants d'un énoncé. On aura donc :

- FN NOM : *Yo soy un hombre*
- FN AC : *Pedro escucha la cancion*
- FN CAU : *Pedro ha sido herido por su culpa*
- FN INS : *Pedro ha sido herido por un obus*
- FN AGE : *Pedro ha sido herido por Juan*
- FN SOC : *Pedro y Juan....*
- FN DAT : *Pedro da una muneca a Juan*
- FN BEN : *Pedro da una muneca para Maria*
- FN FIN :
- FN LOC : *Pedro llevo antes de tu salida*

Le FN par transfert
=====

Au niveau de la fonction il est possible de substantiver des mots ou groupes de mots qui, par nature, ne sont pas destinés à être des substantifs. Pottier distingue les cas suivantes :

a) EN + que

- En situation FN NOM
que hayas ido me extrana
el hecho de que hayas ido

Remarque : Ici il faut bien differentier *el hecho de que* (FN par transfert) de *el hecho que* (FN par nature suivi d'une adjectivation)

- En situation de FN AC
yo creo que esto es cierto
- En situation de cas marginaux
FN LOC : *antes que tu salgas*
FN FIN : *para que tu salgas*
FN CAU/FIN : *para si tu sales*
FN LOC : *cuando tu sales*
FN CAU : *si tu sales*

b) EN + infinitivation

En espagnol n'ayant pas un agent explicite pour l'infinitif, on a une forme générique

- FN NOM : *viajar es siempre agradable*
- FN AC : *me gusta viajar*

c) EN + aspectif

Est déterminé par l'option sémantique de la vision
aceptar inmediatamente
la aceptacion inmediata
el aceptar
que aceptes inmediatamente

d) Nominalisation généralisée

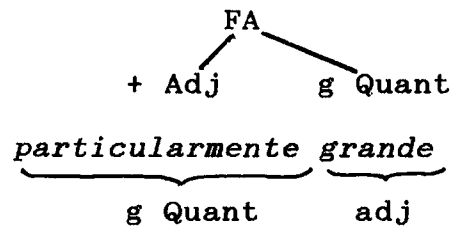
1914 es una fecha celebre
manana empieza hoy

2.2.1.2. LE FONCTEME ADJECTIVAL (FA)

Le FA par nature

=====

La composition d'un FA est la suivante :



Il y a trois types de FA :

- Adjectif suivi d'un FN (cas de FA2)
inclinado a sonar
- Resultant de l'integration d'un Predicat (cas du FA qui fait partie de FN)
el patio cuadrado
- Incident à un verbe (prend *-mente*)
Pedro obra generosamente

Le FA par transfert

=====

C'est le cas de ce que la grammaire traditionnelle appelle relative :

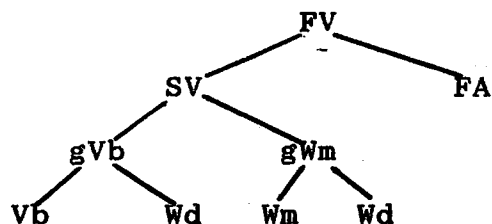
el tren que llega a las 8

Un cas particulier de l'adjectivation c'est l'aposition :

Mozart, musico genial

2.2.1.3 LE FONCTEME VERBAL (FV)

Fonctionnellement le FV est le second élément constitutif d'un énoncé viable. Se présente comme ça :



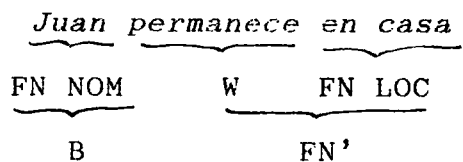
On a deux types de FV :

- FV1 (verbes intransitifs) : *marchar*
- FV2 (suivi d'un ou plusieurs FN) :
llenarse de algo

2.2.2.LES SYNTACTEMES

2.2.2.1. FN <-- FN'

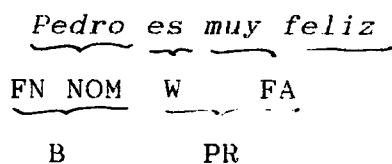
C'est la constitution d'un énoncé ayant comme Base un FN et un FN' qui fait partie du Predicat



Dans le cas des énoncés impersonnels de base 0, on n'aura que le FN'

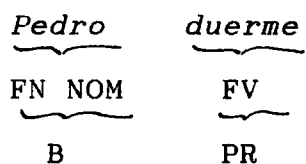
2.2.2.2. FN <-- FA

C'est la constitution d'un énoncé ayant comme Base un FN et comme Predicat un FA



2.2.2.3. FN <-- FV1

C'est le cas de la construction d'un énoncé ayant comme Base un FN et comme Predicat un FV constitué d'un verbe intransitive. L'énoncé a un seul actante.



2.2.2.4. FN <-- FV2

C'est le cas d'un enoncé constitué d'une Base FN
et un Predicat composé d'un FV et d'un ou plusieurs FN

<u>Pedro</u>	<u>come</u>	<u>algo</u>
FN NOM	FV	FN AC
<u>B</u>	<u>PR</u>	

C.CONCLUSION

La théorie de cas qu'on dégage du traité de Pottier est essentielle pour l'analyse automatique de l'espagnol.

La subdivision des mots et des niveaux syntaxiques qu'on vient d'analyser, est une modelisation qui peut être appliquée au traitement automatique de la langue espagnole.

Pottier n'oublie pas qu'un grammairien ne peut pas faire totalement abstraction du niveau de la signification, et à partir de sa théorie on peut exprimer le niveau de contenu d'un modèle par des moyens strictement formels.

Il introduit le problème du sens dans l'étude du langage et constitue sans doute l'un des rares représentants du structuralisme linguistique en sémantique. Reste à l'écart des courants générativistes transformationnelles, occupant une position à part par rapport aux structuralistes classiques.

Pottier représente une voie originale, dont le développement sur bien des points se rapproche des recherches nées dans un cadre poststructuraliste.

Il essaye de limiter le nucléus central des mécanismes linguistiques, pour en faire la base d'autres études sur le langage.

Le but cherché dans ce travail était la recherche des modèles linguistiques de l'espagnol, on voit donc que l'état de la question ce n'est pas vide.

Du fait que la théorie des modèles linguistiques fait partie de la linguistique générale, dans un traité générale on y trouve tous les éléments.

Cela nous amène à remarquer l'insuffisance des techniques documentaires qui n'analysent pas les ouvrages généraux, son centre d'intérêt étant les publications périodiques.

ANNEXE

LISTE D'OUVRAGES CITEES DANS CETTE PARTIE

- (BERR) BERRENDONNER, A. Grammaire pour un ^{aj}analyseur : aspects morphologiques. Document du travail du groupe SYDO 1984
- (POTT-1) POTTIER, Bernard. Linguistica general : teoria y descripcion
- (POTT-2) POTTIER, Bernard. Introduction à l'étude de la morphosyntaxe espagnole : étude structurale
- (RAMB) RAMBURRUN, M. Implémentation en Prolog-Foll d'un analyseur morpho-syntaxique du français écrit : extraction de syntagmes nominaux. Thèse de 3ème cycle. Université Lyon II. 1985

BIBLIOGRAPHIE

INTELLIGENCE ARTIFFICIELLE

Généralités

VERDEJO, Felisa. Analisis predictivo y representacion del conocimiento. *CIL 81*, Convencion Informatica Latina, Barcelone Juin 1981.

VERDEJO, Felisa. Representacion del conocimiento y comprension del lenguaje. *NOVATICA*, vol.5, n° 38-39, p.35-42

LINGUISTIQUE STATISTIQUE ET QUANTITATIVE

KOCK, Josse de. Metodo para una posible automatizacion del analisis estilistico. *Linguistica espanola actual*, 1981, n°3, p.305-336

ROLFE, Oliver W. Morphological frequency : French and Spanisch verbal themes. In *Papers on linguistic and child language*. The Hague : Mouton, 1978, vol.8, p.211-226

MODELES FORMALISES

Généralités

ANDREEWSKI, A., DESI, M., FLUHR, C. Méthodes d'apprentissage pour l'analyse automatique, morphosyntaxique et lexicale sémantique de la langue espagnole. *COLING 82*, 9th International conference on Computational linguistics, Prague 5-10/VII 1982, p.11-16

Modèles morphologiques

MARTI, M.A. Un sistema de analisis morfologico por ordenador. *Procesamiento del lenguaje natural*, Octovre 1986, n°4

MEYA LLOPART, Montserrat. Analisis morfologico como ayuda para la recuperacion de informacion. *Procesamiento del lenguaje natural*, Octovre 1986, n°4, p.91-108

PASTOR, N. *Etude morphologique de l'espagnol : les formants catégorisateurs*. Nancy : Publications linguistiques de la Faculté des Lettres et des Sciences Humaines, 1969

Modèles morphosyntaxiques

BEYL, D.W. *A linguistic Presentation of the PAHO Machine Translation System*. Thèse. Georgetown University. 1978

KLENZ, Ursula. Kontextfreies Parsing angewandt auf elliptische Sätze in spanischen Wetterberichten. - *Sprache und Datenverarbeitung International Journal for languages Data Processing*. Saarbrücken, 1983-1984, n°7, p.21-25

SOPENA, Luis. Grammar of Spanish for User Specialty Languages, TR 82.05.004, IBM Germany, Heildeberg Scientific Center, 1982

SOPENA, Luis. Natural Language Grammars for an Information System. *ACM SIG Information Retrieval Forum*, 1983, vol.17, n°4, p.75-80

SOPENA, Luis. *Aportaciones a la comprension del castellano en un sistema de interrogacion de bases de datos relacionales*. Thèse de doctorat. Université de Pays Basque. 1983

Modèles sémantiques

DAHL, Veronica. *Un système déductif d'interrogation de Bases de données en espagnol*. Thèse. 3ème cycle. Aix-Marseille. 1977

DAHL, Veronica. Translating Spanish into Logic through Logic, *American Journal of Computational Linguistics*, 1981, vol.7, n°3, p.149-163

VERDEJO, Felisa. *Un analizador semantico para un sistema Pregunta-Respuesta en lenguaje natural*. Thèse de Doctorat. Université Complutense de Madrid. 1980

VERDEJO, Felisa. *Etude du langage naturel, simulation d'un robot capable de mener un dialogue en espagnol*. Thèse. 3ème cycle. Paris 6. 1975

Implémentations

GARCIA CAMARERO, E., et al. Un metodo de respuesta a preguntas formuladas a un grafo semantico. Congrès de l'Asociacion Espanola de la Informacion Automatica, Madrid, Octovre 1979

GARCIA CAMARERO, E., et al. Una red semantica para el dominio arqueologico, *Boletin del Centro de Calculo de la Universidad Complutense*, Decembre 1978

GARCIA CAMARERO, E., et al. Red semantica para la gestion de una base de datos linguisticos. *CIL-79*, Convencion Informatica Latina, Barcelone, Juin 1979

Etude comparative

SOPENA, Luis. Sistemas Pregunta-Respuesta en castellano : un resumen comparativo. Convencion Informatica Latina. Barcelone, 1983, p.77-90

TRADUCTION AUTOMATIQUE

DUCROT, J.M. Mise en application et en exploitation operationnelle d'une méthode de traduction automatique de textes documentaires en vue d'accroître leur utilisation dans le monde. DGRST-7371198, 1973

GUZMAN DE ROJAS, I. La informatica en la investigacion de nuestra cultura. *Terminologie et traduction*, 1985, n°2, p.47-62

MEYA LLOPART, Montserrat. *Procesamiento de datos lingüísticos : modelo de traduccion automatica del espanol al aleman*. Madrid : Fundacion Juan March, 1978. 47p. Serie Universitaria

VALLE BRACERO, A., FERNANDEZ, J.A., MORALES FERNANDEZ, R. Separacion automatica de lexemas, sufijos y morfemas y su aplicacion a la traduccion automatica. *Revista Espanola de documentacion cientifica*, 1984, vol.7, n°3, p.185-192

VALLE BRACERO, A., FERNANDEZ GARCIA, J.A. Traduccion automatica de titulos de articulos cientificos del ruso al castellano. *Revista espanola de documentacion cientifica*, 1982, vol.5, n°3, p.231-243

VASCONCELLOS, M. Functional considerations in the postediting of machine-traslated out put : dealing with V(S)O versus SVO. *Computers and traslation*, 1986, vol.1, n°1, p.21-38

ZINGEL, H.J. Experiences with TITUS II. *International Classification*, 1978, vol.5, n°1, p.33-37

TRAITEMENT DE TEXTES

Creation de dictionnaires

CAMPO, Ignacio del, et al. The automatic syntactic analysis as an aid in dictionary making. *COLING* Proceedings of the International conference on Computational Linguistic, Pisa 27-VIII/I 1973, vol.4, n°2, p.161-176

SOPENA, Luis. Linguistica y procesamiento de textos en castellano, *PC WORLD*, 1986, n°8

SOPENA, Luis. Procesamiento de textos en el Centro de Investigacion UAM-IBM. *Procesamiento del lenguaje natural*, 1985, n°3





* 9 5 4 0 9 9 8 *