

**UNIVERSITE PARIS 7**  
**DEPARTEMENT DE RECHERCHES LINGUISTIQUES**  
**LABORATOIRE D'AUTOMATIQUE DOCUMENTAIRE ET**  
**LINGUISTIQUE**

**THESE DE DOCTORAT EN SCIENCES DU LANGAGE**

par

**Mylène BONAN-GARRIGUES**

**METHODE DE PARAMETRAGE DES**  
**DICTIONNAIRES ET**  
**GRAMMAIRES ELECTRONIQUES**  
**Application à des systèmes interactifs en langue naturelle**

Soutenue le 22 janvier 1993 devant un jury composé de :

<b>Dominique PERRIN,</b>	<b>Président</b>
<b>Maurice GROSS,</b>	<b>Directeur</b>
<b>Annibale ELIA,</b>	<b>Rapporteur</b>
<b>Gaston GROSS,</b>	<b>Rapporteur</b>
<b>Laurence DANLOS</b>	
<b>Francis DEBYSER</b>	
<b>Jean-François DEGREMONT</b>	

UNIVERSITE PARIS 7  
DEPARTEMENT DE RECHERCHES LINGUISTIQUES  
LABORATOIRE D'AUTOMATIQUE DOCUMENTAIRE ET  
LINGUISTIQUE

THESE DE DOCTORAT EN SCIENCES DU LANGAGE

par

Mylène BONAN-GARRIGUES

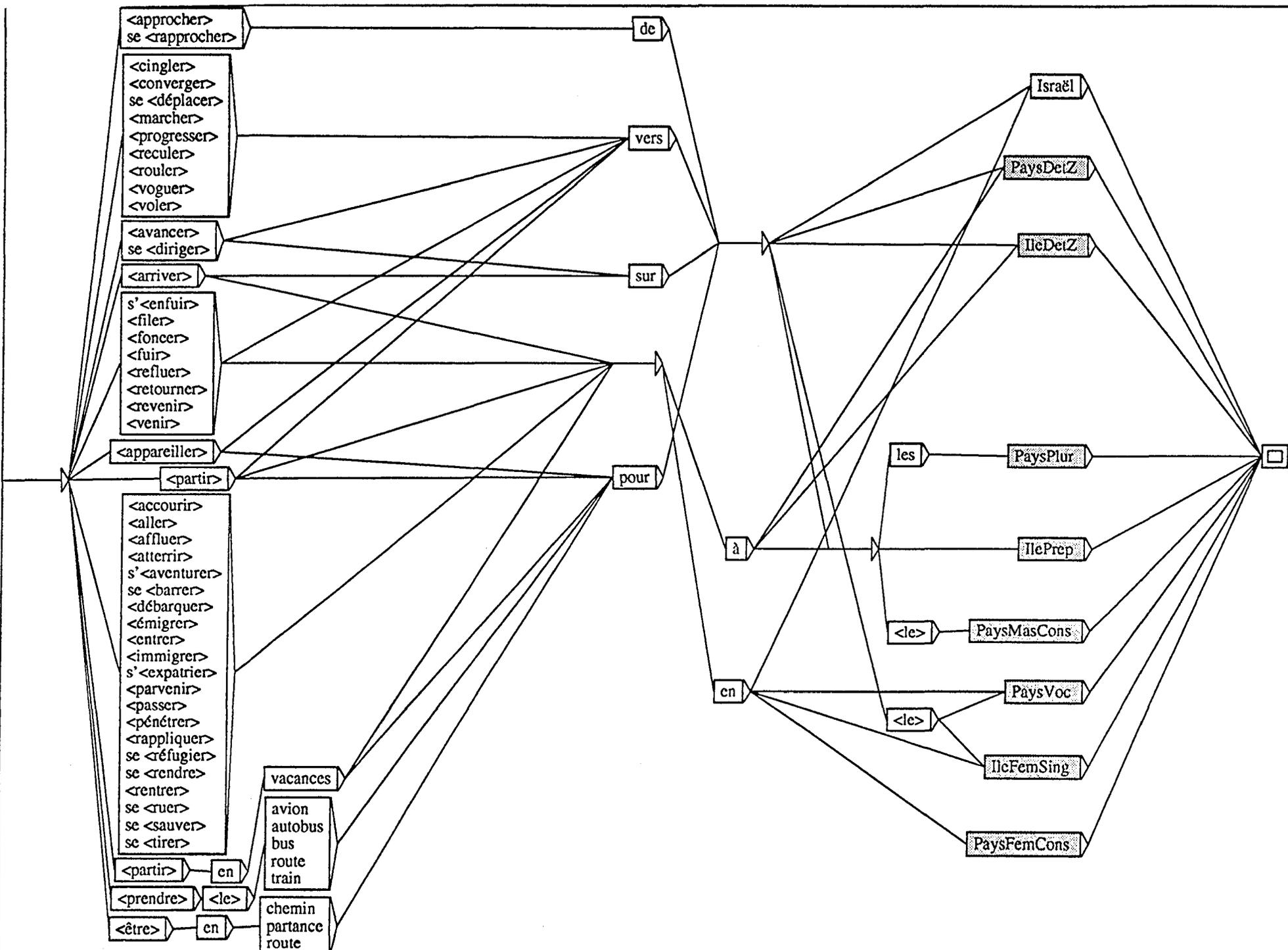
METHODE DE PARAMETRAGE DES  
DICTIONNAIRES ET  
GRAMMAIRES ELECTRONIQUES  
Application à des systèmes interactifs en langue naturelle

Soutenue le 22 janvier 1993 devant un jury composé de :

Dominique PERRIN,	Président
Maurice GROSS,	Directeur
Annibale ELIA,	Rapporteur
Gaston GROSS,	Rapporteur
Laurence DANLOS	
Francis DEBYSER	
Jean-François DEGREMONT	

## Errata

- p.4 Remplacer "auxquels ces dernières s'y réfèrent" par "qui s'y réfèrent".
- p.43 Remplacer "pausable" par "plausible".
- p.53 Remplacer "contenaità" par "contenait à".
- p.59 Remplacer "comter" par "*compter*".
- p.64 Remplacer "aritcles" par "articles".
- p.117 Remplacer "exposée" par "exposées".
- p.129 Remplacer "matériau" par "matériau".
- p.193 Remplacer l'automate NVPrepN par l'automate ci-joint.



## Remerciements

Cette thèse a pu être faite, grâce à une convention entre le CIEP (Centre International d'Etudes Pédagogiques) et le LADL (Laboratoire d'Automatique Documentaire et Linguistique). Je remercie vivement les instigateurs de cet accord : Francis Debyser et Maurice Gross.

Je remercie tous les membres du LADL et du CERIL pour leur esprit coopératif et plus particulièrement David Clemenceau, Blandine Courtois, Emmanuel Roche et Philippe Vasseux.

Je tiens à exprimer ma gratitude à Max Silberztein qui n'a ménagé ni son temps ni sa patience.

Je prie enfin les membres du jury de m'excuser de leur infliger la lecture d'une énième thèse.

# Table des matières

## Introduction

### 1. Partition hiérarchique des dictionnaires électroniques du français

- 1.1. Analyse de l'approche dite objective
  - 1.1.1. L'expérience du français fondamental
  - 1.1.2. Les limites des méthodes statistiques
- 1.2. Approche intuitive et problèmes théoriques
  - 1.2.1. Problématique de la sélection lexicale
  - 1.2.2. La question des deux lexiques
  - 1.2.3. Définition et description
  - 1.2.4. Comprendre un mot a-t-il un sens ?
  - 1.2.5. Inconnus, connus, reconnus : étude de cas
- 1.3. A la recherche du sens
  - 1.3.1. Dictionnaire usuel et lexique interne
  - 1.3.2. Intention signitive et remplissage du sens
  - 1.3.3. Approche phénoménologique de la notion de représentation
- 1.4. La méthode de sélection
  - 1.4.1. Les critères
  - 1.4.2. la méthode
  - 1.4.3. Elaboration de test, analyse et conclusion
  - 1.4.4. Outils et processus de traitement automatique
  - 1.4.5. Résultats et limites
  - 1.4.6. Les prolongements
  - 1.4.7. Les perspectives d'applications

## **2. Dictionnaires électroniques, grammaires locales et applications informatiques en langues**

### **2.1. Les traitement automatique dans la problématique du champ**

2.1.1. Analyse de mots

2.1.2. Analyse de phrases

2.1.3. Analyse de segments d'énoncés dans la langue-cible

2.1.4. Analyse d'énoncés pré-contraints dans la langue-cible

2.1.5. Analyse de réponse en langue maternelle

2.1.6. Les orientations des applications actuelles

### **2.2. Méthodes et outils pour l'élaboration de grammaires locales**

2.2.1. Définition du prédicat

2.2.2. Les modalités et les structures syntaxiques du prédicat

2.2.3. Traitement du matériau de base

2.2.4. Du verbe à la phrase

2.2.5. La sélection d'une préposition : un processus de génération complexe

2.2.6. Réalisation d'automates

**Conclusion**

**Bibliographie**

**Annexes**

## Introduction

Le lexique-grammaire du français représente aujourd'hui un matériau linguistique d'un haut niveau de formalisation. A la différence des matériaux linguistiques traditionnels (dictionnaires et grammaires), il permet la mise en oeuvre des deux grandes familles d'applications informatiques : reconnaissance et génération automatique. Composé de dictionnaires électroniques et de tables syntaxiques qui leur sont corrélées, le lexique-grammaire est associé à un certain nombre d'outils logiciels, eux-mêmes de plus en plus perfectionnés, permettant de les manipuler.

Entrepris il y a déjà longtemps comme un travail théorique de grande envergure, destiné à décrire le fonctionnement de la phrase simple sur l'ensemble de la langue, ce chantier ne sera pas terminé avant longtemps. La masse de matériaux déjà produite permet cependant aujourd'hui d'envisager de les utiliser, sans qu'il soit nécessaire d'attendre que cette tâche considérable soit parvenue à son terme. Par ailleurs, les applications nécessitent que l'on adapte ces matériaux théoriques aux usages ad hoc.

Notre propos était donc d'établir une méthode de paramétrage de ces matériaux en fonction des applications. Nous avons envisagé, dans l'ordre : le problème du paramétrage des dictionnaires électroniques, puis celui des grammaires locales auxquels ces dernières se réfèrent nécessairement.

Dans une première partie, nous verrons par quel cheminement nous sommes parvenue à élaborer une méthode ayant permis de construire une hiérarchie de lexiques électroniques du français à partir du DELAS (Dictionnaire Electronique du LADL pour les mots simples du français). Précisons qu'à notre connaissance, cette partition hiérarchique des mots simples du français n'a jamais été réalisée de cette manière. Nous verrons que

de nombreux prolongements sont à envisager à différents niveaux de la langue, et que les applications en sont variées.

Dans une deuxième partie, nous envisagerons le paramétrage de grammaires électroniques. Les grammaires locales sont en effet destinées à la formalisation d'un point morphosyntaxique particulier. Mais dans quasiment tous les cas, les phénomènes entrent en rapport avec la grammaire complète de la langue. De ce fait, la mise au point d'une grammaire constitue un travail d'une grande complexité. Or, toutes les applications ne demandent pas d'attendre que l'ensemble de la morphosyntaxe du français soit traitée, ni qu'un point donné soit nécessairement traité d'un point de vue exhaustif. Nous montrerons donc que ces grammaires peuvent être élaborées à différents niveaux. S'agissant du traitement de la langue naturelle dans une perspective interactive, nous serons amenée à situer l'utilisation de ces nouveaux outils par rapport aux outils utilisés jusque-là : langages courants de programmation ou aux langages d'auteurs.

Dans nos deux parties, nous nous sommes servie de l'exemple de l'enseignement / apprentissage des langues. Toutefois, les problèmes envisagés et les exemples traités sont communs aux autres domaines d'applications, dès lors qu'ils mettent en oeuvre à un système interactif en langue naturelle.

# 1. Partition hiérarchique des dictionnaires électroniques du français.

## Principes, méthode et résultats

*In alle Wissenschaften kehrt die Not ein,  
letztlich als Not der Methode*<sup>1</sup>

Husserl

Les dictionnaires usuels contiennent à la fois des mots "connus" et des mots "inconnus". Il existe des dictionnaires de mots "inconnus", comme ceux qui ont été élaborés par l'Abbé Prévost (1755) ou par J. Bte Richard de Radonvilliers (1845). On trouve par ailleurs en librairie, des dictionnaires de mots "connus". Mais leur élaboration n'a jamais été entreprise de façon rationnelle. Or la délimitation de couches de vocabulaire, représentant le lexique gradué (du plus au moins essentiel), d'une population ou d'un groupe social, correspond à un nouvel outil conceptuel réclamé aujourd'hui tant dans le domaine pédagogique, que dans celui des industries de la langue (et les projets européens relevant du domaine). Les dictionnaires électroniques constituent des outils formels indispensables à tout traitement automatique de la langue. Les différentes applications ne nécessitent cependant pas d'utiliser des outils aussi volumineux. Il est donc souvent souhaitable, voire nécessaire, de pouvoir effectuer un refaçonnage "sur mesure". Ainsi, des lexiques hiérarchisés, allant du vocabulaire le plus essentiel au moins essentiel, seraient utiles à de nombreuses applications informatisées : élaboration d'interfaces en

---

<sup>1</sup> " Dans toutes les sciences s'insinue la détresse, en dernière analyse comme détresse méthodologique". (*La crise de l'humanité européenne et la philosophie*. Husserl. traduction de P. Ricoeur). Aubier Montaigne. 1977, 1987.

langage naturel pour l'interrogation de bases de données grand public, mise en oeuvre de services télématiques, compression de dictionnaires pour l'analyse automatique de textes, etc.

Dans le domaine pédagogique, aucune réponse satisfaisante n'a été donnée à cette demande sans cesse réitérée. Il existe bien des dictionnaires du français langue étrangère - siglé FLE - (J. Dubois 1978, G. Matoré 1963), mais ceux-ci ne se sont pas donné comme but explicite la représentation du lexique connu d'un groupe de locuteurs. De ce fait, leurs indications méthodologiques restent vagues, voire inexistantes, et les questions théoriques jamais abordées. Or, la délimitation du lexique des mots "connus" d'un ensemble de locuteurs ne va pas de soi. Si certains semblent penser que cette tâche relèverait d'une simple entreprise de nettoyage des dictionnaires usuels (cf. R. Galisson, 1988), les recherches en sciences cognitives indiquent, par l'intérêt qu'elles portent à cette question (T.K. Landauer, 1986), que l'on se trouve devant l'un des points les moins explorés et les plus complexes du domaine de la théorie des connaissances. L'élaboration du lexique commun à un groupe donné, passe par l'exploration du lexique mental d'un individu. Or nous ne connaissons, à l'heure actuelle, pour ainsi dire rien, sur la configuration et le fonctionnement de ce lexique.

Les méthodes employées jusqu'à présent pour tenter d'appréhender un vocabulaire usuel se sont généralement prévalu d'une approche "objective", reposant essentiellement sur le critère de fréquences, le recours à une démarche intuitive étant a priori jugé rétrograde et a-scientifique par les spécialistes du domaine. Paradoxalement, loin de la rejeter, les sciences exactes intègrent certaines intuitions dans leur processus de recherche. Elles ont en effet su distinguer les intuitions reproductibles, recoupables par l'expérimentation, et théorisables, des intuitions non reproductibles qu'elles continuent de rejeter. Nous ferons deux remarques à ce propos :

- Si, dans les méthodes reposant sur les statistiques, les relevés de mots des listes fréquentielles sont bien objectifs, on oublie souvent que la subjectivité n'est pas éliminée pour autant. Elle se trouve déplacée en amont de ce calcul, c'est-à-dire au moment du choix du corpus de textes.

- Il est aujourd'hui possible d'étayer des intuitions de choix de mots, par des expérimentations permettant d'évaluer le taux de couverture d'une sélection

de mots sur des corpus de textes. De nouveaux outils linguistiques et informatiques - dictionnaires électroniques (Bl. Courtois 1989), outils logiciels de navigation dans les textes (M. Silberztein 1989, 1992) - permettent désormais de s'y livrer, à une très grande échelle, et en des temps impensables il y a seulement quelques années.

Avant d'examiner les tenants et aboutissants de la méthode que nous avons développée et utilisée, avec l'aide de tels outils, pour l'élaboration d'une hiérarchie de dictionnaires du français, nous montrerons les limites d'une méthode de sélection lexicale reposant sur une démarche considérée comme "objective", celle du *Français Fondamental*, siglé F.F. Nous nous pencherons ensuite sur deux dictionnaires faisant autorité en matière de fréquence : le *Dictionnaire des fréquences* (1971), siglé DF, réalisé d'après le corpus du *Trésor de la langue française* et le *Dictionnaire de fréquence des mots du français* (J. Baudot 1990). Nous verrons que ce critère de fréquence auquel l'on songe spontanément, s'avère inutilisable pour la sélection de lexiques hiérarchiques d'une langue.

## **1.1. Analyse de l'approche dite "objective" :**

### **1.1.1. L'expérience du *Français Fondamental***

#### **1.1.1.1. Le contexte**

La nécessité de procéder à une limitation du lexique est apparue depuis longtemps, aux enseignants de langue et aux concepteurs de matériaux pédagogiques. Jusqu'à ce que les premières entreprises de ce type offrent des listes de vocabulaire réduites, le choix des mots à enseigner reposait sur l'entière subjectivité de l'enseignant et/ou du concepteur de manuels (souvent réduits à des compilations de textes en majorité littéraires). Des listes arbitraires et indigestes étaient alors proposées aux élèves, avec un profit discutable.

La principale entreprise de délimitation du vocabulaire à des fins d'apprentissage du français pour un public débutant a été l'élaboration du *Français Fondamental*, 1<sup>o</sup> degré (siglé F.F.1) : une liste sélective de 1475 mots (Gougenheim G., Michea R., Rivenc P., Sauvageot A., 1967) parue

avec différents réaménagements entre 1954 et 1959 a été établie selon une méthodologie et des techniques innovatrices à l'époque, combinant :

- l'analyse d'un corpus oral recueilli au magnétophone
- l'établissement de fréquences de mots
- et la recherche de mots "disponibles".

Précisons que ce travail fut conçu comme une première étape répondant à une situation d'urgence et devant mener, par paliers successifs au "français complet" (ibid.,1967). On remarquera que le projet initial visait, comme le nôtre, la réalisation d'un ensemble de dictionnaires gradués des mots du français. Dans les faits, et à la différence de notre expérience, seule une seconde liste d'environ 1900 mots (le F.F. 2) a suivi. Il convient toutefois de noter ici que le F.F. se démarquait radicalement en cela du *Basic English* de Ogden et Richards (conçu en 1928) qui se présentait comme une langue universelle, économe et fermée de 850 mots, dont l'utilisation conduisait non pas à parler anglais, mais à s'exprimer au moyen de ce vocabulaire limité, par des contorsions périphrastiques bien éloignées de la pratique de la langue. "C'est ainsi qu'au lieu de *to ask* on est obligé de dire *to make a request, to put a question* ; au lieu de *to count* : *to get the number of* ; au lieu de *husband* : *married man* , etc." (G. Gougenheim et al., 1967 p. 28).

Soucieuse tout à la fois de mettre fin au règne de l'arbitraire et de travailler sur de la langue parlée, l'équipe du *Français Fondamental* a contribué à ce que l'on envisage désormais l'étude du vocabulaire dans une perspective de mise en pratique et à susciter des questionnements à ce niveau chez les auteurs de méthodes : quel vocabulaire enseigner et pour remplir quels objectifs ? Pourtant, les listes du F.F. sont aujourd'hui désuètes et les auteurs de manuels utilisent, tout en continuant de garder un souci de rentabilité et de fonctionnalité, un vocabulaire choisi de façon pragmatique. Un certain nombre de vocables de base du F.F. se retrouvent nécessairement dans les manuels mais ce matériau ne constitue plus une référence suffisante.

#### 1.1.1.2. Les limites d'une méthode "scientifique"

Cette inadéquation du F. F. aux besoins actuels est liée à plusieurs facteurs :

- le vieillissement inévitable de certains termes (*fourneau, poêle* (masc.), *phono*, par exemple), et son corrélat,
- l'inactualisation de la liste (*aspirateur* ou *vélo* n'y figurent pas, alors qu'ils sont aujourd'hui au moins aussi utilisés que *balai* ou *bicyclette*, qui y figurent).

D'autres raisons, de type méthodologique, expliquent cette désaffection de la part des auteurs de manuels :

- Malgré leur volonté d'adopter une démarche "scientifique" (et d'en écarter tout critère subjectif), en travaillant sur un corpus oral et en s'appuyant sur des indices fréquentiels, les auteurs du F.F. n'ont pas réussi à en extrapoler le vocabulaire nécessaire à la mise en scène des échanges élémentaires de la vie quotidienne. Plusieurs raisons à cela :

- Inadéquation du corpus aux objectifs d'utilisation

Bien que relativement peu directifs, les entretiens destinés à recueillir le corpus oral s'inscrivaient cependant, dans une situation de communication inauthentique : "*Nous priions les personnes de venir chez nous, l'ambiance familière d'une demeure particulière impressionnant moins le sujet que le cadre impersonnel d'un bureau d'étude. Nous l'instruisions sommairement de nos desseins, lui montrions l'appareil et le microphone et engageions la conversation sur un sujet qui lui tenait à coeur: sa famille, sa santé, ses occupations, ses vacances, etc.*". (G. Gougeheim et al., p. 64). Cette situation ne correspondant à aucune des situations d'échanges quotidiens, il n'est pas étonnant que les auteurs de manuels voués eux, par les impératifs méthodologiques de la didactique des langues, au "communicatif" et à l'"authentique" n'aient pas toujours trouvé leur bonheur dans la liste proposée (pourtant après un large remaniement).

Quoi de plus banal, par exemple, que d'aller dans un magasin acheter un vêtement ou une paire de chaussures. Rares sont les méthodes qui ne proposent pas ces situations de réalisations d'échanges langagiers dès le niveau débutant. Or, on ne trouve dans la liste proposée par le F.F.1 "destiné à l'acquisition du vocabulaire nécessaire à la vie quotidienne" (FF1, 1972), ni le mot *taille* entrant dans des énoncés tels que *Quelle taille faites-vous ?* ou *Avez-vous la taille au-dessus ?*, ni le mot *pointure* nécessaire à la compréhension d'énoncés tels que *Nous n'avons pas votre pointure*. On ne trouve pas davantage les mots *croisement* ou *carrefour* nécessaires pourtant à la réalisation et à la compréhension d'énoncés aussi immédiatement fonctionnels que *Après le premier croisement* ou *Au prochain carrefour*. Ces énoncés relèvent pourtant de la catégorie des "notions" et "actes de parole" recensés, plus tard il est vrai, par le Conseil de l'Europe, pour l'acquisition d'un

"niveau-seuil", ou niveau de "compétence générale minimum" (*Un Niveau Seuil*, 1976). On ne trouve pas davantage les mots *horaire, séance, amitié, inviter*, (ni même la symbolique *baguette* française). Mais on trouve les mots *chasse, charbon, chèvre, aiguille, coq, pelle*, etc. dont le caractère d'urgence, pour des échanges usuels, nous échappe. Notons que des mots tels que *séance, amitié* ou *carrefour* sont même absents du F.F.2 (plus particulièrement destiné à la compréhension de l'écrit usuel).

- Les dysfonctionnements de la mise en pratique

Malgré de nombreux ajouts aux listes de fréquence, certains termes de la liste définitive souffrent d'un isolement "fondamental". Ainsi en est-il de la fameuse "*vache fondamentale* dont on ne sait que faire parce qu'on ne peut pas la *traire* et qu'il n'y a pas d'*étable* pour la rentrer ! Il en va de même du *vent*, qui ne *souffle* pas, du *tonnerre* qui ne *gronde* pas..." (R. Galisson, 1971). On pourrait certes rétorquer que les *vaches* peuvent toujours *regarder passer les trains* (du F.F.) et qu'*autant en emporte le vent* ...Toujours est-il qu'en situation d'échange linguistique, on se trouve rapidement bloqué.

Les incohérences apparaissent avec une acuité particulière, si l'on s'en tient aux strictes listes de fréquences issues du corpus oral, avant les remaniements ultérieurs qui furent jugés indispensables (Gougenheim et al., 1967) : il serait par exemple possible de boire du *café* ou du *vin*, mais non du *chocolat* ou du *thé* ; de se servir du *téléphone* ou de la *radio* mais non de la *télé(vision)* ; d'acheter du *pain*, des *oeufs* ou du *poisson*, mais ni de la *viande* ni des *fruits* ni des *légumes* ; de mettre des *chaussures* et une *robe*, mais ni des *chaussettes* ni une *jupe* ni un *pantalon* ; de se voir un *mardi* un *lundi* un *jeudi* un *samedi* ou un *dimanche*, mais ni un *mercredi* ni un *vendredi* ; de trouver que quelqu'un est *sympathique* mais non *antipathique* ; de *dormir* mais non de *rêver*, etc. Ces quelques exemples, que nous pourrions multiplier, mettent en évidence l'un des problèmes les plus gênants d'une liste établie sur des éléments statistiques : le vocabulaire fréquentiel minimum du F.F. ne correspond pas au vocabulaire nécessaire à l'acquisition d'une compétence minimum de communication. Ainsi, le critère de fréquence pourtant appliqué à un corpus oral en vue de "l'acquisition de la grammaire et du vocabulaire de la vie quotidienne" n'a pas permis de sélectionner le vocabulaire nécessaire aux échanges quotidiens les plus élémentaires.

1.1.1.3. L'absence de noms concrets

Fait notable, les auteurs du *F.F.* furent eux-mêmes amenés à constater que "*d'une façon générale, les listes de fréquences ne donnent pas ou ne donnent guère les mots concrets*". (G. Gougenheim et al. 1967). "*Contrairement à ce que l'on croit on ne peut obtenir par la fréquence des mots tels que veston, autobus, timbre, épicier*" (F.F.1, 1972). Ce type de mots fait pourtant partie du bagage linguistique élémentaire de tout locuteur natif, et on voit mal les raisons pouvant justifier qu'on les évacue d'un apprentissage même élémentaire de la langue. Conscients de cette défaillance méthodologique, les auteurs imaginèrent donc d'intégrer, à la liste fréquentielle du F.F.1, une seconde liste de mots reposant sur le critère de "disponibilité". On voulait dire par là que certains mots sont "*toujours à notre disposition même si nous n'avons pas souvent l'occasion de les prononcer*" (G. Gougeheim et ali 1967.)

Le "degré de disponibilité" d'un mot correspondant à sa "*présence plus ou moins immédiate... dans notre mémoire*", il fallut trouver un moyen de provoquer ces associations d'idées, génératrices de mots concrets. La "*méthode des centres d'intérêt*" servit de base à une seconde enquête. Un échantillonnage de population scolaire fut donc invité à inscrire sur papier les 20 noms jugés les plus utiles et se rapportant à 16 centres d'intérêt tels que "les parties du corps", "les vêtements", "la maison", etc.

Outre son caractère directif et fruste - les "associations d'idées affectives" ne furent pas jugées pertinentes (op.cit. p.155) -, cette enquête sur questionnaire instaurait une double rupture méthodologique :

- alors que la liste fréquentielle reposait sur l'exploitation de conversations d'adultes, celle des mots disponibles reposait sur des listes de mots fournies par des élèves. Seules deux "enquêtes restreintes" furent menées auprès d'adultes ;

- alors que le corpus ayant servi à l'établissement des listes de fréquences reposait sur des sujets de conversation orientés mais relativement ouverts, celui qui était destiné à l'établissement des listes de vocabulaire "disponible" reposait sur des sujets préétablis et différents de ceux du premier corpus. Des enquêtes directives, réalisées sous forme de questionnaires, présentaient un choix arbitraire de "centres d'intérêt" ayant visiblement pour but exclusif, de susciter une liste de noms concrets implicitement préétablie. S'agissant plus

d'un découpage conceptuel (les vêtements, les parties du corps, etc.), que d'un véritable choix de centres d'intérêt, correspondant à des besoins langagiers définis, il eût été plus simple de les introduire sans la mise en oeuvre d'un appareillage pseudo-scientifique. Les auteurs signalent d'ailleurs que *"les sondages effectués présentent très peu de différences selon les régions. Ce sont à peu près partout les mêmes mots qui recueillent le plus de suffrages et qui semblent donc avoir le degré de disponibilité le plus élevé"* (G. Gougenheim et al., 1967).

Comme on l'a fait remarquer, *"ces deux ensembles sont-ils susceptibles de se combiner, et n'y aura-t-il pas toujours à regretter un manque d'homogénéité, pour ne pas dire une contradiction ?"* (J. Chaurand et P. Lerat, 1981).

Nous pensons, pour notre part que le principe de la diversification des "enquêtes", eût été loin d'être critiquable, si elle avait, au contraire, été systématisée et érigée en méthode. Les différentes sélections, obtenues selon des points de vue et des critères variés auraient eu comme but explicite de se compléter et auraient sans doute permis de réduire l'arbitraire de la liste proposée

#### 1.1.1.4. La cuisine de listes

Ceci aurait été méthodologiquement d'autant plus défendable, que la Commission chargée de valider la liste n'a pas hésité à se livrer à une série de manipulations justifiées par des critères moraux et/ou idéologiques. Le chapitre "adjonctions à la liste de fréquence" (G. Gougenheim et al. 1967), aligne des séries d'insertions/substitutions de termes, motivées par un ensemble de raisons hétéroclites et très peu linguistiques : ainsi peut-on lire que *"La Commission a introduit quelques mots permettant d'exprimer des notions morales, civiques et culturelles... art, artiste, courage, effort, justice, liberté, libre, paix, poésie, progrès, sculpteur, tableau, vérité"*. Parfois les raisons de cette "cuisine" de liste prêtent à sourire : *"...la Commission a écarté les noms des organes internes (estomac, foie, intestins, poumons, rein) à l'exception de coeur. Elle a jugé que le malade pouvait guider plus utilement le médecin en montrant du doigt la région du corps dont il souffrait qu'en la nommant plus ou moins exactement"*. Sans doute l'influence du docteur Knock ... Les auteurs eux-mêmes n'avaient pas hésité à exclure de leur liste de fréquence des mots comme *copain, gosse* ou *type* jugés vulgaires (op. cit., pp. 207-210).

D'autres modifications liées à la maintenance de la liste (et non issues d'études fréquentielles) ont été décidées par cette même Commission entre l'édition de 1955 et celle de 1959. On peut lire dans l'introduction du F.F.1 " *La Commission a ajouté à la liste des mots de la première édition: 1° des termes de parenté... 2° des termes relatifs aux activités humaines... 3° des termes dont la connaissance est nécessaire dans la vie moderne... 4° Quelques termes de la vie politique... 5° D'autres mots dont l'usage a montré l'utilité... 6° Des interjections... Elle a supprimé quelques mots dont l'utilité n'était pas manifeste à ce degré...*" (FF1 p.14)

On voit que dans le continuum allant des impératifs objectivistes de départ à la liste finale, la méthode a subi des avatars tels, qu'il est impossible de lui accorder un crédit quelconque. Loin de constituer un résultat "scientifique", l'ensemble proposé est issu d'une série d'opérations successives d'insertions et de substitutions arbitraires autour d'une liste-alibi de fréquences. Comme il a déjà été remarqué (P. Rivenc, 1979), sur les 1475 mots proposés par le F.F.1, seuls 712 (soit, 48%) l'ont été par la liste de fréquences ; 240 (soit 16%) l'ont été par les enquêtes de disponibilité et 520, (soit 36%), ont été introduits par la Commission. La part d'empirisme est comme on le voit, considérable. Le F.F.2 présente un ensemble encore moins rigoureux. Outre les mots issus du dépouillement de textes thématiques, ont été introduits ici, les divers reliquats des listes fréquentielles et des enquêtes sur la disponibilité. Sans compter les inévitables retouches hétéroclites de la Commission.

#### 1.1.1.5. Des défaillances formelles

Le flou définitionnel des entrées ne permet pas toujours de savoir quel est le contexte linguistique visé. Ainsi en est-il du mot *poste* qui pourtant apparaît à deux entrées successives, avec les indications (fém.) dans l'une et (masc.) dans l'autre. Or même si nous ne prenons en compte que l'entrée *poste (masc.)*, il est impossible de déterminer sur ces seules indications le référent et la configuration linguistique visés : doit-on se référer à l'un des composés *poste de radio, poste de police, poste de travail, poste à essence* et/ou à son emploi dans des expressions telles que *être fidèle au poste, quitter son poste* etc.? L'apprentissage de ces différentes structures s'adresserait, on le voit, à des niveaux d'apprentissage et à des contextes d'utilisation différents.

De même, en cas d'homographes (tous les cas sont loin d'être envisagés), la liste proposée (mots fréquents et mots disponibles réunis) ne permet pas de savoir à quel sens le mot renvoie. Il est parfois possible de s'en douter, mais il serait important que l'utilisateur en ait une indication systématique. Si le verbe *voler* est présenté à deux entrées successives avec les indications *voler (oiseau, avion)* et *voler (voleur)* ceci est loin d'être le cas pour la grande majorité des verbes. Le verbe *remonter* par exemple fait-il référence au sens de *remonter une pente* ou de *remonter un réveil* ou de *remonter dans sa chambre* ou de *son repas lui remonte* etc. Chaque sens retenu induirait, outre un correspondant différent dans la langue maternelle de l'apprenant, un apprentissage syntaxique approprié (cf. M. Gross 1975, A. Elia 1984).

Ce même flou se retrouve au niveau de l'indication des catégories de discours, mentionnées de façon sporadique. *Faux* par exemple est accompagné de l'indication (*adj.*), tandis que *faible* reste sans précision : s'agit-il du nom, de l'adjectif, des deux ? L'utilisateur n'a aucune réponse.

Enfin, les noms composés n'ont pas été pris en compte en tant que tels. Figurant çà et là de façon sporadique, ils sont loin d'être représentatifs d'un usage même élémentaire de la langue. On sait, aujourd'hui qu'ils ont reçu une description formelle (G. Gross 1986, 1988, 1990 ; M. Silberstein, 1989, 1990, 1993) que leur nombre est sans commune mesure avec celui que les dictionnaires usuels laissent supposer. A titre indicatif, le nombre de noms composés répertoriés dans le DELAC (Dictionnaire électronique des noms composés du LADL) s'élève aujourd'hui à 80.000 ; celui des verbes composés s'élève à 40.000 et celui des adverbes composés à 7.000. (On trouvera dans l'annexe 1 des échantillons des différents dictionnaires électroniques).

Certains mots auraient donc été au moins aussi utiles, à ce niveau, avec leur extension composée. *Feu rouge, feu vert*, pour *feu* ; *carte postale, carte d'identité, carte à jouer*, pour *carte* ; *terrain de foot, de jeu, à bâtir*, pour *terrain* ; *jardin public* pour *jardin*, etc. On aurait mentionné que ces composés peuvent s'employer tronqués, c'est-à-dire limités à leur forme de tête, quand le contexte ou la situation lève l'ambiguïté.

Toutes ces raisons expliquent que nous ayons cherché à explorer une nouvelle voie pour l'extraction des dictionnaires envisagés.

Celles-ci découlent en fait de l'a priori méthodologique, accordant un crédit absolu au décompte de mots dans des textes. Or, comme nous allons le voir, l'examen des principales expériences prouve que les méthodes statistiques sont impuissantes à cerner ce vocabulaire courant, dont on sait pourtant bien intuitivement qu'il existe.

### 1.1.2. Les limites des méthodes statistiques

Deux principales caractéristiques des dictionnaires de fréquences suffiront à montrer, pourquoi il est impossible de se référer au critère de fréquence pour une sélection hiérarchique. Il s'agit de :

#### 1.1.2.1. L'incohérence des regroupements de mots

Le *Dictionnaire des fréquences* (1971), siglé DF, réalisé en vue de l'élaboration du *Dictionnaire du Trésor de la langue française* a porté sur un corpus de deux siècles de littérature française (XIX<sup>ème</sup> et XX<sup>ème</sup> siècles) ayant fourni 71 millions d'occurrences. Or, si l'on examine la table des fréquences décroissantes (DF, t.2) qui va de la fréquence 3.940.365 à la fréquence 1 (ou hapax), on est frappé par le caractère disparate, pour ne pas dire aléatoire des groupements de mots.

Ainsi se retrouvent à la même fréquence 3 par exemple, des mots hétérogènes que nous avons triés dans les deux sous-ensembles ci-dessous :

#### Table de Fréquence 3 (DF):

adurent	<i>archiver</i>
amphictyonique	<i>achetable</i>
amyle	<i>babillage</i>
anagnoste	<i>bagarreur</i>
anhélant	<i>basculement</i>
anspessade	<i>chambouler</i>
balanite	<i>brouillasse</i>
balancine	<i>colmater</i>
batture	<i>chiader</i>
condyloïde	<i>démarreur</i>
contrebatterie	<i>dépanner</i>
coqeron	<i>exonérer</i>
lupercal	<i>malvenu</i>
manrèse	<i>motard</i>

etc.

On comparera à présent le sous-ensemble ci-dessus en italique, à l'ensemble ci-dessous regroupant des exemples pris parmi les 5.000 mots les plus fréquents ; soit, ceux qui, sur les 71.415 mots que compte le DF, ont une fréquence égale ou supérieure à 800. (Le chiffre entre parenthèses indique la fréquence exacte) :

*marquis (4942), tour (20719), ho (21740), chevalier (4180), épée (3656), sabot (1151), vallon (1162), auguste (1170), écurie (1184), altesse (1191), épine (1202), poignard (1250), clocher (1299), fiacre (1261), major (1281), sergent (1313), pasteur (1348), charbon (1401), sabre (1433), louis (1468), prussien (1569) etc.*

On conviendra qu'il est difficile d'admettre un tel écart entre les deux ensembles en question. De même qu'il est difficile d'admettre, que soient considérés comme mots rares, ces mots figurant dans la liste des hapax :

*agresser, agacé, amaigrissant, archicomble, bigleux, banalisation, bidonnant, blocage, bronzage, cafeter, cafouillis, capsule, chatouillis, chronométrage, collimateur, coauteur etc.*

ou ces mots figurant dans la table de fréquence 2 :

*bafouillis, barbiturique, copiable, copinage, commotionné, conasse, couillonner, cramer, déstabilisation, décontraction, décoder, increvable, laxiste, yaourt, marginaliser, occulter, planifier, transférable, muscler, parking, pantouflard, résidentiel, shopping, ulcérant, etc.*

De même, il est remarquable que dans la liste des 900 mots les plus fréquents du DF, regroupés par ordre alphabétique par E. Brunet (1981), on ne trouve pas les mots élémentaires : *cuiller, fourchette, marier, hebdomadaire, hésiter, exactement, doucement, dessous, dessus, direction, dessert, dégoûtant* , etc. On trouve en revanche les mots *église, abbé, roi, palais, comte* , etc.

La même hétérogénéité se retrouve dans la liste des fréquences décroissantes du dictionnaire de J. Baudot (1990, T.2)<sup>2</sup>. On comparera, à titre

---

<sup>2</sup> Le dictionnaire de J. Baudot a pourtant un avantage : certains mots ambigus ont été décomptés séparément, ce qui n'est pas le cas du DF.

d'exemple, les deux ensembles ci-dessous dont tous les mots sont tirés de la fréquence 1:

- Lettre b, fréquence 1 (J. Baudot, T.2 pp.141-144)

balistocardiographie	<i>badigeonner</i>
baronnial	<i>bagarrer</i>
batayole	<i>bague</i>
baryum	<i>bagnole</i>
basicité	<i>balade</i>
bifilaire	<i>banaliser</i>
blandice	<i>banditisme</i>
biauriculaire	<i>baratin</i>
biostrome	<i>bécoter</i>
bouvillon	<i>bibliothécaire</i>
brai	<i>bicoque</i>
brimborion	<i>biscotte</i>
bromatologue	<i>bronzer</i>
brosserie	<i>bizarrie</i>
bruitisme	<i>blagueur</i>
brunante	<i>bougeotte</i>
butylmagnésium	<i>bougrement</i>
bryone	<i>boulot</i>
bijectif	<i>boulangier</i>
brabançon	<i>bourde</i>
bradycardiaque	<i>bredouille</i>
blastocyte	<i>bricole</i>
etc.	

Il apparaît clairement, à l'examen de ces deux dictionnaires qui comptent parmi les références les plus sérieuses dans le domaine, qu'en allant des mots les plus fréquents aux mots les moins fréquents, on ne va pas des mots les plus essentiels aux mots les plus rares.

#### 1.1.2.2. La disparité des différentes listes

D'autre part, les listes présentées par les différents dictionnaires de fréquence ne se recouvrent pas. Il ressort par exemple d'une étude comparative (J.C. Rolland, 1991) que sur les mille premiers mots de 3 dictionnaires de fréquences du français (Français Fondamental (F.F.)1967, D.F. 1971, J. Baudot 1990), seuls les 500 premiers sont communs aux trois listes. En outre, sur ces 500 mots, 150 sont des mots grammaticaux (*le, sur, dans*, etc.). Les 500 autres présentent une disparité dont on trouvera quelques exemples dans le tableau ci-dessous :

	FF	Baudot	DF
absolument	x		
abbé			x
abandonner		x	
billet	x		
bouche			x
adieu			x
style		x	
peau		x	
vertu		x	x
montagne			x
colère			x
			etc.

Cette disparité pose un problème de fond : Si aucune liste n'est superposable à une autre, peut-on en considérer une comme la liste-étalon ? Les tenants des listes de fréquences n'ont toujours pas soulevé la question. L'une des raisons est que les méthodes statistiques opèrent toujours dans le sens constitution de listes de mots à partir de textes, jamais dans le sens validation de la couverture de ces listes sur des textes. Nous présentons une telle approche.

Il serait vain, à notre avis, de prétendre réaliser une liste-étalon servant de référence générale, à partir d'une liste de fréquence. Même si un système technologique futuriste permettait d'enregistrer tous les énoncés, écrits et parlés, de tous les locuteurs d'une langue à un moment M, il est en effet probable, d'une part que certains mots de la langue n'apparaîtraient pas, et

d'autre part que des divergences apparaîtraient avec la liste extraite de la même expérience renouvelée à un moment M'.

Le critère de fréquences ne peut donc servir de base à la constitution de lexiques hiérarchisés de référence.

## **1.2. Approche intuitive et problèmes théoriques**

Contrairement aux méthodes prétendues objectives fondées sur l'exploitation de listes de fréquences, et compte tenu des insuffisances constatées, nous avons décidé d'adopter une démarche résolument opposée, consistant à partir d'un lexique de la langue française contemporaine le plus complet possible, en l'occurrence le DELAS (Dictionnaire électronique des mots simples du français) (cf. Annexe 1), et d'en extraire, "au jugé", les couches essentielles du lexique français, correspondant à la compétence d'un locuteur cultivé, ceci sans a priori sur tel domaine ou tel registre de langue.

Bien que déjà pratiquée, cette approche subjective n'a jamais été systématisée. De ce fait, les fondements méthodologiques des démarches "intuitives" n'ont jamais été clarifiés. Nous mettons à part l'expérience de construction du lexique-noyau des verbes français réalisée par le LADL. La sélection des 12.000 verbes qui le composent, a reposé sur des "intuitions de sens" immédiatement vérifiées par la capacité des sujets parlants, en l'occurrence le groupe de linguistes se livrant à l'expérience, à en produire la liste des propriétés syntaxiques (distributionnelles et transformationnelles) spécifiques (M. Gross, 1975) (cf. Annexe 8). Bien que l'application de cette démarche aux verbes ait produit des résultats justifiant largement les hypothèses de base, il aurait été difficile d'attendre que toutes les catégories du discours soient traitées. Il faut préciser que l'unité de temps de réalisation liée à une telle entreprise est de l'ordre de la décennie... Nous avons donc décidé de mener conjointement à la construction du lexique-grammaire, une expérience de sélection lexicale strictement intuitive, ce qui n'exclura pas, au contraire, une confrontation des deux sélections.

Cette décision d'expérimenter une démarche indépendante de toute préoccupation syntaxique, d'élaborer des critères de choix liés à la seule

subjectivité, et d'en élucider la méthodologie avec toutes ses implications, nous a amenée à aborder une dimension généralement ignorée par les réalisations explicitement ou implicitement subjectivistes, aussi bien que par les tenants des méthodes fondées sur les statistiques : la nécessité de dépasser, dans tout projet de sélection lexicale, le niveau strictement linguistique, et de fonder ces critères sur des présupposés psychologiques et philosophiques explicites. Les premières tentatives d'extraction ont en effet fait surgir une série de questions, nous conduisant progressivement à la conclusion que la mise en oeuvre de quelque méthodologie que ce soit/ serait vouée à l'échec, tant que ne se serait pas exercée cette réflexion en amont.

### **1.2.1. Problématique de la sélection lexicale**

Comment, en effet, définir la compétence lexicale d'un locuteur ? Comment l'évaluer ? Et plus particulièrement, dans notre cas, comment s'auto-évaluer ? On pourrait répondre de façon triviale : la compétence lexicale d'un locuteur correspond aux mots qu'il connaît. Certes, mais

- 1) Qu'est-ce que "connaître" un mot ?
  - 1a) Ceux dont on est capable de donner une définition ?
  - 1b) Ceux que l'on utilise (ou que l'on pourrait utiliser), que l'on sache ou non les définir (connaissance productive) ?
  - 1c) Ceux que l'on comprend mais que l'on n'utilise ou n'utiliserait pas (connaissance réceptive) ?
  - 1d) Les causes de cette "passivité" sont-elles à analyser ?
  - 1e) Si oui, seraient-elles toutes à prendre en compte et comment ?

Rappelons à ce propos que le domaine pédagogique-lexical utilise couramment dans les tests et les études du champs une série de termes se référant à la double notion de compétence réceptive et de compétence productive : vocabulaire actif/passif, compréhension/production, etc. Cette dissociation repose sur l'idée implicite que tout locuteur possède deux lexiques d'importance inégale, celui de la connaissance réceptive étant globalement plus important que celui de la connaissance productive.

- 2) Ces différents points de vue entraînent une seconde série de questions :
  - 2a) Comment définir le seuil d'admissibilité d'une définition de mot ?
  - 2b) Qu'est-ce que "comprendre" un mot ?

2c) Comment évaluer la capacité de production (effective ou potentielle) d'un locuteur ?

3) En supposant acquise la compréhension d'un mot, (que nous entendons ici comme une suite de caractères délimitée par des séparateurs), doit-on vérifier que toutes les acceptions sont connues ou la restitution d'un seul sens suffit-elle ?

3a) Dans ce dernier cas y a-t-il des priorités ?

3b) Par ailleurs, ne peut-il y avoir une compréhension partielle d'un mot ?

3c) Si oui, le mot doit-il être accepté ?

3d) Si oui, à partir de quel seuil ?

### 1.2.2. La question des deux lexiques

Disons tout de suite que la dissociation, généralement considérée comme acquise, entre lexique actif et lexique passif, repose sur des notions mal cernées - les diverses expériences renvoient à des acceptions différentes de ces notions - et un postulat jamais vérifié. L'étude approfondie de J. Melka Teichroew (1989) des techniques de tests (réceptifs et productifs) et de leurs résultats, établit l'arbitraire de ces deux notions : on y voit qu'il n'existe aucune technique capable de "*tester spécifiquement les connaissances lexicales réceptives et productives*", que les résultats des tests "*fluctuent (et souvent dans des proportions énormes) selon les variables sur lesquelles on a bien voulu mettre l'accent*" bref, que "*la validité de conclusions sur l'existence ou la non-existence de la distance RIP, tirée à partir des résultats de tests, n'est pas garantie*".

Ceci rejoint le problème des mots "disponibles" auquel a été confrontée l'équipe du *Français Fondamental*, ou de l'emploi dit potentiel de certains mots. Pourquoi n'utilise-t-on pas tous les mots de notre lexique mental ? Les raisons semblent tenir à un ensemble très complexe de facteurs situationnels, culturels, sociologiques, psychologiques, etc. qui font qu'un mot "connu" d'un locuteur n'est pas "produit". Mais l'une d'entre elles est sans doute d'ordre cognitif : un mot peut être présent en nous selon des modalités infinies, et les conditions de production dépendent en grande partie de celles-ci. De fait, le concept de connaissance d'un mot est beaucoup plus complexe que la simple opposition entre vocabulaire passif et vocabulaire actif tendrait à faire croire. Qu'est-ce que connaître un mot ?

Tentons d'appliquer les critères de définition/description et de compréhension :

### 1.2.3 Définition / description

Tenter de décrire un mot, est ce que nous faisons lorsqu'un interlocuteur nous demande de lui expliquer un terme qu'il ne connaît pas. Cette situation serait-elle reproductible dans le cas d'une expérience ? En d'autres termes, pourrions-nous considérer qu'un mot est connu de nous lorsque nous pouvons le décrire ?

Dès l'application de ce critère, un premier problème de contenu se pose : soit le mot indiscutablement connu de tout le monde, *eau*. Comment un locuteur devrait-il le décrire, si une évaluation reposant sur ce principe le demandait ? En termes descriptifs approximatifs (type : *liquide transparent* etc.), ou en termes scientifiques (type : *H<sub>2</sub>O*) ? Cette question n'est même pas tranchée chez les lexicographes puisque les définitions toujours hétérogènes oscillent entre la technicité extrême (tendance encyclopédique), et la définition naïve (tendance didactique) (cf. J. Melka Teichroew, 1989). Or cette question se pose pour une grande quantité de mots usuels, exemples : *huile, gaz, air, feu, glace, neige, pluie, muscle, os*, etc.

Supposons que l'on décide de s'en tenir à un contenu descriptif "naïf", un deuxième problème se poserait concernant, d'une part, la capacité descriptive du sujet parlant et d'autre part, le seuil d'admissibilité de cette définition. Il suffit de faire l'essai sur des termes aussi usuels que ceux que nous venons d'évoquer, pour se rendre compte de cette double difficulté. Par ailleurs, le seuil d'admissibilité est peut-être susceptible de varier en fonction de la position paradigmatique du mot. Ainsi serait-on probablement plus exigeant si celui-ci se situait à l'intérieur d'une série homogène: suffirait-il par exemple de qualifier *le persil, le cerfeuil, la coriandre, le thym*, etc. de *plantes aromatiques*, ou devrait-on introduire des traits qui les différencient ?

Le problème est bien plus délicat en ce qui concerne les termes abstraits, exemples : *exponentiel, romantisme, accointance, collusion, conjecture*, etc. Les deux premiers mots peuvent donner lieu soit à une définition de spécialiste

des disciplines concernées, soit à une description de l'usage courant. En tout état de cause, les deux définitions seraient également difficiles à formuler par un locuteur (non spécialiste) standard qui par ailleurs serait apte à les utiliser et à les comprendre. D'autre part, le recours à une description référentielle étant, pour tous ces mots, impossible, la définition "naïve" est encore plus malaisée que pour les mots concrets. Il suffit d'avoir été confronté à la nécessité d'expliquer à un enfant des termes aussi courants que *hagard* ou *dégingandé* pour mesurer cette difficulté.

L'aptitude définitoire est, nous le voyons, indépendante de la compétence lexicale et ne peut donc être retenue comme seul instrument de mesure de cette dernière.

#### 1.2.4. Comprendre un mot a-t-il un sens ?

Si la compréhension d'un mot est indépendante de l'aptitude à le définir, qu'est-ce que "comprendre" un mot ? Sans cesse débattue, cette question n'a jamais trouvé de réponse. Les tentatives de définition restent floues et approximatives. Quant aux théories orthosémiques de décomposition du sens des mots en traits fixes (analyse componentielle) ou moins fixes (approche stéréotypique), elles n'ont jamais pu aboutir à des modèles concluants. Hormis quelques cas susceptibles de les illustrer (cf. J. Greimas 1966, B. Pottier 1965), les tentatives d'application ont vite tourné court. Dès lors, on ne s'étonnera pas de la disparité des méthodes de tests de la compétence lexicale: elles traduisent ce même désarroi théorique. Ceci explique également que notre tentative de sélection des mots, selon ce critère de compréhension, ait vite été freiné par des questions sans réponses. Même les mots jugés intuitivement compris posent des problèmes de sélection :

La compréhension d'un mot doit-elle par exemple recouvrir toutes ses acceptions ? Ainsi suffit-il que je donne le sens de *chemin de terre* pour que la sélection du mot *chemin* soit validable, ou dois-je décliner toutes les acceptions du mot puis en vérifier l'exhaustivité (cf. Annexe 2). Les mots les plus usuels (*cachet, bouillon, chemise, bourdon, beurre, charge, chinois, raison, etc.*) présentent une telle variété d'emplois, que ce point de vue s'avère vite irréaliste. Si l'on décide qu'une seule acception suffit, alors se pose la question des degrés : y en a-t-il de plus importantes que d'autres ? On s'aperçoit vite

que toute décision relèverait de l'arbitraire (cf. Annexe.3, les 32 sens du mot *raison*), et ne peut être prise qu'en fonction d'une application particulière.

D'autre part, cette situation de face à face avec une liste de mots bruts (dépourvus de définition et d'exemples), met en évidence un phénomène capital : l'évaluation de la compréhension d'un mot est loin de pouvoir se faire en termes tranchés. Dans de nombreux cas, la compréhension que nous avons d'un mot est partielle ou globale. Tout se passe comme si face à un mot, nous mettions en action un processus d'accommodation ne parvenant pas toujours à son terme et que parfois, un flou de degré variable et impossible à corriger entourait le mot.

Ainsi avons-nous pu personnellement nous rendre compte que nous ne connaissions pas tout-à-fait (ou que nous ne connaissions qu'en gros) le sens de *asphodèle*, *salsepareille*,<sup>3</sup> *lamantin*, *cimaise*, *amadou*, *cirrus*, *césium*, *actuariel*, *lécithine*, *nitrates*, etc. Cette connaissance approximative est d'autant plus difficile à évaluer qu'elle présente également des degrés impossibles à mesurer. Les seuls mots mentionnés ci-dessus offraient pour nous une gamme relationnelle impossible à formaliser : nous savons que l'*asphodèle* est une fleur (mais nous ne saurions dire à quoi elle ressemble). La *salsepareille* évoque pour nous l'idée d'une vague herbe (avec un certain pouvoir magique ?). Pour *lamantin*, nous avons entendu dire que c'était un animal en voie de disparition, et nous l'imaginions vaguement comme une sorte de babouin. De l'*amadou*, nous savons qu'il s'agit d'une matière inflammable, qu'il existait des *briquets à amadou*, mais il nous est impossible de nous représenter cette matière. Le *cirrus* est pour nous un nom de nuage, mais nous ne nous représentons pas sa forme, que nous aurions du mal à distinguer de celle d'un *nimbus* ou d'un *cumulus*. Quant à *cimaise* et *césium*, nous sommes seulement en mesure de les rattacher à des domaines plus ou moins circonscrits (respectivement, architecture et sciences). En ce qui concerne *actuariel* et *lécithine* nous ne pouvons que les situer respectivement dans les composés *taux actuariel brut* et *lécithine de soja*, bien que d'ailleurs nous n'en sachions pas plus sur ces derniers. *Nitrates* est encore particulier, puisqu'il fait partie d'une série de mots (tels que *métabolisme*, *enzymes*, *liposomes* etc.) qu'il est possible de produire dans un contexte étroit, sans en connaître vraiment le sens. Ainsi pouvons-nous être amenée à dire que nous ne mangeons pas de

---

<sup>3</sup> Nous avons pu constater que ce mot est, en revanche, bien connu de nombreux enfants, car la salsepareille est, paraît-il, la nourriture dont les Schtroumps sont particulièrement friands...

produits fumés car ils contiennent des *nitrates* et que les *nitrates* sont cancérigènes. Pourtant nous serions incapable de dire à quoi correspond cet élément, et encore moins de le distinguer des *nitrites*. Nous pouvons énoncer des propos tels que "*le métabolisme d'un fumeur se modifie*" tout en étant incapable d'expliquer en quoi consiste exactement ce processus. Comment gérer des cas de figure aussi hétérogènes et aussi difficilement quantifiables ?

Le critère de compréhension ne permettant pas de gérer ces nuances, il a été impossible de continuer à s'y référer pour notre procédure de sélection.

### 1.2.5. Inconnus, connus, re-connus : étude de cas

Ce que nous constatons en fait, c'est qu'il existe des mots que nous pouvons intuitivement déclarer connus et d'autres inconnus, et ceci, sans autre précision. Par ailleurs, entre les deux existe un large "marais", constitué par des mots n'entrant dans aucune de ces deux catégories, et dont la gamme relationnelle offre une variété telle, que notre rapport à chaque mot semble constituer un cas de figure original. Ni connus, ni inconnus mais cependant re-connus, ils prouvent que la compréhension ne peut s'évaluer en termes de tout ou rien ou que la connaissance d'un mot déborde largement le concept de compréhension.

Outre les mots dont nous pouvons, comme nous l'avons vu ci-dessus, avoir une compréhension approximative mais certaine, bien que difficilement évaluable, se présentent des mots dont nous ne pouvons pas dire que nous les comprenons (dans la mesure où cette compréhension difficilement exprimable semble se situer en deçà des mots) mais dont nous ne pouvons pas dire non plus qu'ils nous sont étrangers. Plus exclusivement liés semble-t-il à la mémoire visuelle auditive ou affective, la réminiscence, même ténue, qu'ils suscitent n'en constitue pas moins un témoin de leur trace dans notre lexique interne<sup>4</sup>.

Précisons toutefois qu'à l'intérieur de cette catégorie de mots non étrangers, puisque "reconnus", certains peuvent encore présenter un degré non nul de compréhension. Dans le continuum allant des mots approximativement connus

---

<sup>4</sup> Nous pensons à ce propos/ que la résistance opposée à la réforme orthographique d'un mot/ relève précisément, d'un réflexe de défense face à un danger menaçant notre intégrité, comme si une fois inscrits en nous, ces mots faisaient partie de notre personne.

aux mots simplement reconnus, la frontière est difficilement traçable. Nous donnerons comme exemple de mots reconnus le cas de souvenirs personnels allant de la rencontre localisée à la vague impression de déjà vu :

#### **la rencontre localisée**

Ce cas s'est produit pour nous, avec des mots tels que *bréhaigne* et *médianoche* (respectivement situés dans les titres des romans, *La Bréhaigne* et *Le médianoche amoureux*), *anachorète* (dans un texte de J.J. Rousseau), *smaragdin* (dans un roman de Nabokov), *nankin* (dans un roman de Stendhal), *astrolabe* (nom d'un restaurant), *macumba* (titre d'une chanson), *codicille* (entendu dans une chanson de Brassens), etc. Comme nous l'avons précisé ci-dessus, la frontière entre mot simplement reconnu et mot approximativement connu n'est pas toujours aisément délimitable. Ainsi nous a-t-il semblé qu'*anachorète*, *codicille* et *nankin* (sur lesquels nous étions en mesure de faire des hypothèses, qui se sont révélées approximativement justes), n'étaient pas pour nous aussi "incompréhensibles" que *smaragdin* ou *bréhaigne*.

#### **la rencontre non localisée**

Nous voulons parler ici des mots dont la "silhouette" visuelle ou sonore nous "dit" quelque chose un peu à la façon d'une personne dont nous serions certain de l'avoir vue quelque part.

Ce cas, moins vérifiable que le précédent, repose cependant sur la certitude que cette rencontre a bien eu lieu (à une ou plusieurs reprises). Correspondaient pour nous à ce cas des mots tels que *coquecigrue*, *chaptaliser*, *ordalie*, *émollient*, *céladon*, *ortolan*, *chamsin*, *impédance*, *dyspnée* etc. Ici encore, certaines hypothèses peuvent permettre d'en dire "quelque chose" ; Nous pouvions par exemple dire que *chaptaliser* avait un rapport avec le vin et *dyspnée*, un rapport avec la respiration (et la plongée ?).

#### **le souvenir vague**

Il existe enfin des mots qu'il nous semble avoir déjà rencontrés sans en être totalement sûr. Des mots tels que *angström*, *letrine*, *alloïne*, *gibelotte*, *dormance*, *daiquiri* (qui n'est pas dans les dictionnaires) etc. ont suscité pour nous ce vague sentiment de déjà vu.

Le degré de "reconnaissance" de ces mots est, on le voit, difficilement quantifiable. Dans la mesure cependant où l'on considère qu'ils sont présents dans notre lexique interne, nous ne pouvons pas leur assigner une valeur zéro. Devrait-on alors les prendre en compte dans une sélection ? Nous insistons sur le fait que l'absence de définition et d'exemple est essentielle dans cette expérience. La lecture d'une définition ou d'un exemple produirait un important effet rétroactif susceptible de transformer notre jugement sur le mot.

Toujours dans le cas de cette confrontation lexicale, il peut enfin se présenter deux autres types de mots :

### **les ressemblances**

Certains mots présentant une similitude graphique partielle, peuvent donner lieu à une hésitation. Exemples : *cautèle* et *cautère* ; *lignite* et *linite* ; *abducteur* et *adducteur* ; *alloïne* et *allantoïne* ; *orfraie* et *orfroi* ; *méthadone*, *méthanal*, *méthane* et *méthanol* ; *orne* et *orme* ; *mélanine* et *mélanite*, etc. De fausses séries pourraient donner lieu à de fausses reconnaissances de mots. Exemples : *castrat*, *castrer*, *castration*, *castrisme*, *castriste* .

### **les dérivés et la connaissance reconstituée**

Il s'agit des mots que nous rencontrons pour la première fois, mais dont nous pouvons inférer le sens après analyse des constituants. Ainsi en a-t-il été pour nous des mots tels que *postembryonnaire*, *postenquête*, *postnéonatal*, *impuberté*, *multimiroir*, *parqueteuse*, *cabochardement*, *bazardisation*, etc. Notons que ce processus de décodage, par application de règles dérivationnelles, joue également dans la compréhension de mots étrangers jamais vus, mais ainsi reconnus (F. Dennighaus, 1976). Le terme "potentiel" appliqué parfois à ce vocabulaire ne doit pas prêter à confusion. Il ne s'agit pas des mots présents en nous mais non utilisés (appelés "disponibles" par l'équipe du F.F. ou également "potentiels" dans certains articles), mais de mots "virtuellement" connus bien que jamais vus.

Que faire de ces mots inconnus mais reconstitués ?

Face à cette prolifération de cas et à la pénurie d'outils conceptuels permettant de les appréhender, la nécessité apparut de différer le processus d'extraction, et de se déplacer du plan linguistique au plan philosophique afin de reconsidérer, de ce point de vue, le problème du sens et de notre rapport aux mots. Il était clair que la problématique ne se situait pas à ce premier niveau, mais qu'elle était à envisager dans le cadre d'une théorie de la connaissance. Seule une réflexion approfondie sur la nature de la relation qui

nous unit à la langue, pourrait donc nous permettre de déboucher sur une méthodologie.

### **1.3. A la recherche du sens**

#### **1.3.1. Dictionnaire usuel et lexique interne**

Précisons tout d'abord qu'un dictionnaire usuel n'est en aucune façon comparable au lexique interne de chaque individu. On voit sans problème à quoi correspond le premier ; on a du mal à se représenter le second. Des différences essentielles les opposent :

**Un locuteur aussi cultivé soit-il n'est pas un lexicologue/graphiste et n'a aucun besoin de l'être.**

De même que tout locuteur utilise la grammaire de sa langue sans nécessairement être en mesure d'en expliciter les milliers de règles qui la fondent, de même, utilisons-nous et comprenons-nous des mots dont nous ne savons pas formuler des définitions.

**Un dictionnaire usuel est à la recherche du sens, un dictionnaire interne le contient.**

Les dictionnaires usuels contiennent des définitions, dont les mots renvoient à d'autres mots, dont les définitions contiennent des mots renvoyant à d'autres mots etc. Renvoyant indéfiniment d'un signe à l'autre, ils se présentent comme d'immenses programmes bouclant sur eux-mêmes (On imagine mal un locuteur apprenant une langue étrangère, avec le seul secours d'un dictionnaire monolingue).

Pouvant donner l'illusion qu'ils contiennent le sens, les dictionnaires usuels ne font que répertorier des valeurs d'usage.

D'un tout autre ordre, notre lexique interne est constitué de l'ensemble des énoncés écrits et oraux auxquels nous avons été exposés (depuis, ou avant même notre naissance - comme on le pense aujourd'hui -), en rapport avec l'ensemble des situations dans lesquelles ils se sont "produits". Notre lexique mental n'est pas constitué de listes de mots, mais d'un vaste texte ouvert et propre à chacun, dans lequel circule le sens. L'analyse des phrases complexes

et du discours par Z.S. Harris, constitue une approche syntaxique de ce mécanisme. Lorsque nous allons chercher un mot, il est clair que nous n'allons pas chercher une définition dans une liste, mais que nous mettons en branle dans notre cerveau, un programme<sup>5</sup> beaucoup plus complexe mettant en relation les occurrences de ce mot dans notre texte, avec l'ensemble des paramètres (psycho-affectifs, ciconstanciels, etc.) qui le lient à notre expérience. Tous les contextes linguistiques et extra-linguistiques dans lesquels nous avons entendu ou produit ce mot sont vraisemblablement réévalués selon un processus mettant en branle l'ensemble de notre système cognitivo-affectif. Le choix d'un mot naît d'un calcul vertigineux mettant en perspective toutes les composantes de notre histoire. Le cas-limite est celui des mots que, consciemment ou inconsciemment, l'"on ne peut plus prononcer", leur charge émotionnelle étant trop forte. Inversement, on sait aujourd'hui qu'un mot peut, par résonance, émouvoir, frapper, déstabiliser, bouleverser voire tuer une personne. La sociologie et la psychanalyse ont, depuis longtemps, montré cette interdépendance du linguistique et de l'historico-affectif.

Par ailleurs, et contrairement à celles d'un dictionnaire, les limites de notre lexique interne ne sont jamais fixées. Soumis à un réaménagement perpétuel, il se transforme sans cesse au gré de l'ensemble des énoncés que nous lisons, entendons ou produisons, dans tous les lieux et dans toutes les situations de notre vie quotidienne. Il est probable qu'une régulation intervient au niveau de la mémoire. Il semble que certains mots (insuffisamment assimilés ?) disparus depuis longtemps de notre actualité pâlisent, tandis que d'autres apparaissent, ou encore resurgissent pour un autre cycle de vie. La mémoire gère-t-elle la rentabilité de notre lexique ? Et de quelle façon ? Ce sont des questions auxquelles il est difficile de répondre aujourd'hui.

La sélection de termes dans une liste lexicale, relève donc de tout autre chose que d'une simple comparaison, avec un objet homologue, rangé dans un coin de notre cerveau, comme le laissent supposer certains modèles à base de "fichiers informatiques". Chaque confrontation avec un mot réactive notre propre texte, lui-même inséré dans l'ensemble de notre expérience. Cette intrication de la langue et de l'expérience, est explicitement postulée par les recherches en intelligence artificielle qui ont mis en évidence, d'une part l'importance de la relation entre "histoires" et processus cognitif (cf.

---

<sup>5</sup> D'aucuns qualifieront ce programme de neuronal

R.C.Schank, 1990), d'autre part l'importance du phénomène d'inférences extralinguistiques dans la communication linguistique (cf. R.C.Schank 1982, 1986).

L'idée que le mot contiendrait le sens en soi relève d'une conception nominaliste fort scolastique de notre rapport aux signes.

### 1.3.2. Intention signitive et remplissement du sens

De Platon à Husserl le caractère arbitraire du signe linguistique a du reste souvent été évoqué par les philosophes. Si les linguistes ont aussi fait état de cette contingence, ils s'en sont toujours tenu à un plan strictement linguistique. Chez Saussure par exemple la description du signe en une entité duelle signifiant/signifié, ou la dichotomie dénotation/connotation constitue un moyen méthodologique limité à la description de la langue comme système. La question du "remplissement" de sens (Husserl, RL III, p. 120) de ce signe insensé et de ses rapports avec l'homme et le monde n'est jamais évoquée. Or, cette question philosophique du sens du signe est au centre de notre problématique : si le sens n'est pas dans le rapport mot-chose, appréhender le sens d'un mot "en soi" a-t-il un sens ? Où et comment le sens se forme-t-il ?

Arbitraires par rapport aux choses, les mots ne se remplissent de sens que par l'*intention de signification (Bedeutungsintention)* d'une conscience visant les choses auxquelles ils sont par là-même associés. (Husserl, RL II,1 p.43). Si c'est la conscience qui fonde le sens, celui-ci n'est pas *dans* les mots mais dans ce qui par eux est signifié à l'intérieur d'un contenu articulé. C'est donc l'*acte significatif* et lui seul qui est fondateur de sens. "*Grâce à cette intentionnalité, le mot signifie quelque chose, l'expression est plus qu'un simple son verbal. Elle est pensée de quelque chose, et, comme telle, se rapporte à l'objet.*" (Husserl, RL II,1 p.102.)

Ceci implique que le mot soit inséré dans un *signifier verbal* mettant en oeuvre un ensemble de lois logiques de différentes natures, dans une combinaison permettant d'enraciner le sens, et que *chaque signification concrète* soit "*un entrelacement de matières et de formes*" (Husserl, RL II,2 p.115). C'est par exemple la combinaison du *matériau pur, homme ou papier*, et d'un matériau syntaxique qui fait effet de sens. "*Les matériaux purs rendent*

*possibles finalement la référence aux choses (Sachbezüglichkeit) par l'intermédiaire de leur mise en forme qui s'opère de degré en degré de sorte que les formations de tout degré présentent, à leur tour, dans les membres de la proposition, des matériaux et des formes relatifs.*" (Husserl, LFLT pp. 392-393).

Le signe in-sensé "en soi" devient donc, grâce à des lois combinatoires extrêmement complexes, média d'un sens qui le contraint et le dépasse à la fois. Ainsi, dans la phrase *L'homme est mortel* le mot *homme* vise-t-il un concept (*être humain*) subsumant la différenciation *homme/femme*, tandis que seul l'objet général *homme* est visé dans la phrase *La femme est l'égale de l'homme*. D'autre part les éléments *est* ou *de* ne recèlent rien en eux d'une référence aux choses. Ces parties de la proposition participent pourtant à la "formation du sens de la référence aux choses de la proposition totale". Le sens de chacune des propositions se réalise au travers d'une "typique analytico-formelle" (Husserl LFLT, p.393) débordant largement l'ensemble du sens de chacune des unités qui les composent. De fait, les discours interprétatifs susceptibles d'explicitier le sens de ces phrases n'arriveraient pas à l'épuiser.

Pour que l'acte de signification soit complet, il faut par ailleurs que soient exercées les lois des "formes possibles de signification" ayant pour fonction de séparer "sens et non-sens (*Unsinn*)" (Husserl RL, II,2 p.85). Ainsi, dans certains cas "la possibilité elle-même de la signification unitaire interdit que certaines significations partielles coexistent entre elles" (Husserl RL, II,2 p. 122). C'est en vertu de ces lois qu'un jugement d'incompatibilité sera porté sur des associations telles que *un rond ou un homme et est..*

La substance morpho-syntaxique elle-même participe à la complétude (*Vollständigkeit*) de la signification en mettant en rapport une pluralité d'éléments s'impliquant réciproquement : c'est déjà le cas des "noms composés" dont la création, liée à une "forme intentionnelle de connexion", conduit à une expression significative originale (Husserl RL, II,2 p.96).

Au niveau de la phrase, ces rapports peuvent parfois rendre ces éléments consubstantiels. Les relations de signification que les mots y entretiennent les uns par rapport aux autres, peuvent en effet être si particulières, que leur statut référentiel même se trouve atteint. Ainsi, dans l'exemple "*ce papier est blanc* le prédicat acquiert, par-delà sa propre teneur matérielle, une relation avec le sujet

*papier*, empiétant, dans le plan de la signification, sur la référence aux choses que comporte celui-ci." (Husserl LFLT, p. 394). Par son actualisation dans cette forme morpho-syntaxique, le mot *blanc*, entrant en rapport avec le mot *papier*, reçoit donc une référence aux choses qui s'étend au-delà de lui-même. Ce *blanc* est donc différent du *blanc* de *cette fleur est blanche*, ou *cette peau est blanche*.

Si donc le sens n'est pas *dans* le signe, mais dans ce qui par lui *se* communique, alors comment l'appréhender en dehors de tout discours ? En d'autres termes, comment appréhender le sens au travers de mots isolés non inscrits dans un projet de sens ? Notre projet de sélection lexicale nous place en effet dans une situation où ne visant aucun objet, nous devons appréhender les signes pour eux-mêmes. Dépouillés de la signification que seul un *acte significatif* peut lui donner, le mot se présente à nous, dans sa pure matérialité. Nous voyons donc pourquoi il est impossible d'opérer une sélection lexicale reposant sur la compréhension du sens d'un mot. Celui-ci est par définition insaisissable.

### **1.3.3 Approche phénoménologique de la notion de représentation**

Différentes propositions de représentation sémantique par décomposition et formalisation du sens ont été élaborées en lexicologie, psychologie et intelligence artificielle. Outre que la notion de représentation y est prise avec des acceptions différentes, elle n'y est pas définie. Par ailleurs, ces propositions n'ont jamais pu aboutir à un modèle fiable, la reproductibilité des différentes expériences, s'avérant toujours limitée.

Nous avons donc décidé de réexaminer la notion de représentation sous un angle philosophique, et plus précisément, de l'envisager dans une problématique de la corrélation, c'est-à-dire d'un ensemble de problèmes liés au rapport sujet-objet. (L'objet étant cela qui est donné, la chose à laquelle on pense, de laquelle on parle, le monde ; le sujet étant le "je", la conscience pour qui elle est donnée, l'homme). Ce positionnement, nous a amenée à introduire la dimension subjective propre à la phénoménologie, et à nous démarquer des modèles formels de représentation sémantique. Nous verrons comment cette nouvelle approche nous a permis de donner une définition opérationnelle de la notion de représentation.

### 1.3.3.1 La dimension subjective

Les modèles formels de représentation sémantique envisagent le mot comme un objet isolable et pourvu d'un sens. "Objectivant" ce sens, ils le séparent arbitrairement du sujet et tentent d'en donner une représentation "dure". Le mot est ainsi artificiellement amputé de sa relation organique aux autres mots, tandis que le statut intersticiel du sens circulant dans la phrase y est ignoré.

Cette façon de procéder revient à tenter subrepticement d'essentialiser le sens. Qu'elles soient référentielles (représentation d'un objet général), ou conceptuelles (représentation d'un concept), ces théories reposent sur le postulat tacitement admis de l'essentialité. Elles présupposent que le sens du mot *blanc* par exemple, dans la phrase *ce papier est blanc*, pourrait être saisi indépendamment du mot *papier* représentant le support sur lequel il s'étale. Or, si nous faisons varier par l'imagination, le prédicat-support sur lequel cette couleur s'étale, le sens du mot *blanc* varie d'un prédicat à l'autre. Si, allant plus loin, nous tentons par ce même procédé de la variation, de séparer cette couleur de tout support, elle devient impossible à penser, im-pensable. On voit que l'essence ou *eidos* en grec, constitue un *horizon* (cf. Husserl, Id. p.141) sans cesse approché mais jamais atteint - un idéal-, vers lequel tendent les différentes actualisations du sens. En d'autres termes, l'*eidos* ne peut s'éprouver que dans une situation vécue. Ce caractère d'inachèvement correspond précisément à ce que l'on a appelé le "style" phénoménologique.

On sait par ailleurs, que les langues n'offrent pas toutes la même gamme de couleurs. Celles-ci relèvent d'un processus de création collective, par un groupe social, à un moment donné. L'origine historique du mot frappe donc de nullité toute démarche essentialiste visant à hypostasier le sens.

Plutôt donc que de nous situer dans la perspective d'une "objectivité idéale", nous avons adopté le point de vue phénoménologique de la congruence du couple sujet/objet .

Chez Husserl en effet, l'indissociabilité du sujet et de l'objet constitue le principe de base de la représentation. L'intentionnalité, ou regard "dirigé" sur, est l'essence même de la conscience. Percevoir, aimer, vouloir, c'est toujours percevoir, aimer, vouloir, quelque chose. A la différence de la conscience

traditionnelle, la conscience intentionnelle n'est pas une entité intérieure existant à côté d'un monde extérieur qu'elle ingurgiterait. Elle est toujours *conscience de*, rayon "dirigé" sur. C'est par cet acte de transcendance que *sujet* et *objet* se trouvent substantialisés en une représentation (nécessairement subjective). Corrélativement, tout objet est *objet pour*. La conscience et le monde forment donc un seul et même tissu, "matérialisé" par la représentation. Celle-ci constitue notre structure de contact avec le réel, notre mode d'être-au-monde. Le *sujet* étant partie prenante de la représentation, c'est tout le flux de nos "vécus" (les *Erlebnisse*), chargé de ses "data d'affection", "de sensation" etc. qui est engagé dans l'intentionnalité. "C'est l'intentionnalité qui caractérise la conscience au sens fort et qui autorise en même temps à traiter tout le flux du vécu comme un flux de conscience et comme l'unité de conscience ." (Husserl, Id. p.283). En ce sens on peut dire qu' il n'y a de monde que subjectif et d'objet qu'intentionnel. "La conscience et son objet forment une unité individuelle uniquement constituée par des vécus " (Husserl, Id. p.122).

Nous ouvrons ici une parenthèse pour préciser que la notion de subjectivité est à prendre à un double niveau, transcendantal et affectif. Le sujet transcendantal englobe l'ego concret au travers duquel se réalisent les actes intentionnels, sous différents modes : perception, sentiment, etc. Tout vécu, ou expérience concrète de la vie, doit être entendu comme un vecteur de la conscience intentionnelle.

La représentation est donc l'*acte intuitif* de la conscience intentionnelle par lequel nous appréhendons *cela* qui est donné. Elle se manifeste, à des degrés divers, dans la perception sous forme de *sensations*, dans l'imagination et la mémoire sous forme de *phantasmes*. Perception, imagination et mémoire peuvent être combinées dans des proportions variables.

Une précision est à apporter ici. Dans le cadre de notre expérience, les mots, nous l'avons vu, se présentent à nous dans leur pure matérialité, et nous n'avons aucune *intention signitive* (ou "intention d'expression"). Dans le contexte de notre expérience, le signe n'est pas en effet, comme dans l'acte significatif, ce au travers de quoi l'objet est visé sans se donner ; il est lui-même visé et perçu. Il ne s'agit toutefois pas d'un simple acte de perception, mais d'un acte de saisie (perception) de la *chose* en vue d'une évaluation. Le jugement est précisément le type même de l'acte à plusieurs rayons ou "visées" par lequel l'objet saisi n'est plus une simple *chose* perçue mais un "*objet*

*intentionnel complet* " sur lequel s'exerce une "double intentio, une double façon d'être dirigé vers " ( Husserl, Id.p.120).

On notera au passage que le concept d'intentionnalité issu de la philosophie scolastique et emprunté par Husserl à Brentano, est considéré aujourd'hui, par les sciences cognitives, comme fondamental dans la question de la représentation : "*Intentionality... is an important aspect of mental phenomena that cognitive science needs to account for. Futhermore, an understanding of intentionality is necessary in order fully to appreciate certain question about mental representation* " (G. Harman, 1988).

C'est également cette dimension subjective de la représentation qui est mise en évidence par Frege. Pour lui en effet, "*la dénotation d'un nom propre est l'objet même que nous désignons par ce nom ; la représentation que nous y joignons est entièrement subjective ; entre les deux gît le sens, qui n'est pas subjectif comme l'est la représentation, mais qui n'est pas l'objet lui-même.* " (G. Frege, 1976). On voit ici encore, que sens et représentation ne peuvent être superposés. Décrite par Frege comme "*un tableau intérieur*" formé du "*souvenir des impressions sensibles et des actions internes et externes auxquelles je me suis livré* ", la représentation "*de l'un n'est pas celle de l'autre*". Variable d'un instant à l'autre, elle ne vaut que hic et nunc (G. Frege 1976)

La notion de subjectivité permet donc de prendre en compte le spectre infini des modes de représentation à chaque conscience. Il est probable que l'image mentale du mot *vérité* apparaîtra avec des modalités différentes d'un individu à l'autre malgré son *faisceau permanent de prédicats* (Ibid. p.440).

Ainsi encore le simple mot *maison* suscitera-t-il des images mentales et des affects différents d'un individu à l'autre, et sans doute, différents d'un moment à l'autre, chez un même individu. Il pourra susciter une image diffuse ou précise, réaliste ou composite, figurant par exemple, une ou plusieurs pièces de la maison d'enfance, son aspect extérieur vu de loin ou de près, sa façade ou son jardin. Ce pourra être l'image d'une maison idéale, d'un appartement, d'un immeuble etc. Le mot pourra également être associé à un ressenti négatif atténuant l'intensité sensorielle de l'image. Il pourra encore renvoyer à une idée liée à l'expression *rentrer à la maison* , etc. Pourtant stricto sensu, une maison n'est ni un jardin, ni une pièce, ni un appartement, etc. On voit que l'approche phénoménologique de la représentation est à la fois, plus ouverte et plus

complexe que la notion de représentation formelle. Elle prend en compte le sujet et son flux de vécu. Ceci explique qu'au-delà de sa structure nécessaire, qui fait que le mot *maison* par exemple, est compris de tout le monde, la représentation d'un mot puisse présenter autant de variantes que d'individus, et même autant de variantes que de manifestations.

On voit comment la notion de représentation, prise dans sa dimension existentielle, permet d'appréhender le mot à évaluer. Le "Quid", ou *faisceau de prédicats essentiels* étant inaccessible, nous pouvons, en revanche, viser le "Quomodo" ou *mode d'apparaître*, lui-même déterminé par le flux du vécu.

### 1.3.3.2 La dimension dynamique

La représentation implique un acte de mémoration, et par conséquent une dimension temporelle de la conscience. L'image représentée naît en effet, d'un processus passif de synthèse mentale, consistant à confronter un présent, à un passé constitué par l'ensemble de nos états de vécus. Ce processus relève d'un calcul d'une complexité et d'une rapidité inconcevables : il suppose qu'à chaque vécu, la conscience, tout en intentionnant l'objet, se saisisse dans le même temps, par un mouvement réflexif, de l'ensemble des vécus (les *Erlebnisse*) passés, pour en unifier le flot et en exprimer une synthèse qu'elle confronte à l'objet, synthèse, elle-même réactualisée, à chaque représentation. D'où le caractère inachevé d'un moi constamment remanié. C'est à ce passé dépassé, et à cet *horizon* indéterminé, que renvoient le célèbre "*je est un autre*" de Rimbaud. C'est également le sens que l'on peut donner à ce passage de Saint-Augustin : "*La mémoire encore garde en un même contenu mes états d'âme, non pas tels toutefois qu'au moment où ils affectent l'âme dans sa substance, mais d'une manière bien différente, selon l'énergie propre de la mémoire. Je ne ressens, en effet aucune joie à me ressouvenir de mes joies de jadis, aucune tristesse à évoquer ma tristesse passée, aucune crainte à me rappeler que j'ai quelquefois eu de la crainte, aucun désir à me remémorer un ancien désir*" (Confessions X-14 (22)).

La psychanalyse et la psychiatrie ont depuis longtemps, mis en évidence cette corrélation entre le passé et le présent. Ainsi admet-on désormais, que même les souvenirs "oubliés" continuent de résonner en nous. "*Ils agissent sur les états représentatifs mais sans le devenir eux-mêmes, provoquent des inhibitions, des pulsions, des résistances, faisant leur oeuvre secrète, tirant sur*

*les ficelles des représentations qui occupent la scène, mais restant dans la pénombre.*" (J. Delay 1942).

Nous faisons ici une digression pour nous interroger sur ce mystérieux processus mental de calcul paramétrique. Ainsi, l'ensemble des énoncés auxquels nous avons été exposés durant notre vie est bien stocké en nous dans une mémoire (l'inconscient ?). La psychanalyse est sous-tendue par ce postulat ; un syndrome psychiatrique, l'ecmnésie, semble le prouver. Pourtant, sauf cas particuliers, il nous est impossible de nous souvenir aussi bien de ces énoncés, que des circonstances de leur production. Il semblerait que notre survie mentale et sociale exigent cet "oubli"<sup>6</sup>. Or, nous nous souvenons des mots ayant constitué ces énoncés, et nous les utilisons de façon pertinente. Il faut donc qu'un processus infiniment complexe "compile" les contextes linguistiques et extralinguistiques de chaque mot, avant d'en extraire une représentation et de reléguer ce matériau linguistique et extra-linguistique dans une zone "interdite". La série des contextes étant illimitée, cette compilation est sans fin. On peut échaffauder une grande quantité de scénarios pour décrire ce processus<sup>7</sup> : ils sont pour le moment tous invérifiables. Les sciences neuronales n'en sont elles-mêmes qu'aux supputations : *"Pourquoi ne pas imaginer que l'entrée en résonance d'objets mentaux au niveau de l'élément "cognitif" se communique au proche voisin "émotionnel" du cortex frontal, y déclenche des rafales d'impulsions qui se propagent jusqu'au système limbique et à l'hypothalamus, et y déterminent une tonalité affective positive de plaisir ou, au contraire, en cas de dissonance, un effet dépressif ?... On comprend... comment un seul mot pourra résonner ou dissoner avec une image de mémoire (et susciter joies ou détresses) "*. (J. P. Changeux, 1983).

Etant donné l'extrême complexité de ce processus, et l'extrême ignorance dans laquelle nous sommes quant à son mécanisme, nous continuerons donc de nous en tenir au plan philosophique.

---

<sup>6</sup> Le syndrome d'ecmnésie (ou hallucination du passé) caractéristique des délires de la mémoire, nous montre ce qui se passe lorsque le torrent d'images jaillies du souvenir submerge la conscience. Le mécanisme d'inhibition nécessaire à l'attention à la vie étant levé, l'adaptation au réel devient impossible. Le malade est inapte à l'action, "dé-socialisé". (cf. J. Delay, 1942).

<sup>7</sup> La modélisation conceptuelle de ce processus appartient à l'intelligence artificielle et au domaine des sciences de la cognition.

De ce point de vue, chaque nouvelle représentation, se constituant en souvenir dès lors qu'elle se produit, vient s'intégrer au flux entier de nos vécus, qu'elle prolonge et transforme. Inversement, la représentation R, d'une série donnée sera toujours différente de la précédente et de la suivante. Il ne s'agit donc pas de l'évolution isolée d'une série de représentations. Chacune d'elles, retentissant sur l'ensemble des représentations passées, c'est toute l'économie de notre système de représentation qui est modifiée, et c'est ce système entier qui effectue une translation dans le temps.

Compte tenu de cette activité incessante, et de cet ajustement permanent de tous les paramètres qui la déterminent, la représentation d'un mot ne peut à notre avis être envisagée que sous un angle dynamique. La représentation d'un mot M, sera donc sensiblement altérée par la représentation M1 du même mot, elle-même altérée par la représentation M2, etc. Cette série ouverte, portée par le mouvement d'une conscience, s'oppose à l'idée d'une représentation absolue représentant la vérité d'un objet.

Cette dimension dynamique s'oppose également à la conception aristotélicienne de l'image mentale statique, considérée comme l'empreinte d'un sceau sur de la cire. *"L'image de mon ami Pierre n'est pas un homunculus, une réduction photographique minuscule, déposée dans un microcosme cérébral, elle est une signification, elle est l'expression de mon ami Pierre. Quand elle se réveille, elle n'est pas à maturité. Elle est floue, elle s'accuse, elle s'estompe ; elle a des péripéties, des vicissitudes, des clartés et des pénombres, des aurores et des crépuscules. Elle vit. Uniforme et multiforme, elle est en continuel devenir "* (J. Delay, 1942).

En raison de cette labilité de l'image, nous ne nous enliserons pas sur le problème de la nature de l'image mentale. Il nous semble difficile, en l'état actuel des connaissances, d'en préciser quoi que ce soit. Depuis Aristote, pour lequel il ne pouvait y avoir de pensée sans image, jusqu'à Wittgenstein pour lequel même une proposition logique est image, ce problème a longtemps été débattu. Il est toujours en suspens dans les sciences cognitives : *"Theorists of mental imagery discuss such questions as whether mental images are more like pictures, descriptions, or complex data-structures "* (G. Harman, 1988).

En tout cas, l'opposition image réaliste / image conceptuelle ne nous paraît pas nécessairement recouvrir l'opposition nom concret / nom abstrait. La dimension subjective de la représentation s'oppose à cette dichotomie et

explique que le mot *justice* par exemple puisse renvoyer à l'image d'une balance, tandis que celle du mot *tapis* puisse renvoyer à une image "conceptuelle". Qu'elle provienne d'un nom concret ou d'un nom abstrait, la représentation, peut, compte tenu de sa dimension subjective, revêtir toute la gamme des modes d'apparaître des plus figuratifs aux plus informels, des plus précis aux plus élagués.

Nous définirons donc la représentation comme une image mentale substantialisant la congruence du couple sujet/objet avec des modes d'apparaître les plus variés. Ainsi conçue, cette notion ne nous permet que de tester la simple présence ou absence d'un mot dans notre lexique-texte "privé" à un moment donné. Bien que modeste, cet objectif présente cependant l'avantage de pouvoir être atteint.

## **1.4. La méthode de sélection**

### **1.4.1. Les critères**

#### 1.4.1.1. la représentation

Le concept de représentation ainsi entendu, nous a donc permis de dépasser les problèmes de sélection liés à l'évaluation de la compréhension d'un mot isolé. Cette dernière ne peut reposer que sur des énoncés, tandis que la représentation telle que nous l'avons définie, se réfère à la stricte présence du mot dans notre lexique interne. Nous avons donc décidé d'en faire un critère préalable, et de considérer qu'à partir du moment où un mot provoquait un "affect", une réminiscence, qu'il suscitait une évocation de quelque nature que ce soit, il serait dans un premier temps sélectionné. Nous dirons donc que, pour nous, la représentation est une **"image mentale - précise ou vague, forte ou faible, émotionnellement positive ou négative - signalant la simple présence du mot dans le lexique mental d'un individu"**. A partir du moment où nous avons adopté cette définition, le problème des degrés de compréhension ou de la variété relationnelle présentés par les mots évoqués en 1.2.4 et 1.2.5 s'est trouvé résolu de facto : ainsi tous les mots évoqués lors de l'exposé de ces deux points ont-ils pu (dans un premier temps) être retenus, ceci sans faire de distinction dans le continuum allant des mots jugés connus (*eau, huile, etc.*) aux mots jugés re-connus, avec ou sans certitude, (*bréhaigne, ortolan, chamsin, dormance, etc.*) en passant par les

mots jugés partiellement connus (*asphodèle, salsepareille, amadou, nitrates, actuariel*, etc.). Ont également été retenus les mots virtuellement connus (comme *multimiroir, postnéonatal, impuberté*, etc.) puisque l'application des règles de la morphologie dérivationnelle prouvait la présence de leur constituants dans notre lexique interne. L'ensemble de ces mots a été affecté de l'indice R1.

N'ont été considérés comme inconnus que les mots présentant un degré zéro de reconnaissance (R0). Exemples : *faine, facule, faiblage, journée, escabelle, grelottière, ancistrodon, exhéréder, kichenotte*, etc. Seuls ces mots ont donc été exclus dans un premier temps. Nous récapitulons ce double tri dans le tableau I ci-dessous :

Tableau I

	R1	R0	
C.	Eau Huile	Faine Facule Faiblage Journée Escabelle Grelottière Ancistrodon Exhéréder Kichenotte	I.
PC.	Asphodèle Salsepareille Amadou Actuariel		
REC.	Bréhaigne Ortolan Chamsin		
RECONS.	Multimiroir Postnéonatal Impuberté		

C. = connus    PC. = partiellement connus    REC. = reconnus  
RECONS. = reconstitué    I. = inconnus

#### 2.4.1.2. La plausibilité

Afin toutefois d'élaguer et de mieux tracer les contours de cet ensemble R1 encore trop imprécis, nous avons été amenée à envisager un second critère, celui de plausibilité. Ce critère, improprement confondu parfois avec les termes de "fréquence" ou de "probabilité" employés par les statisticiens, a consisté pour nous à évaluer intuitivement les possibilités d'apparition du mot dans l'ensemble des énoncés écrits ou oraux auxquels nous sommes habituellement exposée. Ce point de vue nous permettait d'établir un distinguo

entre des mots tels que *chemin*, *définir*, *arbitraire* ou *dithyrambique*, jugés plus plausibles que *salsepareille*, *géhenne* ou *dormance*. On trouvera dans le tableau II une évaluation différentielle d'une série d'exemples de paires reproductibles, les mots de gauche présentant une plus grande plausibilité que ceux de droite :

Tableau II

Evaluation différentielle  
Paires reproductibles

P+	P-
Circonscrire	<i>Circoncire</i>
Angine	<i>Angiome</i>
Veinure	<i>Veinule</i>
Accordéon	<i>Bandonéon</i>
Dragon	<i>Drageon</i>
Dollar	<i>Dirham</i>
Dialoguiste	<i>Dialogueur</i>
Sénile	<i>Sénescent</i>
Poutre	<i>Solive</i>
Chaumière	<i>Chaumine</i>
Odeur	<i>Fragrance</i>
Arriver	<i>Arrimer</i>
Fusil	<i>Arquebuse</i>
Champêtre	<i>Agreste</i>
Mignon	<i>Mignard</i>
Danger	<i>Dangerosité</i>
Epluchage	<i>Equeutage</i>
Miauler	<i>Meugler</i>
Cerisier	<i>Cerisaie</i>
Seriner	<i>Suriner</i>
Pédiatrie	<i>Gériatrie</i>
Mensualisation	<i>Menstruation</i>
Minimiser	<i>Minimaliser</i>
Rempoter	<i>Rempoter.</i>
Inventaire	<i>Eventaire</i>
Masturbation	<i>Onanisme</i>

etc.

Tous les mots sans exception ont été envisagés de ce point de vue. Ainsi les termes dérivés de noms propres (*lacanien*), de noms de lieux géographiques (*bourguignon*) ou de références historiques (*décembristes*) ont-ils été traités de la même façon que le vocabulaire usuel. *Péloponésien*, *cappadocien*, *ombrien* ou *abbeyllien* ont par exemple été jugés moins plausibles que *napoléonien*, *sunnite*, *ubuesque* ou *londonien*.

Par ailleurs, nous avons exercé le critère de plausibilité mot par mot, sans considération du paradigme auquel il appartenait. Ainsi avons - nous jugé *agrafe* ou *agrafeuse* plus plausible qu'*agrafeur* ; *égratigner* plus plausible qu'*égratignoir* ou *égratigneur* ; *porcelaine* plus plausible que *porcelainier* ; *laminer* plus plausible que *laminerie* , etc.

Au cours de ce processus de tri, nous avons été amenée à affiner ce critère en lui adjoignant un indice (plausibilité haute, 1 ; basse, 2 ; ou très faible, 3). Cette nécessité de gradation s'est imposée d'elle-même : le jugement binaire étant, dans de nombreux cas difficile à appliquer, ce procédé nous a permis d'intégrer les hésitations auxquelles ils donnaient lieu. Ainsi des mots tels que *pepsine*, *rempotage*, *crédence*, *djinn*, *quaternaire*, *placette*, *planisphère*, *sorbetière*, *scansion*, *savetier*, ne nous paraissaient pas pouvoir être classés avec des mots tels que *améliorer*, *prendre*, *périodique*, *hebdomadaire*, *progrès*, *place*, *auteur*, *film*, *expansif*, *probable*, *militaire*, *soldat*, etc ; ils ne paraissaient pas non plus devoir être amalgamés à des mots tels que *pseudodébile*, *grammatologie*, *préoedipien*, *soldanelle*, *tensoriel*, *surjeteuse*, *tabaculture*, *coupailler*, *nèpe*, *éclampsie*, *courlieu*, etc. Ils ont donc trouvé naturellement leur place dans une catégorie intermédiaire P2, tandis que les valeurs P1 et P3 ont respectivement été attribuées aux deux autres séries de mots. La série P2 nous a donc paru plus plausible que la série P3 mais cependant moins plausible que la série P1, d'où le tableau III:

Tableau III

P2	P1	P3
Pepsine	Améliorer	Pseudodébile
Rempotage	Prendre	Grammatologie
Crédence	Périodique	Préoedipien
Placette	Hebdomadaire	Soldanelle
Sorbetière	Progrès	Tabaculture
Planisphère	Place	Coupailler
Scansion	Auteur	Nèpe
Djinn	Film	Courlieu

Ce faisant, un bénéfice secondaire est apparu : celui d'une sélection plus décomposée, présentant des couches lexicales hiérarchisées allant de la plus "essentielle" à la moins "essentielle", et par conséquent, apportant de

nombreux avantages du point de vue de l'évaluation et de l'exploitation automatique de textes.

## 1.4.2. La méthode

### 1.4.2.1. La mise en oeuvre

La sélection de chaque mot a ainsi résulté du croisement de l'application de deux critères successifs : la représentation et la plausibilité. Seule la dissociation de ces deux points de vue, intra-subjectif et extra-subjectif, nous a parue apte à garantir une certaine cohérence du processus et une fiabilité normale des résultats.

A chaque mot a donc été attribué, lors d'un premier passage, une valeur R1 ou R0 : R0 pour un degré de représentation nul, R1 pour tout autre cas (allant de la représentation la plus forte à la représentation la plus faible).

Exemples:

-R0: *délétion, vaigrage, martogon, etc.*

-R1: *eau, vin, mentir, collusion, ellipse, synthèse, pléonasme, hypallage, adoubement, aiguière, crédence, méridienne*, aussi bien que des mots à la représentation, pour nous, plus floue (tels que *carénage, balsamine, salsepareille, cimaise, bréhaigne, médianoche* ou *énallage*) ou que des mots reconnus par application de règles dérivationnelles (comme *cabochardement, désucrage, détabouiser.*)

Un deuxième passage sur les mots affectés de la seule valeur R1 a été destiné à trier les mots selon trois niveaux de plausibilité : haute (P1), basse (P2) très faible ou proche de zéro (P3). A l'issue de ce deuxième passage se sont donc trouvés écartés les mots affectés d'un indice de plausibilité P3.. Ainsi le deuxième passage a-t-il donné, pour les mots que nous venons d'évoquer en R1:

-P1: *eau, vin, mentir, collusion, ellipse, synthèse, pléonasme.*

-P2: *nitrates, carénage, salsepareille, amadou, crédence, cimaise, aiguière.*

-P3: *balsamine, hypallage, cabochardement, bréhaigne, médianoche, désucrage, détabouiser, adoubement.*

On remarquera qu'un mot peut donner lieu à une représentation forte et être jugé moyennement ou très peu plausible (exemple : *aiguière, crédence, hypallage*). En revanche il nous paraît peu logique qu'un mot ne suscitant aucune représentation puisse être jugé plausible.

Le résultat de ces deux traitements successifs a donc abouti à la constitution de deux listes représentant respectivement les couches (essentielles) haute (C1), et basse (C2) de notre dictionnaire des mots simples du français, le reliquat constituant logiquement la troisième (C3) :

- les mots R1 / P1 ayant une représentation différente de zéro, et une plausibilité haute (couche haute, appelée C1).
- Les mots R1 / P2 ayant une représentation différente de zéro, et une plausibilité basse (couche basse, appelée C2).

Le filtrage réalisé par le croisement de nos deux critères a donc permis d'éliminer :

- au premier passage, les mots ayant un degré de représentation zéro (R0)
- au deuxième passage, les mots affectés d'une valeur de représentation non nulle (R1) mais présentant un degré de plausibilité très faible ou proche de zéro (R1 / P3). Cet ensemble constitue la couche C3 appelée couche résiduelle. Le tableau IV récapitule cette répartition.

Tableau IV:

**R0**

Délétion  
Vaigrage  
Margoton

**R1**

	<i>P1</i>	<i>P2</i>	<i>P3</i>
Eau	+		
Vin	+		
Mentir	+		
Synthèse	+		
Hypallage			+
Cimaise		+	
Carénage		+	
Désucrage			+

Résultat: Couche C1 = mots R1P1  
Couche C2 = mots R1P2  
Couche C3 = mots R1P3  
mots R0

1.4.2.2. Remarques sur le processus de sélection

### **Le traitement de certaines entrées**

Les listes sur lesquelles nous avons travaillé étant celles du dictionnaire électronique formalisé DELAS, nous nous sommes trouvée devant deux cas particuliers : la systématisation des variantes graphiques et les entrées sémantiquement non autonomes.

En ce qui concerne les variantes graphiques, la plupart du temps, lorsqu'un mot donnant lieu à plusieurs graphies avait été sélectionné sous une forme donnée, toutes les autres graphies ont été sélectionnées dans la même couche lexicale. Exemples: *cashier / kascher / cacher* ; *curry / carry* ; *tzar / czar / tsar* ; *pogrom / pogrome* ; *phantasme / fantasme* ; *pébroc / pébroque* ; *prétontaine / prétantaine*, etc. Les différentes graphies pouvant se trouver éloignées les unes des autres dans le rangement alphabétique, il convient dans ces cas-là de se livrer à des vérifications systématiques. Ces vérifications offrent en même temps une autre occasion de valider la sélection de certains mots, sous un angle différent.

Il est cependant arrivé que certaines graphies nous paraissent nettement moins plausibles que d'autres. A ce moment-là nous avons placé celles-ci dans une couche plus basse. Exemple : *enquiquineur, enquiquiner* (C1) mais *enkikiner, enkikineur* (C2).

Les entrées non autonomes du DELAS sont codées Xinc (cf. Bl. Courtois, 1990). C'est par exemple le cas de *ping* entrant nécessairement dans le composé *ping pong*. Dans ce cas-là, nous avons d'abord cherché à reconstituer l'extension compositionnelle. Si celle-ci était trouvée, le jugement de plausibilité était alors exercé sur l'ensemble de l'expression. Exemples : *ping / pong, parce / que, cahin / caha, nec / plus / ultra, poudre / d' / escampette, couci / couça*, etc.

On notera qu'une entrée non autonome peut entrer en composition avec une ou plusieurs autres entrées non autonomes, ou avec une ou plusieurs autres entrées autonomes. En tout état de cause, il faut s'assurer que le même indice de plausibilité a été attribué aux différents éléments de l'unité sémantique reconstituée.

### **Intra- et extra-subjectivité**

Nous attirons l'attention sur un aspect important de cette sélection. Il faut bien comprendre qu'il s'agit d'une sélection en deux temps, nécessitant un

mouvement inverse. L'application du critère de représentation s'exerce à l'intérieur de soi, tandis que celui de plausibilité exige de se projeter à l'extérieur de soi, et de se situer cette fois, du point de vue de l'ensemble des énoncés écrits ou oraux de notre groupe social. De ce fait, il peut se trouver, nous l'avons vu, des mots qui, tout en suscitant une représentation forte, peuvent être considérés comme devant être affectés d'un degré de plausibilité P2 ou P3. Ainsi connaissons-nous le sens de termes de rhétorique tels que *zeugme*, *hypallage*, *anacoluthie* etc. Or nous leur avons attribué, sans hésitation, un indice de plausibilité P3. En revanche nous avons attribué un indice de plausibilité haut (P1) à *pléonasme*, *métaphore*, *paraphrase*, *euphémisme*, *onomatopée*, etc. car, bien qu'étant des termes de rhétorique, ils nous ont paru susceptibles d'apparaître dans une diversité importante d'énoncés. De la même façon, un amateur de champignons pourra attribuer une valeur de représentation R1 et un indice de plausibilité P3 à des termes tels que *orange*, *cortinaire*, *pholiote*, *pézize*, *indigotier*, *coprin*, *lactaire*, etc. En revanche, il pourra considérer qu'un indice de plausibilité P1 devra être attribué à *cèpe* ou *truffe* et P2 à *amanite*, *bolet*, *chanterelle*, *coulemelle*, etc.

On doit également distinguer la plausibilité intra-subjective de la plausibilité extra-subjective. Il est en effet nécessaire d'évaluer intuitivement cette plausibilité en fonction de l'ensemble du groupe de population visé, ici un public cultivé de niveau universitaire, et non en fonction de sa seule expérience. Nos métiers, nos goûts, nos loisirs, nos centres d'intérêt etc. nous amènent parfois à développer une ou des compétences très pointues (médecine, sport, cuisine, musique, jardinage etc.). Or, même s'il n'est pas toujours facile de trancher, il est possible, dans une certaine mesure, de faire la part de ce qui relève d'une pratique de la langue écrite et orale partagée par le groupe de population visé auquel on appartient, et de mettre de côté ce qui relève d'une exposition à la langue liée à une pratique pointue (professionnelle ou privée). Ainsi, un spécialiste de télématique attribuera-t-il une valeur R1 mais un indice de plausibilité P3 à *multiplexeur* ou *reroutage* même s'il est amené à rencontrer quotidiennement ces termes dans son environnement professionnel. En revanche, il attribuera une valeur R1 et un indice de plausibilité P1 à *minitel*. De la même façon un médecin affectera les mots *électrocorticogramme* ou *électrocoagulation* d'un indice P3 et les mots *électrochoc*, *échographie* ou *électroencéphalogramme* d'un indice P1 (ou P2) etc.

Notons que les centres d'intérêts ne s'excluant pas chez un même individu, on imagine les combinaisons infinies de ces lexiques, chez tout sujet parlant. Ce phénomène d'interpénétration de lexiques "pointus" entre eux d'une part, et avec un lexique généralement partagé d'autre part, impose la figure d'un vaste maillage d'ensembles non-disjoints, eux-mêmes recoupés par des micro-ensembles non-disjoints représentant les lexiques de chaque individu.

Curieusement, les outils informatiques (bases de données relationnelles, hypertextes) évoluent vers ce type de représentation complexe où les objets sont tous interdépendants.

### **Les modifications de la perception du lexique**

Comme on peut s'en douter, des hésitations se produisent nécessairement. Il ne faut d'ailleurs pas espérer établir une sélection en un seul passage. Sept passages, eux-mêmes espacés dans le temps, ont été nécessaires pour l'ensemble des lettres du dictionnaire. Nous avons pu remarquer que la progression, lettre par lettre, sur l'ensemble du lexique français en modifiant la perception intuitive de départ provoquait à plusieurs reprises, une remise en perspective, et entraînait de ce fait, des réajustements rétroactifs de la sélection déjà faite. Les frontières virtuelles de nos trois couches lexicales se sont ainsi déplacées à plusieurs reprises, et nous ont amenée à opérer de nombreuses retouches. Ces frontières virtuelles sont d'ailleurs impossibles à formaliser. Il s'agit plutôt d'"horizons" visés, approchés, mais jamais atteints. C'est en partie pour cette raison qu'un sujet extérieur pourra contester le classement de certains mots. Il n'aura pas eu l'expérience d'une perception globale de l'ensemble du lexique. C'est aussi, une des raisons de l'inefficacité des tests (cf. ci-dessous 1.4.3.). Les sujets testés devraient avoir fait au préalable l'expérience d'une sélection sur l'ensemble du lexique.

Quoi qu'il en soit, un phénomène est apparu tout au long de ces différentes approches de sélection : les hésitations ont toujours joué sur deux couches successives (hésitations entre C1 couche haute, et C2 couche basse ou C2 couche basse, et C3 couche résiduelle), jamais sur les deux couches extrêmes (C1 et C3). Cette constatation étaye l'idée d'un modèle lexical à trois couches, même si ces dernières présentent des fluctuations par rapport à ce modèle théorique.

Par ailleurs, au cours de la sélection, un réflexe de recouplement s'établit entraînant de nombreuses vérifications : le mot est souvent reconsidéré à l'intérieur d'une série, dès lors que l'association avec d'autres mots provoque

cette configuration. La rencontre du mot *valériane* par exemple va nous pousser à vérifier le classement de *badiane*, *passiflore*, *belladone*, etc. et à reconsidérer la série. Ces associations naissent spontanément et sont de toutes natures. Le mot *palimpseste* peut entraîner par exemple la vérification de *palindrome* ou de *palinodie*, ou encore de *palanquin*, du fait d'une certaine ressemblance phonique et graphique de ces mots. De ce fait, la vérification n'a pas toujours le même but. Il peut s'agir tantôt d'harmoniser une série, tantôt de confronter deux jugements éloignés dans le temps.

Les diverses lectures que l'on peut faire en dehors de ce travail suscitent aussi souvent des vérifications.

Ces différentes observations amènent à conclure qu'il faut bannir toute idée de frontière stricte dans le domaine lexical. Aussi nous garderons-nous de présenter nos couches lexicales comme des listes étanches. Il faut plutôt les voir comme oscillant et s'interpénétrant. Du reste seul un modèle "dynamique", aux frontières imprécises, nous paraît pouvoir représenter la réalité lexicale. Il n'en demeure pas moins représentatif de la physionomie générale du bagage lexical d'une population (ici, celui de locuteurs cultivés ayant fait des études supérieures).

#### 1.4.2.3 Les surprises de la langue

Ajoutons que le degré de plausibilité d'un mot sera susceptible de présenter un écart soudain dans la langue. Par ailleurs, il faut s'attendre à voir apparaître quelques écarts dans un même texte. Il faudra en effet toujours compter avec un certain nombre de phénomènes impossibles à maîtriser, que l'on pourrait appeler les "surprises de la langue" : effets-mode (on connaît la fortune du barbarisme *incontournable*), épidémies ou apparitions inattendues. La langue est un corps en mouvement qui échappera sans cesse à nos velléités d'enfermement. Il faudra toujours compter avec un facteur d'imprévisibilité. Journaux et magazines en offrent des exemples frappants tous les jours :

- Le mot *noria* figurant avec une valeur P2 dans notre lexique, s'est par exemple trouvé tout - à - coup suremployé pendant la guerre du Golfe. Une sorte de mini - épidémie dans la presse écrite et parlée a donné lieu à une quantité inhabituelle d'énoncés mentionnant *la noria des avions*, *la noria des camions*, *la noria des bombardiers*, etc.

- Le mot *égoïne*, classé dans la couche basse (C2) de notre lexique s'est trouvé à un moment mis en vedette par l'apparition d'une publicité télévisée sur les *scies égoïnes*.

- Dans un type de texte correspondant, par exemple, au lexique de la couche C1, pourra surgir un mot "inattendu" (de la couche C2 ou C3) : ainsi trouvera-t-on dans l'annexe 4 deux courts textes extraits d'un magazine grand public, le texte d'une critique de spectacle dont tous les mots simples, sauf un, *réticule*, font partie de la couche C1 de notre lexique. Dans le second texte, tous les mots simples sauf un, *tanagra*, font partie de la couche C1. Les deux mots en question, sont classés dans notre couche lexicale C2.

- Des mots comme *cardamome*, *igname*, *pitas*, *nans*, *tzatziki*, etc. que l'on s'attendrait plutôt à voir apparaître dans un livre de recettes de cuisine apparaissent tout à coup en rafales dans un article du *Monde* consacré à l'évolution des moeurs culinaires (cf. Annexe 4). Précisons que nombre d'entre eux (*pitas*, *nans*, etc.) ne figurent d'ailleurs même pas dans pas dans les dictionnaires usuels.

Loin d'être des exceptions, ces exemples illustrent le phénomène le plus mystérieux de l'emploi de la langue : aucune loi ne pourra jamais permettre l'anticipation absolue de l'emploi des mots. Il faut définitivement admettre que n'importe quel terme peut apparaître, à tout moment, dans n'importe quel type de texte.

Afin toutefois de tester notre procédure de sélection nous avons entrepris de réaliser un test - témoin, soumettant à des locuteurs différents une liste brute de mots. Nous décrivons ci-dessous le processus d'élaboration du test. Nous présentons ensuite les conclusions auxquelles nous sommes arrivée.

### **1.4.3. Elaboration de test, analyse et conclusions.**

#### **la mise au point du test**

Il a été nécessaire de tester d'abord les questions posées. Il est en effet difficile de faire comprendre au sujet testé, ce que l'on attend exactement de lui : il faut que le sujet testé puisse très vite saisir qu'il ne s'agit pas de porter un jugement sur son "niveau de vocabulaire", mais d'appréhender grâce à ses indications, le vocabulaire du groupe social auquel il appartient. C'est pour

cette raison que le double critère de représentation et de plausibilité demande d'une part, à être bien dissocié, et doit être d'autre part suffisamment explicite ; il est également bon de préciser que le test doit se faire sans recours à aucun dictionnaire.

Une première version du test a donc été distribuée à une quarantaine de personnes, recouvrant un éventail professionnel diversifié (mais correspondant au public-cible). Cette distribution s'est faite parallèlement à la mise au point progressive des questions. Nous n'avons donc pas distribué ce questionnaire en une fois mais par lots de cinq. Entre chaque distribution le questionnaire subissait les modifications induites par le dernier lot d'entretiens.

Les discussions précédant et/ou suivant l'exécution du test nous ont ainsi permis de préciser de nombreux points, et d'arriver à la version finale (cf. Annexe 5). Outre l'intérêt présenté pour la mise au point des questions, cet exercice de verbalisation contraint à l'élucidation de nombreux aspects encore insuffisamment clairs pour soi-même.

### **Le choix des items**

A la différence des questions, l'échantillon d'items choisi n'a pas subi de modifications. Leur choix avait cependant été dicté par plusieurs préoccupations. Il fallait que les items soient susceptibles :

.de correspondre à un emploi courant (cf. *dactylo*, *bazar* ou *porcelaine*) et plus spécialisé (cf. *lactaire*, *postprandial* ou *monophongue*),

.de présenter des degrés de représentation variés, de façon que les sujets testés perçoivent la nuance entre connaissance et représentation (cf. *musique*, *buis*, *cimaise*, *lactaire*, *carénage*, *entropie*, *épithalame*, *perborate*),

.d'offrir des cas de connaissance virtuelle (cf. *cabochardement*, *bazardisation*),

.de présenter des cas prévisibles de représentation 0 (cf. *délétion*, *déjouement*),

.de présenter des cas générant l'association possible : représentation forte / plausibilité basse (P2 ou P3) (cf. *astragale*, *daguerrotypier*, *lamparo*, *burgrave*, *bazardisation*),

.de présenter des degrés de plausibilité variés (cf. *musique*, *dactylo*, *cocaïne*, *perborate*, *pepsine*, *langoustier*, *damoiseau*),

.de présenter des cas, qui, tout en se trouvant dans le même paradigme, puissent donner lieu à des valeurs de représentation et des indices de plausibilité différents (cf. paradigmes de *cocaïne*, *porcelaine* ou *bazar*).

### **L'analyse des réponses**

Une fois le test mis au point, nous l'avons distribué à un second échantillon de population, correspondant au même public-cible, soit une trentaine de personnes appartenant à des professions différentes (enseignants, professions libérales, ingénieurs.).

L'analyse des résultats a montré que même sur cette population réduite, les écarts étaient si nombreux qu'aucun modèle ne pouvait être élaboré à partir de l'exploitation des résultats ; ces écarts sont en grande partie dus, nous semble-t-il, à la difficulté de la part de sujets non spécialistes, à s'extraire de leur contexte personnel pour effectuer une mise à distance de la langue. Contrairement à notre hypothèse de départ nous sommes donc arrivés à la conclusion que seuls des professionnels (c'est-à-dire des linguistes), habitués à considérer la langue comme un objet d'observation, pouvaient se livrer à une telle activité. Cette expérience montre également qu'une méthode de sélection fondée exclusivement sur l'exploitation statistique de tests à grande échelle, donnerait des résultats probablement inutilisables : les écarts et les aberrations entre individus seraient tels, qu'ils finiraient par noyer les trois couches faisant avorter ainsi toute velléité de modèle.

Nous pensons qu'il serait plus intéressant de tester plutôt les couches déjà séparées. Ce point de vue est fondé sur la constatation suivante : le regroupement de chacun de nos trois ensemble, par les moyens informatiques du système InTex, nous a permis d'avoir une perception inédite de notre sélection. L'homogénéité de chaque couche apparaissant plus clairement une fois les mots de chaque couche triés, rassemblés et imprimés, nous avons mieux pu distinguer les écarts par rapport à chaque ensemble. Il est également plus facile de repérer (et de corriger) les erreurs qui se détachent beaucoup mieux. Le lecteur s'en convaincra aisément en examinant les échantillons donnés dans l'annexe 7.

Si les tests statistiques sur des locuteurs peuvent difficilement constituer un outil de contrôle, il existe en revanche, des moyens informatiques susceptibles de garantir et de perfectionner progressivement une sélection "intuitive", exécutée par un ou plusieurs professionnels. Avant de rendre compte des résultats obtenus par ces moyens, nous expliquerons en quoi ils consistent.

#### **1.4.4. Outils et processus de traitement automatique.**

L'utilisation de la méthodologie que nous avons élaborée n'a pu être envisagée qu'en raison de l'existence de nouveaux outils linguistiques formalisés permettant de l'étayer.

A la différence des matériaux linguistiques traditionnels, les dictionnaires électroniques du LADL sont conçus de façon à servir aux deux grandes familles de traitement automatiques que sont, la génération et la reconnaissance automatiques. De ce fait, ils offraient pour nous un double avantage. Ils nous permettaient :

- en amont, de pouvoir opérer notre sélection sur une base de données tendant à une complétude de la couverture lexicale,

- en aval, d'envisager, grâce à la nature des informations associées à chaque entrée, et au système InTex qui les utilise, la mise à l'épreuve de notre extraction, par traitements automatiques de textes et évaluation de la couverture lexicale par rapport aux différentes couches extraites.

Pour comprendre le principe du traitement automatique, il est nécessaire de bien prendre la mesure de ce qui différencie les matériaux traditionnels des matériaux formalisés. Contrairement à une idée répandue jusque dans les milieux de l'"intelligence artificielle", le simple transfert de dictionnaires - papiers sur support informatique ne peut convenir aux applications informatiques.

#### 1.4.4.1. Dictionnaires usuels et dictionnaires électroniques

##### a) Actualisation et exhaustivité des données

A la différence des dictionnaires du commerce, la taille de ces dictionnaires électroniques n'a pas été limitée par des contraintes financières, commerciales ou pseudo-culturelles (on a vu que le DELAS contenait à lui seul aujourd'hui 80.000 entrées). Tous les registres de langue, du plus familier au plus relevé, ont été pris en compte, à partir du moment où ils se sont trouvés attestés dans la langue écrite ou orale. Soumis en outre à une maintenance constante, commodité qui provient du fait qu'ils se trouvent sur support informatique et qu'un logiciel spécifique a été construit dans ce but (Bl. Courtois, 1990), ils présentent, à tout moment, le dernier état de la langue. Nous avons donc l'avantage de nous trouver devant la base de mots la plus complète et la plus actualisée aujourd'hui. Ceci constitue la condition sine qua non de notre méthodologie. Le principe de l'extraction n'est soutenable que si celle-ci s'exerce à partir d'un matériau offrant des données complètes et actualisées.

Cette raison justifierait déjà, à elle seule, que l'on ne se soit pas livrée à cette opération de délestage à partir d'un dictionnaire usuel. En comparant par exemple le dictionnaire *Lexis* de Larousse (1976) (LS) et le *Dictionnaire du Français* de Hachette (1989) (DF), on constate que certains mots usuels manquent dans l'un comme dans l'autre ; cette absence prouve qu'une date récente de parution ne garantit pas l'actualisation des données. Ainsi ne trouve-t-on dans aucun des deux dictionnaires des mots tels que *acra*, *accessoiriser*, *amochage*, *bintje*, *brandouiller*, *brunch*, *brumisateur*, *casting*, *fréquentiel*, *friqué*, etc. Or, compte tenu de notre objectif, certains de ces mots nous sont beaucoup plus indispensables que des mots comme *imbibition*, *procaryote*, *zancle*, ou *zinzinuler*, qui eux s'y trouvent.

#### b) Les critères de choix

La diversité des tailles des dictionnaires ne correspond pas à un échelonnement allant graduellement des mots les plus faciles aux mots les plus difficiles. Une sélection ne pourrait pas s'appuyer sur une logique de progression garantissant que les mots contenus dans un dictionnaire seraient d'autant plus usuels que sa taille serait plus modeste. Chaque dictionnaire est en effet amené, dans les limites d'une taille établie par des impératifs commerciaux, à ménager un équilibre entre les termes connus de tout sujet parlant (natif), et les mots supposés inégalement connus d'une part, entre les termes usuels et les termes scientifiques et techniques d'autre part. Or, la sélection imposée par des limites de taille impose une marge d'arbitraire s'exerçant aussi bien sur le choix des mots usuels que sur les mots considérés comme plus difficiles. On ne s'étonnera donc pas que les choix des auteurs présentent, du fait de ces contraintes, des variations incompatibles avec notre projet. Une comparaison des entrées de quelques mots rares pris au hasard, sur deux dictionnaires, établit l'aléa de ces choix : Ainsi peut-on trouver les mots *caracal*, *hochequeue*, *échenillage*, *auréomycine*, *anacarde*, *barzoï*, dans le *Nouveau Dictionnaire du Français Contemporain* illustré (NDFC) de Larousse (1980). Or ces mots sont absents du *Dictionnaire du Français Contemporain* (DFC) de Larousse (1966), qui est pourtant plus volumineux que le précédent. De la même façon, on trouve les mots *bugle*, *bugrane*, *bouffette*, *castine*, *harpail*, *cachexie*, *castoreum*, *haschischin*, dans LS et DF, mais on ne trouve ces mots ni dans DFC ni dans NDFC.

L'obligation de se limiter à un certain nombre de mots donne souvent lieu à des choix encore plus aléatoires lorsque ces mots relèvent de processus compositionnels :

En examinant, dans trois des dictionnaires mentionnés, la liste des entrées composées du préfixe (soudé ou non) *anti-*, on remarque qu'aucune des listes n'est superposable à une autre. Ceci est à notre avis lié à deux raisons : d'une part le nombre élevé de mots composés de ce préfixe ne permettait pas de les inclure tous. D'autre part ces dictionnaires se trouvant toujours devant la nécessité de devoir ménager un équilibre entre mots connus et mots peu ou pas connus, n'ont pu faire, là encore, que des choix arbitraires. Ceci donne parfois lieu à des incohérences troublantes. Ainsi dans DF trouve-t-on *dreyfusard* mais non *antidreyfusard*, *dumping* mais non *antidumping*, *fascistes* mais non *antifascistes*, *antisocial* mais non *antigouvernemental*, etc.

Nous avons ici un exemple flagrant d'impossibilité d'utilisation de tels outils, pour des projets reposant sur des objectifs strictement linguistiques. Nous trouvons des mots aussi rarement utilisés que *antiémétique*, *antihistaminique*, *antilacet*, *anti-g*, etc. mais nous ne trouvons pas *antipelluculaire*, *antipoison*, *antigang*, *antijeunes*, etc. Or dans notre lexique, ces derniers mots figurent dans la couche haute, tandis que les premiers sont relégués dans la couche la plus basse. Outre ce manque de fiabilité, les dictionnaires usuels, ne peuvent, du fait même du contenu de l'information liée à chaque entrée, être utilisés dans des applications informatiques. Précisons que le fait que ces dictionnaires soient parfois publiés sur support informatique (cf. *Zyromys* ou *Le Robert* sur CD-ROM) ne change rien au problème. Ce sont de simples dictionnaires-papier transposés sur support optique. Même s'ils sont associés à des logiciels permettant une consultation navigationnelle, celle-ci reste interne au dictionnaire. La nature des informations contenues dans ces dictionnaires dits informatiques, étant rigoureusement identique à celle de leur version-papier, il n'est pas possible de les utiliser pour des traitements automatiques de textes.

#### c) La nature des informations

L'exhaustivité et l'explicitation des informations doit atteindre, pour que la reconnaissance automatique de termes dans un texte soit possible, une systématisation achevée et appliquée aux moindres détails. Ces informations concernent non pas les définitions, dont les dictionnaires électroniques destinés aux applications informatiques n'ont aucun usage, mais les indications destinées à la reconnaissance automatique. Ce processus exige en

effet qu'aucune information ne soit implicite : "un dictionnaire électronique du français devra par exemple comporter une liste complète des adjectifs en -able dérivés des verbes. Comme la dérivation (accrocher-accrochable) est intuitivement régulière, les dictionnaires courants laissent au lecteur le soin d'appliquer la règle. Néanmoins, dans ces dictionnaires on trouve parfois et de manière imprévisible des paires évidentes à constituer comme (abaisser-abaisable) à côté de cas difficiles comme (aimer- aimable) qui eux sont systématiquement énumérés." (M. Gross 1989) Cette absence de règles qui affecte d'autres catégories (adverbes en -ment par exemple) génère un arbitraire inconcevable dans un dictionnaire électronique, dont la fonction consiste à repérer dans un texte quelconque toutes les formes sans exception.

De la même façon, les dictionnaires usuels ne présentent pas systématiquement les différentes graphies de mots. En outre, ces graphies varient d'un dictionnaire à l'autre. Or l'indication systématique de toutes les graphies est indispensable à tout processus de reconnaissance automatique (M. Mathieu-Colas, 1988) : exemples: *czar, tsar, tzar ; casher, cascher, casher ; fromgi, fromegi ; praesidium présidium ; etc.* Le DELAS les a donc systématiquement prises en compte.

C'est pour répondre à cette même fonction de reconnaissance, que les entrées sont assorties, non pas de définitions (inutilisables dans le processus de reconnaissance automatique), mais d'"indications grammaticales codées et formatées utilisables par un analyseur syntaxique" (M. Gross, 1989). Ainsi les 80.000 entrées du DELAS sont-elles toutes associées à une partie du discours et un code flexionnel (cf. Annexe 1). Pour cela, une liste exhaustive des types de flexion de chacune des classes des mots variables (verbes, noms, adjectifs) a été définie, soit : 99 types de verbes et 80 types de noms et adjectifs. Chaque mot du dictionnaire a reçu le code le définissant sur ces bases. Exemples : *chapeau.N3* ou *localiser.V3*. Les mots entrant dans plusieurs classes flexionnelles ont reçu les codes correspondants. Exemple : *lire.N21.V94*

#### 1.4.3.2. L'interaction dictionnaires électroniques-textes

##### a) Les principes

Lors du processus de reconnaissance automatique ces codes sont utilisés de la façon suivante:

Un logiciel de génération automatique de formes fléchies (Bl. Courtois, 1984) engendre, à partir du DELAS, toutes les formes flexionnées de la

langue, soit aujourd'hui 700.000 termes composant le DELAF (dictionnaire électronique des formes fléchies cf. Annexe 1). Les mots des textes sont donc comparés à ce lexique et reconnus sous la forme fléchie où ils apparaissent. Cette dernière est alors automatiquement mise en correspondance avec sa forme lemmatisée (l'entrée du DELAS). A l'issue de ce processus, un sous-dictionnaire du texte est ainsi créé. Les mots de ce sous-dictionnaire sont classés par ordre alphabétique de leur entrée canonique (forme du DELAS) On trouvera dans l'annexe consacrée aux programmes du système InTex (Annexe 6), un exemple de ce programme (programme "creatdic").

#### b) Le problème des ambiguïtés

La reconnaissance automatique demande que l'ambiguïté présentée par de nombreux mots (*le* article et *le* pronom par exemple), puisse être levée. Ce problème ne pourra être résolu que très progressivement. A l'heure actuelle, lorsque pour un mot, plusieurs descriptions sont possibles, elles sont toutes automatiquement attribuées. Le système ne permet pas une désambiguïtation systématique. Cette étape demandera la formalisation préalable, sous forme d'automates, de milliers de grammaires locales (c'est - à - dire de points morphosyntaxiques ou lexico-morphosyntaxique isolés et formalisés au niveau de la langue), pour la plupart d'une grande complexité. Ces travaux sont en cours (Cf. D. Maurel, 1990, E. Roche 1992).

### **1.4.5. Résultats et limites de la sélection**

#### 1.4.5.1. Les résultats

Le résultat de notre sélection a abouti à 48 % des 80.000 mots du total du DELAS. Sur ces 48% (soit près de 38.000 mots), la couche C1 représente environ 30% (soit quelques 24.000 mots) et la couche C2, 18% (soit quelques 14.000 mots). Les 52% restants (soit 42.000 mots environ) constituent donc la couche C3.

Pour la consultation d'échantillons de ces différentes couches, on se reportera à l'Annexe 7.

Rappelons que nous n'avons jusqu'à présent pris en compte que les mots simples. Ces chiffres mettent en évidence la nécessité d'une acquisition lexicale très importante, si l'on veut communiquer en langue étrangère,

autrement qu'à un niveau de survie. Il serait sans doute nécessaire, à la lumière de ces chiffres de reconsidérer la place accordée au lexique dans l'enseignement des langues.

**En faisant l'hypothèse de ces mêmes pourcentages pour l'extraction des mots composés, on peut estimer à plus de 100.000 termes le lexique d'un locuteur cultivé.**

Ces résultats n'offrent toutefois pas le degré de précision que l'on serait en droit de souhaiter. Notre méthode présente en effet des limites qui ne peuvent être facilement dépassées, dès lors qu'il s'agit du traitement du lexique à l'échelle d'une langue. Outre le fait, qu'à ce stade de notre sélection, les mots composés, n'ont pas encore été pris en compte, ces limites sont essentiellement liées au problème de la séparation des sens. Pour cerner les difficultés que nous aurions eu à affronter, si nous nous étions donné comme but une telle entreprise, revenons à notre liste de départ, le DELAS .

#### 1.4.5.2. Le problème de la séparation des sens

Les entrées de ce dictionnaire reposent sur une définition formelle du mot : "*Les mots simples sont des séquences de lettres comprises entre deux séparateurs consécutifs.*" (M. Silberztein, 1989). Que le mot soit unique (comme *anarchie* ou *fabriquer*) ou qu'il ait des homographes, (comme *filer* ou *forfait*), il n'y a donc qu'une seule entrée. Dans un certain nombre de cas, les codes morphologiques associés à chaque entrée distinguent les catégories grammaticales. Ainsi le mot *déjeuner* est accompagné de deux indications (N1, nom masculin qui prend un *s* au pluriel et V3 conjugaison numéro 3). Nous pourrions déduire que ce mot a au moins, deux sens. Pourtant toute une catégorie d'homographes échappe à ces filets morphologiques. Ainsi le mot *rame* n'est accompagné que d'un seul code de flexion, N21, nom féminin qui prend un *s* au pluriel, bien qu'il puisse renvoyer à *rame de bateau*, *rame de papier* ou *rame de métro*, puisque dans les trois cas le mot est féminin. De la même façon, le mot *campagne* qui peut renvoyer à la campagne, opposée à la ville, ou à la campagne au sens militaire de *partir en campagne*, n'est suivie que de l'indication N21. Ainsi encore le mot *cour* n'est suivi que du seul code, N21. Or une *cour d'immeuble* est autre chose qu'une *cour d'assise* ou que la *cour d'un roi*, ou que celle que l'on fait à une dame ou à un homme.

Il aurait fallu pour distinguer ces sens, appliquer la méthode syntaxique utilisée par le LADL, c'est-à-dire recenser et formaliser les propriétés syntaxiques de tous les mots du français. Ce problème est largement résolu pour les verbes, puisque chacune des entrées du lexique-grammaire des verbes français réalisé au LADL, a été déterminée en fonction de ses propriétés syntaxiques (cf. Annexe.8). Les homographes se sont donc trouvés naturellement répartis dans des classes et des tables différentes. Exemple : le verbe *compter*, pour lequel 8 sens ont été distingués, se retrouve comme entrée des 8 tables syntaxiques suivantes :

5	<i>Que Max ait fait cela compte pour Pol</i>
6	<i>Pol compte que Luc sera là</i>
16	<i>Pol compte sur le fait que Luc sera là pour sortir</i>
32PL	<i>Pol compte les oies du troupeau</i>
32NUM	<i>Pol a compté trente oies dans le troupeau</i>
36DT	<i>Pol compte les photocopies à Luc</i>
35R	<i>Max compte sur ses doigts</i>
38R	<i>Pol compte Max parmi ses amis</i>

Ces indications de classes et de tables sont en cours de report dans le DELAS.(C. Leclère, 1990). Désormais chaque verbe du DELAS appartenant au noyau est progressivement accompagné d'une codification morphosyntaxique. Ainsi, l'entrée

*compter, V3*

devient:

*compter, V3 (5, 6, 16, 32PL, 32NUM, 36DT, 35R, 38R).*

indiquant par là que le verbe *compter* de conjugaison V3, entre, pour chacun de ses 8 sens possibles, respectivement, dans les tables syntaxiques 5, 6, 16, 32PL, 32NUM, 36DT, 35R et 38R.

On voit comment ces nouveaux codes constituent les clés d'accès au système de tables syntaxiques. La confrontation de la configuration syntaxique du verbe dans la phrase d'un texte, avec les propriétés syntaxiques définies dans la table, permettront de lever les ambiguïtés liées au problème des homographes.

Ainsi, l'ensemble des propriétés syntaxiques de chaque sens du verbe compter étant différentes, la confrontation de chaque ensemble avec la configuration syntaxique du verbe comter dans un texte permettra de lever les

ambiguïtés de sens. Pour que cette confrontation automatique soit possible, il faudra cependant que le système génère préalablement, par le biais d'automates, toutes les structures correspondant aux propriétés de chaque verbe. (cf. E. Roche, 1992).

Un travail analogue a été fait sur les conjonctions (M. Piot, 1978) et les adverbes (M. Gross, 1990), et est en cours de déroulement sur les autres catégories de mots : Noms (J. Giry-Schneider, 1978, 1982 ; G. Gross, 1990 ; Adjectifs, (A. Meunier 1981). Ces travaux permettront donc à long terme d'apporter à nos couches lexicales les informations formelles permettant la séparation des sens et de donner, par là même, à chacun des mots, une clé d'accès au lexique-grammaire.

Ceci nous amène à revenir sur le critère de représentation envisagé dans son rapport avec les homographes. D'après notre expérience, deux cas peuvent se présenter :

- le mot suscite une représentation nettement prépondérante. Exemples : chemin (de terre) pour *chemin*, ou campagne (opposé à ville) pour *campagne* .

- le mot suscite plusieurs représentations équivalentes. Celles-ci surgissent successivement, et à des intervalles de temps variables selon les mots. Exemple : le mot *noir* peut suggérer différentes représentations liées à la race, à la couleur ou à au sens familier de *café*. Nous avons pu observer que, même de valeur égale, ces représentations semblent ne pas pouvoir être concomitantes. Il y a succession, fût-elle imperceptible, dans le temps. On ne peut par exemple pas penser à la fois à *rame de papier* et *rame de métro*. Il est nécessaire d'exécuter un va-et-vient entre les deux sens. Tous les sens, surtout s'ils sont nombreux, ne viennent d'ailleurs pas à l'esprit, et nous ne nous sommes livrée à aucune vérification, puisque tel n'était pas notre objectif.

Par ailleurs nous avons pu constater, que l'indication grammaticale liée au mot pouvait susciter une représentation rétroactive, que la rencontre avec le mot "brut" ne suffit pas à provoquer. Par exemple le mot *jaune*, peut ne susciter, dans un premier temps, qu'une représentation liée à la couleur. Mais l'indication grammaticale N31 (en plus de A31, N1 et Adve) peut jouer un rôle de rappel, suscitant une seconde représentation, liée au sens de non-gréviste, à laquelle on n'avait spontanément pas pensé. Cette observation semble renforcer l'hypothèse d'un lexique-grammaire mental intégré.

Tous ces cas se ramenant, malgré leur variété, à une représentation R1, ils n'ont pas remis en cause notre méthodologie.

En ce qui concerne le critère de plausibilité, le mot a été envisagé de la même façon que pour les entrées du dictionnaire : soit une suite de caractères donnée, les possibilités d'apparition de cette suite sont-elles fortes ? moyennes ? ou très faibles ?

Là encore, la distinction des sens n'a pas été prise en compte. Nous ne nous sommes donc pas préoccupée de savoir, si par exemple, les mots *garni* ou *rempli*, tous deux accompagnés du double code N, A32, présentaient une plus forte plausibilité comme adjectif que comme nom. (De plus, ces deux mots sont des participes passés passifs de verbes, catégorie que les auteurs de dictionnaire ont du mal à séparer de celle de l'adjectif).

Nous précisons que les méthodes statistiques ne tiennent pas davantage compte des séparations de sens. Ces méthodes n'opèrent que sur des suites de caractères et en général ne sont pas couplées à des procédures de levées d'ambiguïté, qui ne pourraient être que manuelles en l'état actuel des possibilités d'analyse automatique des textes.

#### 1.4.5.3. Evaluation comparative

Nous rendons compte d'une comparaison de nos résultats avec ceux qui ont été fournis par une méthode statistique.

Nous présentons dans les tableaux V et VI deux séries de comparaisons portant respectivement sur les hapax des lettres *a* et *b* du *Dictionnaire de fréquence des mots du français* (J. Baudot, 1990). A l'intérieur de chacune des lettres nous avons distingué deux ensembles constitués respectivement par des exemples de mots figurant en C1 dans notre sélection et des exemples de mots figurant en C3 ou ne figurant pas du tout dans le DELAS.

Tableau V

- Lettre a, fréquence 1 (J. Baudot T.2, pp. 138-141):

C1= mots se trouvant dans la couche C1 du DELAS

C3= mots se trouvant dans la couche C3 du DELAS

\* = mots ne se trouvant pas dans le DELAS

C1	C3 ou *
abêtissement	<i>abhumanisme *</i>
adéquation	<i>ablette</i>
adjuger	<i>acétonitrile *</i>
admonestation	<i>achromatique</i>
admonester	<i>acter</i>
adoré	<i>activographe *</i>
aérer	<i>adipique</i>
aérogare	<i>aérolithe</i>
affligeant	<i>aficion *</i>
afflux	<i>agapologie *</i>
affolant	<i>aigremoine</i>
agencement	<i>aléser</i>
aigri	<i>alluvionnaire</i>
amaigrir	<i>amble</i>
amaigrissement	<i>aminodiacétique *</i>
annoter	<i>ampélopsis</i>
antériorité	<i>anagogiquement *</i>
antisémite	<i>anovulant *</i>
argot	<i>anthracène</i>
apprêté	<i>apostat</i>
attirant	<i>archangélique</i>
atroupement	<i>antigonadotrophine *</i>
aversion	<i>arpentage</i>

Tableau VI

- Lettre **b**, fréquence 1 (J. Baudot, T.2 pp.141-144)

C1= mots se trouvant dans la couche C1 du DELAS

C3= mots se trouvant dans la couche C3 du DELAS

\* = mots ne se trouvant pas dans le DELAS

C1	C3 ou *
badigeonner	<i>balistocardiographie *</i>
bagarrer	<i>barcasse</i>
bague	<i>baronnial</i>
bagnole	<i>baryum</i>
balade	<i>basicité</i>
banaliser	<i>basson</i>
banditisme	<i>batayole</i>
baratin	<i>baudrier</i>
barouf	<i>bénédictité</i>
basané	<i>biauriculaire *</i>
bassine	<i>bifilaire</i>
baudruche	<i>bilharziose</i>
béatitude	<i>bilirubine</i>
bécoter	<i>biostrome *</i>
belliqueux	<i>bistonquette *</i>
bercement	<i>blandice</i>
bévue	<i>bobinage</i>
biberon	<i>boucharder</i>
bibliothécaire	<i>bouvillon</i>
bicoque	<i>brai</i>
bigot	<i>braser</i>
biscotte	<i>brimborion</i>
bistouri	<i>bromatologue *</i>
bizarrerie	<i>brosserie</i>
blagueur	<i>bruitisme *</i>
bougeotte	<i>brunante</i>
bougrement	<i>butylmagnésium *</i>
boulot	<i>bryone</i>
bouilloire	<i>brodeur</i>
bouillotte	<i>bubonique</i>
boulangier	<i>brahmanique</i>
bourde	<i>brabançon</i>
boursoufflé	<i>boyard</i>
bredouille	<i>bradycardiaque *</i>
bricole	<i>blastocyte</i>
bringue	<i>brosserie</i>
brocanteur	<i>blastocyte *</i>
bronchite	<i>bijectif</i>

Même si notre répartition peut donner lieu à certaines discussions, on conviendra qu'une différence d'homogénéité importante apparaît entre les

deux traitements et que notre répartition présente des résultats plus objectifs dans le sens où un consensus de lecteurs peut être obtenu pour les listes intuitives, alors que les listes obtenues automatiquement à partir de corpus offrent une instabilité à laquelle il n'est pas possible de remédier.

#### 1.4.5.4. Evaluation textuelle

Le système InTex permet de valider une sélection lexicale sur de vastes corpus de textes et en des temps limités. A titre indicatif, le dictionnaire et l'étiquetage des mots simples d'une journée de dépêches de l'AFP, soit 1 Méga-octet a été réalisé en une trentaine de minutes. La vitesse d'étiquetage des mots d'un texte est de l'ordre de 200 à 300 mots / seconde.

Nous avons pu évaluer la couverture lexicale par nos trois couches sur un large corpus de textes de sources diverses :

- articles de magazines (Science et Vie, Société Magazine, articles du Monde, du Monde Informatique),
- manuels d'utilisateurs de machines
- textes techniques (Rapports techniques du ministère de la Recherche dans différents domaines)
- textes littéraires (un roman de Jules Verne),
- dépêches de l' AFP,
- etc. (cf. Annexe 9 un échantillon de ces dernières)

L'ensemble représentait d'environ 9 Mo, soit environ 1.328.000 occurrences de mots. Nous précisons que ces textes n'ont pas été choisis selon une typologie précise. Ils présentaient l'intérêt d'être déjà saisis et de se trouver disponibles au LADL. Il n'est en effet pas encore très facile d'obtenir des textes sur support informatique auprès des éditeurs divers, bien que cette pratique de la photocomposition informatisée soit généralisée aujourd'hui).

La constitution du sous-DELAS (ou sous-dictionnaire des mots simples) du corpus a fourni près de 20.000 entrées. Pour la décomposition du processus de traitement automatique (filtres successifs, consultation du DELAF, Dictionnaire électronique des formes fléchies, etc.), on se reportera au rapport de D. Clemenceau, 1992). On pourra y noter que "La recherche des 1.328.000 mots de notre corpus dans le DELAF s'effectue en 13 minutes environ, temps de chargement du dictionnaire compris, sur un PS/2 à base de 1386DX à 25 MHz sous OS/2 1.3. La consultation du DELAF s'effectue donc à une moyenne de plus de 100.000 mots à la minute".

Sur les 20.000 entrées constituant le sous-dictionnaire du corpus :

-15.000 mots correspondent à la couche C1

-2.500 mots correspondent à la couche C2

-2.000 mots correspondent à la couche C3

Le reliquat est constitué de mots non-reconnus par le DELAF. Sur ces mots, cf. D. Clémenceau, 1992.

On trouvera des échantillons des sous-dictionnaires C1, C2 et C3 de ce corpus de textes dans l'annexe 10.

Ces résultats ne peuvent que confirmer l'intérêt de cette méthode subjective et nous encourager à l'approfondir et à la développer.

Notre démarche a donc consisté à prendre un point de vue philosophique pour étayer ensuite la sélection réalisée sur ces bases, par des manipulations pratiques. Cette sélection donnera inévitablement lieu à des discussions. Il est dans la nature même de la méthode subjective de présenter une oscillation dans les jugements. Nous pouvons cependant dire d'ores et déjà, que le taux de polémique sera sans commune mesure avec celui qu'entraîne l'examen des listes de fréquences.

Par ailleurs, la construction d'un lexique demande la mise en oeuvre de critères divers et variés ; nous continuerons donc à procéder à toutes sortes de comparaisons. C'est précisément la diversité des approches, et la redondance dans les procédures de sélection des mots qui réduit l'arbitraire d'une liste. (cf. J. Sinclair, 1987).

#### **1.4.6. Les prolongements**

Nous envisageons de poursuivre progressivement ce travail dans plusieurs directions complémentaires :

**Partition de la couche lexicale C1 en 2 ou 3 sous-couches, toujours selon la même méthode.**

Nous voudrions, entre autres, arriver par une série de va-et-vient entre textes et sous-dictionnaires à déterminer un lexique-seuil, en d'autres termes

de cerner un taux garanti de couverture lexicale. Il restera à ajuster ce taux de couverture avec le taux lexical minimum exigé par une communication langagière

**Elaboration de corpus de textes jugés subjectivement de niveau lexical 1, 2, ou 3 et mise en correspondance de ces corpus avec les couches lexicales correspondantes.**

La confrontation de tels lexiques, et de textes jugés subjectivement de même niveau, devrait être un des facteurs contribuant à valider à la fois la sélection des lexiques et celle des textes. Les méthodes statistiques arguent de l'"objectivité" de leur méthode par l'utilisation de corpus de textes. Outre le fait que ces corpus ne permettent pas d'extrapoler le vocabulaire "courant" d'une langue, ils n'échappent pas à la subjectivité présidant au choix des textes. Nous pensons que la sélection délibérément subjective de textes devrait permettre d'arriver à de meilleurs résultats. L'interaction de la double subjectivité s'exerçant au niveau des lexiques et des corpus de textes devrait paradoxalement permettre d'atteindre des résultats plus "objectifs", par exemple à des listes de fréquences bien plus cohérentes et plus stables que les statistiques brutes proposées aujourd'hui.

**Partition hiérarchique des noms composés.**

Nous pensons dans un premier temps comparer les noms de notre couche C1 avec ceux qui entrent en composition dans les noms-composés recensés dans le DELAC (Dictionnaire des noms composés) Cette liste peut facilement être établie automatiquement. Il faudra ensuite procéder, comme pour les mots simples à une répartition manuelle en couches hiérarchiques. Ce même travail devra être fait pour les adjectifs les adverbes et les verbes composés. Nous présentons dans l'annexe 11 quelques exemples de sélection de noms composés de type NAdj.

**Elaboration d'automates de variantes d'expressions**

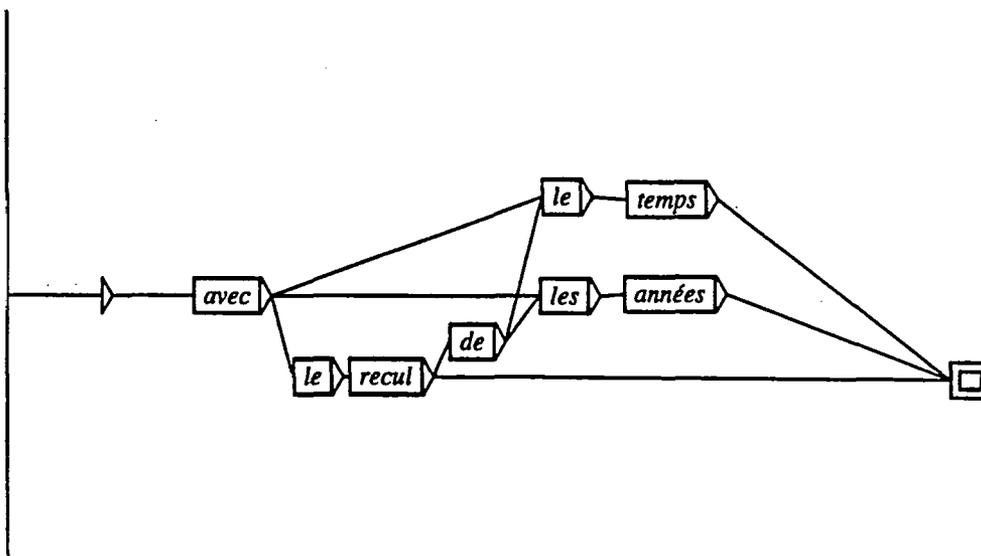
Un grand nombre d'unités sémantico-lexicales apparaissent comme des ensembles constitués de variantes. Ces ensembles peuvent et doivent être formalisés si l'on veut que le processus de reconnaissance automatique puisse s'exercer sur l'ensemble des variantes.

Une méthode naturelle de représentation des variantes lexicales et des abréviations de termes techniques consiste à utiliser des graphes d'automates finis (DAG : Directed Acyclic Graphs). Chaque chemin du

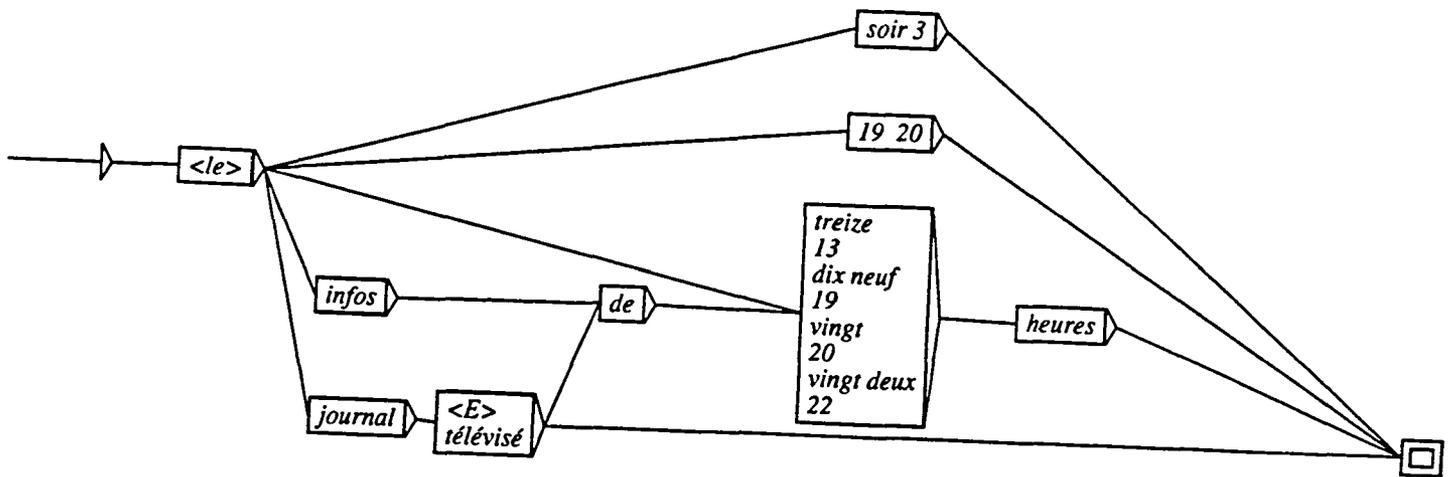
graphe, parcouru de gauche à droite, correspond à une expression possible. Le symbole <E> est un "zéro" qui correspond à des abréviations ou effacements. L'intérêt de cette représentation est que l'ensemble des expressions d'un graphe est déclaré comme comportant des formes équivalentes d'un point de vue de leur reconnaissance automatique dans des textes. L'indexation d'un texte regroupera donc en une seule unité les expressions d'un même automate qui sont détectées dans le texte. Les outils du système InTex permettent (comme nous le verrons de façon plus détaillée en 2.2) de formaliser de tels ensembles

Exemples :

*Avec le (recul du) temps* sera représenté par l'automate suivant :



*Le ((journal (télévisé) de) 20 heures* constituera un des parcours d'un automate plus complexe, consacré à l'ensemble des variantes de cette expression. :



### Elaboration de niveaux morphologiques

Le DELAF, constitue le dictionnaire de l'ensemble des formes fléchies des mots simples du français, il est nécessairement généré dans sa totalité lors du processus de reconnaissance automatique. Or, les formes générées ne sont pas également pertinentes. Le fait d'avoir à les produire toutes, est donc un facteur de ralentissement et de lourdeur. En appliquant le même principe de sélection qu'au DELAS, on obtiendra un outil plus léger, plus maniable et donc plus efficace dans les applications.

De fait, toutes les formes ne présentent pas le même degré de plausibilité d'apparition. Ainsi les verbes *rire*, *naître*, ou *paître*, figurant dans la couche C1 du DELAS (sous la forme canonique), auront des formes fléchies réparties dans différentes couches du DELAF : (je) *ris*, (nous) *naïssons* ou (ils) *paissent*, présentent en effet un degré de plausibilité plus grand que (que je) *risse*, (que nous) *naquissions* ou que (je) *pais*. Il faudra donc établir un second niveau de sélection, enchâssé dans celui des couches du DELAS.

### Elaboration de niveaux syntaxiques

Le même travail de sélection sera également à effectuer au niveau des propriétés syntaxiques attribuées à chacun des verbes dans les tables du lexique-grammaire construit au LADL. Toutes les propriétés ne présentent pas le même intérêt, pour des applications. Cette construction étant théorique, certaines propriétés ont été forcées. Ainsi que le remarque F. Debyser (s. d.), dans un ouvrage traitant de l'utilisation pédagogique de ces tables : "Alors que M. Gross a volontairement forcé la grammaticalité, c'est-à-dire qu'il a préféré

admettre, pour rendre compte dans le recueil exhaustif des données, des phrases qu'un puriste tâillon pouvait considérer comme douteuses, inusuelles ou "tirées par les cheveux", nous n'avons retenu, toujours dans la perspective de l'application pédagogique que des constructions ou des exemples dont l'acceptabilité ne peut être contestée."

Nous voyons donc qu'à ce niveau encore, le travail consistant à adapter des données théoriques aux applications, permettra de mettre ces dernières en oeuvre plus rapidement et plus efficacement. La simple partition du DELAS présente déjà des avantages dans plusieurs domaines d'applications.

### **1.4.7. Les perspectives d'applications**

#### **1.4.7.1. Aide à l'indexation documentaire automatique**

Grâce à cette hiérarchisation lexicale, il sera possible, d'améliorer le processus d'indexation automatique de textes spécialisés. Avec de tels textes, les mots courants ne doivent pas constituer des entrées de l'index. L'extraction préalable et automatique des mots simples et composés correspondant à la couche C1, permettra d'obtenir une liste plus pertinente des mots indexés. Lors du processus d'indexation, les items correspondant au vocabulaire courant (couche C1) pourront donc être automatiquement supprimés.

Bien entendu, ceci n'éliminera pas les retouches nécessitées par le problème de la séparation de sens (surtout pour les mots simples), qu'il faudra encore effectuer, si l'application le demande, à la main. Dans certaines applications, il ne sera peut-être nécessaire de le faire que pour certains mots. Prenons l'exemple de textes techniques d'ordre informatique, la première urgence sera peut-être de lever l'ambiguïté de mots tels que *micro*, *serveur*, *bus*, *puce*, *cartouche*, *souris*, etc. encore que dans des textes spécialisés, l'indexation de termes aussi courants ne serait sans doute pas toujours pertinente.

Précisons à ce propos, que certains logiciels lexicologiques ou documentaires distribués dans le commerce ou réalisés par des enseignants permettent de créer des dictionnaires de texte, sans dictionnaire : un logiciel

recupère tous les mots du texte et les classe sous forme alphabétique. Le problème est qu'il est très difficile ensuite de se livrer aux opérations d'ajustement "manuelles", car l'utilisateur se trouve noyé devant les innombrables séries de formes fléchies (exemple : *ai, ont, eut, aurait, avait, auront*, etc. au lieu de la forme canonique *avoir*). De plus, les mots composés ne sont pas systématiquement pris en compte. En outre, le lexique ainsi créé ne contient aucune information permettant d'envisager un quelconque prolongement d'utilisation. L'impossibilité devant laquelle se trouvent ces logiciels de mettre ces formes fléchies en correspondance avec les formes lemmatisées, et l'absence de toute référence aux codes qui doivent accompagner ces dernières ne permettent pas non plus d'envisager à plus long terme une mise en rapport avec une description syntaxique, (comme avec les tables du lexique-grammaire), elle-même nécessaire à l'amélioration de la reconnaissance automatique. Les systèmes "KWIC" d'indexation de documents basés sur les fréquences d'occurrences de mots dans les textes présentent les mêmes limitations. Dans les deux cas, les dictionnaires électroniques hiérarchisés, permettent une lemmatisation et une structuration des termes qui devraient améliorer notablement les résultats obtenus.

#### 1.4.7.2. Lexiques pédagogiques

Dans le domaine pédagogique, la confrontation de nos couches lexicales avec des textes techniques permettra d'extraire le vocabulaire technique spécifique, et d'en apprécier sa proportion, par rapport au vocabulaire usuel de notre lexique de référence (C1 et/ou C2). Cette application sera particulièrement appréciée dans l'enseignement des langues en milieu professionnel. On s'est aperçu en effet que contrairement à une idée répandue, les problèmes de compréhension d'un texte technique en langue étrangère provenaient moins des termes techniques (souvent transparents) que de la langue usuelle (A. Phal, 1971). On pourra donc, pour un corpus de textes de spécialité donné, délimiter ces deux types de vocabulaire et donner la liste des mots correspondant à l'un et à l'autre.

Inversement, il sera possible de fournir des textes correspondant à un ou plusieurs lexiques donnés. On évaluera par exemple, la couverture lexicale d'un corpus de textes A, avec le sous-dictionnaire extrait d'un corpus de textes B. Cette possibilité sera appréciée en langue où la demande de textes

correspondant à certains niveaux de vocabulaire, ou offrant des exemples de réemploi, est restée jusqu'à présent sans réponse.

Si, par ailleurs, l'application demandait l'illustration des entrées par des exemples, il serait également possible de les prélever automatiquement sur le corpus de textes, puisque le programme de "concordances" du système InTex permet de repérer toutes les occurrences d'un mot.(cf. Annexe 6).

Un autre type de demande pourra également être satisfait, celui du "vocabulaire d'intérêt" ou vocabulaire thématique. F. Debyser (1977) montre comment depuis "le choc en retour du Niveau 2" (à la suite des tentatives méthodologiques d'élaboration d'un Niveau 2 pour l'enseignement du français, langue étrangère), méthodologues et didacticiens se sont détachés de l'idée de niveau, pour privilégier la "définition des besoins langagiers des apprenants", la "recentration sur l'enseigné", le "développement de l'expression et de la créativité", l' "utilisation du document authentique", le choix de "thématiques pertinentes", etc.

D'où la nécessité de lexiques spécifiques, voire "sur mesure", répondant à la diversité des besoins de l'apprenant. Sous la poussée des orientations mentionnées, l'idée d'un lexique monolithique et perdurable, valable pour tout apprenant, en tout lieu et en tout temps, a vécu. Le problème est que, d'une part la demande de tels lexiques n'a jamais pu être satisfaite, et que d'autre part les notions de vocabulaire d'intérêt ou vocabulaire thématique n'a jamais été très clairement définie. La démarche consistant à définir des objectifs à partir de typologies de textes correspondant au public-cible, permettra de cerner ces zones d'intérêt et d'en extraire le vocabulaire.

Il sera également possible de combiner l'approche par niveaux et l'approche par centres d'intérêts, puisqu'il sera possible de déterminer pour un corpus thématique donné, la part de vocabulaire correspondant à chacune de nos couches lexicales. On pourra ainsi répondre au souhait resté jusqu'à présent incantatoire, de voir les listes du Français fondamental actualisées, et prolongées par une multiplicité de listes de vocabulaire, correspondant à une diversification des objectifs. *"On parviendra peut-être ainsi à maintenir le nécessaire équilibre entre les apports du vocabulaire général commun, servant de référence minimale, et les vocabulaires d'intérêt... variable selon les publics"*. (P. Rivenc 1979)

Rien n'interdira d'ajouter, à l'intérieur de ces lexiques-textes spécifiques, des indications de fréquences.

Ces lexiques-textes pourront également être adaptés et actualisés à volonté puisque le support informatique sur lequel ils se trouvent couramment aujourd'hui, et le système InTex le permettent.

#### 1.4.7.3. Aide au marquage lexicologique

Les dictionnaires usuels sont confrontés, nous l'avons vu, à la nécessité d'opérer des choix dans les vocabulaires spéciaux. Ce point constitue l'un des noeuds de difficultés les plus importants de cette discipline. A la difficulté de la sélection des termes de spécialité s'ajoute, entre autres, le problème toujours en suspens, des indications de domaines.

Une étude minutieuse du traitement dictionnaire du vocabulaire de la psychiatrie par M. Coutier (1988), met en évidence, à partir de l'examen de cinq dictionnaires grand public, entre autres défauts :

- *"l'hétérogénéité dans le traitement de ces marques par les divers dictionnaires et les contradictions propres à chacun d'eux"*. Ainsi, le nombre de domaines est-il variable d'un dictionnaire à l'autre. De *acoustique* à *zootechnique* on trouve : "283 (domaines) dans Lex., 239 dans H, 167 dans QF, 165 dans PL et 152 dans PR.",

- l'absence d'harmonisation dans la *"représentativité du vocabulaire de la psychiatrie"* dans ces cinq dictionnaires d'usage "grand public". Sur échantillonnage constitué de termes allant de *psy* à *psychopathologique*, 46 termes figurant dans les revues spécialisées et la presse, sont sporadiquement représentés dans les dictionnaires d'usage et pas du tout *"répertoriés dans les dictionnaires spécialisés"*,

- l'absence de normalisation dans la distribution des marques.

On constate que certains termes peuvent bénéficier d'indicateurs différents d'un dictionnaire à l'autre :

Exemples :

*"psychasthénie* : **Didact.** (PR), **Psychiatr.** (Lex., H), **Psychol.**(QF).

*psychédélique*: **Psychiatr.** (PR,H), **Physiol.** et **Psychol.** (QF)  
etc."

Cette absence de cohérence peut se retrouver à l'intérieur d'un même dictionnaire :

Exemple :

"dans H: **Méd.** pour *psycho-analeptique* et *psychodysleptique*, mais **Pharm.** pour *psycholeptique*"

Ces défauts se trouvent aggravés par d'autres problèmes qui viennent s'y greffer. Notamment :

- l'interférence de domaines conjoints : **neurobiologie, psychanalyse et philosophie** par exemple.

- Le passage de certains termes de spécialité dans le domaine public. Certains deviennent courants (*psychanalyse, psychiatre, psychopathe*) voire familiers (*parano, schizo*). Ils n'en conservent cependant pas moins leur usage spécialisé. Or, l'articulation entre ces deux plans ne se fait pas non plus de façon cohérente.

Cet ensemble de problèmes amène l'auteur à conclure, à la "*nécessité et à l'urgence de dépouillements plus systématiques (avec l'aide de l'informatique) de textes spécialisés et de vulgarisation, de monographies sur des domaines encore mal ou peu explorés, travaux sur lesquels pourraient s'appuyer les dictionnaires pour rendre compte des vocabulaires spéciaux et de leur usage*".

On voit que l'élaboration automatique de sous-dictionnaires de textes (dont le corpus serait constitué par des spécialistes de la discipline) pourraient constituer une aide précieuse dans le système de marquage des domaines. La généralisation du support informatique, permet également d'envisager une maintenance automatique afin de tenir compte de la prolifération terminologique due à l'évolution accélérée des spécialités.

Les conséquences de ces nouvelles possibilités de traitement automatique sont susceptibles d'atteindre à plus long terme, les fondements même de l'épistémologie lexicale.

Rien ne s'opposera désormais à ce que l'on envisage de réaliser le lexique du "fait divers", du "spectacle", de l'"hôtellerie", des "jeux télévisés", des "échecs" du "rodéo", etc. En réunissant un corpus de textes correspondant à l'objectif de l'application lexicale, on obtiendra automatiquement le lexique ad hoc.

Rien n'interdira non plus que l'on se livre, à l'aide d'opérateurs logiques, à des hybridations de lexiques. Il sera tout à fait possible d'obtenir le lexique du "fait divers" et de l'"hôtellerie", ou le lexique commun au "fait divers" et à l'"hôtellerie".

Ces manipulations pourront en outre donner lieu à des croisements entre lexiques spécifiques de la langue standard, ou à des croisements entre lexiques spécialisés, ou à des croisements entre lexique(s) spécifiques de la langue standard et lexique(s) spécialisées. Ainsi de nouvelles espèces lexicales du

type lexique de la "psychiatrie" ET lexique de l'"automobile" ET lexique du "fait divers" - pourront-elles être créées.

Le territoire lexical traditionnellement réparti entre langue d'usage et langues de spécialités est destiné à être sillonné en tous sens par une multiplicité de lexiques spécifiques. Cette configuration en structure maillée nous paraît du reste mieux représenter la réalité lexicale. De même que les lexiques-textes (internes) intra- et inter-individu(s) se recoupent, à un niveau microcosmique en un ensemble infini de classes non disjointes, de même, le lexique-texte que constitue la langue peut-il être représenté, à l'échelle macrocosmique, comme un enchevêtrement inextricable de lexiques-textes spécifiques.

Cette représentation du lexique en classes non-disjointes s'oppose à la classification en sous-classes disjointes à laquelle les dictionnaires usuels figés sur leur support unidimensionnel essaient de se conformer d'autant plus mal que la langue, animée d'une énergie continue, échappe à tout carcan classificatoire.

C'est donc l'adaptabilité des dictionnaires électroniques qui nous a permis d'en redistribuer les entrées de façon hiérarchique ; et c'est cette hiérarchisation qui permettra, comme nous venons de le voir au travers des champs d'application envisagés, d'en faciliter leur exploitation et d'apporter une aide "à façon".

Nous allons voir à présent que les grammaires électroniques qui manipulent ces lexiques offrent les mêmes capacités adaptatives.

## **2. Dictionnaires électroniques, grammaires locales et applications informatiques en langues**

*"- Mais alors, soupira Doug, selon vous, les oiseaux parleraient.*

*- Eh bien alors... baptisez-les ! "*

Vercors

### **2.1. Le traitement automatique dans la problématique du champ**

L'utilisation de l'informatique dans l'enseignement/apprentissage des langues s'inscrit dans la problématique née de la convergence de l'informatique, de la linguistique et de la didactique des langues. C'est la difficile articulation entre ces trois domaines qui explique les possibilités et les limites de l'utilisation des technologies avancées ainsi que les orientations des applications actuelles

La maîtrise d'une langue donnée est liée à l'acquisition d'une compétence de communication, autrement dit, à la capacité à produire et à comprendre un nombre infini d'énoncés de cette langue.

On peut s'étonner qu'il faille énoncer une telle évidence. Rappelons simplement qu'il y a une trentaine d'années encore, enseigner une langue étrangère consistait à enseigner la grammaire d'une part, et des listes de vocabulaire d'autre part, les deux domaines restant bien étanches. Résultat : une fois plongé en "milieu naturel", l'élève se trouvait devant l'impossibilité de demander un ticket de métro, même après 6 ou 7 années d'apprentissage.

Depuis, l'influence de la linguistique puis de la sociolinguistique (américaine surtout) ont progressivement transformé cet enseignement. Au carrefour de ces différentes sciences, un champ disciplinaire nouveau s'est ouvert, celui de la didactique des langues. De nouveaux champs continuent d'y converger, psycholinguistique ou sciences de la communication par exemple.

Au coeur de cette évolution un concept-clé est né : celui de compétence de communication. On redécouvrait par là qu'apprendre une langue ce n'était pas seulement apprendre un système formel de règles ou des listes de termes, mais que c'était surtout apprendre à "parler" en langue étrangère. Les termes "approches" ou "méthodes communicatives" ont depuis fait fortune. Il est peu de manuels ou de méthodes qui ne s'y réfèrent. Désormais, tout projet d'apprentissage s'inscrit dans la relation saussurienne locuteur-message-destinataire. Nous verrons plus bas, pourquoi les exercices élaborés dans cette perspective sont, pour le moment, difficilement compatibles avec un traitement linguistique automatique.

En situation de classe, le professeur qui a pour tâche de développer cette compétence de communication est capable de contrôler la production écrite et orale de tout apprenant, tant au niveau de sa correction strictement linguistique (orthographe, morphosyntaxe) qu'au niveau de la substance de son contenu (sémantique) et de sa pertinence fonctionnelle (pragmatique). Il est également, en principe, susceptible de produire des énoncés écrits et oraux destinés à l'acquisition d'une aptitude à la compréhension d'énoncés écrits et oraux. L'ordinateur est loin de pouvoir se substituer au professeur. Mais on sait aussi qu'il peut constituer pour ce dernier un auxiliaire nouveau susceptible d'un renouvellement dans sa pratique de classe.

Quelles sont les possibilités et les limites de l'ordinateur dans le domaine de l'enseignement/apprentissage des langues ? Comment et à quels niveaux dictionnaires et grammaires électroniques peuvent-ils intervenir ? C'est ce que nous nous proposons de montrer.

La seule façon de communiquer avec l'ordinateur est d'emprunter le canal de l'écrit. Nous écartons de notre propos le cas particulier de l'interaction non verbale - par le biais de menus ou de symboles par exemple - qui ne permet

aucun échange langagier.<sup>8</sup> L'utilisation du canal écrit suppose donc que l'utilisateur va produire des phrases que l'ordinateur pourra contrôler, d'un triple point de vue morpho-syntaxique, sémantique et pragmatique. S'agissant de l'apprentissage d'une langue, la production linguistique suscitée pourra viser des objectifs dissociés, si bien que ce triple contrôle pourra parfois avoir à s'exercer de façon non cumulative. Cet avantage - du point de vue du traitement informatique de la réponse - ne peut cependant être exploité que de façon limitée, car nous verrons que le niveau que l'ordinateur peut le plus facilement traiter, n'est pas celui qui répond le mieux aux approches préconisées actuellement par la didactique des langues.

Nous examinerons donc tour à tour les possibilités d'analyse automatique de productions, au niveau du mot, de la phrase et de l'énoncé. Ce faisant, nous examinerons les problèmes spécifiques à la production en langue cible dans le contexte d'un exercice de langue. Nous aborderons enfin, toujours dans le même contexte, le cas de l'analyse de réponses "ouvertes" en langue maternelle.

Lors de ce parcours, nous serons également amenée à montrer, d'une part que le terme de "traitement informatique" renvoie à des outils et des systèmes de traitement d'ordre radicalement différent, d'autre part que le "traitement automatique" peut, grâce à l'adaptabilité des dictionnaires et grammaires électroniques, s'exercer à différents niveaux.

### **2.1.1. Analyse de mots**

L'analyse d'un texte écrit passe nécessairement par la "reconnaissance" de ses unités minimales : les mots. Par "mot" nous entendons "une association entre une forme et son sens. La partie formelle du mot simple est alors une suite de caractères comprises entre deux séparateurs consécutifs" (M. Gross, 1989) ; (pour les mots-composés, cf. G. Gross 1988; M. Mathieu-Colas 1988; M. Silberztein 1989)

---

<sup>8</sup> C'est le cas de certaines applications en vidéodisque notamment (cf. M. Garrigues, 1985, 1987).

Dans les logiciels éducatifs, la "reconnaissance" des mots produits dans la langue-cible, repose sur une programmation préalable où chacun des mots produits a été anticipé par le concepteur.

La reconnaissance non pas "programmée", mais automatique des mots d'un texte non déterminé, dans une langue donnée, suppose que l'on dispose d'outils linguistiques dont la formalisation préalable - et adaptée à une utilisation informatique - a été réalisée au niveau d'une langue entière. Dans ce cas-là, l'anticipation des réponses et le traitement de celles-ci par une programmation ad hoc ne serait plus nécessaire. Voyons ce que ce principe implique au niveau de la reconnaissance et de la correction d'un mot.

#### 2.1.1.1 Reconnaissance de mots simples et composés

Le principe des "correcteurs orthographiques" repose sur la comparaison automatique de chacun des mots d'un texte donné avec la liste de termes (ou dictionnaire) plus ou moins importante qu'ils contiennent. Lorsque la comparaison aboutit, il y a "reconnaissance". Si elle n'aboutit pas il y a présomption de faute. Il est donc facile d'en déduire que plus le dictionnaire sera exhaustif et précis, moins le logiciel laissera passer de fautes.

Si l'on s'avisait cependant de transférer de simples dictionnaires usuels sur support informatisé le résultat serait très décevant. (cf. M. Gross 1989). Comme nous l'avons vu dans notre première partie, aucun dictionnaire usuel n'est exhaustif, aucun ne prend en compte, en tant que tels, les mots composés du français ; aucun ne tient compte des variations graphiques des mots simples et composés (M. Mathieu-Colas 1987), et aucun ne contient, de façon systématisée, les codes indispensables à la reconnaissance automatique de l'ensemble des formes fléchies du français.

Pour le français, les dictionnaires électroniques DELAS et DELAC, remplissent les conditions d'exhaustivité et de formalisation nécessaires aux deux grandes classes d'applications informatiques : reconnaissance et génération automatique. Le DELAS contient aujourd'hui 80.000 entrées, le DELAC autour de 100.000.

Le logiciel de flexion automatique élaboré au LADL (B. Courtois 1989-1990) permet en outre d'engendrer automatiquement à partir du DELAS toutes les flexions des verbes, noms et adjectifs du français, soit environ 700.000 termes correspondant au DELAF, dictionnaire électronique des formes fléchies. D'autres logiciels (M. Silberztein 1990) permettent également

d'engendrer, à partir du DELAC, toutes les formes fléchies des noms composés (le DELACF) soit 150.000 termes environ. Il s'y ajoute d'autres lexiques comme celui des adverbes composés (en général invariables) qui contient plus de 6.500 entrées. Par ailleurs les adjectifs composés font l'objet d'une autre forme de description (commune aux verbes simples, et aux verbes composés) : le lexique-grammaire (M. Gross et C. Leclère 1990).

Cf. Annexe 1 pour des échantillons de ces dictionnaires.

L'ensemble des outils logiciels du système InTex (cf. Annexe 6), sont couplés avec la consultation de ces dictionnaires et permettent, entre autres de repérer automatiquement dans n'importe quel texte, tous les mots simples et noms composés qu'il contient, quelle qu'en soit leur flexion, de remonter à leur forme canonique dans le DELAS ou le DELAC et de retrouver les informations morphologiques et flexionnelles qui leur sont associées. Ainsi, la "reconnaissance" du mot simple *défauts* (DELAF) renverra automatiquement à la forme canonique *défaut* (DELAS) et restituera, sous forme d'"étiquette" les différentes informations la concernant. Ici : *défauts (défaut N1: mp)*. En cas d'ambiguïté (*le* pronom et *le* article, par exemple), toutes les possibilités sont envisagées dans l'étiquette.

Dans le texte ci-après (Figure 1), nous avons procédé, à titre d'exemple, à la reconnaissance automatique des noms composés. On peut, en effet, sélectionner la catégorie de mots que l'on souhaite rechercher (noms simples ou composés, adjectifs, adverbes, etc.)

N.B. Lors du processus de reconnaissance automatique, *bonne grosse* a été retenu comme nom composé, ce qu'il peut être dans un autre contexte ; une autre procédure est nécessaire pour reconnaître le substantif-tête *engueulade* et le modifieur figé *des familles* .

#### 2.1.1.2. Analyse des erreurs

S'agissant d'apprentissage d'une langue étrangère, on serait en droit d'attendre, que tout processus de reconnaissance automatique sur la production écrite d'un apprenant, détecte dans le même temps, les erreurs faites par ce dernier et lui envoie un feed-back même sommaire (comme un "bip" sonore) lui permettant de se corriger.

Les erreurs possibles d'un apprenant étranger sur un mot peuvent être de différentes natures ; elles peuvent relever de la forme (orthographe d'usage ou

1  
 2 L'exaspération dans le couple  
 3  
 4 ( ELLE 18 Janvier 88)  
 5  
 6 ...  
 7 Ah, l'exaspération... Dans cette bizarre association  
 8 appelée couple et constituée par deux <1êtres humains1> dont le  
 9 moteur commun est, théoriquement, l'amour qu'ils se portent  
 10 l'un à l'autre, c'est ce que chacun légitimement ressent  
 11 devant les défauts trop criants du conjoint (qui par un  
 12 bizarre <2effet-boomerang2> peut en retour être exaspéré par  
 13 l'exaspération de la <3partie adverse3>). Les motifs en sont  
 14 extrêmement variés et vont du plus infime (" il n'est jamais  
 15 foutu de reboucher le dentifrice") au plus intime ("tu m'as  
 16 encore donné des <4coups de pied4> dans ton sommeil...), du plus  
 17 trivial ("Il fait du bruit en mangeant") au plus profond  
 18 ("Elle veut toujours avoir le dernier mot"). L'effet reste  
 19 cependant le même: une forte irritation qu'on dissimule ou  
 20 qu'on exhale selon l'humeur et les circonstances. Ce qui  
 21 donne au mieux un grommellement rentré, au pire une <5scène de  
 22 ménage5> (avec tous les risques que cela comporte) et entre les  
 23 deux, toute une gamme d'échanges délicats et verbaux allant  
 24 des quelques hurlements de routine à la <6bonne grosse6>  
 25 engueulade des familles. On remarquera au passage l'emploi  
 26 presque systématique des adverbes "toujours", "jamais"  
 27 "encore" qui reviennent comme un refrain obsédant dans la  
 28 bouche de l'exaspéré.  
 29 D'après un sondage réalisé par la Sofres et publié en  
 30 1984 par "Madame Figaro", 23 % des français se disputent à  
 31 cause des manies du conjoint. C'est même la troisième cause  
 32 de discorde dans un couple derrière l'éducation des enfants  
 33 et les <7problèmes d'argent7>. " Elle est toujours dans la <8salle  
 34 de bains8> quand j'ai besoin de me laver", "Elle reste trois  
 35 heures au téléphone", "Il ne range jamais ses affaires".  
 36 Autant de petites phrases assassines, de minuscules motifs de  
 37 <9prises de bec9>, dus en général à des divergences fondamentales  
 38 sur la manière dont l'autre appréhende le quotidien."  
 39  
 40 ...  
 41  
 42 M. Fitoussi  
 43  
 44

\*\*\* 44 lignes, 16 phrases, 9 mots composés, 320 mots simples \*\*\*

Index :

bonne/grosse (AN:un:Nfs) 24  
 coups/de/pied (NDN:un:Nmp) 16  
 effet-boomerang (NN:un:Nms) 12  
 êtres/humains (NA:un:Nmp) 8  
 partie/adverse (NA:la:Nfs) 13  
 prises/de/bec (NDN:une:Nfp) 37  
 problèmes/d'argent (NDN:un:Nmp) 33  
 salle/de/bains (NDN:une:Nfs) 33  
 scène/de/ménage (NDN:une:Nfs) 21

orthographe grammaticale), ou de la valeur d'emploi (déterminant ou temps impropre, inadéquation lexicale par exemple).

### **Erreurs portant sur la forme des mots**

Considérons l'un des exercices de langue les plus triviaux : celui du texte à "trous" destiné à susciter la production morphologique, morphosyntaxique (accord de verbes, d'adjectifs, emploi des prépositions etc.) et/ ou lexicale (réemploi du vocabulaire appris, etc.) d'un seul mot.

Soit la consigne et l'exercice suivants :

Exercice n°1

#### **9. Un film à la télé. Accordez l'adjectif :**

*J'aime bien les (vieux) comédies (musical) (américain).*

*Hier soir, à la télé, j'ai vu « La Brune et la Rousse ». C'est l'histoire de deux (jeune) New-Yorkaises. Nancy et Doris. Nancy est (petit) et (brun). Elle a de (beau) yeux (bleu). Doris est (différent) : c'est une (grand) (roux) aux yeux (vert). La musique est (merveilleux) et les comédiennes sont très (bon). C'est une comédie (amusant). Mais je ne vais pas vous raconter le film !*

(*Sans Frontières 2*, M. Verdelhan-Bourgade, M. Verdelhan, P. Dominique p.8)

Pour la première réponse, un apprenant étranger serait susceptible de produire des erreurs telles que *vieile(s)*, *vielle(s)*, ou *vielle*. Pour que l'ordinateur les détecte, il serait possible d'élaborer, avec un langage de programmation (langage évolué tels que Basic, Pascal, etc. ou langage d'auteur tels que Arlequin, Duo, Tencore, etc.) un programme informatique consistant à entrer les données propres à cet exercice (réponses justes, anticipation des réponses fausses, feed-back, etc.) et à comparer chaque mot produit par l'apprenant au mot "juste" correspondant. Cette solution est fort coûteuse puisqu'elle demande à l'auteur une programmation préalable de l'exercice, et qu'en outre, elle ne permet pas de réutiliser celle-ci pour un autre exercice de même type comprenant d'autres mots.

Penchons-nous, à présent, sur le processus de correction automatique. Ici, l'exercice serait saisi (en tout ou en partie) par l'apprenant au traitement de texte, lui-même muni d'un correcteur orthographique, et corrigé automatiquement pendant la saisie. Aucune programmation ne serait donc nécessaire. Pour bien comprendre ce processus, il est nécessaire de distinguer deux types de fautes : les fautes d'usage et les fautes grammaticales.

### **les fautes d'usage**

Soit les fautes *vieile* ou *vielle* :

Il ne s'agit pas là de fautes d'accord (liées au non respect de règles grammaticales) mais de fautes d'usage. Ces erreurs liées toutefois à la méconnaissance de la langue sont à distinguer des fautes de frappe. Comme nous l'avons dit, le principe des "correcteurs orthographiques" repose, pour l'orthographe d'usage, sur une procédure complexe de comparaison automatique de chaque mot produit, avec la liste de termes (ou dictionnaire) qu'il contient.

Pour corriger les fautes, le programme doit faire des hypothèses. Lorsque la comparaison aboutit sans ambiguïté, il y a "reconnaissance". Si elle n'aboutit pas, il y a présomption de faute. La présomption de faute conduit à des hypothèses sur la nature de l'erreur.

Premier problème :

Dans le cas de la première faute, *vieile*, le correcteur pourra proposer en correction,

- *vielle* (si le dictionnaire a tendu à l'exhaustivité, il contiendra le mot *vielle* qui existe). Hypothèse de faute : une lettre tapée à la place d'une autre (*i* au lieu de *l*).

- *vielle*. Hypothèse de faute : une lettre omise, (*l*)

- *virile*. Hypothèse de faute : une lettre tapée à la place d'une autre (*e* au lieu de *r*)

Deuxième problème :

Dans le cas de la deuxième faute ; *vielle*, le correcteur ne proposera rien puisque le mot *vielle* existe dans le dictionnaire. De la même façon, la réponse *vairs* pour *verts* ne serait pas considérée comme une faute, puisque le mot *vair* existe.

Pour détecter cette faute, faudrait donc, que le processus de correction repose sur un système beaucoup plus intelligent, que celui de simples algorithmes de permutations de lettres, comme c'est le cas des vérificateurs et correcteurs actuels du commerce.

Certains outils de reconnaissance et de vérification automatiques développés au LADL permettent d'améliorer considérablement cette correction :

- le programme "creatdic" du système InTex permet, nous l'avons vu, d'extraire des dictionnaires DELAS et DELAC, les sous-dictionnaires d'un ou plusieurs textes donnés (saisis sur support informatique). L'exécution extrêmement rapide est produite par une simple commande.

Dans le cas qui nous occupe, une fois l'extraction faite à partir du texte de l'exercice, *vielle(s)* et *vair(s)* ne figureraient plus dans ce sous-dictionnaire. N'étant plus "reconnus" ils seraient donc signalés comme "fautes" lors de l'exécution de l'exercice par l'apprenant.

- le système de phonémisation automatique élaboré par E. Laporte (1988) permet de calculer la phonémisation (représentation phonétique théorique) du mot, de la mettre en relation avec le ou les mot(s) ayant la même phonémisation et se trouvant dans le DELAP-FH (dictionnaire électronique pour la phonémique-formes fléchies, homonymes cf. Annexe 1) L'utilisation d'un tel outil permet d'éviter, comme c'est le cas souvent, les corrections aberrantes par exemple, la proposition de correction *virile* dans le cas qui nous occupe.

Par ailleurs, les fautes d'usage peuvent aussi avoir pour effet la production de mots ne correspondant à aucune graphie admise. Toujours dans le même exercice, l'apprenant pourrait écrire *américainnes* pour *américaines* ou *rouxe* pour *rousse*. Dans ce cas -là il y a échec de la consultation du dictionnaire et détection immédiate de l'erreur, puisque ces mots n'existent pas.

En plus de la détection de l'erreur, des corrections pourraient toutefois être proposées grâce au même système de phonémisation automatique. Toujours selon le même processus, *américainnes* serait phonémisé en / *ameriken* / puis rapporté à l'entrée phonétique du DELAP-FH *amerik n* qui propose "*américaine, es*". Le système ayant prévu l'éventuelle phonémisation du *x* en *s* (à cause de mots comme *dix* et *six*) "*rousse, es*" serait également proposé à l'utilisateur.

Précisons que "vérification" et "correction" sont deux phases distinctes du processus de correction orthographique (E. Laporte, M. Silberztein 1989). Les "correcteurs orthographiques" usuels se limitent souvent à une vérification elle-même peu performante dans le cas d'un apprentissage de la langue. Les produits du commerce partent en effet plutôt du principe qu'une erreur de frappe a eu lieu et fondent leur système de correction sur des algorithmes de combinatoire des lettres.

### **Les fautes grammaticales**

Toujours dans le même exercice, soit la réponse *vieille* (singulier) pour *vieilles* (pluriel). Dans le cas de cette troisième faute possible, il s'agit d'orthographe grammaticale relevant d'autres procédures de correction. Pour que cette faute puisse être corrigée automatiquement, il faut que le correcteur comporte toutes les formes fléchies des mots ainsi qu'une grammaire formalisée de l'accord du nom et de l'adjectif en français. Signalons que toutes sortes de complications supplémentaires pourraient apparaître avec d'autres configurations linguistiques. Signalons par exemple celui du syntagme *de vieilles pieds-noirs*. Il faudra que le correcteur contienne le dictionnaire des mots composés fléchis, et qu'une première passe vérifie l'accord de *vieilles* avec *pieds-noirs*, qui, contrairement au mot *pied* isolé peut être féminin.

Les fautes grammaticales sont très inégalement repérées par les "correcteurs orthographiques" en usage aujourd'hui ; si l'on s'en tient au plan morphologique strict, il faut déjà, nous l'avons vu, que le dictionnaire contienne toutes les formes fléchies du français. L'accès au DELAF permet par exemple de repérer automatiquement des erreurs telles que "(il) mangait " puisque cette forme n' existe pas dans le système verbal et que l'absence de "e" est en réalité une erreur orthographique d'usage.

Le cas de "*ils mangeait*" ou de "*je les voit* " est plus complexe puisqu'il s'agit, pour détecter la faute localisée dans la forme verbale, de mettre cette dernière en rapport avec un autre "mot", le pronom sujet, dont elle est parfois séparée par d'autres particules préverbales telles que le pronom *les* . Il s'agit là déjà d'un problème morphosyntaxique. Cette erreur peut toutefois être repérée par le vérificateur orthographique du LADL : ainsi le recours à l'automate des particules préverbales (D. Perrin 1989) permet-il de vérifier systématiquement l'accord du verbe avec son pronom sujet quel que soit leur éloignement et par

conséquent d'en déceler la non conformité. Nous reviendrons sur cet automate en 2.1.2.2.

Pour d'autres phénomènes morphosyntaxiques locaux (grammaire des accords entre certains déterminants et noms, ou accord des noms et des adjectifs par exemple), les mêmes outils de vérification orthographique du LADL permettent de mettre en oeuvre une grammaire locale, et de repérer les fautes en rapport avec cette grammaire dans toute production écrite en français.

Un phénomène syntaxique local ne veut en effet pas dire qu'il est localisé dans un exercice, mais qu'il est circonscrit dans l'ensemble du système de la langue. D'où la généralisation possible de traitement automatisé du phénomène. Grâce aux mêmes outils, il est cependant possible de créer, selon les besoins de l'application, des grammaires visant non pas la langue entière mais un exercice ou un ensemble d'exercices, et par conséquent, un système de repérage de fautes généralisable à un ensemble d'exercices construits sur le même modèle.

Si l'on reprend l'exercice présenté ci-dessus : il porte, on le voit, sur l'accord d'un certain nombre d'adjectifs, placés avant ou après le nom, ou après le verbe *être*. Les cas d'accord noms-adjectifs sont donc variés. On peut, après avoir répertorié l'ensemble des cas de figures de l'exercice, les décrire, à l'aide des catégories ou des mots figurant dans le DELAS ; cette opération est destinée à établir la liste des configurations ou "expression rationnelle" suivante :

```
<ADJ-ms>/<NOM-ms>+
<ADJ-mp>/<NOM-mp>+
<ADJ-fs>/<NOM-fs>+
<ADJ-fp>/<NOM-fp>+
<NOM-ms>/<ADJ-ms>+
<NOM-mp>/<ADJ-mp>+
<NOM-fs>/<ADJ-fs>+
<NOM-fp>/<ADJ-fp>+
<ADJ-ms>/<ADJ-ms>+
<ADJ-mp>/<ADJ-mp>+
<ADJ-fs>/<ADJ-fs>+
<ADJ-fp>/<ADJ-fp>+
<NOM-ms>/<être-3s>/<ADJ-ms>+
<NOM-fs>/<être-3s>/<ADJ-fs>+
<NOM-mp>/<être-3p>/<ADJ-mp>+
<NOM-fp>/<être-3p>/<ADJ-fp>
```

On voit qu'il a suffi de recenser et de décrire les configurations morphosyntaxiques visées dans l'exercice. Ainsi tous les cas d'associations noms-adjectifs ne répondant pas à l'un des cas de figures intégrés dans l'"expression rationnelle" ci-dessus ne seraient pas reconnus et seraient donc rejetés comme faux. Ce serait le cas de *vieille comédies* puisque l'association < NOM-fs> / < ADJ - fp > ne figure pas dans la liste des configurations linguistiques décrites.

On remarquera que la description des cas de figures peut se faire en termes de catégories (ADJ, NOM, etc.) accompagnées ou non d'indications de genre, de nombre ou de personne (m, f, s, p, 3 p, etc.) ou de mots précis (*être*) eux-mêmes accompagnés ou non d'une description flexionnelle. Quant aux noms propres, ils peuvent facilement être introduits dans le dictionnaire avec les caractéristiques que l'on souhaite leur associer (catégorie, genre, nombre, etc.). Ainsi aurions-nous pu introduire le nom propre Nancy accompagné des marques féminin et singulier, "Nancy N: fs", dans le sous-dictionnaire de l'exercice. L'accord de *petite*, ou de tout autre adjectif, aurait ainsi pu être vérifié automatiquement.

Le fait que les catégories et/ou les mots de l'"expression rationnelle" renvoient aux catégories et mots de dictionnaires exhaustifs (DELAS, DELAC) eux-mêmes générateurs de formes fléchies (DELAF, DELACF) permettrait de faire faire une série illimitée d'exercices reposant sur les associations décrites dans l'"expression rationnelle" ; il suffirait, dans le cas qui nous occupe de faire saisir le texte des exercices par les apprenants ; Ils seraient automatiquement corrigés.

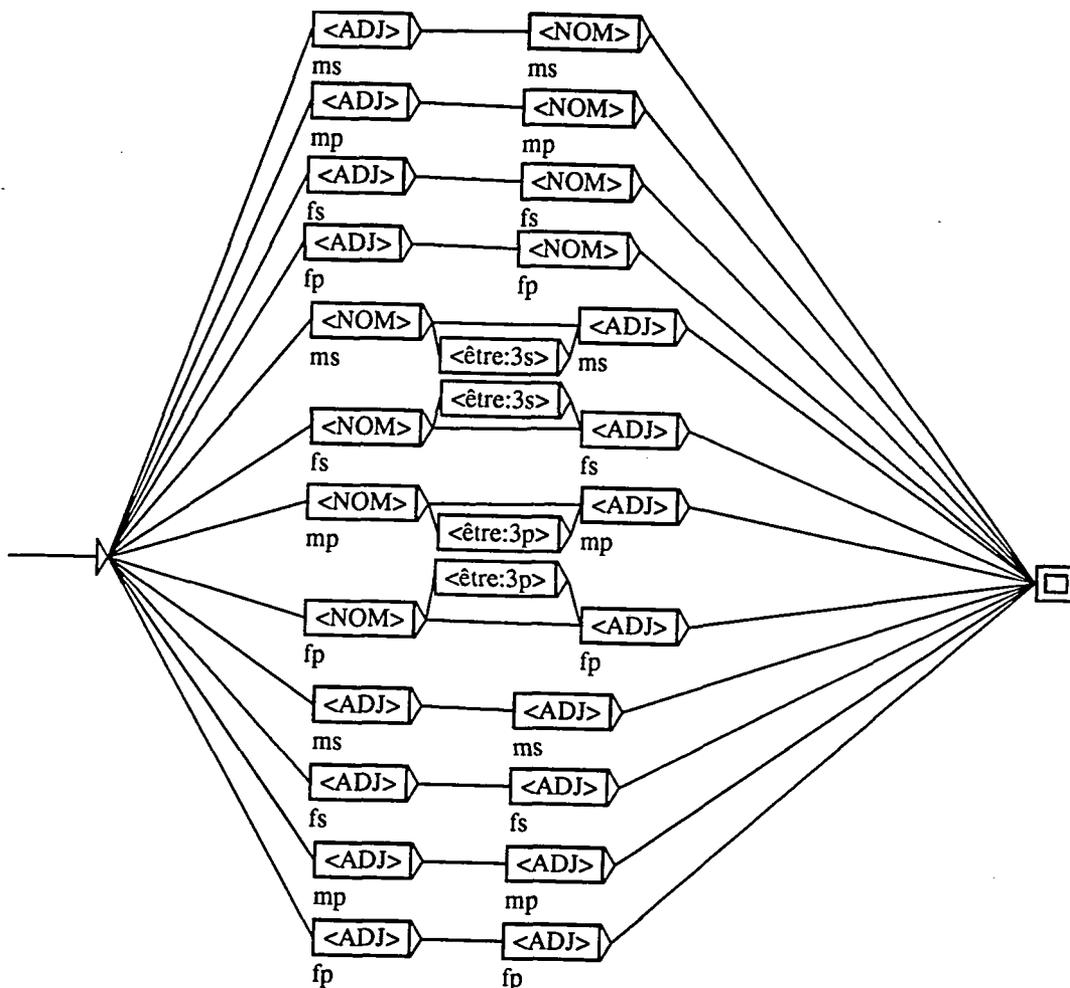
Lorsque l'élaboration de l'"expression rationnelle" est représentative de l'ensemble du phénomène envisagé dans la langue, l'éventail des exercices possibles sur ce point devient alors véritablement infini.

Aucun langage de programmation ne permettrait de faire l'économie de l'entrée de données (morphologiques, lexicales, etc.). Le recours à ce type d'outil exige en effet qu'on les introduise en machine, avec le système de codification souvent très complexe propre à chacun d'eux, sans parler de la programmation nécessaire à l'exécution de chaque exercice.

Ajoutons qu'il est possible de vérifier empiriquement, avant la mise en service des exercices la validité d'une expression rationnelle en procédant sur

un corpus de textes (sans limite de longueur) à la reconnaissance automatique des exemples couverts par cette expression.

Ajoutons encore que l'éditeur d'automates du système InTex permet aujourd'hui de créer ces automates avec une grande facilité, dès lors bien sûr (comme nous le verrons en 2.2.) que la formalisation du point grammatical visé a été faite. Ainsi l'"expression rationnelle" (ou automate) de notre exercice peut être éditée par la création d'un graphe au moyen d'une simple manipulation d'objets sur écran. Le graphe est alors immédiatement converti en automate de reconnaissance dans les textes, par une simple commande du système.



Nous nous focaliserons en 2.2 sur le processus de création de grammaires locales plus élaborées.

### Erreurs portant sur la valeur d'emploi des mots

Penchons-nous à présent sur le cas d'un exercice à trous destiné à solliciter l'emploi du subjonctif :

#### Exercice n° 2

Complétez les phrases suivantes avec les formes correctes des verbes donnés entre parenthèses.

- 1 Il veut que vous ..... une heure plus tôt, si c'est possible. (*venir*)
- 2 J'exige que ce travail ..... fait pour demain. (*être*)
- 3 Comment voulez-vous que j'..... à Paris sans voiture? (*aller*)
- 4 Je regrette que nous ne ..... pas vous donner satisfaction. (*pouvoir*)
- 5 Il n'est pas possible qu'il ..... à Paris sans que nous le ..... (*venir - recevoir*)
- 6 Bien que je ne ..... que quelques mots d'italien, j'ai compris ce qu'il voulait dire. (*savoir*)
- 7 Il est indispensable que vous ..... et que vous ..... Sinon, je crains que vous ne ..... une dépression nerveuse. (*dormir - se reposer - faire*)
- 8 Qu'il ..... son manteau s'il a froid! (*mettre*)
- 9 Je ne pense pas que ce livre vous ..... (*plaire*)
- 10 Il faut que je vous ..... ce que je pense, bien que je ne ..... pas à ce que nous ..... (*dire - tenir - se fâcher*)

(*C'est facile à dire*, G. Capelle, F. Greillet p.78)

L'apprenant étranger est souvent confronté au problème de l'emploi du subjonctif qu'il a souvent du mal à distinguer de l'emploi de l'indicatif. Les erreurs commises sur ce type d'exercice seraient essentiellement des erreurs d'emploi de mode. Dans la première phrase, l'erreur la plus probable serait donc *venez* pour *veniez*.

Si cet exercice était proposé sur ordinateur, il faudrait que soit reconnue la bonne réponse *veniez*, et qu'elle soit distinguée de la mauvaise réponse *venez*.

La seule utilisation d'un langage de programmation ne permettrait pas d'automatiser le traitement informatique de cet exercice. Comme dans les

exercices portant sur le système morphologique de la langue, il serait en effet indispensable d'introduire dans le programme les données formelles et lexicales à ce traitement. Or, les règles des grammaires (traditionnelles et modernes) concernant les valeurs d'emploi sont très loin de permettre une quelconque formalisation. L'utilisation de langages de programmation pour l'élaboration d'exercices nécessite, on l'a vu, l'entrée de données ponctuelles. Ce travail est déjà très fastidieux lorsque l'on utilise des données comportant des règles formalisables. Or, le concepteur d'exercices se trouverait ici confronté de surcroît à l'absence totale de règles qu'il devrait formaliser lui-même.

D'où la quasi inexistence d'exercices de ce type sur ordinateur. Des tentatives ont bien été faites mais il s'agit dans ces cas-là d'un bricolage informatique dans lequel chaque réponse d'apprenant, préalablement anticipée, est traitée au coup par coup avec, dans le meilleur des cas, un feed-back métalinguistique aussi abondant que rebutant. Comme il n'existe pas de règle, "l'ordinateur" disserte laborieusement sur chaque différence ou erreur d'emploi.

Dans notre exemple, le traitement automatique de la faute envisagée, nécessite des outils de très grande envergure. Le processus de reconnaissance et de vérification automatique dont nous avons parlé, accepterait en effet le mot *soit* aussi bien que le mot *est*, puisque ces formes existent dans la morphologie du système verbal français. Donc, seule l'intervention d'un analyseur syntaxique très puissant reposant sur un traitement formel de l'ensemble des règles syntaxiques du système verbal français permettrait de différencier l'emploi de l'indicatif de l'emploi du subjonctif.

Ce travail de formalisation, nous l'avons dit, a été réalisé par M. Gross et son équipe : le fonctionnement des 12.000 verbes du français analysé selon une grille de 600 propriétés syntaxiques a donné lieu à l'examen de 7.200.000 possibilités), (M. Gross 1975 ; J.P. Boons, A. Guillet, C. Leclère, 1976 a,b ; A. Guillet, C. Leclère 1992).

Un travail consistant à présent, à générer automatiquement la totalité des structures syntaxiquement correctes de chacun des verbes de cette grille, est en cours. Réalisé par E. Roche 1992, il permettra de comparer l'ensemble des structures générées de chaque verbe, à l'environnement syntaxique du verbe produit. Ainsi, l'environnement syntaxique de la réponse 1 de l'exercice n° 2, pourra-t-il être mis en rapport avec la propriété *que Psubj* du verbe *vouloir* (cf.

Figures 2 et 3, ci-après). Seuls des analyseurs syntaxiques très complexes (cf. aussi, M. Salkoff, 1973, 1979) permettront donc à long terme de corriger ce type d'erreur.

On voit que le traitement automatique d'un simple mot envisagé dans le contexte de l'apprentissage d'une langue pose encore des problèmes que seuls des outils linguistiques formalisés à un haut niveau pourront progressivement résoudre.

Nous allons examiner en détail les contraintes auxquelles se heurtent la mécanisation de la détection des erreurs dans une phrase. Nous prendrons comme exemple des erreurs couramment commises dans l'apprentissage d'une langue (en l'occurrence, le français langue étrangère). Mais l'analyse présentée est plus générale et la plupart des contraintes signalées affectent aussi bien la correction orthographique des fautes grammaticales que la post-édition humaine de traductions automatiques.

### **2.1.2. Analyse de phrases**

L'analyse automatique de la langue ne peut s'arrêter à l'analyse de chacun des mots d'une phrase envisagé séparément. Comme nous venons de commencer à le voir dans ce qui précède, ceux-ci ont en effet entre eux des rapports - intra et interphrastiques - très complexes qu'il importe également de prendre en compte.

#### **2.1.2.1 La phrase comme unité discursive minimum**

Depuis Saussure, le statut de la phrase a donné lieu à de nombreux débats théoriques entre linguistes ; sa définition toujours fluctuante relève d'une complexité conceptuelle liée à l'histoire même de la linguistique et de la logique.

Pour les besoins de notre propos, nous commencerons par envisager la phrase comme "unité discursive minimale". Le repérage de ces unités ou phrases dans un texte ne peut pas se faire très simplement.(M. Silberztein 1989). Adoptons provisoirement le principe suivant : "Sont considérés comme





marques d'arrêt : le point, le point d'interrogation, le point d'exclamation, les points de suspension parfois" (D. Maingueneau 1976)

Prenons à présent l'exemple de l'accord du participe passé. Tout le monde se souvient de l'apprentissage des règles qui régissent son accord (auxiliaire *avoir* ou *être*, participe passé situé avant ou après le complément d'objet direct, etc.). On pourrait donc penser que la formalisation de ces règles, et par conséquent la vérification de leur application, ne devrait pas trop poser de problèmes, même s'il s'agit d'un travail complexe (V. Le Kien, 1987). Pourtant si la phrase contenant le participe passé se trouve dans un texte même minimal, la vérification automatique de son accord devient souvent impossible.

Imaginons ce simple fragment de texte : *Jeanne est arrivée avec les garçons, la chienne et les saucisses. Elle les a piqué(e)s puis a vite mis de l'eau à bouillir.* Pour définir et vérifier l'accord de *piquer*, il est nécessaire de savoir si l'antécédent de *les* correspond à un nom féminin pluriel, ou à un nom masculin pluriel, ou à deux (ou plusieurs) noms coordonnés masculins (singuliers ou pluriels), ou à deux (ou plusieurs) noms féminins (singuliers ou pluriels), ou à deux (ou plusieurs) noms dont l'un au moins est masculin (singulier ou pluriel). Les pronoms personnels ont souvent, comme ici, une valeur "anaphorique" ; c'est-à-dire qu'ils reprennent au cours du discours, sous une autre forme, un segment de phrase déjà énoncé. Or aucune marque ne permet de repérer automatiquement ce segment, qui peut se situer plus ou moins en amont dans le texte et nécessiter, dans les cas d'ambiguïté, la mise en oeuvre d'opérations hypothético-déductives à partir d'indices contextuels et/ou intertextuels et/ou extralinguistiques.

Si l'on se reporte à notre exemple, simple en apparence, comment faire en sorte que le système de vérification automatique se réfère à *saucisses* plutôt qu'à (*garçons* et *chienne* et *saucisses*) ? De fortes présomptions pèsent en faveur de *saucisses*, mais les présomptions ont très peu de chances de se laisser enfermer un jour dans une description formalisée. C'est certainement à ce niveau que l'analyse automatique profonde des textes a le moins d'espoir de déboucher. La lisibilité même d'un texte se fait au travers d'un décodage de son système référentiel. Voici, toujours à propos de la fonction anaphorique du pronom personnel, un exemple tiré de *Madame Bovary* :

*Plus d'un montrait avec orgueil son humble médaille, et sans doute, revenu chez lui près de sa bonne ménagère, il l'aura suspendue en pleurant aux murs discrets de sa chaumine.*

Nous rapportons naturellement le pronom *l'* au mot *médaille* et non au mot *ménagère* qui pourtant le précède immédiatement. Sur quels critères de formalisation un système pourrait-il le faire automatiquement ? Il est impossible d'y répondre aujourd'hui.

#### 2.1.2.2 La phrase comme modèle de compétence

La phrase peut être aussi envisagée "comme une suite d'éléments syntaxiques minimaux enchaînés (ou *concaténés*)" (N. Ruwet 1967). On considère alors que "La phrase relève de la compétence et l'énoncé de la performance." (ibid.,)

Plus abstraite que dans l'acception précédente, la phrase représente ici un modèle de fonctionnement de la langue déconnecté de toute implication contextuelle.

Certains exercices de langue ont pour objet la production de "phrases", ayant pour but de fixer cette compétence. Ainsi les "exercices de grammaire" en offrent-ils plusieurs variétés, dont la "transformation" et la "reconstitution". Bien qu'occupant aujourd'hui une place modeste dans les activités langagières de la classe de langue, ce type d'exercice reste pratiqué. (L'épistémologie du savoir grammatical et son articulation avec la didactique des langues posent des problèmes que nous signalons, mais que nous ne pouvons aborder dans ce cadre (cf. H. Besse, R. Porquier 1984).

Voyons donc comment pourrait s'effectuer le traitement informatique d'un exercice transformationnel ne suscitant qu'une seule réponse acceptable

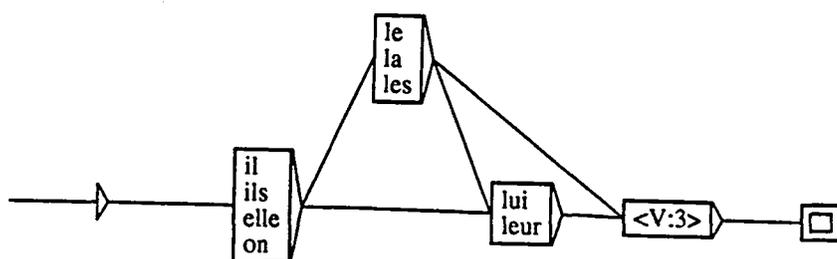
#### **Une seule bonne réponse possible**

Soit le point syntaxique suivant : l'ordre des pronoms préverbaux en français. Cet ordre contraint est fixé par des règles difficiles à intérioriser pour un apprenant étranger. Ce processus demande donc souvent quelques exercices de fixation dont l'exemple suivant pourra donner une idée :

Soit la phrase: La secrétaire a donné les clefs du bureau à Jeanne, à partir de laquelle l'apprenant serait invité à opérer une transformation destinée à lui

faire utiliser les pronoms compléments de la 3ème personne. Réponse attendue: *Elle les lui a données* .

Dans un exercice de ce type, il serait possible de vérifier automatiquement l'ordre des pronoms en recourant à l'utilisation de l'automate des particules préverbaux évoqué plus haut (en 2.1.1.2) puisqu'il représente un modèle générique pour la langue entière. Nous avons vu que le système InTex permet d'élaborer des graphes générant instantanément des automates, susceptibles de reconnaître n'importe quelle structure syntaxique. Le graphe permettant de vérifier l'ordre des pronoms à la troisième personne, dans une phrase élémentaire, serait donc :



Il ne serait en revanche pas possible de détecter de façon automatique une erreur de genre ou de nombre sur l'un des pronoms. L'automate contrôlant la production de l'apprenant dans le cas envisagé, signalerait l'inacceptabilité de *Elle lui les a données* mais non celle de *Elle le lui a donné* puisque cette dernière phrase est acceptable du point de vue de la langue, mais erronée en regard de la phrase de référence de l'exercice. Nous nous trouvons en fait ici encore, devant un problème d'anaphore.

Si l'on voulait, dans le cadre de cet exercice transformationnel, absolument corriger toutes les fautes d'accord, en plus de l'ordre des pronoms, il serait nécessaire d'avoir recours à un langage d'auteur.

Les langages d'auteurs permettent, de façon inégale selon leur qualité, des traitements à façon. Rappelons qu'un langage d'auteur est un langage informatique dont les fonctionnalités sont destinées à faciliter les applications pédagogiques. (P. Barril et al. 1985 ; Epi, 1983).

Le détail du processus de traitement informatique, à l'aide d'un de ces langages, pourra donner une idée du degré de précision nécessaire à la mise en oeuvre d'un des cas d'analyse de réponse les plus simples.

Une méthodologie sérieuse implique une phase préalable de prétests sur papiers (sur environ 150 apprenants correspondant au public-cible). Ces prétests serviront à établir une typologie des erreurs destinée à définir la grille d'analyse et de traitement.

### **Grille d'analyse**

Production		Rétroaction sur :	
elle	lui les		erreur sur l'ordre des pronoms
elle	les lui	a donné donnée	erreur d'accord en nombre
		a donnés	erreur d'accord en genre
elle	les lui	a données	réponse correcte
autre	—	—	niveau d'affinement d'analyse à évaluer selon les pré-tests sur le public cible
—	autre	—	
—	—	autre	

La deuxième phase est le précodage (ou codage "naïf") de la maquette. Il est nécessaire à ce moment-là de connaître les fonctionnalités offertes par le langage informatique (langage évolué ou langage-d'auteur) qui sera utilisé. Il est donc indispensable que le concepteur de l'exercice les maîtrise, même s'il ne participe pas à la programmation proprement dite. Compte tenu de ces fonctionnalités et de la typologie des réponses du public-cible prétesté, il lui appartient également de décider jusqu'à quel degré d'affinement d'analyse il faudra aller. Nous avons utilisé ici les fonctionnalités offertes par le langage d'auteur Arlequin - Diane (N. Rodriguez, D. Valentin 1985). Les éléments de syntaxe retenus pour notre exercice, et liés à quelques unes de ses nombreuses fonctionnalités, se définissent comme suit :

le symbole  $\hat{\uparrow}$

permet de reconnaître une phrase commençant ou se terminant obligatoirement par un mot ou groupe de mots. Cette possibilité permet de baliser une prévision de réponse.

le symbole  $\leftarrow$

définit un caractère de lien entre deux mots. Ces derniers constitueront alors un groupe de mots.

L'opérateur SV (suivi de)

est un "et" coordonné permettant de reconnaître deux éléments de réponse (mots, groupes de mots ou modèles) dans un ordre déterminé mais sans que la distance qui les sépare soit nécessairement définie.

L'opérateur SVS (strictement suivi de)

permet de reconnaître deux éléments de réponse dont l'un est immédiatement suivi de l'autre (dont il n'est séparé que par des caractères séparateurs).

Voici le codage naïf de l'exemple qui nous occupe à l'aide des éléments de syntaxe que nous venons de définir.

PRÉVISIONS	CAS COMMENTÉS
1-« l Elle – les – lui – a – données ↑ » 2-« Elle – lui – les »	Bonne réponse Erreur sur l'ordre des pronoms
3-« Elle – les – lui – a » SVS « donné, donnée »	Erreur d'accord en nombre de l'auxiliaire
4-« Elle – les – lui – a – donnés »	Erreur d'accord en genre de l'auxiliaire
5-Autres cas	Non reconnus

Ce précodage dit naïf ne présente qu'une étape préalable - mais indispensable - à la programmation proprement dite. Il constitue en quelque sorte l'interface entre la conception et la programmation.

Il peut varier selon les fautes que l'on juge utile de prendre en compte, et les facteurs de risque que l'on a évalués, après examen des prétests.

Ce même exercice pourrait donc être codé de plusieurs autres façons. En voici une seconde :

PRÉVISIONS	CAS COMMENTÉS
1-« l Elle – les – lui – a – données »	Bonne réponse
2-« lui – les »	Erreur sur l'ordre des pronoms
3-« a » sv « donné, donnée »	Erreur d'accord en nombre de l'auxiliaire
4-« a – donnés »	Erreur d'accord en genre de l'auxiliaire
5-Autres cas	Non reconnus

Au moment du précodage de la maquette, une série de choix se pose au concepteur à plusieurs niveaux :

- Dans le cas d'un cumul de fautes, faut-il présenter un diagnostic simultané ou successif de chacune d'elles ?

La seconde option est en principe à éviter pour des raisons pédagogiques : il est préférable que l'apprenant corrige ses fautes une par une, et ergonomiques : l'écran serait trop surchargé.

- Quelle hiérarchisation des fautes établir ?

Si l'on a choisi le diagnostic successif, le système exécute la première prévision rencontrée comme vraie : on pourra donc choisir, soit l'ordre syntagmatique (dans l'exemple envisagé, le système sortirait dès qu'il aurait rencontré autre chose que *elle* par exemple), soit un ordre prédéterminé d'importance croissante des fautes (par exemple : ordre des pronoms, puis accord du participe, etc.)

Il faudrait, en tout état de cause se préoccuper simultanément, lors du codage, de la logique algorithmique du traitement de ces prévisions de réponse. Il serait incohérent par exemple de faire corriger l'accord du participe passé avant les erreurs commises sur l'ordre ou l'accord des pronoms, pour une réponse du type *Je la ai* (ou *je l'ai*) *donné*.

- Jusqu'à quel point affiner l'analyse des erreurs ?

Dans le cas qui nous occupe, aurait-on dû par exemple envisager et traiter de façon spécifique une erreur sur le pronom-sujet ? Une erreur sur le (ou l'absence de) pronom objet direct ? Une erreur sur l'auxiliaire *avoir* ? Une faute d'orthographe sur *donn(er)* ? etc.

L'opportunité des traitements affinés est liée à l'analyse des prétests sur papier effectués sur le public-cible et aux fonctionnalités du langage utilisé.

### **Plusieurs bonnes réponses possibles**

Soit l'exercice dans lequel l'apprenant serait invité à construire une phrase, sur un modèle imposé, à l'aide des éléments donnés suivants :

*hier / aller / restaurant*

La phrase attendue serait : *hier je suis allé(e) au restaurant*

Le traitement automatique d'une telle phrase supposerait que l'on ait élaboré un certain nombre de grammaires locales :

- la grammaire locale de l'accord du déterminant et du nom
- la grammaire locale du prédicat sémantique correspondant au déplacement d'un lieu à un autre par un sujet humain (cf. notre partie 2.2).
- la grammaire locale de l'accord du participe passé, avec toutefois cette limite due à la fonction anaphorique du pronom (ici le problème du pronom sujet : renvoie -t-il à un *Nhum* féminin ou à *Nhum* masculin ?)

Il supposerait en outre, qu'une grammaire locale mettant en relation les adverbes et les temps des verbes ait été réalisée. Il supposerait aussi bien sûr la présence d'un correcteur orthographique susceptible de vérifier l'orthographe lexicale des mots.

Malgré l'ampleur de la tâche, on peut s'attendre à ce que l'on arrive très progressivement à réaliser l'ensemble de ces traitements.

Grâce aux mêmes outils de traitement automatique, et comme nous l'avons vu pour l'exercice n°1, il y aurait toutefois la possibilité de réaliser ces grammaires, non pas au niveau de la langue, mais au niveau d'un ensemble d'exercices proposés. Il suffirait d'étudier l'ensemble du corpus des exercices et de réaliser les automates correspondant aux différentes grammaires des points visés. Celles-ci seraient indiscutablement plus simples, compte tenu du niveau de langue visé, que des grammaires visant l'ensemble des productions d'une langue.

De plus, le système des grammaires locales étant modulaire, chacune d'entre elles pourrait être réutilisée, réadaptée et recombinaée avec d'autres grammaires selon les types d'exercices proposés.

Le recours à un langage d'auteur serait beaucoup plus artisanal et nécessiterait, comme on peut le voir ci-dessous, un traitement propre à chaque phrase de chaque exercice. Pour la phrase envisagée, les différentes étapes du traitement seraient donc l'élaboration d'une grille d'analyse, le précodage de la maquette, et la programmation proprement dite.

La grille d'analyse pourrait se présenter ainsi :

Production					Rétroaction sur
hier	je suis	allé(e)	au	restaurant	bonne réponse
	j'a —				erreur d'auxiliaire
		all — (≠ allé(e))			erreur sur le participe passé
			φ φ + le		oubli de la prépo- sition
			à + φ		préposition sans article
			à + le		non contraction de la préposition et de l'article
			chez		erreur de préposi- tion
h - r (≠ hier)			r — r— (≠ restaurant)		erreur d'orthogra- phe
autres					autres cas rejetés

Avec le même langage d'auteur, cette dernière pourrait donner lieu au précodage suivant :

PRÉCODAGE MAQUETTE	
1- («   hier — je — suis » SVS « allé, allée » SVS « au — restaurant   » ) ou ( «   je — suis » SVS « allé, allée » SVS « hier » SVS « au restaurant   » ) ou ( «   je — suis » SVS « allé, allée » SVS « au — restaurant — hier   » )	Bonne réponse
2- « a\$*\$ — all\$*\$ » [5]	Erreur d'auxiliaire
3- « suis — all\$*\$ » ET NON ( « suis » SVS « allée, allé » )	Erreur sur le participe passé
4- « all\$2\$ — rest\$*\$ » OU « all\$2\$ — le — rest\$*\$ »	Absence de préposi- tion
5- « all\$2\$ — a — rest\$*\$ »	Absence de l'article
6- « all\$2\$ — a — l\$2\$ »	Non contraction
7- « chez — \$*\$ — rest\$*\$ »	Erreur de préposition
8- « h\$*\$r\$*\$ » ET NON « hier » OU « rest\$*\$ » ET NON « restaurant »	Erreur d'orthographe
9- autres	Réponses rejetées

Ce précodage nécessitait à certains moments de ne prendre en compte que des "squelettes de mots". Ce langage permet, grâce à une utilisation particulièrement optimisée des "jokers" (\$), de définir des tolérances de caractères très variées. Ici :

$\$\$$  = Ce signe tolère l'absence ou la présence d'un caractère au plus (quel qu'il soit). Le codage "maison $\$\$$ " permettrait de reconnaître *maison* ou *maisons* .

$\$n\$$  = Ce signe tolère l'absence ou la présence de n caractères au plus (quels qu'ils soient). Par exemple: "maison $\$3\$$ " permettrait de reconnaître *maisonnée* mais aussi *maison* ou *maisons* mais non *maisonnées*

$\$*\$$  = Ce signe tolère l'absence ou la présence d'un nombre indéterminé de caractères (quels qu'ils soient) dans un mot. Le codage "chant $\$*\$$ " permettrait de reconnaître toutes les formes des verbes *chanter et chantonner* ainsi que les noms *chant , chants, chanterelle, chantier* etc.

Le précodage pourrait là aussi donner lieu à différentes variantes. La bonne réponse pourrait par exemple être codée de façon plus "lâche" mais aussi plus risquée :

" Je suis" SVS "allé,allée" SV "au restaurant" ET "hier "

Ce modèle aurait l'avantage de regrouper toutes les variantes combinatoires, mais l'inconvénient de pouvoir aussi laisser passer, autour des éléments de bonne réponse, des éléments parasites risquant de compromettre la pertinence du commentaire rétroactif. Ce serait le cas par exemple de : *Je suis allé avec le voiture au restaurant hier*. Le nombre de phrases correctes étant cependant dans ce cas-ci fini et limité on peut raisonnablement envisager de "verrouiller" la bonne réponse, comme le montre le codage de la maquette.

Bien qu'il demande de redéfinir pour chaque phrase un système d'analyse particulier modélisant l'ensemble des bonnes réponses et des erreurs prévues, ce genre d'exercice reste réalisable sur ordinateur (avec un langage d'auteur performant). Il vise, on le voit, l'apprentissage du système morphosyntaxique de la langue, apprentissage dont il est difficile de se passer complètement. Les

exercices de grammaire ("à trous", structurels voire structuraux) sont du reste présents, même dans les méthodes les plus "progressistes" et bien que les auteurs, par réaction à l'enseignement traditionnel, répugnent souvent à les présenter pour ce qu'ils sont.

Cette utilisation de l'ordinateur ne soulève cependant pas l'enthousiasme des spécialistes du domaine. La didactique des langues largement représentée aujourd'hui par les approches ou méthodes communicatives réserve en effet, une place relativement modeste aux exercices phrastiques. Désormais, si "La connaissance des formes et des structures reste indispensable en langue étrangère", sa mise en oeuvre se fait "toujours au service d'un projet de sens" (F. Debyser 1985).

### 2.1.3 Analyse de segments d'énoncés dans la langue-cible

Les implications méthodologiques des approches communicatives préconisées par la didactique des langues ont, en effet, corrélativement conduit les concepteurs de matériel pédagogique, à recourir le moins possible aux phrases isolées, et à proposer de travailler sur des échanges ou des micro-conversations. Après s'être exclusivement centré sur l'acquisition d'une **compétence**, l'enseignement des langues vise aujourd'hui l'acquisition d'une **compétence de communication**.

Il est nécessaire de préciser ici que le concept de **compétence de communication**, employé en didactique des langues est à distinguer du concept de **compétence** tel qu'il est employé par les linguistes dans le couple compétence/performance. La notion de compétence de communication pose aux spécialistes du domaine un problème de définition toujours en suspens. Disons, pour simplifier, que la **compétence de communication** communément entendue représente la capacité à gérer un échange langagier alors que la **compétence** des linguistes représente "le savoir linguistique implicite de sujets parlants" (N. Ruwet, 1967). Dans la mesure où la **performance** représente au contraire "l'actualisation ou la manifestation de ce système dans une multitude d'actes concrets toujours différents" (ibid., 1967), elle serait plus proche de la compétence de communication. Cette dernière impliquant toutefois l'intériorisation d'un "savoir linguistique", nous dirons

pour notre part que la compétence de communication englobe, outre des compétences extra-linguistiques, la dichotomie compétence/performance.

Cependant, en privilégiant les modèles de performance, même dans l'acquisition d'une compétence strictement linguistique, la didactique des langues favorise un porte à faux fatal au traitement informatique. Chaque fois que cela est possible, l'unité d'apprentissage est constituée par un énoncé plutôt que par une phrase. Or, à la différence de la phrase, le champ des énoncés possible relève par définition de l'infini. Cette orientation didactique explique en grande partie, que l'apparition de l'ordinateur, ne se soit pas accompagnée d'un transfert massif des exercices des manuels en usage, sur support informatique.

Un simple exercice à trou, conçu dans une perspective communicative pose en effet un problème d'analyse de réponse d'autant plus insoluble que l'ambiguïté que nous venons d'évoquer est grande.

Soit l'exercice-papier n° 3 (Figure.4, ci-dessous) :

(*Premiers Exercices de Grammaire* G.D. de Salins, S. Dupré Latour.)

Il s'agit évidemment pour les auteurs de faire pratiquer l'accord de l'adjectif possessif. Cette présentation (en modèles de performances) est certes plus plaisante qu'une simple série de phrases hors contextes. Le problème est que dans quasiment chaque cas un grand nombre de segments d'énoncés ne répondant pas à l'objectif de l'exercice serait fonctionnellement acceptables :

Dans le premier échange (Au restaurant), *ce steak et ces frites* conviendrait aussi bien que *mon steak et mes frites* ; *le steak et les frites* serait également acceptables.

Dans le deuxième échange (Au garage), le champ des réponses possibles serait encore plus large: *est-ce que la voiture que j'ai amenée hier est prête ?* , *est-ce que l'auto de ma femme est prête ?* etc., seraient aussi acceptables que *est-ce que ma voiture est prête ?*

Tout se passe comme si les auteurs proposaient de travailler sur des énoncés tandis que l'apprenant devait les considérer, au moment de répondre, comme des phrases isolées, destinées à exercer sa compétence (linguistique) sur le possessif.

**Madame Saule reçoit des amis**

- Je vous présente mon fils, ma mère, mon amie Catherine et mes amis
- Voilà ma nouvelle villa! canadiens.
- Mes autres invités sont en retard.

**Entendu en passant****1. Au restaurant:**

- Garçon! ..... steack et ..... frites! J'attends depuis une demi-heure!
- J'arrive tout de suite, Monsieur!
- Garçon! alors ..... café!
- Tout de suite, Madame!

**2. Au garage:**

- Est-ce que ..... est prête?
- Pas encore Monsieur, revenez ce soir.
- Allô! c'est bien la station Shell?
- Oui Monsieur.
- ..... camion est en panne, est-ce que vous pouvez venir?

**3. Chez le cordonnier:**

- Bonjour Monsieur! ..... sont prêtes?
- Mais oui Madame, les voilà!

**4. Dans la rue:**

- Quelle heure est-il Monsieur, s'il vous plaît?
- J'ai neuf heures moins dix mais ..... retardé.
- Oh! Je suis en retard! ..... n'a pas sonné ce matin!

Le fait d'envisager de transférer ces exercices sur ordinateur met en évidence ce porte-à-faux : Jusqu'à quel point la médiatisation d'exercices papiers sur support informatique devrait-elle anticiper les productions imprévues mais fonctionnellement et linguistiquement correctes ? Quel sort faudrait-il réserver à cette production. Quel feed-back renvoyer à l'apprenant ?

L'exercice morphosyntaxique ne pose on l'a vu aucun problème de traitement automatique dès lors qu'on le propose comme tel ; il en va tout autrement si cet objectif se trouve confondu (même pour des raisons didactiques) avec un exercice d'expression ou de communication introduisant la dimension sémantique.

Il est en effet beaucoup plus difficile de mettre en oeuvre sur ordinateur des exercices touchant à ce qui dans la grammaire relève du sens (emplois des déterminants, valeurs aspectuelles des temps, etc.). Le travail linguistique préalable à une quelconque formalisation n'a pas encore été fait. Si les exercices sur papier peuvent entretenir la confusion entre ces deux plans (formes-sens) sans trop de conséquence, il est clair que l'ordinateur exige une rigueur sans faille :

- Travaille-t-on sur des formes ou travaille-t-on sur du sens (ou sur les deux à la fois) ? Il faut se déclarer et assumer les conséquences de ses positions en introduisant en machine les informations pertinentes aux deux approches.

- Que faire des productions linguistiquement correctes ne répondant pas à l'objectif de l'exercice ? Quel feed-back renvoyer à l'apprenant ?

- En admettant que l'on décide d'accepter ces dernières, cela supposerait que tout le travail nécessaire à l'analyse automatique de la morphosyntaxe du français ait été réalisé. C'est un énorme travail de formalisation dont nous venons de donner une idée dans l'exemple de l'exercice n°2.

Il n'est cependant pas déraisonnable de penser que ce premier niveau de contrôle pourra alors s'exercer avec de plus en plus d'efficacité.

Si l'on admet par ailleurs, que toute phrase syntaxiquement bien formée est porteuse de sens, tout contrôle syntaxique revient à exercer un contrôle sémantique. Le système n'a donc pas à extraire ou formaliser le sens, c'est l'hypothèse adoptée par la traduction automatique. La formalisation du sens est du reste aujourd'hui impossible. Aucun des projets visant cette réalisation n'est viable.

Si donc tout contrôle syntaxique recouvre un contrôle sémantico-syntaxique, toute phrase syntaxiquement correcte sera acceptée. On ne se préoccupera pas par exemple, de savoir si la phrase *la terre est bleue comme une orange* est sémantiquement acceptable, puisqu'elle est syntaxiquement bien formée. De fait, chaque lecteur peut l'interpréter comme il l'entend.

On pourrait se contenter de ce niveau d'analyse automatique. Le problème est cependant plus compliqué dans le cas de l'apprentissage des langues.

#### **2.1.4 Analyse d'énoncés précontraints dans la langue-cible**

De nombreux exercices pratiqués aujourd'hui exigent la mise en rapport des énoncés de deux locuteurs, et l'évaluation de la pertinence fonctionnelle de l'un ou de l'autre. C'est le cas des exercices dits conversationnels : dialogues à une entrée, ou canevas de micro-conversations (cf. F. Debyser 1973, 1974). Le contrôle de ces exercices nécessiterait la mise en rapport de deux énoncés entre eux afin d'évaluer la pertinence de l'énoncé produit dans le cadre de l'échange proposé. A l'analyse morpho-syntaxique complète, devrait donc s'ajouter un niveau d'analyse encore plus complexe, celui de l'analyse de l'implicite linguistique et extra-linguistique. Le contrôle du niveau pragmatique ne peut faire l'économie de cette double analyse.

On comprend que les exercices conçus pour le support informatique, avec le souci d'une production langagière contrôlée par ordinateur, et inscrite dans un projet communicatif soient relativement rares. La combinaison de ces trois paramètres : approches communicatives, support informatique, production écrite contrôlée, revient en effet à placer le concepteur devant un cas de très haute complexité de traitement du dialogue homme-machine en langage naturel:

Examinons un exemple-type de ce genre d'exercice. Il s'agit d'un dialogue à une entrée destiné à faire travailler l'apprenant sur la cohérence textuelle et à susciter, entre autres, la mise en oeuvre de la fonction déictique des pronoms (personnels, possessifs, etc.)

Exercice n°4 (Figure 5, ci-dessous)



Imaginez : Alain et Bernard se parlent dans la rue.  
Complétez leur conversation :

1. *Alain* : — .....  
*Bernard* : — Il va tout saccager dans leur appartement !
2. Alain n'est pas du tout d'accord avec ce commentaire de Bernard : comment peut-il justifier que l'appartement ne sera pas saccagé ?  
*Alain* : « .....  
..... »
3. Maintenant, Bernard s'intéresse à cette histoire et veut avoir des informations supplémentaires !  
*Bernard* : — .....  
*Alain* : — Il est adorable ! Ses pattes et sa queue sont noires, et il a les oreilles et le corps tout blancs !  
*Bernard* : — ..... ?  
*Alain* : — Mais non ! c'est un caniche nain !
4. Maintenant, Alain va donner une information surprenante à Bernard.  
*Alain* : — .....  
.....  
*Bernard* : — Dans leur lit ? mais c'est de la folie ! c'est Inimaginable !
5. Alain veut prouver que finalement cette situation est tout à fait normale :  
*Alain* : « .....  
..... »

*Exercices de grammaire* G. D. de Salins, S. Dupré Latour p. 18-19

Malgré le caractère pré-contraint du dialogue, on peut constater qu'il est impossible de circonscrire l'ensemble des énoncés possibles pour chacune des entrées.

D'autre part, si l'on répondait à la première entrée : *Luc est retourné chez sa soeur*, cet énoncé serait morphosyntaxiquement (et sémantiquement) acceptable. Il est fonctionnellement inadéquat. Pour qu'un système puisse renvoyer ce feed-back, il faudrait qu'il épuisse la série de contextes inférés par le verbe *saccager*. Or, en plus du contexte de l'exercice où le saccage est exercé par un animal, on peut imaginer une série infinie de situations justifiant la production de cet énoncé : Luc est un drogué. En revenant chez sa soeur, il va recommencer à tout saccager. Ou bien, Luc est un malade mental, il sort de l'hôpital, il retourne chez sa soeur, on peut prévoir que ses crises de démence vont recommencer et qu'il va tout saccager, etc. Quel système pourrait détecter ces contextes implicites et les dysfonctionnements pragmatiques ? Il est tout-à-fait impossible de le dire aujourd'hui. Contrairement au discours tenu par certains, les recherches à ce niveau ne sont pas près d'y arriver. L'analyse de la pertinence d'un énoncé relève des recherches en linguistique formelle et en intelligence artificielle. Comme le montrent les travaux de Z. S. Harris 1968, et de R.C. Schank 1986, 1990, une quantité colossale d'informations linguistique et extra-linguistique implicites entrent en jeu dans l'inférence du sens d'un énoncé.

Des essais de simulation de dialogues précontraints ou d'échanges conversationnels plus ou moins brefs, ont cependant parfois été tentés sur ordinateur. Mentionnons à ce propos la célèbre simulation de dialogue avec un psychanalyste *Eliza* qui, n'ayant pas pour objectif l'enseignement-apprentissage d'une langue, s'est révélée rapidement infructueuse dans ce domaine.

L'une des tentatives les plus intéressantes est le type de logiciels réalisés à l'université du Québec à Chicoutimi (F. Labelle 1988) : des jeux d'aventure au cadre situationnel bien précis, auxquels sont liés des objectifs linguistiques spécifiques. L'exécution de chacun de ces jeux permet à l'apprenant d'exercer sa compétence linguistique sur un point précis. L'un de ces logiciels, *La Banquière* (ibid.,88) vise par exemple la maîtrise des formes interrogatives. Le joueur, qui doit découvrir le meurtrier de cette femme, est ainsi conduit à poser des questions à chacun des six suspects. Ces questions sont suscitées par de courts fragments de textes, imposant pour que l'histoire continue, le recours à

une demande d'information. Cette production astucieusement précontrainte est cependant suffisamment ouverte pour que l'apprenant puisse formuler l'énoncé de son choix, tout en répondant à l'objectif de l'exercice.

Dans le module ci-dessous (ibid., 88) on peut voir qu'un nombre non négligeable d'énoncés interrogatifs fonctionnellement pertinents sont susceptibles d'être acceptés :

**Module 5 : C'était le téléphone. J'ai répondu. Une femme voulait parler à la Banquière.**

**Formes acceptables :**

Qui ?

Qui était cette dame (femme, personne) (qui voulait parler à la Banquière) ?

Qu'est-ce que cette dame (femme, personne) voulait ?

Est-ce que vous avez reconnu sa (la) voix (de cette dame, femme, personne) ?

Pourriez-vous nous dire ... ? etc.

La programmation en LISP de ce module est présentée ainsi (ibid., 1988) :

**Schémas de phrase :**

INTIND qui SV : 6

INTIND INT c'était : 6

INT SV1 : 6

c'était INT : 6

INTIND1 SN : 6

INTIND2 SN3 : 6

de INT SAGIR : 6

quel ETRE SN : 6

pourquoi : 6

INT2 V1 DET N : 6

que voulait-elle : 6

qu'est-ce qu'elle V1 : 6

INT SV1 : 6

INT2 DET N V1 : 6

qui est-ce qui voulait parler à la Banquière : 6

**Variables locales :**

INT = qui, quelle N

INT2 = que, qu'est-ce que

INTIND1 = connaissez-vous, connaissez-vous, est-ce que vous connaissez, est-ce que vous connaissez

INTIND2 = avez-vous reconnu, est-ce que vous avez reconnu

SV = était SN2, est SN2

SV1 = était-ce, était-elle

SV2 = qui voulait parler à la Banquière, NIL

SN = le nom de SN2, son nom, SN2

SN2 = DET N SV2

SN3 = la voix de SN2, sa voix, cette voix, SN2

DET = cette, la

N = femme, dame, personne

V1 = voulait

ETRE = est, était

SAGIR = s'agissait-il, s'agit-il

**Variables globales :**

INTIND = VIND savoir, est-ce que VIND2 savoir, savez-vous, est-ce que VIND3, POUV2 PROIND dire

VIND = pourrions-nous, pourrais-je, nous voudrions, nous aimerions, nous désirons, j'aimerais, je désirerais, je désire

VIND2 = nous pourrions, je pourrais

PROIND = me, nous

VIND3 = vous savez, vous POUV PROIND dire

POUV = pouvez, pourriez

POUV2 = pouvez-vous, pourriez-vous

En cas d'erreur, l'apprenant reçoit un feed-back différencié selon qu'il s'agit d'une erreur orthographique, syntaxique ou fonctionnelle. Ainsi une erreur orthographique provoquera-t-elle d'autres suggestions (ibid.,1988) :

---

QUESTION : Quel es  
Je ne comprends pas ce mot : es  
Voulez-vous dire : est ?

QUESTION : Quel est le nom de la fame ?  
Je ne comprends pas ce mot : fame  
Voulez-vous dire : femme, dame ?

---

Il serait possible et plus rentable de créer ici un ensemble d'automates, constituant la grammaire de l'ensemble des formes interrogatives induites par les différentes unités de l'"enquête". On pourrait également y introduire un certain nombre de grammaires d'accord (Accord Déterminant-Nom, Ordre des particules pré-verbales, etc.). Le tout serait joint à l'ensemble des outils du vérificateur / correcteur orthographique du LADL. La seule limite serait ici encore d'ordre pragmatique. On ne pourrait contrôler la pertinence fonctionnelle de l'énoncé produit.

Il serait toutefois intéressant de réaliser un tel produit afin de tester l'opportunité d'une analyse de ce genre. Il est possible que dans un certain nombre de cas, la quantité peu élevée d'énoncés fonctionnellement inadéquats, montre que l'on pourrait se passer du niveau de l'analyse pragmatique. L'intervention humaine, en l'occurrence l'enseignant, pourrait alors pallier les défaillances relevant de ce niveau.

Soulignons qu'il est d'autant plus difficile de concilier autonomie d'énonciation pour l'apprenant et anticipation des erreurs donnant lieu à une rétroaction que le champ de l'activité paraphrastique de l'énonceur est large.

Examinons le traitement détaillé d'un échange précontraint (destiné à l'apprentissage du français langue étrangère) que nous avons réalisé, à l'aide du langage d'auteur Arlequin-Diane :

Le cadre situationnel donné dans la langue-cible était le suivant:

Mme BARROT est arrivée à Paris en 1970. Nous sommes en 1985. Elle y habite encore.
--

L'instruction donnée (toujours en langue-cible) était : "Répondez à sa place en utilisant la première personne du singulier :"

L'échange précontraint par le cadre situationnel et l'instruction ci-dessus avait pour but la production de tournures temporelles utilisant "*ça fait... depuis... il y a...* " avec les temps de verbes correspondants. Il se présentait ainsi :

- *Vous habitez Paris depuis longtemps ?*

- *Oui.....*

Le noyau paradigmatique des énoncés corrects les plus standards d'un locuteur français a dans un premier temps été établi de la façon décrite ci-après, en Figure 6 . La modélisation du noyau des bonnes réponses envisagé dans le corpus ci-dessus a été obtenu selon la démarche suivante :

- Etablissement de modèles "nommant", soit les syntagmes revenant dans plusieurs énoncés, soit les syntagmes longs. Ce langage présente l'avantage (pour des applications en langues) de pouvoir établir des modèles et surtout de pouvoir les nommer afin de les recombinaer à la façon d'un automate. Cette fonctionnalité, que beaucoup d'autres langages d'auteur ne présentent pas, permet de "ramasser" un nombre de réponses relativement important. Nous avons donc établi et nommé, pour les syntagmes envisagés ci-dessus, les modèles suivants :

modèle DF  
modèle QZ  
modèle AN  
modèle CF  
modèle YV  
modèle VE  
modèle LI  
modèle AR  
modèle YA

**Enoncés sans verbe**

[oui]

- très longtemps

• depuis ça fait il y a	DF	(très) longtemps 15, quinze ans une quinzaine d'années	QZ
-------------------------------	----	--	----

- depuis 70, 1970  
AN

**Enoncés avec verbe**

**Verbes duratifs**

Structure « ça fait/il y a »

[oui]	il y a ça fait	CF	15, quinze ans une quinzaine d'années	QZ	que	j'y habite j'y suis j'y vis	YV
						j'habite à Paris ici Paris	LI

•	il y a ça fait	CF	15, quinze ans une quinzaine d'années	QZ	que	je vis je suis	VE	ici à Paris	LI
---	-------------------	----	--	----	-----	-------------------	----	----------------	----

Structure « depuis »

oui	j'y habite j'y suis j'y vis	YV			depuis	15, quinze ans une quinzaine d'années	QZ
	j'habite je vis je suis	VE	ici à Paris Paris	LI		1970, 70	AN

**Verbe non duratif**

- **Ordre 1**

je suis arrivée	AR	à Paris ici	LI	en 1970, 70	AN		
j'y suis arrivée	YA			il y a ça fait	CF	quinze ans une quinzaine d'années	QZ

- **Ordre 2**

il y a ça fait	CF	quinze ans une quinzaine d'années	QZ	que je suis arrivée	AR	à Paris ici	LI
				j'y suis arrivée	YA		

- Modélisation de l'ensemble des bonnes réponses prévues

« très — longtemps »  
ou  
DF SVS (« très — longtemps » ou « longtemps ») ou (QZ)  
ou  
« depuis » SVS AN  
ou  
CF SVS QZ SVS « que » SVS YV ou ((j ← habite SVS (LI ou Paris))  
ou CF SVS QZ SVS « que » SVS (VE ou LI)  
ou  
YV ou ((j — habite SVS (LI ou « Paris »)) ou (VE SVS LI)  
SVS « depuis » SVS (QZ ou AN)  
ou  
((AR SVS LI) ou YA)) SVS AN ou (CF SVS QZ)  
ou  
CF SVS QZ SVS « que » SVS (AR SVS LI) ou YA

Le modèle ci-dessus est représenté par un codage naïf. Le détail de la programmation proprement dite, demanderait d'expliquer le langage tout entier. Il est surtout important de souligner ici que la modélisation obtenue à partir de la typologie prévue ne prend en compte qu'une liste finie d'énoncés corrects.

- Intégration des erreurs et des écarts

Le problème se complique singulièrement si, comme cela ne manque jamais de se produire, les pré-tests sur papier font apparaître un certain nombre d'erreurs de langue que l'on envisage de corriger, ou des énoncés corrects présentant un écart par rapport au paradigme prévu.

A titre d'exemple, on trouvera en Figure 7 l'échantillonnage des réponses recueillies par une série de pré-tests sur papier, portant sur trois classes de débutants et faux débutants en FLE - français, langue étrangère - (une classe à la Sorbonne, Paris IV, une classe à Censier, Paris III, et une classe à l'université d'Albany à New-York) :

Ce premier tri fait surgir immédiatement un certain nombre de remarques et de problèmes à partir desquels des choix sont obligatoirement à faire lors du traitement de chaque ensemble trié :

- Enoncés fonctionnellement incorrects :

1-Énoncés corrects

## 1.1 sans verbe

*depuis 15 ans*  
*depuis quinze ans*  
*il y a 15 ans*  
*ça fait 15 ans*  
*depuis 1970*

## 1.2 avec "être"

*je suis ici depuis longtemps*  
*je suis là depuis 15 ans*  
*j'y suis depuis quinze ans*

## 1.3 avec "arriver"

*je suis arrivée en 1970*  
*je suis arrivée ici en 1970*  
*j'y suis arrivée il y a 15 ans*  
*il y a quinze ans que je suis arrivée à Paris*

## 1.4 avec "habiter"

*il y a quinze ans que j'habite à Paris*  
*il y a quinze ans que j'habite Paris*  
*j'habite Paris depuis quinze ans*  
*j'habite à Paris depuis quinze ans*  
*j'habite ici depuis 15 ans*  
*j'y habite depuis quinze ans*  
*j'y habite depuis 15 ans*  
*j'y habite depuis 15 années*  
*j'y habite depuis longtemps*  
*j'y habite déjà depuis 15 ans*  
*ça fait quinze ans que j'habite Paris*  
*ça fait 15 ans que j'habite ici*

2 Énoncés linguistiquement incorrects

*j'habite ici pour 15 ans*

*pour 15 ans*

*pour quinze années*

*pendant quinze ans*

*j'habite pendant quinze ans alors je connais pas mal la ville*

*j'y habit il y a 15 ans*

*j'habit depuis quinze ans*

*j'avais habiter à Paris depuis 1970*

*j'y habite ici déjà 15 années*

*j'habite depuis quinze ans*

*j'habite Paris depuis quenze ans*

*j'habite à Paris depuis quinz ans*

*presque j'y habite depuis quinze ans*

*c'est pas encore longtemps, c'est simplement quinze ans*

3- énoncés fonctionnellement incorrects

*ça fait 6 mois que j'habite Paris*

*je suis venu à Paris il y a 6 mois*

*je suis arrivé à Paris en mai dernière*

*non 15 ans n'est pas long quand on a 100 ans*

*j'habite Paris depuis 1 an*

Faut-il différencier lors du traitement informatique, les énoncés fonctionnellement incorrects mais linguistiquement corrects, des énoncés fonctionnellement et linguistiquement incorrects ?

Compte tenu de la masse linguistique qu'il reste à traiter par ailleurs, cette différenciation de traitement alourdirait considérablement la tâche et le coût (machine et temps-homme).

Une solution consiste à envoyer à l'apprenant un feed-back suffisamment général, pour qu'il couvre les deux cas de figure, et à lui demander de faire un autre essai. Mais dans ce cas, celui-ci n'a pas d'information sur la correction linguistique de son énoncé.

#### - Énoncés linguistiquement incorrects

Les erreurs jouant sur des énoncés diversifiés, leur mise sous forme canonique s'avère particulièrement complexe. En effet, un élément peut être linguistiquement incorrect dans un énoncé mais linguistiquement correct s'il s'inscrit dans un autre énoncé.

Ainsi dans le cadre de cet échange, la réponse *j'habite Paris pendant 15 ans* révèle une confusion avec *depuis* (faute courante chez des anglo-saxons débutants). Mais *pendant* est correct s'il est précédé du même verbe au passé composé. La réponse *J'ai habité Paris pendant 15 ans* suppose cependant que l'on n'habite plus Paris, ce qui est contraire aux paramètres situationnels donnés

Jusqu'à quel point va-t-on affiner l'analyse des erreurs ? Va-t-on par exemple différencier les erreurs portant sur l'emploi des temps des erreurs portant sur la forme ? Plus l'analyse des erreurs sera affinée, plus elle demandera de travail, plus le programme sera lourd et complexe. Il faudra donc décider de sélectionner un certain nombre d'erreurs et de sacrifier le traitement de certaines autres.

#### - Énoncés corrects

Si l'on compare la modélisation de départ et la liste des énoncés corrects des pré-tests, on remarquera que certains écarts n'avaient pas été prévus: *habiter + 15 années*, *depuis longtemps*, et *déjà*.

Deux solutions sont alors possibles :

- réaménager la modélisation des bonnes réponses de façon à récupérer ces trois éléments,
- desserrer cette modélisation afin de permettre à ces écarts - mais aussi à des écarts incorrects, et c'est le risque - de s'y insérer.

Outre l'énorme quantité de travail qu'ils demandent d'investir dans la phase d'anticipation et de codage des réponses, les exercices de type conversationnel font toucher du doigt les possibilités, mais aussi les limites d'utilisation de l'ordinateur dans le traitement de la langue en contexte d'apprentissage.

On voit ici pourquoi l'utilisation d'outils informatiques tels que des langages-auteurs est totalement à déconseiller dans le cas du cumul de traitements aussi complexes. Bien que l'exercice soit informatisé, l'analyse de la production de l'apprenant relève d'un traitement on ne peut moins automatique. La production langagière dans la langue cible accroît en effet la complexité algorithmique déjà non négligeable de l'analyse de réponse. On se trouve en effet dans ce cas-là devant la nécessité de contrôler la chaîne linguistique entière de chaque phrase ou énoncé, ce qui relève d'une gageure avec l'utilisation de moyens aussi sommaires

L'attitude la plus raisonnable nous semble reposer aujourd'hui sur l'alternative suivante :

- Soit l'on attend qu'une quantité suffisante de grammaires locales aient été développées, de façon à pouvoir procéder à un traitement automatique de la morphosyntaxe. Le jour où un traitement de ce niveau arrivera à des résultats satisfaisants, il faudra encore évaluer, dans le cas des exercices de langue communicatifs, la possibilité de se passer du dernier niveau d'analyse, celui de l'analyse sémantico-pragmatique.

- Soit l'on se sert des outils de traitement automatique existants, pour réaliser des dictionnaires et des grammaires adaptés à un corpus d'exercices donnés.

Il est paradoxalement plus facile de mettre en oeuvre un dialogue ouvert visant une production linguistique en langue maternelle. Dans ce dernier cas, on considère en effet la langue comme acquise, ce qui évacue les problèmes de correction linguistique.

### 2.1.5 Analyse de réponse en langue maternelle

L'utilisation de certains langages d'auteurs permet d'analyser, de façon toujours artisanale, des messages d'apprenant relativement longs, formulés de façon totalement libre, en langue maternelle.

Les réponses en langue maternelle sont requises, à notre connaissance, par deux types de logiciels :

a) des logiciels dont l'objectif n'est pas l'apprentissage d'une langue.

Nous développons un exemple ci-dessous.

b) des logiciels dont l'objectif vise l'apprentissage d'une langue étrangère. Dans ce dernier cas, une phrase est par exemple proposée en anglais, et une consigne (en anglais ou en français) demande à l'apprenant français d'observer un phénomène métalinguistique et de donner une réponse en français. Ce type de logiciels a vite été en butte à toutes sortes de critiques, car il ne correspondait en rien aux approches communicatives.

Par ailleurs, s'agissant, de l'apprentissage d'une langue étrangère, un dialogue apprenant-machine en langue maternelle ne nous paraît absolument pas rentable linguistiquement pour l'apprenant, même s'il s'exerce à propos d'un travail de conceptualisation grammaticale sur la langue-cible.

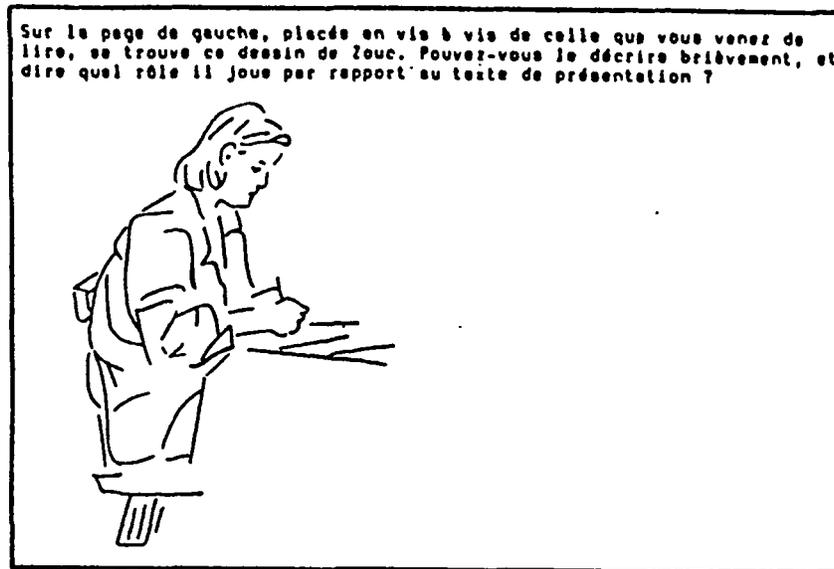
On verra en outre que, tout comme pour l'exemple décrit ci-dessous, l'élaboration de tels logiciels demande au concepteur un investissement disproportionné, par rapport à l'objectif visé.

Seul le premier type nous paraît donc intéressant à considérer. L'analyse des réponses de l'apprenant (en langue maternelle) se fait ici par repérage, regroupement, et mise en facteur d'ensembles de mots-clés. Bien que relativement fruste, - comme dans le programme *Eliza*, l'ordinateur ne "comprend" rien -, cette technique d'analyse donne des résultats, si elle a été réalisée dans un cadre méthodologique satisfaisant (prétests sur papier auprès d'un public représentatif notamment).

Nous présentons, à titre d'exemple, un échantillon de dialogue ouvert réalisé en langage Arlequin-Diane (N. Rodriguez, D. Valentin 1985 ). Ces auteurs donnent dans leur ouvrage sur ce langage, l'exemple d'un travail de lecture dirigée sur ordinateur à partir d'un texte d'Hervé Guibert *Zouc par*

Zouc. L'utilisateur vient de lire une page de texte, puis sur un deuxième écran représentant Zouc assise apparaît la question suivante :

"Sur la page de gauche, placée en vis à vis de celle que vous venez de lire, se trouve ce dessin de Zouc. Pouvez-vous le décrire brièvement, et dire quel rôle il joue par rapport au texte de présentation ?"



D'après le dessin de R. Montaudon, dans *Zouc, Zouc*, Balland, 1978.

Cette question ouverte s'adresse à des apprenants dont on considère a priori qu'ils maîtrisent leur langue. L'analyse anticipée des réponses ayant préalablement donné lieu à des prétests sur papier, il aura toutefois été facile de repérer d'éventuelles fautes d'orthographe susceptibles non pas d'être corrigées - tel n'est pas ici le propos - mais prises en compte pour le repérage des termes-clefs. Ces tolérances orthographiques sont rendues possibles par l'utilisation des "jokers" (cf. plus haut).

Les prétests ont porté ici sur une soixantaine d'élèves. Leurs réponses ont été décomposées en éléments de sens eux-mêmes classés en 4 sous-ensembles thématiques pertinents :

- Sous-ensemble 1: "c'est Zouc qui écrit"

Cet ensemble regroupe des éléments phrastiques tels que : "Zouc tient un stylo, elle est en train d'écrire, on la voit écrire, elle est représentée écrivant elle-même" etc.

- Sous-ensemble 2 : "rôle d'Hervé Guibert"

Cet ensemble regroupe des éléments phrastiques tels que: "Hervé Guibert n'intervient pas, le journaliste n'a été qu'un transcripteur, l'interviewer n'a pas fait de retouches, elle écrit seule, il n'y a pas réécriture." etc.

- Sous-ensemble 3 : "concentration"

Cet ensemble regroupe les éléments phrastiques se rattachant à ce thème " Zouc est sérieuse, triste, anxieuse. Elle a une attitude mélancolique. Elle est concentrée, absorbée, pensive . Elle n'est pas gaie. Le dessin imprécis évoque une personnalité qu'on a du mal à cerner " etc.

- Sous-ensemble 4 : "spontanéité"

Cet ensemble regroupe des éléments tels que "Zouc est spontanée, Zouc n'est pas compliquée, il lui suffit d'une chaise, d'une table et d'un stylo. Il n'y a pas de censure" etc.

Ce regroupement (cf. ibid.85) que nous venons de présenter ici de façon très simplifiée, a donné lieu à l'élaboration de modèles, eux-mêmes combinables entre eux à l'aide des opérateurs "et" et "ou".

Pour l'ensemble "concentration" par exemple, 3 modèles regroupant les mots-clés de cette classe ayant un sens équivalent ont été élaborés et regroupés sous un terme générique :

Modèle DESSIN : "DESSIN, TRAIT\$\$, VISAGE, ATTITUDE"

Modèle SERIEUX : "SERIEU\$\$, TRISTE\$\$, ANXIE\$\$, MELANCOLI\$\$, REFLECHI\$\$, CONCENTR\$\$, ABSORBE\$\$, PENSI\$\$".

Modèle AMBIGU : "RENFERME\$\$, AMBIGU\$\$"

Modèle GAI : " GAI\$\$, JOYEUS2\$"

La possibilité (offerte par ce langage d'auteur) de nommer ces différents modèles permet de les réutiliser pour établir un ou plusieurs autres modèles. Ici, le modèle (M3) du sous-ensemble 3, combine les modèles décrits ci-dessus avec l'un des modèles du sous-ensemble 1 (ZOUK) et l'un des modèles du sous-ensemble 2 (NEGATION) (non décrits ci-dessus).

La modélisation finale de ce sous-ensemble de réponses se présente donc ainsi :

Modèle M3 : (ZOUK ou DESSIN) et (SERIEUX ou AMBIGU ou (NEGATION et GAD))

La mise en facteur de ces sous-ensembles permet d'élaborer des architectures souples, complexes (si nécessaire) et aptes à regrouper un grand nombre de prévisions de réponses libres.

L'idée de vouloir cerner le sens d'une réponse en langue maternelle, à l'aide d'outils aussi artisanaux, nous paraît cependant peu réaliste. Bien que cette démarche donne des résultats, il est difficile d'envisager la généralisation de son utilisation pour des réponses libres, fussent-elles en langue maternelle. En effet, l'analyse préalable au traitement informatique d'une réponse (dont nous n'avons donné ici qu'un exemple partiel et simplifié) n'est opérationnelle que pour celle-ci. La quantité de travail nécessaire au traitement de chacune d'elles est donc d'autant plus démesurée qu'elle n'est pas réinvestissable pour l'analyse d'une autre réponse libre. Rappelons en outre, que le traitement de l'ensemble des réponses de chaque exercice est lié à un public-cible très circonscrit.

Il serait plus réaliste d'envisager l'analyse de réponses en langue maternelle, dans le domaine professionnel, et dans un cadre très circonscrit : réponse à des questionnaires pour des enquêtes pointues dans des domaines de spécialité (assurance, banque, etc.) par exemple. On adapterait à ce moment-là, dictionnaires électroniques et grammaires locales à tel ou tel type d'enquête. Il serait d'ailleurs peut-être intéressant d'utiliser ces produits, comme matériau authentique, dans le processus d'enseignement/apprentissage des langues. A l'heure actuelle, la réalisation de ce type de produit n'est pas encore généralisée.

#### **2.1.6. Les orientations des applications actuelles**

Prenant acte des expériences menées ces dix dernières années, les concepteurs ont abandonné l'idée de faire jouer par l'ordinateur le rôle d'un répétiteur plus ou moins limité ou au contraire d'un interlocuteur à part entière ; ils s'attachent plus lucidement aujourd'hui à tirer parti de ses fonctionnalités

spécifiques, pour fournir des outils simples et attrayants, susceptibles d'aider et de renouveler les pratiques de classe.

D'où l'apparition de "logiciels outils" ou "programmes auteurs" que l'on peut grosso modo regrouper en deux grandes classes, même si la frontière n'est pas toujours aussi nette :

- Dans la première classe peuvent se situer les produits destinés à fournir au professeur une sorte de "patron d'exercice" dans lequel il peut entrer ses propres données et dont toute la partie fastidieuse (programmation, gestion automatique des réponses) lui est évitée. Modeste d'ambition, ce type de produit peut s'avérer aussi fonctionnel que facile à insérer dans des processus pédagogiques variés, à condition qu'il soit bien conçu.

Sa double caractéristique est d'être "ouvert" en ce qui concerne l'introduction des données de l'exercice, et d'éviter l'écueil des lourdes analyses de réponses des apprenants.

Nous situons dans cette catégorie les logiciels d'"élaboration de Q.C.M." (questionnaires à choix multiples), type *A Votre Avis* ou *Echolanges* ou de petits exercices à réponses courtes ou même les logiciels de "reconstitution de texte" (type *Storycorner*).

Soulignons que ce type de produit - du fait qu'il est indépendant des contenus - donne généralement lieu à des versions destinées à l'apprentissage de différentes langues (français, anglais, allemand, espagnol généralement).

- Dans la deuxième classe, se situent les produits destinés à *susciter* et *stimuler* diverses activités langagières : lecture / écriture, production orale, observation de faits de langue, etc. Ces produits se présentent soit comme des "moules" destinés à recevoir et à gérer - mais non à *contrôler* - la production écrite des apprenants (type *Multiscénario* pour la création de récits arborescents), soit comme des utilitaires destinés à capturer et/ou manipuler et/ou traiter des données provenant d'autres supports tels que Minitel, ou des recherches d'élèves (type *Poeteach* pour une génération automatique de poèmes "à la manière de" ou *Concordteach* pour la réalisation de concordances sur des textes capturés). Pour une description détaillée de ces logiciels on se reportera à M. Garrigues (1989).

Il faut ajouter à ces deux types de produits relativement limités, les *logiciels professionnels* tels que logiciels de *traitement de texte* ou logiciels *hypertextes* de plus en plus utilisés dans le contexte de l'apprentissage des langues.

Nous ne développerons pas ici les possibilités d'intégration de tels outils dans divers processus pédagogiques. Quel que soit ce processus, l'apprenant sera souvent amené de toute façon, à produire de l'écrit. Les enseignants souhaiteraient donc que les correcteurs orthographiques insérés dans ce type d'outils soient mieux adaptés aux fautes d'un apprenant, et plus particulièrement à celles d'un apprenant étranger.

Il nous semble donc que l'élaboration de grammaires locales, correspondant à une typologie des fautes les plus caractéristiques faites par un apprenant étranger, permettrait de répondre progressivement à ce besoin. En les intégrant au correcteur orthographique du système InTex, et en les couplant avec la couche lexicale C1 extraite du DELAS, on obtiendrait ainsi un vérificateur - correcteur orthographique dédié à l'apprentissage du français, langue étrangère de Niveau 1.

Ces grammaires étant ajustables et modulables, rien n'interdirait qu'on les emploie pour de simples exercices sur objectifs spécifiques, qu'ils soient strictement morpho-syntaxique ou communicatifs.

Nous nous proposons précisément de montrer comment les outils linguistiques et logiciels destinés au traitement automatique, permettent de réaliser de telles grammaires.

## **2.2. Méthode et outils pour l'élaboration de grammaires locales :**

L'application que nous proposons est celle du prédicat : "Mouvement de déplacement d'un sujet humain par rapport à un lieu désigné"

Les expressions les plus simples de ce prédicat sont des phrases comme *Jo va dans sa chambre / à son bureau).*

Or pour un apprenant étranger, l'utilisation adéquate des prépositions en français, même dans des phrases aussi élémentaires est redoutable de difficulté. Le traitement de cette difficulté a fait l'objet de nombreux travaux, des plus scientifiques aux plus appliqués. Dans les ouvrages de grammaire, il est limité à un simple inventaire assorti d'exemples venant parfois illustrer le sens présumé des différentes prépositions, ce qui rend son enseignement très

ingrat pour le professeur, et particulièrement pour le professeur de français langue étrangère qui ne trouve exposée que de façon éparse les difficultés auxquelles il est confronté.

D'une manière générale, si certaines prépositions font constitutionnellement partie du verbe (*convenir de, s'attendre à, être confronté à, etc.*), dans de nombreux autres cas, et en particulier celui des phrases de mouvement, leur emploi est strictement dépendant de l'association d'un verbe et d'un nom. Jamais les manuels de grammaire n'envisagent le problème du point de vue de cette entité syntaxique complexe. Or, ce phénomène met en évidence l'ampleur du problème. On comparera :

*Jo va dans la chambre / Jo va \*à la chambre et*  
*Jo va dans la cuisine / Jo va à la cuisine*

En général, les méthodes et les manuels de grammaire destinés à l'apprentissage du français langue étrangère, fournissent en annexe la liste des principaux verbes du français régissant une extension prépositionnelle. Leur formalisation se réduit à un ordre alphabétique. Quant au point infiniment plus complexe du choix d'une préposition à l'intérieur d'une combinaison verbe-nom, il n'est qu'à peine évoqué, ce qui escamote les problèmes. Dans le cas des phrases de déplacement, ce problème est toujours réduit à un exemple : l'emploi de la préposition avec un nom de pays. C'est l'éternel chapitre *aller en Italie, venir du Japon, etc.* Or, si tous ces manuels reproduisent inlassablement ce même exemple, ce n'est pas un hasard : il est bien le seul à présenter une régularité permettant de présenter quelques règles. L'inconvénient de cet unique exemple, est qu'il entretient implicitement l'idée, que l'emploi de la préposition répond à des règles formelles simples. En fait, le professeur ne trouve, pas plus qu'ailleurs, les réponses aux difficultés posées par les fautes de ses élèves.

En tout état de cause, il serait tout à fait impossible de réaliser des automates visant la correction automatique au moyen du seul matériau fourni par ces manuels, qu'ils soient destinés au français langue maternelle ou langue étrangère, et les données tirées des savantes études traditionnelles ne sont pas suffisantes non plus.

Nous avons donc choisi de nous pencher sur le prédicat sémantique (M. Gross, 1981) dans lequel se situe précisément cet exemple partout ressassé, et de montrer la complexité extrême de ce sujet, dès lors qu'on l'aborde avec

la rigueur nécessaire à un travail de formalisation. Celle-ci est en effet l'étape préalable et indispensable au traitement automatique de tout fait linguistique.

Afin de montrer en quoi consiste cette formalisation, et à l'aide de quels outils il est possible d'arriver à des résultats, nous avons explicité la démarche conduisant à la réalisation de grammaires locales. Nous avons développé celle-ci jusqu'à la réalisation de quelques modules issus de notre prédicat de départ. Celui-ci demande donc à être soigneusement circonscrit.

Dans le cas de notre application, les critères syntaxiques et les classements utilisés au LADL pour la réalisation des tables de verbes (sujet humain / non humain, verbes de mouvement, verbes à trois arguments, etc.) nous ont permis d'élaborer et de sérier des scénarios et ce faisant, d'arriver à clarifier une situation des plus confuses.

### **2.2.1. Définition du prédicat**

Notre prédicat sémantique, se rapporte donc aux deux scénarios complémentaires suivants :

**a) Un sujet humain effectue un déplacement d'un lieu d'origine à un lieu de destination.**

Dans les phrases exprimant ce prédicat, peut ou non, intervenir une préposition locative:

*Jo a regagné Paris*

*Jo est arrivé à Londres*

*Jo sort de sa chambre*

*Jo se dirige vers la porte*

Il arrive, selon les verbes, que l'un de ces deux lieux soit implicite. On a donc des phrases comme :

*Jo va de Paris à Lyon*

*Jo va à Lyon*

*Jo vient de Paris*

Elles correspondent cependant au même scénario. Il arrive également qu'une étape du trajet soit précisée :

*Jo va de Paris à Lyon, par Dijon*

**b) Un sujet humain reste sans se déplacer dans le lieu où il est.**

*Jo habite un hôtel particulier*

*Jo reste à la maison*

*Jo est dans l'Yonne*

*Jo demeure dans le quatorzième arrondissement*

Nous précisons que notre étude ne prendra pas en compte les scénarios suivants, voisins du scénario a) :

**c) Un sujet humain effectue un déplacement dont aucune des bornes n'est définie :**

*Jo parcourt le jardin*

*Jo sillonne le pays*

*Jo fait des kilomètres dans la région*

**d) Un sujet non-humain se déplace d'un lieu à un autre.(avec les mêmes variations de nombre de compléments qu'en a) :**

*L'eau s'écoule (de la gouttière) dans le caniveau*

*La rumeur se propage dans toute la ville*

*Le colis est arrivé à destination*

*Le ballon s'envole vers le ciel*

**e) Un sujet non-humain effectue un déplacement dont aucune des bornes n'est définie :**

*La rivière serpente dans la vallée*

*L'air circule dans la pièce*

**f) Un sujet humain déplace un objet ou un humain d'un lieu à un autre :**

A la différence des verbes à deux arguments utilisés dans les scénarios précédents, les verbes servant à exprimer les déplacements des objets ou des humains causés par un sujet humain, sont des verbes à trois arguments, et la cause est en position de sujet

-déplacement d'un objet :

*Jo range la vaisselle dans le placard*

*Jo sort le beurre du frigo*  
*Jo prend les livres sur l'étagère*  
*Jo déplace le piano de sa chambre au salon*

-déplacement d'un humain :

*Jo a envoyé sa fille en Angleterre*  
*Jo a retiré son fils de ce lycée*  
*Jo a renvoyé les élèves chez eux*  
*L'administration a muté Jo de Paris en Province*

**g) Un sujet humain fait une action :**

L'action peut ou non comporter des mouvements ou des déplacements, mais ils sont implicites et secondaires par rapport à l'action exprimée.

*Jo travaille dans le seizième*  
*Jo dort dans sa chambre*  
*Jo court dans le parc (= Jo fait de la course dans le parc)*  
*Jo balaye l'entrée*

**h) un sujet humain exécute un mouvement du corps sans déplacement :**

*Jo s'affale dans le canapé*  
*Jo s'abat sur son lit*  
*Jo s'effondre sur un banc*  
*Jo se vautre dans l'herbe*

En ce qui concerne le scénario b de notre prédicat, nous ne prendrons pas en compte le scénario voisin suivant :

**i) un sujet non-humain est statique :**

*Le beurre est dans le frigo*  
*La salon ouvre sur le jardin*  
*Son portefeuille sort de sa poche*  
*Le chemin descend vers la mer*

Nous n'avons écarté ici écarté les scénarios c, d, e, f, g, h, i, voisins de notre prédicat sémantique, que pour des raisons pratiques de délimitation de l'étude. Nous précisons qu'il sera toujours possible de les traiter par la même

méthode. La construction d'automates pouvant se faire en effet, de façon très modulaire, on pourra élaborer et intégrer par étapes des grammaires et des sous-grammaires locales élaborées pour ces scénarios. Celles-ci viendront ainsi perfectionner et complexifier progressivement l'automate de base correspondant à l'ensemble de ces scénarios. Le caractère à la fois artisanal et démesuré de ce travail pourra paraître étonnant. Il est nécessaire de donner à un ordinateur les mêmes connaissances linguistiques qu'un correcteur humain. Il n'y a aucune autre façon de procéder pour améliorer la correction automatique des fautes d'un apprenant, et tout traitement automatique de la langue en général.

### **2.2.2. les modalités et les structures syntaxiques du prédicat**

Soit donc les éléments prédicatifs **TRAJET (Nhum, Nloc)**. Il est possible d'envisager deux modalités élémentaires de déplacement :

- l'orientation **ALLER (Nhum, Nloc)**.
- l'orientation **VENIR (Nhum, Nloc)**.

Examinons de quelles façons, elles peuvent être traduites en constructions du français. Considérons la première :

#### **2.2.2.1. ALLER (Nhum, Nloc)**

A cette modalité de déplacement correspondent deux structures syntaxiques:

##### **2.2.2.1.1. La structure syntaxique *N0 V N1*.**

On a donc la préposition zéro, ce que l'on note : *Prép =: E*.

Deux ordres sont possibles :

- *N0 =: Nhum, N1 =: Nloc*

*Jo a atteint le sommet*

*Jo regagne sa chambre*

- *N0 =: Nloc, N1 =: Nhum*

*Ce film attire les foules*

(= *Les foules vont à ce film*)

*Cette exposition a reçu 1000 visiteurs*

(= *1000 visiteurs sont allés à cette exposition*)

2.2.2.1.2. La structure syntaxique est *N0 V Prép N1*.

Celle-ci est susceptible de se réaliser avec un grand nombre de prépositions :

Prép =: à	<i>Jo va à la mairie</i>
Prép =: en	<i>Jo va en banlieue</i>
Prép =: dans	<i>Jo va dans sa chambre</i>
Prép =: sur	<i>Jo va sur la place</i>
Prép =: sous	<i>Jo va sous la tente</i>
Prép =: vers	<i>Jo va vers la voiture</i>
Prép =: devant	<i>Jo va devant la porte</i>
Prép =: derrière	<i>Jo va derrière la maison</i>
Prép =: contre	<i>Jo va contre le mur</i>

Nous signalons la structure à deux prépositions *de ... à*, qu'il conviendrait de prendre en compte dans une grammaire locale, visant à formaliser l'ensemble du prédicat.

Nous mentionnons également l'emploi possible de de locutions telles que *jusqu'à, jusqu'au bout de, depuis.etc.*

Dans cette modalité (prépositionnelle ou non prépositionnelle) de déplacement, *Nloc = NlocD* (destination).

#### 2.2.2.2 VENIR (*Nhum, Nloc*)

A cette modalité de déplacement correspondent deux structures syntaxiques :

2.2.2.2.1 La structure syntaxique est *N0 V N1*.

On a donc, Prép =: E

Deux ordres sont possibles :

*N0 =: Nhum, N1=: Nloc*

*Jo quitte la maison*

*Jo abandonne sa cachette*

*N0 =: Nloc, N1=: Nhum*

*Le métro vomit ses voyageurs*

*(= Les voyageurs sortent du métro)*

*Le train déverse ses voyageurs*

*(= Les voyageurs descendent du train)*

#### 2.2.2 2.2. La structure syntaxique est *N0 V Prép N1*

L'éventail des prépositions actualisables est nettement moins large pour cette modalité prépositionnelle de déplacement, *Nloc = NlocS* (source). Il se limite à *de* (accompagné ou non d'une seconde préposition) et *depuis*.

*Prép =: de*            *Jo vient du (de le) bistrot*  
                          *Jo vient de la campagne*  
                          *Jo vient de chez la voisine*  
                          *Jo vient des (de les) environs*  
                          *Jo vient d'en (de en) face*

*Prép =: depuis*        *Jo arrive depuis son bureau*

Cette dissymétrie entre le grand nombre de prépositions actualisables avec *Nloc = NlocD* (destination) et le nombre restreint de prépositions réalisables avec *Nloc = NlocS* (source) n'est, en fait, qu'apparente. Nous avons probablement affaire, dans un grand nombre de cas à un processus d'écrasement par la préposition *de* d'une seconde préposition aux formes variées. Si nous comparons

*Jo va dans la chambre*

*Jo vient de sa chambre,*

il est possible d'interpréter la structure sous-jacente de cette seconde phrase comme :

*Jo vient de dans sa chambre.*

La préposition *dans* subirait dans ce cas un effet d'écrasement. La préposition *chez* ne la subit pas :

*\*Jo vient de la voisine*

Un certain nombre de faits linguistiques étayent cette hypothèse :

- Une tournure populaire garde, avec les verbes de mouvement (à deux ou à trois arguments) les deux prépositions, comme par besoin de rendre explicite, par ce moyen, une combinaison de rapports qu'une seule préposition ne traduit qu'imparfaitement . Il s'agit de la tournure du type :

*Elle l'a sorti de dedans son sac.*

elle-même d'ailleurs souvent prononcée avec un *d* géminé pour *dans* :

*Elle l'a sorti de d'dans son sac.*

On retrouve également cette emphase syntaxique dans la tournure populaire interrogative :

*De d'où (est-ce qu') il sort celui-là ?)*

Il arrive que cette tournure populaire soit acceptée, même à un niveau de langue correct, lorsqu'il s'agit de traduire, par l'explicitation même de cette combinaison de rapports, un sentiment d'exaspération .Ainsi l'injonction faite à un enfant sautant sur une table :

*Descends de cette table!*

revêt une nuance d'insistance plus grande avec l'utilisation de la tournure populaire à deux prépositions :

*Descends de sur + dessus cette table!*

On remarquera en outre que l'explicitation des deux prépositions est dans certains cas obligatoire, notamment pour lever toute ambiguïté. La même injonction faite à un enfant se trouvant non pas *sur* mais *sous* une table ne pourrait être que :

*Sors de sous + dessous cette table!*

*Sors de cette table!* aurait un autre sens.

On constate, en fait, que dans l'utilisation du couple *de Prép =: de*, tous les cas de figure sont possibles :

a) Maintien de *de Prép* obligatoire :

*Sors de sous cette armoire!*

*\*Sors de cette armoire!*

(autre sens = Jo est dans l'armoire)

*\*Sors sous cette armoire!*

Ce cas lui même peut donner lieu à des tolérances d'ambiguïté, lorsque les interprétations des phrases de base conduisent à des situations référentielles quasiment identiques :

*Sors de sous tes couvertures!*

(= *Jo est sous les couvertures*)

*Sors de tes couvertures!*

(= *Jo est dans les couvertures*)

La différence entre *être sous les couvertures* et *être dans les couvertures* étant infime, les emplois de *de* ou de *de sous* sont quasiment interchangeables en regard de la phrase de référence.

b) Maintien obligatoire de *de Prép*

Ecrasement de *Prép* (tournure populaire) :

*Jo vient de chez le coiffeur*

*Jo vient du coiffeur* (tournure pop.)

c) Maintien de *de Prép*.

Ecrasement de *Prép* :

*Enlève ce chat de sur mes genoux!*

*Enlève ce chat de mes genoux!*

*\*Enlève ce chat sur mes genoux!*

d) Ecrasement de *Prép*.

Maintien de *de Prép* (tournure populaire) :

*Il l'a sorti de la poubelle*

*Il l'a sorti de (de)dans la poubelle* (tournure pop.)

*\*Il l'a sorti dans la poubelle (sauf avec un sens différent)*

e) Ecrasement de *de*.

Maintien de *de Prép* :

*Quelqu'un l'a pris de sous mon oreiller*

*Quelqu'un l'a pris sous mon oreiller*

*\*Quelqu'un l'a pris de mon oreiller*

f) Ecrasement de *de*.

Ecrasement de *Prép*

Maintien de *de Prép*.

*Prends le sucre sur l'étagère*

*Prends le sucre de l'étagère*

*-Prends le sucre de sur l'étagère*

Des exemples historiques attestent également cet emploi de doubles prépositions. "On trouve à toutes les époques de l'histoire de la langue, des *groupes de prépositions* pour rendre des *combinaisons de rapports* :

*De vers Ardenne vit venir uns leuparz*

*(Chanson de Roland, v. 728),*

De la région d'Ardenne (exactement : *de du côté d'Ardenne*) il vit venir un couple de léopards.

*Il trait esraument .j. anel*

*D'a son doit*

*(Fabliau de Constant du Hamel, v. 106-107),*

(le forestier) tire vivement un anneau *d'à* son doigt.

Corneille ... écrit, en vers dans une comédie :

Ils ont vu tout cela *de sur* une éminence

*(Suite du Menteur, v. 1118)."*

(F. Brunot et C. Bruneau, 1964).

Quoi qu'il en soit, nous ne prendrons en compte pour les besoins de notre formalisation, que la forme actualisée des prépositions. Mais on voit déjà qu'il est difficile d'établir des règles générales. en fait, les cas devront être examinés au coup par coup.

Envisageons à présent le scénario b) du prédicat, celui de l'absence de déplacement d'un humain. A ce scénario correspond une seule modalité : la modalité ETRE, elle-même réalisable selon trois structures syntaxiques possibles.

### 2.2.2.3 ETRE (Nhum, Nloc).

A cette modalité correspondent trois structures syntaxiques :

#### 2.2.2.3.1 La structure syntaxique est $N0 V N1$

Prép =: E

Deux ordres sont possibles :

$N0$ =: *Nhum*,  $N1$ =: *Nloc*

*Dix familles occupent cet immeuble*

*Jo habite Paris*

$N0$ =: *Nloc*,  $N1$ =: *Nhum*

*Cet immeuble abrite 10 familles*

*Cette malle renferme un corps*

#### 2.2.2.3.2 La structure syntaxique est $N0 V Prép N1$

*Jo est au bistrot*

*Jo habite dans le quatorzième arrondissement*

#### 2.2.2.3.3. La structure syntaxique est :

*il y a Nhum Prép Nloc*

*Il y a un enfant dans cette pièce*

Nous récapitulons ci-dessous l'architecture générale de l'automate correspondant à notre prédicat :

## Architecture générale des formes du prédicat

### 1 Déplacement

#### 1.1 ALLER (Nhum, Nloc,)

##### 1.1.1. NO V N1

##### 1.1.1.1 Nhum V Nloc

*Jo regagne sa chambre*

##### 1.1.1.2 Nloc V Nhum

*Ce film attire les foules*

##### 1.1.2. Nhum V Prép Nloc

*Jo va dans sa chambre*

#### 1.2 VENIR (Nhum,Nloc,)

##### 1.2.1. NO V N1

##### 1.2.1.1. Nhum V Nloc

*Jo quitte la maison*

##### 1.2.1.2. Nloc V Nhum

*Le train déverse les voyageurs*

##### 1.2.2. Nhum V Prép Nloc

*Jo vient du bistrot*

### 2. Absence de déplacement

#### ETRE (Nhum, Nloc,)

##### 2.1 Nhum V Nloc

*Jo habite Paris*

##### 2.2 Nhum V Prép Nloc

*Jo est à la maison*

##### 2.3. Il y a Nhum Prép Nloc

*Il y a un homme dans l'armoire*

Nous ne développerons pas ici toutes les branches, ni même une branche complète de cet automate. Mais cette structure permet déjà de situer sémantiquement et syntaxiquement les automates que nous avons développés en exemple (cf. ci dessous, 2.5 ) ainsi que tous ceux qui s'inscrivent dans ce cadre. Elle constitue en outre un modèle pour le développement ultérieur d'autres automates du champ (cf. scénarios c, d, e, f, g, h,i)

### 2.2.3. Traitement du matériau de base

L'existence de matériaux de base formalisés à un niveau théorique apporte une aide considérable à la réalisation de grammaires formelles et d'automates destinés aux applications. Ils permettent d'envisager leur construction dans des conditions et des délais raisonnables. Il serait peu réaliste d'envisager la création de tels outils, en partant de zéro. Ils nous ont permis, nous l'avons vu, de dégager une structure cohérente à partir de scénarios établis à l'aide de critères formels éprouvés. Ils offrent en outre l'intérêt de réunir des corpus très complets.

### 2.2.3.1 Le corpus

L'existence d'un lexique - grammaire du français nous a permis en effet de travailler sur des corpus de verbes déjà constitués et classés.

Nous avons donc utilisé comme base de départ les matériaux constitués des tables syntaxiques suivantes :

Pour la modalité prépositionnelle :

-Table T2 = table syntaxique des verbes de mouvement (M. Gross, 1975),

-Tables T35L et T 35ST (J.P. Boons, A. Guillet, C. Leclère, 1976) contenant respectivement des verbes de mouvement et des verbes à aspect statique.

-Table T38L1

-Tables des phrases en *être Prép* (L. Danlos, M. Gross)

*Etre au fin fond de*

*Etre à proximité de*

-Tables des phrases figées

*Faire un saut à Loc*

*Faire une excursion à Loc*

*Prendre le train pour Nloc*

Pour la modalité non-prépositionnelle :

- Table 36L1 (A. Guillet, C. Leclère 1992)

### 2.2.3.2 La sélection

Un premier travail a consisté à sélectionner les entrées de ce matériau théorique. Nous avons donc parcouru ces tables en vue de répartir les verbes selon nos couches C1, C2, C3 (cf. Chap. 1) et procédé à l'élimination d'un certain nombre d'entre eux pour des raisons diverses :

S'agissant d'automates destinés à corriger une langue courante, nous avons éliminé les verbes présentant une faible plausibilité d'apparition dans le cadre de notre prédicat. C'est par exemple le cas de *amerrir* (table T2, couche C2), *crawler* (table T2, couche C2), *pérégriner* (table T2 couche C2) ou *pagayer* (table T2, couche C2). N'ont donc été gardés que les verbes correspondant à la couche C1 de notre dictionnaire.

Ont également été écartées, les entrées dont le sens visé dans la table présentait une faible plausibilité. Rappelons en effet que les tables verbales ont

opéré les séparations de sens (cf. *glisser* sur une tache d'huile, et *glisser* un papier à Jo), tandis qu'au niveau de notre sélection lexicale, cette séparation n'a pas encore été faite. Nous avons donc écarté les verbes *s'enfourner* ou *s'insinuer*, qui pourtant correspondent à la couche C1 sous leur forme d'entrée (*enfournier*, *insinuer*) dans le DELAS : la structure *Nhum s'enfourne dans Nloc* ou *Nhum s'insinue dans Nloc* nous paraît en effet présenter une plausibilité moins grande que *Nhum enfourne N1 dans Nloc* (*Jo enfourne du charbon dans le poêle*), ou *Nhum insinue que P* (*Jo insinue que je mens*). Or ces sens constituent des entrées différentes dans d'autres tables. Par ailleurs les structures qui leur sont liées ne correspondent pas toujours aux paramètres définis dans notre prédicat. C'est également le cas de *couler* dont le sens correspond, dans T2, à *decouler au fond* pour un sujet humain. Ce sens ne nous paraît pas avoir un indice de plausibilité élevé dans le cadre de notre prédicat. Le verbe *couler* est pourtant classé en C1 dans notre DELAS, car la séparation entre les divers sens de *couler*, n'est pas encore réalisée à ce niveau.

Certaines tables (cf. T35L), contiennent des verbes acceptant un sujet humain, des verbes qui ne l'acceptent pas et d'autres encore, qui acceptent à la fois, un sujet humain et non-humain. On a donc, des verbes marqués "+" à la propriété syntaxique *N0 =: Nhum*, des verbes marqués "+" à la propriété *N0 =: N-hum*, et des verbes marqués + à la propriété *N0 =: Nhum* et à la propriété *N0 =: N-hum*.

Ainsi trouve-t-on les verbes *découcher* et *marcher* marqués seulement "+" à la propriété *Nhum*, les verbes *suinter* et *valser* marqués seulement "+" à la propriété *N-hum*, et les verbes *bifurquer* et *cheminer* marqués "+" aux deux propriétés. Or, les verbes concernant notre propos devant obligatoirement accepter la propriété *N0 =: Nhum*, ceux qui ne l'acceptent pas, ont été éliminés. (exemple : *dégouliner*). En ce qui concerne les verbes acceptant les deux propriétés, nous les avons au départ retenus. Cet ensemble a toutefois subi d'autres ajustements. Nous avons en effet écarté des verbes qui, tel le verbe *serpenter* (couche C1), sont dotés de la double marque *Nhum* et *N-hum*, mais présentent une plus grande plausibilité avec *N-hum*. La table T 35 L lui attribue une marque positive (+) à la propriété *N0 =: Nhum*. On comparera :

*La rivière serpente dans la vallée.*

*La manifestation serpente dans les rues*

- Un troisième type d'entrées a également été écarté : celles pour lesquelles la propriété syntaxique *Prép Nloc* (cf. T35L) qui leur est attribuée, présente une faible plausibilité. Rappelons en effet que les propriétés théoriques attribuées aux différents verbes, présentent des degrés de plausibilité divers au niveau d'une actualisation langagière. Ainsi, *Il claudique* ou *Il clopine vers la sortie* nous paraissent des actualisations peu plausibles. A ce niveau, des phrases comme *Il est parti en claudiquant* ou *en clopinant vers la sortie*, nous paraissent plus plausibles. De même, *Il se débine* (employé intransitivement), nous paraît plus plausible que *Il se débine vers sa chambre*.

On pourrait croire qu'une fois ce travail de sélection achevé, il ne resterait plus qu'à regrouper les verbes correspondant à la structure *Nhum V Nloc* d'un côté (groupe 1), et ceux correspondant à la structure *Nhum V Prép Nloc* de l'autre (groupe 2). Or, comme le montre la suite de ce travail d'élaboration d'un automate visé par le prédicat, il est nécessaire de descendre à un niveau de complexité et de détails que les grammaires sont loin de nous laisser supposer.

#### 2.2.4. Du verbe à la phrase

Il est en effet nécessaire, ensuite :

- d'établir, un tableau distributionnel des prépositions utilisées avec chacun des verbes appartenant au groupe 2,
- de constituer un sous-dictionnaire des noms concrets, puisque *Nloc* ne peut être que concret. Le système de manipulations de textes InTex permet, de connecter de tels lexiques à l'automate. (Signalons que le marquage *Nconcret / Nhum* a été réalisé pour la catégorie *NAdj* des mots composés M. Zilberstein, version 3., et qu'elle est en cours pour les noms du DELAS). Pourtant, s'agissant d'un travail destiné à une application pratique d'apprentissage de la langue, à savoir la réalisation d'automates pour la reconnaissance de structures de phrases acceptables (et au rejet des autres), ce marquage ne suffirait pas, car de nombreuses difficultés surgissent .

##### 2.2.4.1 La compatibilité verbe-nom

On constate que les verbes, même s'ils entrent dans la même catégorie syntaxique, ne sont pas tous en relation avec les mêmes noms, même si ces derniers sont tous concrets. Tout en étant syntaxiquement correctes, ces associations peuvent aller en effet, de l'incompatibilité sémantique totale à une compatibilité présentant des degrés variés de plausibilité (selon des cas de figure plus ou moins particuliers).

Exemples :

Plausibilité proche de zéro :

*monter à la cave, descendre au grenier*

(Nous avons pourtant connaissance d'un immeuble où les "caves" sont à l'étage supérieur)

*se barrer dans le frigo*

*monter sur les spaghettis*

*grimper en Italie*

Plausibilité moyenne à faible:

*entrer dans le frigo*

(envisageable dans une boucherie ou un restaurant de collectivité)

*marcher sur les spaghettis*

(des spaghettis sont tombés par terre)

*grimper sur sa tante*

(un bébé grimpe sur sa tante qui le tient par les bras)

*enjamber la baignoire*

(dans le sens de la largeur, si elle n'est pas contre un mur).

La plausibilité peut varier en fonction de l'évolution des objets. Ainsi pouvons-nous dire aujourd'hui, *Le beurre est dans la porte*, puisque les portes des réfrigérateurs le permettent.

L'élaboration d'automates devra donc prendre en compte ces niveaux de plausibilité et évaluer les rapports de chaque verbe avec chaque nom de chaque grammaire locale, elle-même définie en terme de prédicat sémantique, couvrant lui-même différents champs sémantiques. On verra dans l'automate "Pays" (cf. 2.2.6) que nous n'avons pas retenu certains verbes de mouvement, faisant pourtant partie de notre couche lexicale C1, parce que l'association avec un nom de ce champ (**Pays** ou **Ile**) présentait un faible degré de plausibilité, voire une incompatibilité totale :

-*s'enfoncer en Italie* nous paraît peu plausible. On s'enfonce plutôt dans la jungle ou dans une forêt,

-*marcher sur*, au sens de *aller vers*, est tout à fait incompatible avec un nom d'île, et moins plausible avec un nom de pays (*marcher sur l'Italie*) qu'avec un nom de ville (*marcher sur Rome*) ; ce dernier emploi étant lui-même plus acceptable si le sujet est une armée.

- *monter* et *descendre* s'emploient plutôt avec un nom de ville qu'avec un nom de pays, encore que le cas de ces deux verbes nous semble différent. On comparera *monter à Paris* et (?)*monter en Belgique*  
et *descendre à Toulouse* et *descendre en Espagne*

L'incompatibilité est parfois extrêmement subtile. Prenons l'exemple du verbe *envahir* :

*Les invités ont envahi la place*

*la terrasse*

*le jardin*

*la pièce*

\**le plancher*

\**le sol*

\**le carrelage*

\**le tapis*

\**la moquette*

*La foule a envahi la plage*

*la rue*

*le trottoir*

*la pelouse*

\**le macadam*

\**les pavés*

\**le sable*

\**l'herbe*

Au regard de ces exemples, le verbe *envahir* admet donc des noms incluant le sol dans un espace à trois dimensions (*plage, jardin*, etc.), mais non des noms précisant la nature de ce sol (*sable, pavés, moquette*, etc.), qui eux, ne renvoient qu'à un espace à deux dimensions.

Pourtant, on constatera sans rire, que quelques êtres, à condition qu'ils soient vraiment minuscules (à partir des *souris* ?) et qu'ils ne volent pas (sauf

les *fourmis volantes*), permettent la combinaison de *envahir* avec un nom d'espace à deux dimensions :

*\*Les chats ont envahi le plancher*

*\*Les enfants ont envahi la moquette*

*(?)Les souris ont envahi le sol*

*Les fourmis ont envahi le sable*

*Les puces ont envahi le plancher*

*\*Les mouches ont envahi le sol*

*Les cafards ont envahi le carrelage*

On voit donc ici, que les simples marques distinguant dans un dictionnaire électronique (auquel l'automate peut se reporter), les noms abstraits des noms concrets, serait loin de suffire à une telle application. Les noms concrets devraient encore être regroupés en une multitude de sous-ensembles susceptibles d'être mis en relation avec chacun des verbes concernés.

#### 2.2.4.2 La compatibilité verbe-préposition-nom

En admettant qu'un travail de bénédictin ait permis de constituer un répertoire de tous les noms sémantiquement compatibles avec chacun des verbes concernés par notre prédicat,<sup>9</sup> il ne serait encore pas possible d'établir des règles formelles d'emploi de prépositions.

A titre d'exemple nous nous sommes livrée à une série d'associations extrêmement simples à partir du verbe *être* et d'une série de noms faisant partie d'un champ sémantique réduit et bien délimité : celui des différents lieux habituellement aménagés dans un appartement ou une maison. Nous avons mis en corrélation ce verbe avec le paradigme de ces lieux. Le tableau ci-dessous (Figure 8) met en évidence un certain nombre d'irrégularités qu'a priori on ne s'attendrait pas à voir surgir, dans un champ d'application aussi réduit et aussi homogène.

Il s'agit dans notre exemple de répondre à la question : *Où est Jo ?* en considérant que le lieu est un espace intérieur d'habitation. Nous donnons dans la figure 8 des sorties de réponses dont l'acceptabilité ou l'inacceptabilité semblent impossibles à prévoir.

---

<sup>9</sup> Cf. Travaux de G. Gross sur les classes d'objets (1992).

## Figure 8

*Où est Jo ?*

*à la cuisine / dans la cuisine*

*au salon / dans le salon*

*à la salle de bains / dans la salle de bains*

*à la salle à manger / dans la salle à manger*

*\*à l'entrée / dans l'entrée*

*\*à la chambre / dans la chambre*

*?à la terrasse / \*dans la terrasse / sur la terrasse*

*\*au toit / \*dans le toit / sur le toit*

*au balcon / \*dans le balcon / sur le balcon*

*aux toilettes / dans les toilettes*

*?au séjour / dans le séjour*

*au grenier / dans le grenier*

*à la cave / dans la cave*

*au jardin / dans le jardin*

*à l'appartement / dans l'appartement*

*?au cellier / dans le cellier*

*à la maison / dans la maison*

*à l'étage / \*dans l'étage*

*\*à l'escalier / dans l'escalier*

*\*au palier / \*dans le palier / sur le palier*

Nous ferons plusieurs remarques :

a) Les inacceptabilités émaillant ce paradigme échappent à toute rationalisation. Pourquoi est-il possible de dire *Jo est à la cuisine, au salon, à la salle à manger*, et tout à fait impossible de dire *Jo est à la chambre* ? il est impossible de répondre.

b) Lorsque pour un même nom deux prépositions ou plus sont acceptables, elles ne semblent pas présenter le même degré de plausibilité. Ainsi à la question posée dans notre exemple, nous avons spontanément répondu en premier : *Il est à la cuisine* mais *Il est dans la salle de bains*.

*Jo est à la salle à manger (P-)*

Selon l'application, ou l'évaluation du degré d'écart de pausibilité, on décidera ou non de ramener ce cas au cas précédent.

c) *à* et *dans* présentent des nuances de sens différentes :

*Jo est à l'appartement*

*Jo est dans l'appartement*

Selon l'application, on décidera de la pertinence de l'un ou de l'autre. Si les deux sens sont pertinents à l'application, il faudra envisager les deux phrases, mais il sera impossible de distinguer le bon emploi de l'un ou de l'autre, par un apprenant éventuel.

d) *à* et *dans* présentent des nuances de sens différentes, mais un indice plausibilité différent dans l'absolu, ou dans le champ sémantique traité. Ce cas ne se présente pas dans cette application. Mais on a par exemple :

*partir à la campagne*

*partir en campagne*

Si le champ sémantique traité est celui des "lieux de nature", il sera recommandé de ne pas faire figurer *en campagne*, qui précisément pourra constituer une erreur de la part de l'apprenant, et qu'il sera utile de déceler.

Le cas que nous venons d'envisager est loin d'être exceptionnel. On peut répéter l'expérience avec des champs sémantiques encore plus restreints et avec d'autres verbes, on se heurtera toujours à des irrégularités empêchant d'établir des règles formelles.

Prenons l'exemple du verbe *aller* et de trois séries de noms appartenant respectivement à trois champs sémantiques et répondons à la question : *Où va Jo ?*

Les réponses à cette question mettent en évidence dans le premier groupe, la disparité des prépositions, et dans le troisième groupe, l'arbitraire de l'usage d'une même préposition.

*Où va Jo?*

G1

*sur / dans l'avenue*

*sur la chaussée*

*sur la place*

*sur l'esplanade*

c) L'alternance possible des prépositions recouvre une différence de sens parfois si subtile qu'aucune explication ne réussit à l'épuiser. Ainsi semble-t-il que *dans Nloc* suggère que la personne est occupée à faire quelque chose dans ce lieu. *Il est à la cuisine, Il est aux toilettes* suggère que la personne est en train d'utiliser ces lieux, tandis que *Il est dans la cuisine, Il est dans les toilettes* suggère que la personne est dans ces lieux pour une activité quelconque : réparer quelque chose, nettoyer, se cacher, etc. Cependant, il n'est pas sûr que nous respectons systématiquement cette nuance dans tous nos énoncés. La frontière séparant par exemple, *Il est au salon* de *Il est dans le salon* nous paraît bien ténue. Avec *salle de bains*, cette nuance semble même fonctionner en sens inverse. On comparera :

*Il est dans la salle de bains* et

*Il est à la salle de bains*

d) Dans certains cas pourtant, il y a une différence d'emploi bien nette.

L'alternance *dans l'appartement / à l'appartement* par exemple présente deux sens bien différenciés. *Il est à l'appartement* sous-entend qu'il s'agit d'un appartement en cours d'aménagement et que la personne est allée y faire des travaux ou apporter quelque chose. *Il est dans l'appartement* signifie simplement que la personne est à l'intérieur d'un appartement quelconque.

On voit, que toute tentative de formalisation, même à un niveau aussi élémentaire, devra examiner les combinaisons *Etre Prép Nlieu* une par une. Sauf pour les cas bien tranchés, il faudra en outre renoncer à distinguer, dans le cas de l'application envisagée ci-dessus, l'emploi des prépositions *à* et *dans*. Ils seront prévus par l'automate, mais non différenciés au niveau de l'emploi. Plusieurs stratégies sont cependant à envisager pour l'élaboration de tout automate. Si l'on prend par exemple l'alternance *à/dans*, dans l'application qui nous occupe, quatre cas de figure peuvent se présenter :

a) *à* et *dans* présentent des nuances de sens impossibles à distinguer de façon claire. On acceptera les deux prépositions dans l'automate :

*Jo est au salon*

*Jo est dans le salon*

b) *à* et *dans* présentent des nuances de sens difficiles à distinguer, mais avec un niveau de plausibilité différent :

*Jo est dans la salle à manger (P+)*

*sur le boulevard*  
*sur / dans le chemin*  
*sur la route*  
*sur le terre-plein*  
*dans l'impasse*  
*dans la voie sans issue*  
*dans la rue*  
*dans l'allée*  
*sur le trottoir*  
*sur la chaussée*  
*sur le quai*  
*sur le pont*  
*au / dans le jardin)*

## G2

*à la voiture*  
*\*à l'auto*  
*\*au vélo*  
*\*à la moto*  
*(?)au train*  
*au métro*  
*\*au bus*  
*\*au car*  
*(?)à l'avion*  
*au bateau*

On voit que les éventuelles explications tentant de généraliser l'emploi de *à* par le sens (exemple: *voiture* = trois dimensions, *moto* = deux dimensions) ne parviendraient pas à gérer un champ d'application aussi minuscule.

Lorsque, comme dans le troisième regroupement sémantique envisagé ci-dessous (G3), deux prépositions peuvent entrer en alternance, on remarquera d'une part, que ces deux prépositions n'alternent pas régulièrement, d'autre part qu'elles ne présentent pas une alternance de sens homogène correspondant à l'alternance des deux prépositions.

G3

*Où va Jo?*

*en ville* (Nloc source = périphérie ou banlieue)

*à la ville* (Nloc source = campagne)

*en banlieue*

*\*à la banlieue*

*en mer* (= en bateau)

*à la mer* (=au bord de la mer)

*en campagne* (=sens militaire)

*à la campagne* (=dans une localité de campagne)

*à la montagne* (=dans une localité de montagne)

*En montagne* (=pour une marche, expédition, etc.)

On mesurera la faiblesse des ouvrages grammaticaux, comme outils de formalisation :

**Les inventaires rassemblent pêle-mêle des syntagmes ou des phrases très hétéroclites :**

Ainsi trouve-t-on dans un paragraphe traitant de la "localisation extrinsèque au moyen d'un complément indirect", la série d'exemples suivants (C. Brunot, 1927, p.422) :

a) Il joue au billard tous les jours **au café du Commerce**

b) Il part **pour Paris**

c) Pierre n'est pas **dans son atelier**

d) Une ouvrière **en chambre**

e)le travail **à domicile**

etc.

Or, en a) il s'agit d'un verbe d'action accompagné d'un complément scénique (cf. notre scénario g). En d) et e) il s'agit de noms composés. Seuls les compléments des phrases b) et c) sont des compléments de localisation, encore que l'un exprime un mouvement (complément directionnel, modalité ALLER) et l'autre l'absence de mouvement (complément de lieu, modalité ETRE). C'est toujours ce point de vue approximatif qui conduit l'auteur à noter au chapitre III (Les mouvements, p. 430) :

"L'idée de mouvement n'est pas exprimée... souvent l'idée de la position particulière dans l'espace subsiste seule, et on néglige le mouvement. Par ex. : *entrez dans le salon ; - prenez du sucre dans le sucrier ; ...* Il faut observer

ici que la direction surtout n'a souvent aucunement besoin d'être marquée ... On dit : *aller en Amérique, entrer dans la tranchée*, comme : *être dans la tranchée*."

Les exemples choisis, soit pour montrer l'irrationalité de l'emploi d'une préposition, soit au contraire, pour justifier une explication par le sens, ne sont jamais envisagés d'un point de vue syntaxique et systématique :

Ainsi C. Brunot remarque-t-il, à propos de certaines classifications reposant sur l'évolution historique du sens des prépositions : "...l'enchaînement logique des sens n'a pas cette logique rigoureuse, que des classifications rigoureuses prétendent lui donner... Ce n'est pas seulement d'une époque à l'autre que des différences de sens se constatent. De nos jours, on écrit : *donner un coup de poing dans le dos*, mais : *une gifle sur la figure*. On dit *se promener dans l'avenue*, ou *sur l'avenue*, mais non *dans les boulevards*, ni *sur la rue*" (pp. 413-414).

Au contraire, G. Gougenheim (1963) note dans un paragraphe traitant de l'emploi des prépositions *dans /en /sur* (p.310) :

"L'opposition *dans /en /sur* est une opposition de sens : *dans/en* marquent l'intériorité, *sur* la superposition. Si le concept est considéré comme un contenant, on emploie *dans /en* ; s'il est considéré comme une surface, on emploie *sur*.

...Voies publiques : sont considérés comme des contenant les concepts de *rue, avenue, allée, chemin* : *dans une rue, dans un chemin*, comme des surfaces les concepts de *route, cours, boulevard, chaussée, place, pont* : *sur une route, sur une place*"

L'auteur envisage bien ici un champ sémantique, mais comme il s'agit pour lui de justifier le sens présumé des deux prépositions, il distingue deux sous-ensembles sémantiques très discutables dans lesquels *avenue* serait un contenant et *boulevard* une surface. Il suffit d'envisager ces mêmes mots dans diverses configurations syntaxiques : expressions figées ou phrases, pour constater que ces mêmes exemples accepteraient, et même demanderaient la préposition attribuée au sous-groupe inverse.

Exemples :

*Il y a un monde fou sur l'avenue*

*Il y a un monde fou dans l'avenue*

*sur le chemin de l'école*  
*\*dans le chemin de l'école*

*Cet appartement donne sur la rue*  
*\*Cet appartement donne dans la rue*  
(sauf si l'appartement est de plein-pied avec la rue)

**Aucun des ouvrages de grammaire n'envisage l'ensemble du paradigme prépositionnel pour un exemple donné.**

Ainsi, dans l'ouvrage en question, pour ce champ sémantique seules les prépositions *dans*, *en* et *sur* sont envisagées. Or, on ne voit pas pourquoi la préposition *à* ne serait pas à prendre également en compte :

*Cette rue conduit à une place*  
*Vous arriverez à un carrefour*

On trouve un peu plus loin, toujours pour justifier la même alternance de sens des deux prépositions *dans* et *sur*, l'exemple de la série "espaces découverts" : " ... *dans une cour, dans un jardin... sur le territoire de à côté de dans le territoire de*", etc. Or dans cette dernière série l'alternance *à/dans* (à côté de *sur*), qui n'est pas même évoquée, serait plus pertinente notamment pour les deux premiers exemples.

En répondant à la question : *Où sont les enfants ?* , on aurait :

<i>dans une cour</i>	<i>dans un jardin</i>	<i>dans un parc</i>
<i>dans la cour</i>	<i>dans le jardin</i>	<i>dans le parc</i>
<i>*à la cour</i>	<i>au jardin</i>	<i>au parc</i>
<i>*à une cour</i>	<i>*à un jardin</i>	<i>*à un parc</i>

On peut constater que chaque élément de cet ensemble sémantiquement homogène, fonctionne différemment. On remarquera notamment que :

- *à la cour* n'est pas acceptable ici (= à la cour du roi)
- *à un parc* n'est pas acceptable, sauf s'il s'agit d'un *parc d'attraction*
- *au jardin* peut renvoyer à un jardin privé ou public, tandis que *dans le jardin* renvoie à un jardin privé
- *au parc* renvoie à un parc public et *dans le parc* à un parc privé.

On voit donc l'utilité des grammaires locales ; elles demandent d'envisager tout phénomène de langue dans la chaîne syntaxique minimale constituée par la phrase.

#### 2.2.4.3 Le paradigme verbal

Pour en revenir à notre processus, une fois franchie l'étape consistant à définir, pour un verbe donné, toutes les associations *Prép Nloc* d'un champ sémantique donné, serait-il possible de procéder en sens inverse et d'attribuer le paradigme *Prép Nloc* d'un champ sémantique établi pour un verbe, à d'autres verbes du même champ ?

Le tableau établi plus bas (figure.9), à partir du verbe *aller* et du champ sémantique généralement abordé dans les manuels de français langue étrangère (déplacement dans une ville ou un pays), montre précisément que la distribution des prépositions varie selon le verbe employé, même lorsque, comme c'est le cas ici, les verbes évoquent le déplacement de *Nhum* dans le même sens.

Une digression est ici nécessaire à propos de la marque "sc" figurant à certains endroits du tableau. Le segment *Prép Nloc*, représente dans notre prédicat un complément "directionnel":

- (a) *Jo sort de sa chambre.* (directionnel d'origine)
- (b) *Jo rentre dans sa chambre* (directionnel de destination)

Or, il se trouve que dans certains cas, un complément de nature différente a exactement la même structure syntaxique, nous voulons parler du complément "scénique" :

- (c) *Jo marche dans sa chambre* (scénique)

Dans ce dernier cas en effet, la chambre constitue le décor à l'intérieur duquel se déroule le procès évoqué par le verbe, plutôt que la destination ou l'origine du procès.

La traduction en anglais de ces différentes phrases nécessiterait en l'occurrence l'utilisation de prépositions différentes :

- (a) *Jo comes from his room*

(b) *Jo goes to his room*

(c) *Jo walks in his room*

Pour la distinction de ces trois types de compléments avec l'ensemble des verbes de mouvements, d'action, et des verbes statiques on se reportera à J.P. Boons, A. Guillet, C. Leclère 1976).

Nous présentons en figure 9, un tableau utilisant des verbes à complément directionnel de destination, entrant donc dans la catégorie des verbes exprimant la modalité ALLER du déplacement. Les variations observées dans ce tableau montrent que celles-ci peuvent même jouer sur un seul item du paradigme des noms d'un champ sémantique. Un exemple extrêmement sommaire suffira à illustrer ce propos :

Figure 9

Verbe	à l'école	vers l'école	pour l'école
<i>aller</i>	+	+	-
<i>se rendre</i>	+	-	-
<i>partir</i>	+	+	+
<i>foncer</i>	+	+	-
<i>filer</i>	+	+	-
<i>s'acheminer</i>	-	+	-
<i>se casser</i>	+	-	-
<i>se barrer</i>	+	-	-

Il ressort de tout ceci, que la réalisation d'un automate relevant de ce prédicat exige en réalité, que l'on descende du niveau de la phrase (ou niveau théorique), au niveau de l'énoncé (ou niveau de l'actualisation langagière). Ce processus méthodologique tendrait à prouver que la "génération" naturelle d'une préposition dans un acte langagier met en branle un mécanisme cognitif d'une grande complexité.

### 2.2.5. La sélection d'une préposition : un processus de génération complexe

D'un point de vue méthodologique, la sélection de la modalité syntaxique (prépositionnelle / non prépositionnelle) peut être décidée à partir du seul verbe, et donc dès le niveau théorique des tables syntaxiques : Ainsi peut-on décider de la modalité non-prépositionnelle entraînée par le choix du verbe *atteindre* (*atteindre Paris*). Mais comme le montre le processus plus complexe de la sélection de la préposition, le choix de la modalité syntaxique se situe en réalité, plus en aval dans la génération " naturelle" de la phrase.

Le processus d'élaboration d'un automate, dont nous donnons un récapitulatif ci-dessous en figure.10, tend en effet à laisser penser que le parcours menant de l'intention signitive à l'acte significatif, comporte une série d'aiguillages successifs menant à l'"output" final ou énoncé. Les différentes étapes de la formalisation de l'emploi de la préposition dans le cadre de notre prédicat montrent au moins 4 niveaux de discrimination enchâssés les uns dans les autres, allant du plus abstrait au plus concret. Ce n'est qu'au terme de ce schème que la préposition fournit sa forme actualisée dans une réalisation langagière (cf. figure 11) :

- Niveau 1 : Modalité de déplacement.

ALLER vs VENIR

- Niveau 2 : Emergence du champ sémantique visé

"Pays", "Espace privé", "Espace public ouvert" etc.

- Niveau 3 : Choix du verbe et de la modalité syntaxique

*V Prép* (modalité syntaxique prépositionnelle) ; *V* (modalité syntaxique non-prépositionnelle). Cette répartition est déjà effectuée dans les tables de verbes.

- Niveau 4 : choix de la préposition par la mise en rapport du verbe et du nom, et production de la combinaison sémantico-linguistique *Verbe Nloc* ou *Verbe Prép Nloc*.

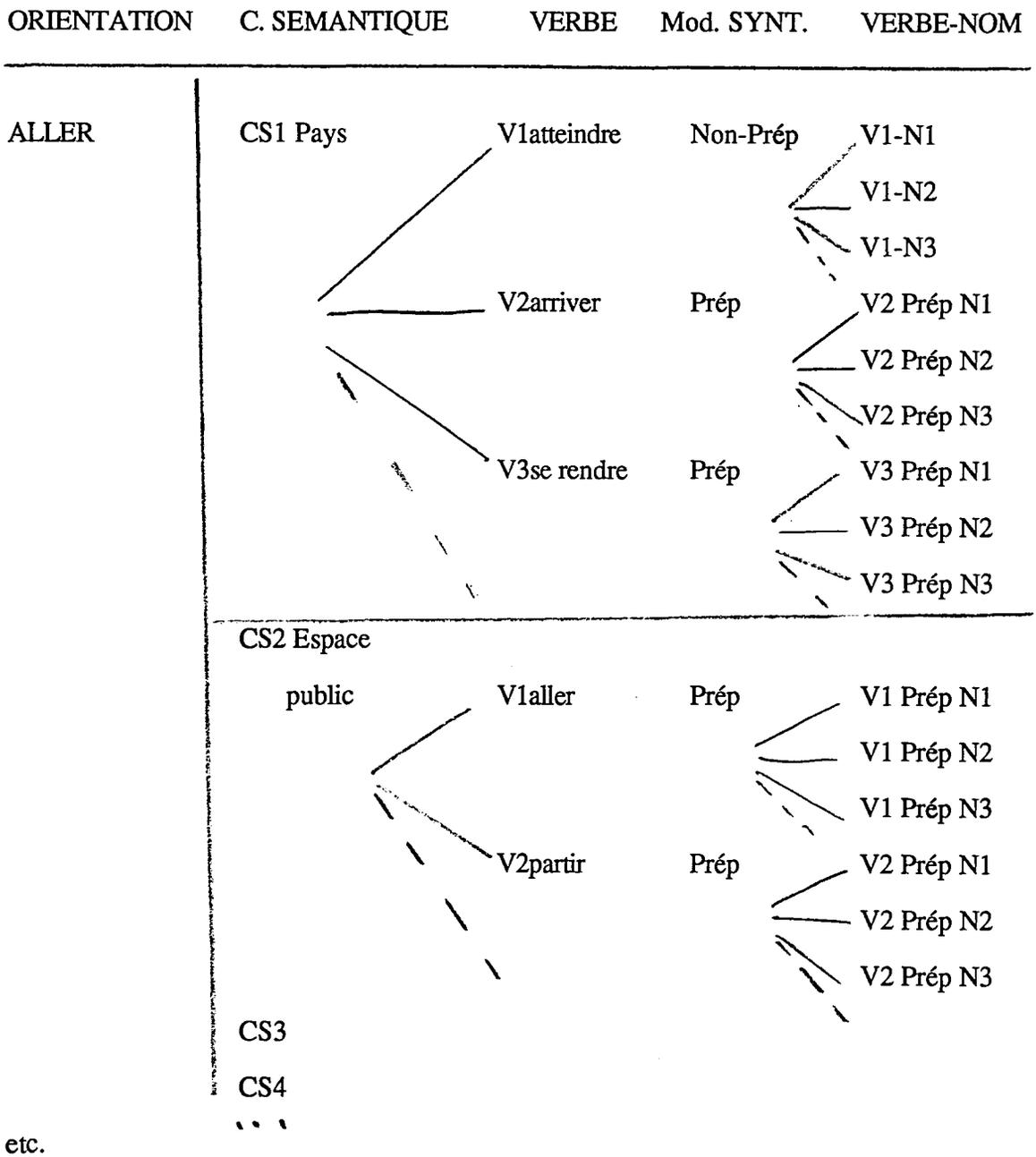
Figure 10

Récapitulatif de la méthodologie

CHAMP	VERBES	PREPOSITIONS	PREP-NOM
CS1 Espace privé	V1 Aller	à, dans, vers, sur	<p>à / dans</p> <ul style="list-style-type: none"> <li>la cuisine</li> <li>le salon</li> <li>les toilettes</li> <li>le jardin</li> <li>la cave</li> <li>le grenier</li> <li>le cellier</li> </ul> <p>dans</p> <ul style="list-style-type: none"> <li>le séjour</li> <li>l'entrée</li> <li>la chambre</li> <li>l'escalier</li> <li>le couloir</li> </ul> <p>à / dans</p> <ul style="list-style-type: none"> <li>l'appartement</li> <li>la maison</li> </ul> <p>sur</p> <ul style="list-style-type: none"> <li>le toit</li> <li>le palier</li> </ul> <p>dans / sur</p> <ul style="list-style-type: none"> <li>la véranda</li> </ul> <p>sur / à</p> <ul style="list-style-type: none"> <li>la terrasse</li> </ul> <p>sur / à / dans</p> <ul style="list-style-type: none"> <li>le balcon</li> </ul>
CS2	V1		
CS3	V1		
etc.			

Figure 11

Processus de génération de la préposition dans le prédicat défini



Comme nous l'avons expliqué à propos du problème du sens, tout "acte signitif" est porté par une "intention", qui s'enracine dans une combinaison de matériaux lexico-syntaxiques. C'est à notre avis, autour de cette "intention", que des blocs (?) sémantico-syntaxiques verbaux et nominaux émergent et s'ajustent jusqu'à faire coïncider un verbe et un nom, et déterminer du même coup la préposition contribuant à l'actualisation de l'intention signitive. On voit ici que les informations fournies par les grammaires sont loin de correspondre au type de données nécessaires à la reconstitution d'un tel processus. On mesure également l'ampleur de la tâche qui reste à faire, puisque précisément tout travail de représentation formelle nécessite de circonscrire ces "blocs", et d'établir les corrélations syntaxico-sémantiques qui les unissent.

Cette théorie portant sur un mécanisme cognitif est à l'heure actuelle, comme tout ce qui concerne ce domaine, invérifiable. Il reste néanmoins, que ce mécanisme théorique est en cohérence avec le processus de formalisation de données syntaxiques empiriques, en vue de la reconnaissance de productions langagières.

Il reste également que dans le domaine de l'intelligence artificielle, le processus menant à la génération d'une action ("Aller chez le dentiste", "Prendre l'avion", etc.), été décrit par R.C. Schank comme une suite d'opérations discriminatoires menant, par étapes, à partir d'une représentation générale, abstraite et constamment remaniée, elle-même élaborée à partir des expériences auxquelles nous sommes confrontés, au but (ou *goal*) final. Nous "descendons" alors des schémas de représentation les plus abstraits (les *MOPs* ou *memory organisation packet*, voire les *Méta-MOPs*) aux schémas les plus concrets (les *scripts*) en passant par les *scenes*. (R.C. Schank, 1982) pour aboutir à la réalisation d'une action. Bien que R.C. Schank ne s'intéresse qu'à l'aspect extra-linguistique du processus cognitif, nous pensons que ce schème discriminatoire, menant par paliers de la représentation la plus abstraite à la réalisation concrète concerne également la production strictement linguistique. Le but du processus (ou *goal*), n'est plus une action ou un schéma conceptuel d'énoncé, mais la production d'un énoncé authentique. Ces deux processus infra-langagier et intra-langagier sont du reste difficilement imaginables déconnectés l'un de l'autre. Dans tout déroulement de *scripts*, ou de *scènes*, s'inscrivent inévitablement des déroulements "de *scripts* syntaxiques".

Bien qu'il soit difficile actuellement de donner des preuves dans le domaine de la connaissance, tout pousse à penser en tout cas, que ce schéma général de mémorisation menant d'un stockage des connaissances et des expériences (linguistiques et extra-linguistiques) sous forme abstraite, à un décompactage permettant de cheminer dans un arbre discriminatoire à X niveaux et à X branches, et de générer un acte (linguistique ou extra-linguistique) paraît vraisemblable.

Ce point de vue rejoint les observations que nous avons faites au niveau du lexique : de même que dans notre lexique interne, le sens d'un mot est extrait d'un traitement mystérieux de l'ensemble des occurrences et des contextes dans lesquels nous l'avons lu ou entendu, de même, les grammaires par lesquelles ces mots entrent en relation de façon "acceptable" sont-elles, pensons-nous, stockées de façon "codée" ou "compactée" pour s'explicitier à chaque production linguistique.

#### **2.2.6. Exemples d'automates**

Nous présentons deux réalisations d'automates correspondant aux données suivantes :

1° automate : ALLER / PREP / PAYS, ILES

scénario a)

modalité de déplacement ALLER

champ sémantique PAYS, ILES

modalité syntaxique prépositionnelle

2° automate : ALLER / PAYS, ILES

scénario a)

modalité de déplacement ALLER

champ sémantique PAYS, ILES

modalité syntaxique non-prépositionnelle

Le premier automate, ou graphe, correspond à la structure **N V Prép N**

Le second automate, ou graphe, correspond à la structure **N V N**

Les cases grises correspondent aux **sous-graphes** des noms de pays ou d'îles. (On se reportera à l'annexe 12 pour la consultation de l'ensemble de ces graphes).

## 1) NVPrép N

### **PaysDetZ**

regroupe les noms de pays à déterminant 0. Exemple : *Panama*

### **PaysPlur**

regroupe les noms de pays au pluriel .Exemple : *Etats-Unis*

### **PaysMascCons**

regroupe les noms de pays masculins commençant par une consonne.

Exemple : *Brésil*

### **PaysVoc**

regroupe les noms de pays commençant par une voyelle. Exemple : *Irak*.

On remarquera qu'*Israël* a un fonctionnement à part. On comparera :

*aller en Irak / partir pour l'Irak / approcher de l'Irak*

*aller en Israël / \*partir pour l'Israël / \*approcher de l'Israël*

*Cuba* fonctionnerait comme *Israël*, mais il s'agit d'un nom de pays qui est aussi une Île, et nous avons regroupé ces cas -là dans les îles. *Israël* serait-il le seul pays à fonctionner comme une île ?

### **PaysFemCons**

regroupe les noms de pays féminins commençant par une consonne.

Exemple : *Belgique*

### **IleDetZ**

regroupe les îles qui ne prennent pas, ou peuvent ne pas prendre de déterminant. Exemple : *Madagascar, Malte*

### **IlePrep**

regroupe une série d'îles demandant le déterminant, fonctionnant avec une préposition autre que *dans*.

### **IleFemSing**

regroupe les noms d'îles au féminin. Exemple : *Sardaigne*

Une série d'îles fonctionnent avec la préposition *dans* lorsqu'elles entrent en composition avec certains verbes. On trouvera donc un automate annexe **NVDansile** regroupant lui-même trois sous-graphes (**IleDans1**, **IleDans2** et **IleDans3**) correspondant à 3 des groupes d'Îles dont le fonctionnement est particulier à chacun.

On pourra noter que l'automate **IlePrep** diffère de **IleDans3** par la différence d'emploi suivante :

*Il va à la Réunion*

*\*Il va dans la Réunion*

## 2) N V N

On verra que dans ce second automate le nombre de verbes est beaucoup plus restreint tandis que les sous-graphes restent les mêmes que dans l'automate **N V PrepN**.

## Conclusion

Ces premiers résultats ouvrent des pistes prometteuses, notamment à partir de l'établissement de nos couches lexicales. Nous comptons donc étendre et développer ce travail dans deux directions complémentaires :

- l'affinement et multiplication de lexiques de références pour les mots simples, les mots composés et les expressions figées. La combinaison de ces sous-lexiques de référence et des lexiques de spécialité qui seront construits en fonction des applications augmenteront la précision et la souplesse du paramétrage.
- l'application de tous les lexiques à des textes permettra l'élaboration d'un corpus de textes standards et spécialisés.

Notre projet est de constituer progressivement une **banque lexicale du français** susceptible de fournir à tout utilisateur différents types de lexiques et de corpus de textes destinés au traitement automatique. Cette banque pourra servir de support au développement de tels lexiques pour d'autres langues (allemand, anglais, espagnol, italien), par le biais du réseau de laboratoires européens Relex. Les matériaux qui y sont développés correspondent en effet aux normes du LADL, qui constitue leur centre ressource. Nous pourrions alors créer un **fond lexical européen** pour le traitement automatique des langues naturelles. Du fait de ce réseau les méthodes et les matériaux produits seront donc harmonisés.

Parallèlement, nous nous efforcerons de multiplier les expériences de confrontation dictionnaires / textes. Ceci devrait nous amener à long terme à la constitution d'un corpus représentant le **texte de référence du français** (cf. J. Sinclair 1987).

D'autre part, il sera possible de se servir, à l'occasion d'une application pointue, des outils étudiés dans notre deuxième partie. On pourrait par exemple envisager, que la détermination du lexique d'une application s'accompagne du développement de grammaires locales ad hoc. De telles

grammaires permettront, entre autres, de lever les ambiguïtés nombreuses que les systèmes d'étiquetage automatique engendrent inévitablement. L'analyse de la langue du système interactif mis en oeuvre dans l'application n'en serait que plus performante.

## Bibliographie

Abbé Prévost, 1755. *Manuel Lexique ou Dictionnaire portatif des mots françois dont la signification n'est pas familière à tout le monde*. Didot. Paris

Baudot, Jean 1990. *Dictionnaire de Fréquence des Mots du Français*. Université de Montréal. Publié en 1992 : *Fréquences d'utilisation des mots en français écrit contemporain*. Presses de l'université de Montréal. 431p.

Barril, P. ; de La Passardière, B. ; Le Calvez F. ; Madaule, F. ; Poc, M-M.; Urtasun M. 1985. *Etude comparative de systèmes et langages d'auteur*. Rapport MASI n° 73. Laboratoire Méthodologie et Architecture des Systèmes Informatiques. CNRS UA 18. Université P. et M. Curie. Paris.

Besse H. et Porquier R. 1984. *Grammaires et didactiques des langues* Paris. Hatier- Crédif.

Boons, Jean- Paul ; Guillet, Alain ; Leclère Christian 1976 a. *La structure des phrases simples en français. I Constructions intransitives*, Genève : Droz

Boons, Jean-Paul ; Guillet Alain ; Leclère Christian 1976 b. *La structure des phrases simples en français, II Constructions transitives*, Paris : Rapport de recherches du LADL, n°6

Brunet, E. 1981 *Le vocabulaire français de 1789 à nos jours*. Slatkine. Genève.

Brunot, Ferdinand 1926. *La Pensée et la Langue*. 2ème éd. Masson et Cie éd. Paris

Changeux, J.P. 1983 *L'homme neuronal*. Fayard

Chaurand, J. et Lerat P. 1981 "Français fondamental et français d'aujourd'hui." *Le Français dans le Monde* n°159. Hachette.  
Clémenceau, David 1992. "Analyse d'un corpus de référence". Rapport LADL / Université de Paris 7

Courtois, Blandine 1989-1990. *Dictionnaire électronique du LADL pour les mots simples du français* ; DELAS VO6/2, Rapport technique du LADL n°20, Université Paris 7 : LADL

Courtois, Blandine 1990. "Un système de dictionnaires électroniques pour les mots simples du français", *Langue française* n° 87, Paris : Larousse

Courtois, Blandine et Silberztein, Max éds. 1990. Dictionnaires électroniques du français, *Langue française* n°87, Paris : Larousse.

Coutier, Martine. 1988. "Vocabulaires spéciaux et dictionnaires d'usage. Vocabulaire de la psychiatrie". *BULAG* (Bulletin de linguistique générale et appliquée) n° 14-1988. Université de Franche-Comté. Besançon.

Danlos, Laurence 1985. *Génération automatique de textes écrits* Paris: Masson.

- Danlos, Laurence 1980. *Représentations d'informations linguistiques : constructions "N être Prép X"*, Thèse de 3ème cycle, LADL, Université Paris 7.
- Debyser, Francis 1973, 1974. "Simulation et réalité dans l'enseignement des langues vivantes". *Le Français dans le Monde* n° 97 (Juin 1973) et n° 105 (Juin 1974), Paris : Hachette
- Debyser, Francis. 1977. "Le choc en retour du Niveau 2". *Le Français dans le monde*. n° 133. Nov-déc 1977. Hachette. Paris.
- Debyser, Francis 1987. "De l'imparfait du subjonctif aux méthodes communicatives", *Le Français dans le Monde* n° 209, Paris : Hachette
- Debyser, Francis (s.d.). *Les constructions verbales* . BELC. Paris
- Dictionnaire des Fréquences*. Etudes statistiques sur le vocabulaire français. Vocabulaire littéraire des XIXème et XXème siècles. (4 vol.). (1971). CNRS. Centre de recherche pour un trésor de la langue française, Nancy. Didier
- Delay, Jean 1942 *Les maladies de la mémoire*. Presses Universitaires de France
- Dennighaus, Friedhelm 1976 "Der kontrollierte Erwerb eines potentiellen Wortschatzes im Fremdsprachunterricht", in *Praxis des Neusprachlichen Unterrichts* 23,1.
- Dubois, Jean 1978, 1979 *Dictionnaire du français langue étrangère*. Niveau 1 (1978), Niveau 2 (1979). Larousse. Paris
- Elia, Annibale. 1984 *Le verbe italien, Les complétives dans les phrases à un complément*, Bari-Paris : Schena-Nizet.
- Frege, Gottlob 1976 *Ecrits logiques et philosophiques* (traduction de Claude Imbert). Editions du Seuil. Paris
- Le Français fondamental (1° et 2° Degrés)* 1972. Institut National de Recherche et de Documentation Pédagogiques. Paris.
- Galisson, Robert. 1971. *Inventaire thématique et syntagmatique du français fondamental*. Larousse. Paris
- Galisson, Robert 1988 "... le vocabulaire revient !". *Reflète* n° 27. Alliance française/Crédif/Hatier. Paris.
- Garrigues, Mylène éd. 1989. *Logiciels éducatifs de français. Recensement analytique* Sèvres. CIEP.
- Garrigues, Mylène. 1987. *Vidéodisque et enseignement des langues*. Mémoire de DESS "Applications de la télématique". Université de Paris 7.
- Gougenheim, G., Rivenc, Michéa et Sauvageot 1967. *L'élaboration du français fondamental*. Didier. Paris.
- Gougenheim, G 1962 *Système grammatical de la langue française*. éditions d'Artrey. Paris;

- Greimas, Algirdas J. 1966 *Sémantique structurale*. Larousse. Paris
- Gross Gaston, 1990 "Définition des noms composés dans un lexique-grammaire". *Langue Française* n°87. Septembre 1990. Larousse.
- Gross, Gaston 1988. "Degré de figement des noms composés" *Langages* N° 90, Paris: Larousse.
- Gross Gaston, 1992 *Classes d'objets et description des langues*. Rapport LLI. Université de Paris 13.
- Gross, Maurice 1975. *Méthodes en syntaxe* Paris : Hermann
- Gross, Maurice 1981. "Les bases empiriques de la notion de prédicat sémantique, Formes syntaxiques et prédicats sémantiques", A. Guillet et C. Leclère éd., *Langages* N° 63, Paris : Larousse.
- Gross, Maurice 1989. "Qu'est-ce qu'un mot ?", *Le Français dans le Monde* N° spécial Recherches et Applications : Août-Septembre 1989, Paris : Hachette.
- Gross, Maurice 1989. "La construction de dictionnaires électroniques". *Annales des Télécommunications*, Tome 44, N° 1-2
- Gross, Maurice, 1990. *Grammaire transformationnelle du français. 3- Syntaxe de l'adverbe*, Asstril, Paris.
- Gross, Maurice et Leclère, Christian 1990. *Présentation des classes du lexique grammaire des verbes simples et composés*, Rapport de Recherches du LADL.
- Guillet, Alain et Leclère, Christian. 1992. *La structure des phrases simples en français*. Constructions transitives locatives. Droz, Genève-Paris.
- Harman, Gilbert, 1988. *Philosophical issues in cognitive science*. CSL report 14. February 1988. Cognitive Science Laboratory. Princeton University.
- Husserl, Edmund *Idées directrices pour une phénoménologie* traduit de l'allemand par Paul Ricoeur. Gallimard 1950
- Husserl, Edmund *Recherches logiques*. (3 tomes). Traduction par H. Elie, A.L. Kelbel et R. Schérer. PUF, 1969-1974.  
 - Tome 1 : *Prolégomènes à une logique pure*  
 - Tome 2 : *Recherches pour la phénoménologie et la théorie de la connaissance*  
 - Tome 3 : *Elements d'une élucidation phénoménologique de la connaissance*
- Husserl, Edmund *Logique formelle et transcendantale*. PUF Paris. 1974
- Labelle, François. 1988. "Micro-ordinateur : pour un enseignement interactif". *Le Français dans le Monde*. N° spécial Recherches et applications. Août-septembre 1988, M. Garrigues éd. Paris : Hachette.

Landauer, Thomas K. 1986. "How much do people remember ? Some estimates of quantity of learned information in long-term memory" *Cognitive Science* 10, 477-493.

Laporte, Eric 1988. "La reconnaissance des expressions figées lors de l'analyse automatique", *Langages* N° 90, Paris: Larousse.

Laporte, Eric 1988. *Méthodes algorithmiques et lexicales de phonétisation du français*, thèse de doctorat, Université Paris 7 : LADL

Leclère, Christian 1990. "Organisation du lexique-grammaire des verbes français", *Langue française* N°87, Paris: Larousse.

Le Kien, Van 1987. *Génération automatique de l'accord du participe passé*. thèse de doctorat, Université Paris 7: LADL

Maingueneau, Dominique 1976. *Initiation aux méthodes de l'analyse du discours* Paris: Hachette Université.

Mathieu-Colas, Michel 1988. "Variations graphiques des noms composés dans le Petit Larousse et le Petit Robert", *Linguisticae Investigationes*, Vol. I, N° 2, Amsterdam- Philadelphie: J. Benjamins Pub. Co.,

Mathieu-Colas, Michel 1990. "Orthographe et Informatique: remarques sur l'établissement d'un dictionnaire électronique de variantes graphiques", *Langue française* N°87, Paris: Larousse.

Matoré, Georges 1963. *Dictionnaire du vocabulaire essentiel*. Larousse. Paris

Maurel, Denis, 1989. *Reconnaissance de séquences de mots par automates. Adverbes de date du français*. Thèse de doctorat. Université de Paris 7. Paris LADL

Melka Teichroew, Francine J. 1989 *Les notions de réception et de production dans le domaine lexical et sémantique. Etude exploratoire*. Publications Universitaires Européennes. Editions Peter Lang SA, Berne.

*Un Niveau-Seuil*. Coste Daniel et al. Conseil de l'Europe. Strasbourg 1976.

Phal, André. 1971 *Vocabulaire général d'orientation scientifique*. Crédif. Didier. Paris.

Perrin, Dominique 1989. "Automates et algorithmes sur les mots", *Annales des Télécommunications* , Tome 44, N° 1-2

Pottier B, 1967. "Présentation de la linguistique. Fondements d'une théorie", *Tra Li Li* 5 (I) Strasbourg.

*Pour une introduction aux langages d'auteur 1983 : EPI*. Supplément au bulletin trimestriel. Numéro spécial. Septembre 83. Association Enseignement public et Informatique. Kremlin Bicêtre.

Richard de Radonvilliers J. Bte 1845 *Enrichissement de la langue française. Dictionnaire de mots nouveaux*. Léautey. Paris

Rivenc, Paul. 1979. "Le français fondamental vingt-cinq ans après". *Le Français dans le monde* n° 148 Oct. 1979. Hachette. Paris

Roche, Emmanuel 1992 "Une représentation par automate fini des textes et propriétés transformationnelles de verbes". *Linguisticae Investigationes* vol. 16 : 1 John Benjamins. Amsterdam Philadelphie.

Rolland, Jean Claude, 1971 Dictionnaire de base du FLE. CIEP.

Rodriguez, Nicole avec la collaboration de Valentin, Danièle 1985 : *Concevoir un didacticiel avec Arlequin* Paris: Cedic- Nathan.

Ruwet, Nicolas 1966 *Introduction à la grammaire générative* Paris : Plon

Salkoff, Morris 1973. *Une grammaire en chaîne du français* Paris : Dunod.

Salkoff, Morris 1979. *Analyse syntaxique du français: Grammaire en chaîne* Amsterdam - Philadelphie: J. Benjamins Pub. Co.

Schank, Roger C. 1982 *Dynamic memory. A theory of reminding and learning in computers and people*. Cambridge University Press. New-York.

Schank, Roger C. 1986 *Explanation Patterns. Understanding mechanically and creatively*. Lawrence Erlbaum Associates, Publishers. Hillsdale, New Jersey.

Schank, Roger C. 1990. *Tell me a story. A new look at real and artificial memory*. Macmillan Publishing Company. New-York

Silberztein, Max 1989. *Dictionnaires électroniques et analyse lexicale automatique*, thèse de doctorat, Université Paris 7, Paris: LADL

Silberztein, Max 1990. "Le dictionnaire électronique des mots composés", *Langue française* N° 87, Paris: Larousse.

Silberztein, Max. 1992. *Dictionnaires électroniques et analyse automatique de textes : le système INTEX*. Université de ParisVII-Institut Blaise Pascal-LADL.

Sinclair, J. 1987 *Cobuild English Language Dictionary*. Collins ELT. Londres, Glasgow.



activateur biologique	un/N1/A31/ms; - +	activité psychologique	une/N21/A31/fs; - +
activateur embryonnaire	un/N1/A31/ms; - +	activité publique	une/N21/A46/fs; - +
activateur enzymatique	un/N1/A31/ms; - +	activité réflexe	la/fs; - -
activation corticale	une/N21/A76/fs; - +	activité respiratoire	la/fs; - -
activation électrique	une/N21/A31/fs; - +	activité salariée	une/N21/A32/fs; - +
activation neutronique	une/N21/A31/fs; - +	activité séismique	une/N21/A31/fs; - +
activation thermique	une/N21/A31/fs; - +	activité sexuelle	une/N21/A40/fs; - +
activisme royaliste	un/N1/A31/ms; - +	activité sismique	la/fs; - -
activiste basque	un/N31/A31/ms; + +	activité sociale	une/N21/A76/fs; - +
activité aérienne	une/N21/A41/fs; - +	activité solaire	une/N21/A31/fs; - +
activité agricole	une/N21/A31/fs; - +	activité solaire	la/fs; - -
activité antivirale	une/N21/A76/fs; - +	activité spatiale	une/N21/A76/fs; - +
activité apériodique	une/N21/A31/fs; - +	activité spécifique	une/N21/A31/fs; - +
activité architecturale	une/N21/A76/fs; - +	activité sportive	une/N21/A38/fs; - +
activité artistique	une/N21/A31/fs; - +	activité statutaire	une/N21/A31/fs; - +
activité associative	une/N21/A38/fs; - +	activité stellaire	une/N21/A31/fs; - +
activité audiovisuelle	une/N21/A40/fs; - +	activité syndicale	une/N21/A76/fs; - +
activité bancaire	une/N21/A31/fs; - +	activité tectonique	une/N21/A31/fs; - +
activité biologique	une/N21/A31/fs; - +	activité téléphonique	une/N21/A31/fs; - +
activité cardiaque	la/fs; - -	activité tellurique	une/N21/A31/fs; - +
activité cérébrale	la/fs; - -	activité théâtrale	une/N21/A76/fs; - +
activité cinématographique	une/N21/A31/fs; - +	activité universitaire	une/N21/A31/fs; - +
activité commerciale	une/N21/A76/fs; - +	activité urbaine	la/fs; - -
activité comportementale	une/N21/A76/fs; - +	activité vénale	une/N21/A76/fs; - +
activité créatrice	une/N21/A36/fs; - +	activité virale	une/N21/A76/fs; - +
activité culturelle	une/N21/A40/fs; - +	activité volcanique	une/N21/A31/fs; - +
activité cynégétique	une/N21/A31/fs; - +	activité volontaire	la/fs; - -
activité diplomatique	une/N21/A31/fs; - +	activités culturelles	des/fp; - -
activité diurne	une/N21/A31/fs; - +	activités manuelles	des/fp; - -
activité économique	une/N21/A31/fs; - +	activités parascolaires	des/fp; - -
activité éditoriale	une/N21/A76/fs; - +	activités professionnelles	des/fp; - -
activité électrique	une/N21/A31/fs; - +	activités sportives	des/fp; - -
activité enzymatique	la/fs; - -	actographie qualitative	une/N21/A38/fs; - +
activité fantasmatique	une/N21/A31/fs; - +	actographie quantitative	une/N21/A38/fs; - +
activité ferroviaire	une/N21/A31/fs; - +	actrice dramatique	une/N36/A31/fs; - +
activité générale	la/fs; - -	actualité cinématographique	la/fs; - -
activité hospitalière	une/N21/A42/fs; - +	actualité culturelle	une/N21/A40/fs; - +
activité immobilière	une/N21/A42/fs; - +	actualité discographique	la/fs; - -
activité industrielle	une/N21/A40/fs; - +	actualité économique	la/fs; - -
activité informatique	une/N21/A31/fs; - +	actualité éditoriale	une/N21/A76/fs; - +
activité Intellectuelle	la/fs; - -	actualité judiciaire	une/N21/A31/fs; - +
activité islamique	une/N21/A31/fs; - +	actualité littéraire	la/fs; - -
activité journalistique	une/N21/A31/fs; - +	actualité médicale	une/N21/A76/fs; - +
activité laitière	une/N21/A42/fs; - +	actualité météorologique	une/N21/A31/fs; - +
activité libidinale	la/fs; - -	actualité musicale	une/N21/A76/fs; - +
activité littéraire	la/fs; - -	actualité politique	la/fs; - -
activité ludique	une/N21/A31/fs; - +	actualité sociale	la/fs; - -
activité magmatique	une/N21/A31/fs; - +	actualité sonore	une/N21/A31/fs; - +
activité magnétique	une/N21/A31/fs; - +	actualité sportive	la/fs; - -
activité médicale	une/N21/A76/fs; - +	actualités cinématographiques	les/fp; - -
activité ménagère	une/N21/A42/fs; - +	actualités régionales	les/fp; - -
activité mentale	la/fs; - -	actualités télévisées	les/fp; - -
activité métabolique	une/N21/A31/fs; - +	acuité auditive	la/fs; - -
activité militaire	une/N21/A31/fs; - +	acuité sensorielle	une/N21/A40/fs; - +
activité militante	une/N21/A32/fs; - +	acuité stéréoscopique	la/fs; - -
activité missionnaire	une/N21/A31/fs; - +	acuité tactile	la/fs; - -
activité moléculaire	une/N21/A31/fs; - +	acuité visuelle	la/fs; - -
activité musculaire	une/N21/A31/fs; - +	adage acrobatique	un/N1/A31/ms; - +
activité navale	une/N21/A32/fs; - +	adage seul	un/N1/A32/ms; - +
activité nocturne	une/N21/A31/fs; - +	adaptation agricole	une/N21/A31/fs; - +
activité nucléaire	une/N21/A31/fs; - +	adaptation audiovisuelle	une/N21/A40/fs; - +
activité occupationnelle	une/N21/A40/fs; - +	adaptation auditive	la/fs; - -
activité onirique	la/fs; - -	adaptation chromatique	une/N21/A31/fs; - +
activité optique	une/N21/A31/fs; - +	adaptation cinématographique	une/N21/A31/fs; - +
activité organique	la/fs; - -	adaptation convergente	une/N21/A32/fs; - +
activité périodique	une/N21/A31/fs; - +	adaptation enzymatique	une/N21/A31/fs; - +
activité pétrolière	une/N21/A42/fs; - +	adaptation génétique	une/N21/A31/fs; - +
activité phagocytaire	une/N21/A31/fs; - +	adaptation hormonale	une/N21/A76/fs; - +
activité phonatoire	une/N21/A31/fs; - +	adaptation individuelle	une/N21/A38/fs; - +
activité piscicole	une/N21/A31/fs; - +	adaptation littéraire	une/N21/A31/fs; - +
activité politique	une/N21/A31/fs; - +	adaptation musicale	une/N21/A76/fs; - +
activité porteuse	une/N21/A35/fs; - +	adaptation nerveuse	la/fs; - -
activité portuaire	une/N21/A31/fs; - +	adaptation télévisée	une/N21/A32/fs; - +
activité professionnelle	une/N21/A40/fs; - +	adaptation théâtrale	une/N21/A76/fs; - +
activité psychique	la/fs; - -	adaptation visuelle	la/fs; - -

### III. DELAF

#### Exemple de génération de formes flechées pour le dictionnaire électronique DELAF

%château,château.N3 :Nms	%vendaient,vendre.V67 :IIIm3p
%châteaux,château.N3 :Nmp	%vendis,vendre.V67 :IPa1s :IPa2s
%maison,maison.N21 :Nfs	%vendit,vendre.V67 :IPa3s
%maisons,maison.N21 :Nfp	%vendimes,vendre.V67 :IPa1p
%maison,maison.A80 :Ams :Afs :Amp :Afp	%vendites,vendre.V67 :IPa2p
%sec,sec.N2S :Nms	%vendirent,vendre.V67 :IPa3p
%sec,sec.A4S :Ams	%vendrai,vendre.V67 :IFu1s
%sèche,sec.A4S :Afs	%vendras,vendre.V67 :IFu2s
%secs,sec.A4S :Amp	%vendra,vendre.V67 :IFu3s
%sèches,sec.A4S :Afp	%vendrons,vendre.V67 :IFu1p
%sec,sec.ADV :Inv	%vendrez,vendre.V67 :IFu2p
%soigneux,soigneux.A63 :Ams :Amp	%vendront,vendre.V67 :IFu3p
%soigneuse,soigneux.A63 :Afs	%vende,vendre.V67 :SPr1s :SPr3s
%soigneuses,soigneux.A63 :Afp	%vendes,vendre.V67 :SPr2s
%invariablement,invariablement.ADV :Inv	%vendisse,vendre.V67 :SIIm1s
%vendre,vendre.V67 :Inf	%vendisses,vendre.V67 :SIIm2s
%vendant,vendre.V67 :Pant	%vendit,vendre.V67 :SIIm3s
%vendu,vendre.V67 :PPms	%vendissions,vendre.V67 :SIIm1p
%vendue,vendre.V67 :PPfs	%vendissiez,vendre.V67 :SIIm2p
%vendus,vendre.V67 :PPmp	%vendissent,vendre.V67 :SIIm3p
%vendues,vendre.V67 :PPfp	%vendrais,vendre.V67 :CPr1s :CPr2s
%vends,vendre.V67 :IPr1s :IPr2s :Imp2s	%vendrait,vendre.V67 :CPr3s
%vend,vendre.V67 :IPr3s	%vendrions,vendre.V67 :CPr1p
%vendons,vendre.V67 :IPr1p :Imp1p	%vendriez,vendre.V67 :CPr2p
%vendez,vendre.V67 :IPr2p :Imp2p	%vendraient,vendre.V67 :CPr3p
%vendent,vendre.V67 :IPr3p :SPr3p	%circuler,circuler.V3U :Inf
%vendais,vendre.V67 :IIIm1s :IIIm2s	%circulant,circuler.V3U :Pant
%vendait,vendre.V67 :IIIm3s	%circulé,circuler.V3U :PPms
%vendions,vendre.V67 :IIIm1p :SPr1p	%circule,circuler.V3U :IPr1s :IPr3s
%vendiez,vendre.V67 :IIIm2p :SPr2p	:SPr1s :SPr3s :Imp2s

### IV. DELACF

#### Les premières entrées du dictionnaire des formes DELACF

a/capella,PC	à/altitude/moyenne,PCA	à/bon/droit,PAC
a/contrario,PC	à/aucun/prix,PDETC	à/bon/écient,PAC
a/fortiori,PC	à/bâbord,PC	à/bon/prix,PAC
a/majuscule,NA :ms	à/bas/prix,PAC	à-propos,PN :ms :mp
a/majuscules,NA :mp	à/hus/régime,PAC	abaissée/d'aile,NDN :fs
a/minuscule,NA :ms	à/hasse/altitude,PAC	abaissées/d'aile,NDN :fp
a/minuscules,NA :mp	à/bâtons/rompus,PCA	abaissement/volontaire,
a/posteriori,PC	à/belles/dents,PAC	NA :ms
a/priori,PC	à/blanc,PC	abaissements/volontaires,
a/privatif,NA :ms	à/bloc,PC	NA :mp
a/privatifs,NA :mp	à/bon/compte,PAC	abaisseur/de/la/lèvre,NDN :ms
		abaisseur/du/sourcil,NDN :ms

V. DELAP-H

babillard, babijard\*, N32.A32  
 balourd, balurd\*, N1.N32.A32  
 bambochard, banboSard\*, N32.A32  
 banlieusard, banli0zard\*, N32.A32  
 bâtard, batard\*, N1.N32.A32  
 bavard, bavard\*, N32.A32  
 béquillard, bekijard\*, N32.A32  
 binoclard, binoklard\*, N32.A32  
 braillard, brajard\*, N32.A32  
 briard, briard\*, N1.N32.A32  
 cabochar, kaboSard\*, N32.A32  
 cafard, kafard\*, N1.N32.A32  
 camard, kamard\*, N32.A32  
 camisard, kamizard\*, N32.A32  
 campagnard, kanpaNard\*, N32.A32  
 capitular, kapitylard\*, N32.A32  
 chamoniard, Samoniard\*, N32.A32  
 chançard, Sansard\*, N32.A32  
 combinard, konbinard\*, N1.N32.A32  
 communard, komynard\*, N32.A32  
 conard, konard\*, N32.A32  
 connard, konard\*, N32.A32  
 cossard, kosard\*, N32.A32  
 couard, kuard\*, N32.A32  
 crevard, kr'vard\*, N32.A32  
 débrouillard, debrujiard\*, N32.A32  
 1 = dreyfusard, drefyziard\*, N32.A32  
 2 = dreyfusard, drefyziard\*, N32.A32  
 égrillard, egrijard\*, N32.A32  
 faiblard, feblard\*, N32.A32  
 fêtard, fetard\*, N32.A32  
 flemard, flemard\*, N32.A32  
 flemmard, flemard\*, N32.A32  
 fouinard, fuinard\*, N32.A32  
 froussard, frusard\*, N32.A32  
 fuyard, fyiard\*, N32.A32  
 gaillard, gajard\*, N1.N32.A32  
 geignard, ZeNard\*, N32.A32  
 gueular, g0lard\*, N1.N32.A32  
 guignard, giNard\*, N1.N32.A32  
 jobard, Zobard\*, N32.A32  
 lombard, lonbard\*, N1.N32.A32  
 montagnard, montaNard\*, N32.A32  
 mouchard, muSard\*, N1.N32.A32  
 mulard, mylard\*, N32.A32  
 musard, myzard\*, N32.A32  
 nullard, nylard\*, N32.A32  
 paillard, pajard\*, N32.A32  
 pantouillard, pantuflard\*, N32.A32  
 papelard, pap'lard\*, N1.N32.A32  
 patriotard, patriotard\*, N32.A32  
 pécinard, pécnard\*, N32.A32  
 pénard, pécnard\*, N32.A32  
 piaillard, piijard\*, N32.A32  
 picard, pikard\*, N1.N32.A32  
 pillard, pijard\*, N32.A32  
 pinçard, pinsard\*, N32.A32  
 pinchard, pinSard\*, N32.A32

pleurard, pl0rard\*, N32.A32  
 pleurnichard, pl0rniSard\*, N32.A32  
 pochard, poSard\*, N32.A32  
 politicard, poliikard\*, N32.A32  
 poupard, pupard\*, N1.N32.A32  
 prétentiard, pretansiard\*, N32.A32  
 revanchard, r'vanSard\*, N32.A32  
 rigolard, rigolard\*, N32.A32  
 ringard, ringard\*, N1.N32.A32  
 rondouillard, rondujard\*, N32.A32  
 rossard, rosard\*, N32.A32  
 roublard, rublard\*, N32.A32  
 savoyard, savujard\*, N32.A32  
 snobinard, snobinard\*, N32.A32  
 soiffard, suafard\*, N32.A32  
 sorbonnard, sorbooard\*, N32.A32  
 soflard, sulard\*, N32.A32  
 sourd, surd\*, N32.A32  
 tocard, tokard\*, N1.N32.A32  
 toquard, tokard\*, N1.N32.A32  
 traînard, trenard\*, N1.N32.A32  
 trouillard, trujard\*, N32.A32  
 tubard, tybard\*, N32.A32  
 vachard, vaSard\*, N32.A32  
 vantard, vantard\*, N32.A32  
 veinard, venard\*, N32.A32  
 vétilard, vetijard\*, N32.A32  
 bégard, begard\*, N32  
 broussard, brusard\*, N32  
 cagoulard, kagulard\*, N32  
 cambrouard, kanbruzard\*, N32  
 charognard, SaroSard\*, N32  
 chauffard, So'fard\*, N32  
 chevillard, S'vijard\*, N32  
 clochard, kloSard\*, N21.N32  
 cumulard, kymylard\*, N32  
 démerdard, demerdard\*, N32  
 déveinard, devenard\*, N32  
 grognard, groNard\*, N32  
 huitard, hyitard\*, N32  
 maquisard, makizard\*, N32  
 paniquard, panikard\*, N32  
 pendard, pandard\*, N32  
 pistard, pistard\*, N32  
 renard, r'nard\*, N1.N32  
 richard, riSard\*, N32  
 salonard, salonard\*, N32  
 salonnard, salonard\*, N32  
 salopard, salopard\*, N32  
 scribouillard, skribujard\*, N32  
 smicard, smikard\*, N32  
 smigard, smigard\*, N32  
 taulard, to'lard\*, N32  
 thésard, tezard\*, N32  
 tôlard, to'lard\*, N32  
 vicillard, vicjard\*, N32

lourd, lurd\*, N1.A32.ADVE  
 babillard, babijard\*, N32.A32  
 balourd, balurd\*, N1.N32.A32  
 bambochard, banboSard\*, N32.A32  
 banlieusard, banli0zard\*, N32.A32  
 bâtard, batard\*, N1.N32.A32  
 bavard, bavard\*, N32.A32  
 béquillard, bekijard\*, N32.A32  
 binoclard, binoklard\*, N32.A32  
 blafard, blafard\*, A32  
 braillard, brajard\*, N32.A32  
 briard, briard\*, N1.N32.A32  
 cabochar, kaboSard\*, N32.A32  
 cafard, kafard\*, N1.N32.A32  
 cagnard, kaNard\*, N1.A32  
 camard, kamard\*, N32.A32  
 camisard, kamizard\*, N32.A32  
 campagnard, kanpaNard\*, N32.A32  
 capitular, kapitylard\*, N32.A32  
 chamoniard, Samoniard\*, N32.A32  
 chançard, Sansard\*, N32.A32  
 combinard, konbinard\*, N1.N32.A32  
 communard, komynard\*, N32.A32  
 conard, konard\*, N32.A32  
 connard, konard\*, N32.A32  
 cossard, kosard\*, N32.A32  
 couard, kuard\*, N32.A32  
 crevard, kr'vard\*, N32.A32  
 criard, kriard\*, A32  
 débrouillard, debrujiard\*, N32.A32  
 discord, diskord\*, A32  
 1 = dreyfusard, drefyziard\*, N32.A32  
 2 = dreyfusard, drefyziard\*, N32.A32  
 égrillard, egrijard\*, N32.A32  
 faiblard, feblard\*, N32.A32  
 fêtard, fetard\*, N32.A32  
 flemard, flemard\*, N32.A32  
 flemmard, flemard\*, N32.A32  
 flottard, flotard\*, N1.A32  
 fouettard, fuettard\*, A32  
 fouinard, fuinard\*, N32.A32  
 froussard, frusard\*, N32.A32  
 furibard, fyribard\*, A32  
 fuyard, fyiard\*, N32.A32  
 gaillard, gajard\*, N1.N32.A32  
 geignard, ZeNard\*, N32.A32  
 goguenard, gog'nard\*, A32  
 gourd, gurd\*, A32  
 gueular, g0lard\*, N1.N32.A32  
 guignard, giNard\*, N1.N32.A32  
 hagar, hagar\*, A32  
 intramontagnard, intramontaNard\*, A32  
 jobard, Zobard\*, N32.A32  
 lombard, lonbard\*, N1.N32.A32  
 mignard, miNard\*, A32  
 mochar, moSard\*, A32  
 montagnard, montaNard\*, N32.A32  
 mouchard, muSard\*, N1.N32.A32

gloire. *Le chemin de la vertu, de la perfection, de la perte.* Spécialt. *Il a su trouver le chemin de son cœur, il a su toucher le cœur de cette personne, s'en faire aimer.* Expr. *Il est à la croisée des chemins, il doit faire un choix décisif. Suivre les chemins battus, s'attacher aux usages établis, manquer d'originalité.* Prov. *À chemin battu, il ne croit point d'herbe, il n'y a point de profit à s'occuper d'une affaire dont trop de gens se mêlent. Tous les chemins mènent à Rome, divers moyens conduisent à la même fin.*

II. Étendue que l'on parcourt pour avancer, pour aller d'un lieu à un autre; itinéraire conduisant à un lieu déterminé. I. Distance, trajet. *La ligne droite est le plus court chemin d'un point à un autre. Le chemin est plus court par mer que par terre. Nous avons fait le chemin à pied, à cheval, en voiture. S'arrêter à mi-chemin, à moitié chemin. Le chemin m'a paru court, long. Faire beaucoup de chemin, parcourir une longue distance. Vous avez deux heures de chemin jusqu'à la ville. Passer son chemin, continuer de marcher sans se laisser arrêter. Poursuivez votre chemin. Nous avons rebroussé chemin à cause de la tempête. Rencontrer quelqu'un en chemin. Chemin faisant, nous avons bavardé.* Expr. fam. *Un bout de chemin. Voir Bout.*

Expr. fig. *Il a encore bien du chemin à faire, bien des efforts, bien des progrès. Il serait dommage de vous arrêter à mi-chemin, en si bon chemin, d'abandonner une entreprise dont la réussite paraît assurée. Faire son chemin, faire du chemin, obtenir de l'avancement, s'enrichir, etc. Il a su faire son chemin.* Par ext. *L'idée avait peu à peu fait son chemin, elle se répandait, gagnait les esprits.* Par anal. PHYS. *Chemin optique d'un trajet donné, distance que la lumière parcourrait dans le vide pendant le temps qu'elle met à effectuer un trajet donné.*

2. Itinéraire, voie que l'on doit suivre. *Demander, perdre, retrouver son chemin. Je me suis trompé de chemin. Une avalanche nous a contraints à changer de chemin. Chercher son chemin sur une carte. Je passerai à la poste, c'est sur mon chemin. Ne me reconduisez pas, je connais le chemin. Le chemin des écoliers, voir Écolier.* Par ext. Passage. *Un arbre abattu nous barrait le chemin. Se frayer un chemin dans la neige, dans la forêt vierge. S'ouvrir un chemin dans la foule.* Expr. fig. *L'affaire est sur le bon chemin, en bon chemin. Il veut faire fortune, mais n'en prend guère le chemin. Montrer le chemin aux autres, faire ce que d'autres peuvent imiter, montrer l'exemple. Suivre des chemins différents. Il faut le remettre dans le bon chemin, le ramener dans le droit chemin.* Fam. *Aller son chemin, aller toujours son chemin, aller son petit bonhomme de chemin, ne pas se laisser détourner de la conduite qu'on a choisi de tenir. Aller droit son chemin, procéder avec sincérité, loyauté, sans artifice. Prendre des chemins de traverse, agir d'une façon dissimulée. Il ne faut pas y aller par quatre chemins, il faut parler, agir franchement, sans détours, sans ambages. Couper, barrer le chemin à quelqu'un, se mettre sur le chemin de quelqu'un, chercher à le gêner, à faire obstacle à ses initiatives. Trouver une pierre, des pierres en son chemin, trouver des empêchements, des obstacles à ce qu'on a dessein de faire. Aplanir le chemin à quelqu'un, lui éviter toute difficulté, tout obstacle. Chemin de velours, voie facile, agréable, par opposition à *Chemin difficile, malaisé. Il est arrivé à la fortune par un chemin de velours.* Avec une nuance de menace. *Il me trouvera sur son chemin, je saurai m'opposer à lui. Qu'il ne se trouve pas sur mon chemin!* Pop. Surtout dans des phrases négatives. *Ça en prend, ça n'en prend pas le chemin, cela semble, ne semble pas devoir aboutir.**

3. RELIG. *Le chemin de la croix, le chemin que le Christ parcourut en portant sa croix, jusqu'au Calvaire. Un chemin de croix, suite de quatorze bas-reliefs ou tableaux placés dans une église ou dans un lieu de pèlerinage, et représentant diverses scènes de la Passion. Faire un chemin de croix, son chemin de croix, s'arrêter et prier devant chacune de ces quatorze stations et, fig., vivre une série d'épreuves et de souffrances. Le chemin de Dumas, par allusion à la conversion de saint Paul, instant, circonstance où s'opère de façon subite et imprévue une conversion.* Par ext. Ad-

hèsion brusque et totale à une doctrine philosophique, morale, politique, etc.

\* CHEMIN DE FER n. m. I. Voie formée par deux rails métalliques parallèles, sur lesquels roulent les trains. *Les stations, la gare d'un chemin de fer. Le chemin de fer suit le fleuve.* (Aujourd'hui, on dit aussi *Voie ferrée.*)

2. Moyen de locomotion utilisant des véhicules équipés d'essieux guidés et portés par deux rails métalliques. *Prendre le chemin de fer. Voyager en chemin de fer. Expédier des marchandises par chemin de fer. Le chemin de fer métropolitain et, ellipt., le métropolitain ou le métro.*

3. Entreprise, administration qui exploite ce réseau ferré. *Une compagnie de chemin de fer. La Société nationale des chemins de fer français* (par abréviation SNCF). *Consulter l'indicateur des chemins de fer.*

4. TECHN. Outil de tailleur de pierre formé de lames d'acier dentelées.

5. JEUX DE CARTES. Variété de baccara.

## CHEMIN n. m.

I. Petite voie utilisée pour aller d'un lieu à un autre. I. Voie d'intérêt local qui relie, à la campagne, des champs, des fermes, des hameaux, des villages. *Un chemin de terre. Un chemin uni, pierreux, raboteux, fangeux. Tracer un chemin. Les chemins creux du bocage, encaissés entre des talus ou bordés de haies. Un chemin carrossable. Un chemin muletier. Un chemin forestier. Prendre un chemin de traverse, qui coupe à travers la campagne. Au carrefour des chemins. Le chemin du château, de la rivière, celui qui y mène. Chemin privé, chemin communal. Chemin vicinal, qui va d'une commune à l'autre et qui est sous la dépendance directe de l'autorité municipale. Les chemins ruraux. Chemin de halage, qui suit le bord d'une rivière ou d'un canal, et qui servait naguère au halage des bateaux. Chemin de douane, sentier suivi par les douaniers en bordure de mer ou le long d'une frontière. Anciennt. Grand chemin ou chemin royal, chemin large et bien entretenu qui reliait les grandes villes. Un voleur, un bandit de grand chemin, qui attaquait et dévalisait les voyageurs sur les routes et, fig., un homme sans honnêteté, sans aucun scrupule. Expr. *Aller par les chemins, vagabonder au hasard. Être par voies et par chemins, être toujours en voyage.* — ARCHIT. MILITAIRE. *Chemin de ronde, passage protégé par un parapet qui court intérieurement le long d'une enceinte. Chemin couvert, chemin à ciel ouvert établi sur la contrescarpe, entre le parapet et le fossé, et où les défenseurs étaient à couvert des projectiles des assiégeants.**

2. Par anal. *Chemin d'escalier, bande de tapis posée sur les marches d'un escalier. Chemin de table, bande d'étoffe ou de broderie dont on orne toute la longueur de la table.* — MÉCAN. *Chemin de roulement, surface profilée sur laquelle se déplace un organe mobile.*

3. Fig. Moyen conduisant à une fin. *Le chemin de la liberté. Les chemins de la connaissance, de la*

LA SIGNIFICATION DES DIFFÉRENTS MOTS RAISON :

- (1) Jo a raison de partir
- (2) Jo a (une + des) raisons de partir  
Jo a ses raisons de partir  
Il n'y a pas de raison que Jo parte  
Le départ de Max a (une + plusieurs) raisons  
Le départ de Max est sans raison(s)  
Une des raisons de son départ est que Max reste  
Max a toutes (E + les) raisons de penser que Jo restera
- (3) Que Jo parte n'est pas une raison pour que Max reste
- (4) Jo a toute sa raison  
Jo (a perdu + est revenu à) la raison  
Max ramène Jo à la raison
- (5) Max s'est rendu aux raisons de Jo
- (6) Max s'est fait une raison, il ira travailler
- (7) Ce texte (n'a ni + est sans) rime ni raison
- (8) Jo a eu raison de toutes les difficultés
- (9) Max donne raison à Jo dans ce conflit
- (10) Max a demandé raison à Jo de sa conduite  
Jo rendra raison
- (11) Max entendra raison
- (12) La raison de cette progression (arithmétique + géométrique) est
- (13) Max écoute la voix de la raison
- (14) Jo a fait un mariage de raison
- (15) Jo a l'âge de raison
- (16) Jo a une raison (d'être + de vivre)
- (17) Jo a une raison sociale
- (18) La raison d'état impose cette mesure injuste
- (19) Max invoque (la raison pure + la raison du plus fort)
- (20) (partir) (à + avec + non sans) raison
- (21) (partir) à juste raison
- (22) (partir) comme de raison
- (23) (dormir) plus que de raison
- (24) (agir) à plus forte raison
- (25) (partir) pour des raisons de santé
- (26) (partir) à tort ou à raison
- (27) (partir) en raison du brouillard
- (28) (payer les gens) en raison de leurs efforts
- (29) (arriver) à raison de six à l'heure
- (30) (agir) en raison (directe + inverse) des interventions
- (31) (agir) pour la bonne raison que Jo part
- (32) (agir) raison de plus pour que Jo parte



## FRÉQUENCE FREGOLI

avary met en scène Bernard Haller dans le rôle de Fregoli. Ou plutôt: les rôles de Fregoli, cet homme qui conquiert la célébrité, au début du siècle, en changeant de costume soixante fois par heure. Le père de Fregoli s'opposait pourtant à sa vocation de comédien. Jusqu'au jour où il vit arriver une jeune femme en larmes: « Leopoldo! Où est Leopoldo? Votre fils m'a séduite et refuse de reconnaître son enfant! » Et de tomber évanouie dans les bras de Fregoli père, lequel part chercher des sels. Quand il revint, c'était Leopoldo qui

demandait: « Crois-tu maintenant que je suis capable de jouer la comédie? » Il la jouera. En occupant dix loges à lui tout seul, avec ses 800 costumes, ses 1 200 perruques, et une technique d'enfer que Bernard Haller devra retrouver pour ses 25 changements de costume. Fregoli procédait en deux secondes et demie avec l'aide de trois habilleurs. Ses redingotes avaient des emmanchures très larges, ses robes s'ouvraient dans toute leur longueur et se fermaient au moyen de deux grandes agrafes. Les parties accessoires, comme le tablier de la femme de chambre ou le réticule de la mondaine,

étaient cousues aux vêtements, et les chapeaux fixés aux perruques. Comme, de la coulisse, Fregoli ne cessait de parler, le public ne s'apercevait pas du temps qui s'était écoulé. C'était, selon lui, le plus difficile: « Parler avec une voix lente et posée, alors que mes mouvements sont follement précipités. » Autant de difficultés que Bernard Haller devra affronter, avec l'aide de moyens modernes, tels que le ruban Velcro et la machine à culotter: le pantalon est plié au sol, en accordéon, on y pose ses pieds, et un ressort remonte le tout. Il lui faudra un peu de temps avant de dire, comme Fregoli, « j'ai trois secondes pour me changer bien tranquillement ». S. O. Théâtre national de Chaillot, salle Jean Vilar, du 5 avril au 29 juin. Tél. 47 27 81 15.

PHOTOS JENNY ECOFFIER, JEAN THALANCE.



MARIE-PIERRE DUROS

### ELLE SE DÉCARCASSE

La dévoreuse de pistes a un corps de tanagra. Championne de France de cross-country (6 kilomètres), championne du monde du 3 000 mètres en salle à Séville, Marie-Pierre Duros confirme avec un bel aplomb: « Mais je n'ai jamais douté. J'avais choisi de mener cette finale, car j'avais la certitude de pouvoir l'emporter. » Sitôt dit, sitôt fait. Yeux clairs, taches de rousseur, la petite (1,60 m, 45 kg) Bretonne de Saint-Brieuc courrait à 10 ans avec les garçons, parce que les filles n'étaient pas de son niveau. Elle sera quinze fois championne de France de cross-country et sur piste, et médaille d'argent aux championnats d'Europe juniors en 1988. « C'est en seniors que je voulais briller; avant ça ne compte pas. » Ah, bon? A 24 ans, elle travaille au conseil général des Côtes d'Armor et c'est encore avec les garçons qu'elle préfère faire son entraînement. Quand on sait qu'elle est capable de courir à la vitesse de 2,57 mn au km, ça ne doit pas être facile de jouer avec elle au « suivez-moi jeune homme ».

J. C. Buggeat, Tavernier/Siis Sport.

*Elle* Avril 1991

**E**n entrée, des bricks aux œufs cachers frits maison dans des feuilles de pâte toutes préparées ou bien un assortissement de hors-d'œuvre grecs en barquettes, tzatziki aux herbes, tarato aux aubergines, feuilles de vigne farcies, salades à la feta et, bien sûr tarama, servis avec des blinis russes ou des pitas libanais et arrosés d'un riesling chinois Huedong ; comme plat principal, un chili con carne mexicain en boîte ou un chicken sathay thaïlandais et un biryani indien de légumes réchauffés au micro-ondes, servis avec des nâns et accompagnés, par exemple, d'un ruby de Magarach russe, dense et profond ; enfin, pour terminer, des litchis frais, quelques loukoums d'origine yougoslave ou bien des apfel strudels, faits à l'aide de feuilles enséchées, le tout avec par un vin blanc sud-africain et ponctué par quelques rasades d'une liqueur de sorgho chinoise ou d'un soda à base de soja de Singapour : tel est l'un des menus planétaires que l'on peut aujourd'hui quotidiennement composer à partir des denrées en vente dans une grande surface tout à fait anonyme du quinzième arrondissement de Paris. Le tour du monde dans une assiette au coin d'un rue banale et à tout petits prix.

Etonnante évolution, tout de même, si l'on songe qu'il y a seulement dix ans, il fallait traverser tout Paris pour trouver de la banale mozzarella italienne, une bouteille d'ouzo grec ou bien encore une Michelob yankee ! Les frontières culinaires semblent abolies : si l'on ajoute aux grandes surfaces les épicerie de luxe et les échoppes spécialisées, il paraît aujourd'hui possible de se procurer sur-le-champ toutes les denrées nécessaires à l'élaboration de

n'importe quel plat exotique, d'Afrique, du Moyen et de l'Extrême-Orient ou d'ailleurs.

La cardamome, la coriandre, le curcuma ; l'igname, la patate douce, le citron vert ; le raz el hanout, la harissa, l'origan et la menthe fraîche : tous ces produits et épices si difficiles naguère à trouver sont devenus presque banals. Le voyage dans l'assiette s'est généralisé, popularisé et démocratisé. Et le métissage est entré en force dans le lieu le mieux gardé des identités nationales et régionales : la cuisine. Les grandes firmes agroalimentaires ont d'ailleurs été plus rapides que les sociologues à déceler le phénomène et à en profiter : on ne compte plus aujourd'hui les soupes chinoises ou les riz indiens et thaïlandais pré-cuits en sachet, les currys tout préparés, les sauces en boîte ou en sachet déshydraté, aujourd'hui presque aussi français que le gaulois cassoulet.

Et Paris n'est pas la seule touchée par cette évolution. A Londres, ultime réminiscence d'un vaste empire, la cuisine indienne est devenue la seconde - la seule même, ironisent les anglophobes - cuisine du terroir. A Berlin, immigration oblige, les kebabs turcs et les cevapcici yougoslaves taillent des croupières aux traditionnelles wurst teutoniques. Et ainsi de suite : de même que l'on parle, en musique, de *sono mondiale* pour caractériser le fantasti-

## Table



que mixage actuel de genres planétaires, c'est presque parler aujourd'hui *world cooking*, tant l'effacement des cuisines est a-

**C**ela ne va pas sans le surgissement de variétés inédites : c'est Essentines concoctant, « nues de son célèbre bouche », à partir de provenances diverses, inouïes, les ménagères

Annexe 5

Questionnaire du test

Evaluer la liste de mots ci-jointe selon les **deux critères** suivants:

1- Le critère de **représentation** (valeurs: 1 / 0).

\_ Le mot évoque, même vaguement ( *impression de déjà vu, souvenir diffus* etc.), quelque chose en vous, (c. e. vous disposez d'une "représentation" de ce mot):

vous marquerez:.....1

\_ Le mot n'évoque rien en vous:

vous marquerez:.....0

2- Le critère de **plausibilité** (valeurs: +++ / ++ / +):

Dans tout ce que vous lisez et entendez habituellement, les possibilités d'apparition du mot, sont:

\_ fortes, vous marquerez:.....+++

\_ moyennes à faibles, vous marquerez:.....++

\_ très faibles ou proches de zéro, vous marquerez:.....+

\_ Le critère de plausibilité est indépendant de la connaissance que vous pouvez avoir d'un mot. Exple: vous pouvez connaître le sens d'*hypallage* et juger que sa plausibilité d'apparition est très faible. Vous marquerez, dans la colonne "plausibilité":+.

\_ Si vos activités professionnelles ou privées vous exposent à un vocabulaire très pointu, vous ferez une estimation intuitive de la plausibilité d'apparition de ce vocabulaire auprès d'un public cultivé de niveau études supérieures (non nécessairement spécialiste) Exple: vous êtes dans l'informatique: vous attribuerez un indice de plausibilité +++ à *ordinateur, logiciel* ou *listing*, un indice ++ à *interface, bit, ou octet*; et un indice + à *randomiser, progiciel* ou *pixel* (même si ces derniers termes sont très employés dans ce domaine).

MOT	REPRESENTATION	PLAUSIBILITE
	Valeur : 0 / 1	Valeur +++ / ++ / +
Exemples		
Eau	1	+++
Salsepareille	1	++
Leproïde	1	+
Eluvion	0	
Burgrave		
Monophtongue		
Damoiseau		
Délétion		
Déjouement		
Elephantesque		
Lactaire		
Labrador		
Palanquin		
Astragale		
Buis		
Entropie		
Cimaise		
Carénage		
Epithalame		
Lamparo		
Dansomètre		
Postprandial		
Amadou		
Dague		
Dandinette		
Metastase		
Pepsine		
Perborate		
Porcelaine		
Porcelainier		
Porcelanique		
Bazar		

Bazardage  
Bazarder  
Bazardisation  
Langouste  
Langoustier  
Langoustière  
Langoustine  
Langoustiner  
Cabochar  
Cabocharnement  
Cocaïne  
Cocainique  
Cocainiser  
Cocainisme  
Cocainomane  
Cocainomanie  
Musique  
Musiquer  
Musiquette  
Danger  
Dangereusement  
Dangerosité  
Daguerréotype  
Daguerréotyper  
Daguerréotypiste  
Dactylo  
Dactylocodage  
Dactylocoder  
Dactylocodeur  
Dactylodiasrophie  
Dactylogramme  
Dactylographe  
Dactylographiquement

## Annexe 6

### Le système INTEX

#### Les Programmes de reconnaissance lexicale :

Ces programmes développés par M. Silberstein permettent, à partir de tout texte, sans limite de taille, et dans des délais extrêmement rapides<sup>1</sup> de :

- Calculer la liste des mots du texte  
Programme inverse

- Créer le dictionnaire du texte  
Programme creatdic

-Etiqueter les mots du texte  
Programme étiqu

-Rechercher les mots composés du texte  
Programme compo

- Elaborer des graphes et de les reconnaître  
Editeur d'automates et programme reconnait

- Chercher des concordances de mots ou groupes de mots  
Programme concord

---

<sup>1</sup> La vitesse d'étiquetage d'un mot est de l'ordre de 200 à 300 mots /seconde.

PARIS - Valéry Giscard d'Estaing estime que "la ratification du traité de Maastricht contribuerait à rétablir la confiance".

"Son rejet aggraverait la crise du système monétaire européen et risquerait de l'étendre à d'autres monnaies".

RENNES - André Lajoinie a estimé jeudi que "c'est un mensonge de prétendre que la menace d'une victoire du «non» au referendum est à l'origine de la déconfiture de la livre ou de la lire" sur les marchés financiers.

Il a indiqué que cet argument était "une pression des partisans du «oui» sur les électeurs".

PARIS - Le 6<sup>ème</sup> prix Radio-France Nancy-Lorraine de la communication 1992 a été attribué, dans le cadre du Livre sur la Place, à Jean Ferniot pour son recueil de mémoires "Je recommencerais bien", paru aux éditions Grasset.

Il y tire le bilan d'une vie et près d'un demi-siècle de journalisme.

PARIS - Michel Sapin a indiqué que "le peuple français a (vait) aujourd'hui une responsabilité comme rarement il en a eu dans son histoire".

La crise du SME montre "l'importance pour le monde entier de la décision des Français".

LONDRES - L'indice Footsie des cent grandes valeurs a gagné 105,6 points (4,4%) à 2.483,9.

Certains investisseurs espèrent une baisse supplémentaire des taux d'intérêt maintenant que le gouvernement n'a plus à maintenir la parité de la livre.

PARIS - Le promoteur immobilier Christian Pellerin a été inculpé jeudi de "délit d'initié".

Il est soupçonné d'avoir spéculé sur des actions de la Société Générale en 1988, avant que le raid de Georges Pébereau, le PDG de la société Marceau Investissements, sur la Société Générale ne soit connu du public.

PARIS - Rapports du Quinté+ disputé jeudi à Maisons-Laffitte dans le Handicap de la Tamise (3<sup>ème</sup> course -Tous partants)

QUINTE+ 11-2-17-9-3. Rapports pour 10 F:

Dans l'ordre : 791.250,00

Ordre différent : 15.825,00

Générale ne soit connu du public.

PARIS - Rapports des Tiercé/Quarté+ à Maisons-Laffitte dans le Handicap de la Tamise (3<sup>ème</sup> course -Tous partants)

TIERCE 11-2-17 Rapports pour 6 F:

Dans l'ordre : 4.104,00

Ordre différent : 820,80

QUARTE+ 11-2-17-9 Rapports pour 8 F:

Dans l'ordre : 105.404,80

Ordre différent : 3.609,60

PARIS - 39 universitaires lancent jeudi un appel à voter "non" dimanche au referendum.

Ils constatent que le traité de Maastricht "entraînerait les Nations et les individus dans un processus irréversible de soumission à des instances que ni les Nations ni les Individus ne pourront contrôler".

PARIS - Pierre Bérégovoy a assuré que "le système monétaire européen (SME), sans l'Union Economique et Monétaire, est plus fragile".

"Les spéculations qui testent la solidité du SME depuis plusieurs jours se multiplieraient si le 'non' devait par malheur l'emporter" dimanche.

MOSCOU - Les discussions multilatérales organisées dans le cadre de la conférence sur la paix au Proche Orient se sont achevées jeudi sans qu'une nouvelle date ait été fixée pour une prochaine rencontre.

Des propositions variées ont été faites pour améliorer le climat de confiance.

TEL-AVIV - Le chef de la diplomatie chinoise a promis que la Chine ne vendra plus d'armes aux pays du Moyen-Orient, a annoncé le Conseil israélien.

"Nous ne prendrons aucune initiative susceptible de nuire au processus de paix".

PARIS - Pierre Bérégovoy a estimé que le "débat a été pollué par des calculs politiques dans le camp de l'opposition".

Pour lui, dans le camp de l'opposition, "on règle des comptes, au lieu d'en rendre à la France et à sa jeunesse".

PARIS - L'Imprimerie nationale indique avoir imprimé et livré, en deux mois de temps, 45,5 millions de

versions intégrales du traité de Maastricht, 47 millions de notices explicatives, 198 millions de bulletins de vote, 44 millions d'enveloppes, 410.000 affiches et 240.000 procès-verbaux.

PARIS - Le dollar, après avoir atteint 5,22 F et 1,52 DM dans la matinée à la faveur des remous du SME, a reflué dans la journée.

Estimé par la Banque de France à 5,1155 F, contre 5,1195 F mercredi, le billet vert se traitait en fin de journée à 5,0850 F et 1,4860 DM.

TOULOUSE - Le temps s'améliorera vendredi sur le nord-ouest, mais sera orageux ailleurs.

Les températures maximales iront de 20 à 27 degrés du nord-ouest au sud-est.

COPENHAGUE - La Banque centrale du Danemark est intervenue jeudi après-midi pour soutenir la couronne danoise.

Elle était tombée à son cours plancher au sein du système monétaire européen.

ROME - Le gouvernement italien a décidé la création d'un impôt de 0,7% sur le patrimoine des sociétés dans le cadre du plan de réduction du déficit budgétaire pour 1993, a annoncé le ministre italien des travaux publics Francesco Merloni.

WASHINGTON - La crise que traverse le SME "n'est pas seulement le reflet des peurs du marché, mais la conséquence de déséquilibres fondamentaux plus profonds (notamment en Italie et en Allemagne)", a déclaré jeudi le FMI.

"Le marché veut du concret et pas seulement des promesses".

PEAGE-DE-ROUSILLON - Louis Mermaz a affirmé jeudi que "l'agriculture française s'est développée grâce à l'Europe".

"Si la Communauté devait demain du fait d'un vote négatif des Français perdre sa vitalité, les agriculteurs français seraient frappés de plein fouet".

NEW YORK - Le dollar "reste neutre" dans cette tempête monétaire qui frappe l'Europe.

Hormis ses gains face à la livre, le dollar était en léger recul face à l'ensemble des autres grandes monnaies. Il cotait ainsi vers 10H00 locales 5,0960 FF contre 5,1090 FF.

VALENCE - 80 gendarmes mobiles sont intervenus jeudi après-midi avec des gaz lacrymogènes pour disperser une

manifestation d'agriculteurs qui bloquaient les voies de l'autoroute A7 près de Pont-de-l'Isère (Drôme).

Les agriculteurs avaient auparavant déversé près de 40 tonnes de pêches sur la RN7.

ROME - Le gouvernement italien a adopté jeudi une série de mesures économiques pour réduire le déficit budgétaire de 1993.

Il a notamment décidé l'augmentation de l'âge de la retraite et le blocage pendant un an des départs à la retraite

LE VATICAN - Le Saint-Siège a proposé un plan en huit points pour mettre fin au conflit en Bosnie-Herzégovine.

Il prévoit notamment "un véritable blocus naval", complété si nécessaire "par la fermeture de l'espace aérien de la Bosnie-Herzégovine à tous les vols, à l'exception de ceux de l'ONU".

LAUSANNE - Un Français vivant en Suisse a mis au point et commercialise une montre destinée à lutter contre le sida en facilitant l'usage du préservatif.

La montre est munie d'un préservatif caché sous le bracelet.

A BORD D'AIR FORCE ONE - Le porte-parole de la Maison Blanche Marlin Fitzwater a exprimé jeudi sa confiance dans l'avenir du dollar après la tempête monétaire enregistrée en Europe.

"L'économie américaine est toujours la plus forte du monde".

LONDRES - Le chômage s'est fortement accru en Grande-Bretagne le mois dernier, à presque 10% de la population active.

Le nombre des demandeurs d'emploi a augmenté de 47.400 en août à 2,807 millions, portant le taux de chômage à 9,9%.

## Programme Inverse

A  
a  
à  
accru  
achevées  
actions  
active  
adopté  
aérien  
affaire  
affiches  
affirmé  
âge  
aggraverait  
agriculteurs  
agriculture  
ailleurs  
ainsi  
AIR  
ait  
Allemagne  
améliorer  
améliorera  
américaine  
an  
ancien  
André  
annoncé  
août  
appel  
après  
argument  
armes  
assuré  
atteint  
attribué  
au  
aucune  
augmentation  
augmenté  
aujourd  
auparavant  
autoroute  
autres  
aux  
avaient  
avait  
avant  
avec  
avenir  
AVIV  
avoir  
baisse  
Banque  
Barril  
Bérégovoy  
bien  
bilan  
billet  
Blanche  
blocage  
blocus  
bloquaient  
BORD  
Bosnie  
bracelet  
Bretagne  
budgétaire  
bulletins  
c  
caché  
cadre  
calculs  
camp  
ce  
cent  
centrale  
Certains  
cet  
cette  
ceux  
chambre  
chef  
Chine  
chinoise  
chômage

# Programme creatdic

a. .N2:ms:mp.avoir.V1:F3s  
 à. .PREP:Inv  
 accru. .A32:ms. .N1:ms,accroître.V73:Kms  
 achevées,achevé.A32:fp,achever.V6:Kfp  
 actions,acter.V3:1fp:S1p,action.N21:fp  
 active. .N2S:fs,actif.A38:fa,actif.N38:fs,activer.V3:P1s:P3s:S1s:S3s:Y2s  
 adopté. .A32:ms. .N32:ms,adopter.V3:Kms  
 aérien. .A41:ms. .N1:ms  
 affaire. .N21:fs,affairer.V3:P1s:P3s:S1s:S3s:Y2s  
 affiches,affiche.N21:fp,afficher.V3:P2s:S2s  
 affirmé,affirmer.V3:Kms  
 âge. .N1:ms  
 aggraverait,aggraver.V3:C3s  
 agriculteurs,agriculteur.N36:mp  
 agriculture. .N21:fs  
 ailleurs. .ADV:Inv  
 ainsi. .ADV:Inv  
 air. .N1:ms  
 ait,avoir.V1:S3s  
 améliorer. .V3:W  
 améliorer,améliorer.V3:F3s  
 américaine. .N21:fs,américain.A32:fs,américain.N32:fs  
 an. .N1:ms  
 ancien. .A41:ms. .N1:ms. .N41:ms  
 annoncé,annoncer.V4:Kms  
 août. .N1:ms  
 appel. .N1:ms  
 après. .ADV:Inv. .PREP:Inv  
 argument. .N1:ms  
 armes,arme.N21:fp,armer.V3:P2s:S2s  
 assuré. .A32:ms. .N32:ms,assurer.V3:Kms  
 atteint. .A32:ms,atteindre.V64:Kms:P3s  
 attribué,attribuer.V3:Kms  
 au. .DET:ms. .PREP:Inv  
 aucune,aucun.A32:fs,aucun.DET:fs,aucun.PRON:fs  
 augmentation. .N21:fs  
 augmenté,augmenter.V3:Kms  
 aujourd. .XINC  
 auparavant. .ADV:Inv  
 autoroute. .N21:fs+. .N1:ms  
 autres,autre.A31:mp:fp,autre.N1:mp,autre.PRON:mp:fp  
 aux. .DET:mp:fp. .PREP:Inv  
 avaient,avoir.V1:I3p

conférence. .N21:fs  
 confiance. .N21:fs  
 conflit. .N1:ms  
 connu. .A32:ms. .N2S:ms,connaître.V69:Kms  
 conseil. .N1:ms  
 conséquence. .N21:fs  
 constatent,constater.V3:P3p:S3p  
 contre. .ADV:Inv. .N1:ms. .N2S:ms. .PREP:Inv,contrer.V3:P1s:P3s:S1s:S3s:Y2s  
 contribuerait,contribuer.V3:C3s  
 contrôler. .V3:W  
 correctionnel. .A40:ms  
 cotait,coter.V3:I3s  
 couronne. .N21:fs,couronner.V3:P1s:P3s:S1s:S3s:Y2s  
 cours. .N2:ms:mp,cour.N21:fp,courir.V31:P1s:P2s:Y2s  
 course. .N21:fs,courser.V3:P1s:P3s:S1s:S3s:Y2s  
 création. .N21:fs  
 crise. .N21:fs  
 D. .CR:500  
 d. .N2:ms:mp. .PREP:Inv,du.DET:ms:fs:mp:fp  
 danoise,danois.A61:fs,danois.N61:fs  
 dans. .PREP:Inv  
 date. .N21:fs,dater.V3:P1s:P3s:S1s:S3s:Y2s  
 de. .PREP:Inv,du.DET:mp:fp  
 de. .N1:ms  
 débat. .N1:ms,débattre.V68:P3s  
 décidé. .A32:ms,décider.V3:Kms  
 décision. .N21:fs  
 déclaré. .A32:ms,déclarer.V3:Kms  
 déconfiture. .N21:fs  
 déficit. .N1:ms  
 degrés,degré.N1:mp  
 demain. .ADV:Inv  
 demandeurs,demandeur.N35:mp,demandeur.N37:mp  
 demi. .(PFX):Xin. .ADV:Inv. .A80:ms:fs:mp:fp. .N1:ms  
 départs,départ.N1:mp  
 depuis. .ADV:Inv. .PREP:Inv  
 dernier. .A42:ms. .N42:ms  
 les,un.DET:mp:fp  
 déséquilibres,déséquilibre.N1:mp,déséquilibrer.V3:P2s:S2s  
 destinée. .N21:fs,destiner.V3:Kfs  
 lieux. .A80:ms:fs:mp:fp. .DETQ:mp:fp. .N2:ms:mp  
 devait,devoir.V41:I3s  
 développée. .N21:fs,développé.A32:fs,développer.V3:Kfs

avait,avoir.V1:I3s  
 avant. .ADV:Inv. .A80:ms:fs:mp:fp. .N1:ms. .PREP:Inv  
 avec. .ADV:Inv. .PREP:Inv  
 avenir. .N1:ms  
 avoir. .N1:ms. .V1:W  
 baisse. .N21:fs,baisser.V3:P1s:P3s:S1s:S3s:Y2s  
 banque. .N21:fs,banquer.V3:P1s:P3s:S1s:S3s:Y2s  
 bien. .ADV:Inv. .A80:ms:fs:mp:fp. .N1:ms  
 bilan. .N1:ms  
 billet. .N1:ms  
 blanche. .N21:fs,blanc.A47:fs,blanc.N47:fs  
 blocage. .N1:ms  
 blocus. .N2:ms:mp  
 bloquent,bloquer.V3:I3p  
 bord. .N1:ms  
 bracelet. .N1:ms  
 budgétaire. .A31:ms:fs  
 bulletins,bulletin.N1:mp  
 c. .N2:ms:mp  
 caché. .A32:ms,cacher.V3:Kms  
 cadre. .N1:ms. .N31:ms:fs,cadrer.V3:P1s:P3s:S1s:S3s:Y2s  
 calcul,calcul.N1:mp  
 camp. .N1:ms  
 ce. .DET:ms. .PRON:ms3s:mp3p  
 cent. .A80:ms:fs:mp:fp. .A31:ms:fs. .DETQ:mp:fp. .N1:ms. .N2:ms:mp  
 centrale. .N21:fs,central.A76:fs  
 certains. .PRON:mp,certain.A32:mp,certain.DET:mp,certain.N1:mp  
 cet,ce.DET:ms  
 cette,ce.DET:fs  
 ceux,celui.PRON:mp  
 chambre. .N21:fs,chambrier.V3:P1s:P3s:S1s:S3s:Y2s  
 chef. .N1:ms. .N31:ms:fs  
 chine. .N1:ms. .N21:fs,chiner.V3:P1s:P3s:S1s:S3s:Y2s  
 chinoise,chinois.A61:fs,chinois.N61:fs,chinoiser.V3:P1s:P3s:S1s:S3s:Y2s  
 chantage. .N1:ms  
 climat. .N1:ms  
 comme. .ADV:Inv. .CNJS:CSu  
 commercialise,commercialiser.V3:P1s:P3s:S1s:S3s:Y2s  
 communauté. .N21:fs  
 communication. .N21:fs  
 complété,compléter.V7:Kms  
 comptes,compte.N1:mp,compter.V3:P2s:S2s  
 concret. .A43:ms. .N1:ms

déversé. .A32:ms,déverser.V3:Kms  
 différent. .A32:ms. .N1:ms  
 dimanche. .N1:ms  
 diplomatie. .N21:fs  
 directeur. .A36:ms. .N36:ms  
 discussions,discussion.N21:fp  
 disperser. .V3:W  
 disputé. .A32:ms,disputer.V3:Kms  
 dollar. .N1:ms  
 du. .DET:ms. .PREP:Inv  
 économie. .N21:fs  
 économique. .A31:ms:fs. .N2S:ms. .N21:fs  
 économiques,économique.A31:mp:fp,économique.N21:fp  
 écrit. .A32:ms. .N1:ms,écrire.V92:Kms:P3s  
 éditions,éditer.V3:1fp:S1p,édition.N21:fp  
 électeurs,électeur.N36:mp  
 elle,il.PRON:fs3s  
 emploi. .N1:ms  
 exporter. .V3:W  
 en. .ADV:Inv. .PREP:Inv. .PRON:Inv  
 enregistrée,enregistrer.V3:Kfs  
 ensemble. .ADV:Inv. .N1:ms  
 entier. .A42:ms. .N1:ms  
 enveloppes,enveloppe.N21:fp,envelopper.V3:P2s:S2s  
 espace. .N1:ms. .N21:fs,espacer.V4:P1s:P3s:S1s:S3s:Y2s  
 espèrent,espérer.V7:P3p:S3p  
 est. .A80:ms:fs:mp:fp. .N2S:ms,être.V2:P3s  
 estime. .N21:fs,estimer.V3:P1s:P3s:S1s:S3s:Y2s  
 estimé,estimer.V3:Kms  
 et. .CNJC:CCo  
 étaient,étayer.V14:F3p:S3p,être.V2:I3p  
 était,être.V2:I3s  
 été. .N1:ms,être.V2:Kms  
 étendre. .V67:W  
 eu,avoir.V1:Kms  
 européen. .A41:ms. .N41:ms  
 exception. .N21:fs  
 explicatives,explicatif.A38:fp  
 exprimé,exprimer.V3:Kms  
 f. .N2:ms:mp  
 fabriqué,fabriquer.V3:Kms  
 face. .N21:fs,facier.V4:P1s:P3s:S1s:S3s:Y2s  
 facilitant. .A32:ms,faciliter.V3:G

Programme etiq

PARIS -	pari	N1:mp	européen	européen	A41:ms
Valéry	?		et	européen	N41:ms
Giscard	?		risquerait	et	CNJC:CCo
d'	d	N2:ms:mp	de	risquer	V3:C3s
	d	PREP:Inv		de	PREP:Inv
	du	DET:ms:fs:mp:fp		du	DET:mp:fp
Estaing	?		l'	l	N2:ms:mp
estime	estime	N21:fs		le	DET:ms:fs
	estimer	V3:P1s:P3s:S1s:S3s:Y2s	étendre	le	PRON:ms3s:fs3s
que "	que	ADV:Inv	à	à	V67:W
	que	CNJS:CSu	d'	d	PREP:Inv
	que	PRON:ms:fs:mp:fp		d	N2:ms:mp
la	la	N2:ms:mp		du	PREP:Inv
	le	DET:fs	autres	de	DET:ms:fs:mp:fp
	le	PRON:fs3s		autre	A31:mp:fp
ratification	ratification	N21:fs	monnaies"	autre	N1:mp
du	du	DET:ms	RENNES -	monnaie	PRON:mp:fp
	du	PREP:Inv	André	monnayer	N21:fp
traité	traité	N1:ms	Lajoinie	renne	V14:P2s:S2s
de	traiter	V3:Kms	?	?	N1:mp
	de	PREP:Inv			
	du	DET:mp:fp	a	a	N2:ms:mp
Maastricht	?		estimé	avoir	V1:P3s
contribuerait	contribuer	V3:C3s	jeudi	estimer	V3:Kms
à	à	PREP:Inv	que "	jeudi	N1:ms
rétablir	rétablir	V18:W		que	ADV:Inv
la	la	N2:ms:mp	c'	que	CNJS:CSu
	le	DET:fs	est	que	PRON:ms:fs:mp:fp
	le	PRON:fs3s		c	N2:ms:mp
confiance". "	confiance	N21:fs	un	est	A80:ms:fs:mp:fp
Son	son	DETP:ms3s:fs3s		est	N2:ms
	son	N1:ms	mensonge	être	V2:P3s
rejet	rejet	N1:ms	de	un	A32:ms
aggraverait	aggraver	V3:C3s		un	DET:ms
la	la	N2:ms:mp	prétendre	un	DET:ms
	le	DET:fs	de	un	N1:ms
	le	PRON:fs3s		un	PRON:ms
crise	crise	N21:fs	de	un	N1:ms
du	du	DET:ms	prétendre	de	PREP:Inv
	du	PREP:Inv	que	du	DET:mp:fp
système	système	N1:ms		prétendre	V67:W
monétaire	monétaire	A31:ms:fs		que	ADV:Inv
	que	CNJS:CSu	livre	livre	N1:ms
	que	PRON:ms:fs:mp:fp		livre	N21:fs
la	la	N2:ms:mp	ou	livrer	V3:P1s:P3s:S1s:S3s:Y2s
	le	DET:fs	de	de	CNJC:CCo
	le	PRON:fs3s	la	du	PREP:Inv
menace	menace	N21:fs	d'	la	DET:mp:fp
	menacer	V4:P1s:P3s:S1s:S3s:Y2s		le	N2:ms:mp
d'	d	N2:ms:mp	lire"	le	DET:fs
	d	PREP:Inv	sur	lire	PRON:fs3s
	du	DET:ms:fs:mp:fp		lire	N21:fs
une	une	N21:fs	les	sur	V94:W
	un	A32:fs	marchés	sur	(PN):Xin
	un	DET:fs	financiers.	sur	A32:ms
	un	PRON:fs	Il	sur	PREP:Inv
victoire	victoire	N21:fs	a	le	DET:mp:fp
du +	du	DET:ms	indiqué	le	PRON:mp3p:fp3p
	du	PREP:Inv	que	marché	N1:mp
non+	non	ADV:Inv		financier	A42:mp
	non	N2:ms:mp	cet	financier	N42:mp
au	au	DET:ms	argument	il	PRON:ms3s
	au	PREP:Inv	était "	a	N2:ms:mp
referendum	referendum	N2:ms:mp	une	avoir	V1:P3s
	referendum	XINC	pression	indiqué	A32:ms
est	est	A80:ms:fs:mp:fp	des	indiquer	V3:Kms
	est	N2:ms	partisans	que	ADV:Inv
	être	V2:P3s	du +	que	CNJS:CSu
à	à	PREP:Inv	oui+	que	PRON:ms:fs:mp:fp
l'	l	N2:ms:mp	sur	ce	DET:ms
	le	DET:ms:fs		argument	N1:ms
	le	PRON:ms3s:fs3s		être	V2:I3s
origine	origine	N21:fs		une	N21:fs
	originer	V3:P1s:P3s:S1s:S3s:Y2s		un	A32:fs
de	de	PREP:Inv		un	DET:fs
	du	DET:mp:fp		un	PRON:fs
la	la	N2:ms:mp		un	N21:fs
	le	DET:fs		un	DET:mp:fp
	le	PRON:fs3s		un	A32:mp
déconfiture	déconfiture	N21:fs		un	N32:mp
de	de	PREP:Inv		du	DET:ms
	du	DET:mp:fp		du	PREP:Inv
la	la	N2:ms:mp		oui	ADV:Inv
	le	DET:fs		oui	N2:ms:mp
	le	PRON:fs3s		sur	(PN):Xin

## Programme concord

- André Lajoinie a estimé jeudi que 'c'est un mensonge de prétendre que la me  
e d'une victoire du «non» au referendum «est à l'origine de la déconfiture de  
1  
anciers. Il a indiqué que cet argument «était une pression des partisans du  
ncy-Lorraine de la communication 1992 a «été attribué, dans le cadre du Livre  
s  
omoteur immobilier Christian Pellerin a «été inculpé jeudi de «delit  
d'initié».  
«inculpé jeudi de «delit d'initié». Il «est soupçonné d'avoir spéculé sur des  
stissements, sur la Société Générale ne «soit connu du public. PARIS -  
Rapports  
dre différent : 15.825,00 Générale ne «soit connu du public. PARIS -  
Rapport  
, sans l'Union Economique et Monétaire, «est plus fragile». «Les  
spéculations  
férence sur la paix au Proche Orient se «sont achevées jeudi sans qu'une  
nouvel  
ées jeudi sans qu'une nouvelle date ait «été fixée pour une prochaine  
rencontre  
encontre. Des propositions variées ont «été faites pour améliorer le climat  
de  
erre Bérégovoy a estimé que le «débat a «été pollué par des calculs  
politiciens  
lorera vendredi sur le nord-ouest, mais «sera orageux ailleurs. Les  
température  
de 20 à 27 degrés du nord-ouest au sud-est >COPENHAGUE - La Banque  
centra  
NHAGUE - La Banque centrale du Danemark «est intervenue jeudi après-midi pour  
s  
our soutenir la couronne danoise. Elle «était tombée à son cours plancher au  
s  
NGTON - La crise que traverse le SME «n'est pas seulement le reflet des peurs  
mé jeudi que «l'agriculture française s'est développée grâce à l'Europe».  
«Si  
sa vitalité, les agriculteurs français «seraient frappés de plein fouet».  
NEW  
is ses gains face à la livre, le dollar «était en léger recul face à  
l'ensemble  
90 FF. VALENCE - 80 gendarmes mobiles «sont intervenus jeudi après-midi avec  
directeur et journaliste du Monde. Ils «étaient poursuivis par M. Paul  
Barril.

tant l'usage du préservatif. La montre «est munie d'un préservatif caché s.  
trée en Europe. «L'économie américaine «est toujours la plus forte du mond.  
rte du monde». LONDRES - Le chômage s'«est fortement accru en  
Grande-Bretagne

## Annexe 7

### Extraction des couches lexicales

N.B. Il est dans la nature même de la méthode subjective de présenter une oscillation de jugement sur un certain nombre de mots. A plusieurs mois de distance et en revoyant notre sélection, nous nous sommes interrogée par exemple sur un certain nombre de mots. Ainsi au vu de l'échantillon de la lettre A de la couche C1 nous nous sommes demandée si les mots *amazonien*, *ambidextre*, *ancillaire*, *anémie*, *anagramme* ne pourraient pas figurer dans la couche C2. Cependant, au regard des mots de la couche C2, on verra que la décision n'est pas évidente à prendre.

Nous pensons qu'en fonction des applications au traitement automatique, les utilisateurs procéderont aux ajustements ad hoc des couches, si besoin.

alto,,N1.N31 /4	amèrement,,ADV /4	ampoulé,,A32 /4
altruisme,,N1 /4	américain,,N1.N32.A32 /4	amputation,,N21 /4
altruiste,,N31.A31 /4	américaine,,N21 /4	amputé,,N32.A32 /4
alu,,N1 /4	américanisation,,N21 /4	amputer,,V3 /4
aluminium,,N1 /4	américaniser,,V3 /4	amusant,,N2S.A32 /4
alvéole,,N1.N21 /4	américano,,N1.XINC /5	amusement,,N1 /4
amabilité,,N21 /4	amertume,,N21 /4	amuser,,V3 /4
amadouer,,V3 /4	ameublement,,N1 /4	amuseur,,N35 /4
amaigri,,A32 /4	ameuter,,V3 /4	amygdale,,N21 /4
amaigrir,,V18 /4	ami,,N32.A32 /4	anachronique,,A31 /4
amaigrissant,,A32 /4	amiable,,A31.(NAN) /4	anachroniquement,,ADV /4
amaigrissement,,N1 /4	amical,,A76 /4	anachronisme,,N1 /4
amalgamant,,A32 /5	amicale,,N21 /4	anagramme,,N21 /4
amalgame,,N1 /4	amicalement,,ADV /4	anal,,A76 /4
amalgamer,,V3 /4	amidonner,,V3 /4	anale,,N21 /5
amande,,N21 /4	amie,,N21 /5	analogie,,N21 /4
amandier,,N1 /4	amincissant,,A32 /4	analogique,,A31 /4
amant,,N32 /4	amincissement,,N1 /4	analogue,,N1.A31 /4
amarante,,N21.A80 /4	amiral,,N76.A76 /4	analphabète,,N31.A31 /4
amarre,,N21 /4	amitié,,N21 /4	analphabétisme,,N1 /4
amarrer,,V3 /4	amnésie,,N21 /4	analysable,,A31 /4
amas,,N2 /4	amnésique,,N31.A31 /4	analysant,,N32 /4
amasser,,V3 /4	amnistie,,N21 /4	analyse,,N21 /4
amateur,,N31.A31 /4	amnistié,,N32.A32 /4	analysé,,N31 /4
amateurisme,,N1 /4	amnistier,,V3 /4	analyser,,V3 /4
amazone,,N21 /4	amocher,,V3 /4	analyseur,,N1.A35 /4
amazonien,,N41.A41 /4	amoindrir,,V18 /4	analyste,,N31 /4
ambages,,N22P /4	amonceler,,V9 /4	analytique,,N21.A31 /4
ambassade,,N21 /4	amoncellement,,N1 /4	analytiquement,,ADV /4
ambassadeur,,N36 /4	amont,,N1.A80 /4	ananas,,N2 /4
ambiance,,N21 /4	amoral,,A76 /4	anar,,N31.A31 /4
ambient,,A32 /4	amoralité,,N21 /5	anarchie,,N21 /4
ambidextre,,N31.A31 /4	amorçage,,N1 /4	anarchique,,A31 /4
ambigu,,N1.A50 /4	amorce,,N21 /4	anarchiquement,,ADV /4
ambiguïté,,N21 /4	amorcer,,V4 /4	anarchisme,,N1 /4
ambitueusement,,ADV /4	amorphe,,A31 /4	anarchiste,,N31.A31 /4
ambitieux,,N63.A63 /4	amorti,,N1.N32.A32 /4	anathème,,N1.N31.A31 /4
ambition,,N21 /4	amortir,,V18 /4	anatomie,,N21 /4
ambivalence,,N21 /4	amortissable,,A31 /4	anatomique,,A31 /4
ambivalent,,A32 /4	amortissement,,N1 /4	ancestral,,A76 /4
ambre,,N1.A80 /4	amortisseur,,N1.A35 /4	ancêtre,,N31 /4
ambré,,A32 /4	amour,,N1.N6 /4	anchois,,N2 /4
ambulance,,N21 /4	amouracher,,V3 /4	ancien,,N1.N41.A41 /4
ambulancier,,N42.A42 /4	amourette,,N21 /4	anciennement,,ADV /4
ambulant,,N32.A32 /4	amoureuseusement,,ADV /4	ancienneté,,N21 /4
âme,,N21 /4	amoureux,,N63.A63 /4	ancillaire,,A31 /4
améliorable,,A31 /4	amovible,,A31 /4	ancrage,,N1 /4
améliorant,,A32 /4	amphithéâtre,,N1 /4	ancre,,N21 /4
amélioration,,N21 /4	amphitryon,,N1 /4	ancrer,,V3 /4
améliorer,,V3 /4	amphore,,N21 /4	andalou,,N1.N33.A33 /4
amen,,N2.ADV /4	ample,,A31 /4	andouille,,N21.A80 /4
aménageable,,A31 /4	amplement,,ADV /4	andouillette,,N21 /4
aménagement,,N1 /4	ampleur,,N21 /4	âne,,N1.N39 /4
aménager,,V5 /4	ampli,,N1 /4	anéantir,,V18 /4
amende,,N21 /4	amplifiant,,A32 /4	anéantissement,,N1 /4
amène,,A31 /4	amplificateur,,N1.A36 /4	anecdote,,N21 /4
amener,,V6 /4	amplification,,N21 /4	anecdotique,,A31 /4
aménité,,N21 /4	amplifié,,A32 /4	anecdotiquement,,ADV /4
amenuiser,,V3 /4	amplifier,,V3 /4	anémie,,N21 /4
amer,,N1.A42 /4	ampoule,,N21 /4	anémié,,A32 /4

endiguer, V3 /4  
 endimanché, A32 /4  
 endimancher, V3 /4  
 endive, N21 /4  
 endoctrinement, N1 /4  
 endoctriner, V3 /4  
 endolori, A32 /4  
 endommagement, N1 /4  
 endommager, V5 /4  
 endormant, A32 /4  
 endormi, A32 /4  
 endormir, V26 /4  
 endosser, V3 /4  
 endroit, N1 /4  
 enduire, V91 /4  
 enduit, N1 /4  
 endurance, N21 /4  
 endurent, A32 /4  
 endurci, A32 /4  
 endurcir, V18 /4  
 endureissement, N1 /4  
 endurent, V3 /4  
 énergétique, N21.A31 /4  
 énergie, N21 /4  
 énergique, A31 /4  
 énergiquement, ADV /4  
 énergumène, N31 /4  
 énervant, A32 /4  
 énervé, N32.A32 /4  
 énervement, N1 /4  
 énerver, V3 /4  
 enfance, N21 /4  
 enfant, N1.N31 /4  
 enfantement, N1 /4  
 enfanté, V3 /4  
 enfantillage, N1 /4  
 enfantin, A32 /4  
 enfer, N1 /4  
 enfermer, V3 /4  
 enfermer, V3 /4  
 enfilade, N21 /4  
 enfiler, V3 /4  
 enfin, ADV /4  
 enflammé, A32 /4  
 enflammer, V3 /4  
 enflé, A32 /4  
 enfler, V3 /4  
 enflure, N21 /4  
 enfoiré, A32 /4  
 enfoncé, A32 /4  
 enfoncement, N1 /4  
 enfoncer, V4 /4  
 enfouir, V18 /4  
 enfourcher, V3 /4  
 enfoutisme, N1(XN) /7  
 enfoutiste, N31.A31(XA) /7  
 enfreindre, V64 /4  
 enfuir, V27 /4  
 enfumé, A32 /5  
 enfumer, V3 /4  
 engagé, N32.A32 /4  
 engageant, A32 /4  
 engagement, N1 /4  
 engager, V5 /4  
 engageance, N21 /4  
 engendrement, N1 /4  
 engendrer, V3 /4  
 engin, N1 /4  
 englober, V3 /4  
 engloutir, V18 /4  
 engloutissement, N1 /4  
 engluer, V3 /4  
 engoncer, V4 /4  
 engorgement, N1 /4  
 engorger, V5 /4  
 engouement, N1 /4  
 engouer, V3 /4  
 engouffrer, V3 /4  
 engourdi, A32 /4  
 engourdir, V18 /4  
 engourdisant, A32 /5  
 engourdissement, N1 /4  
 engrais, N2 /4  
 engraissement, N1 /4  
 engraisser, V3 /4  
 engranger, V5 /4  
 engrosser, V3 /4  
 engueulade, N21 /4  
 engueuler, V3 /4  
 enguirlander, V3 /4  
 enhardir, V18 /4  
 énième, N31.A31 /4  
 énigmatique, A31 /4  
 énigmatiquement, ADV /4  
 énigme, N21 /4  
 enivrant, A32 /4  
 enivrement, N1 /4  
 enivrer, V3 /4  
 enjambée, N21 /4  
 enjambement, N1 /4  
 enjamber, V3 /4  
 enjeu, N3 /4  
 enjôler, V3 /4  
 enjôleur, N35.A35 /4  
 enjoliver, V3 /4  
 enjoué, A32 /4  
 enjouement, N1 /4  
 enjouer, V3 /4  
 enlacement, N1 /4  
 enlacer, V4 /4  
 enlaidir, V18 /4  
 enlaidissant, A32 /5  
 enlaidissement, N1 /4  
 enlevé, A32 /4  
 enlèvement, N1 /4  
 enlever, V6 /4  
 enlissement, N1 /4  
 enliser, V3 /4  
 enneigé, A32 /4  
 enneigement, N1 /4  
 ennemi, N32.A32 /4  
 ennui, N1 /4  
 ennuyant, A32 /4  
 ennuyé, A32 /4  
 ennuyer, V13 /4  
 ennuyeux, A63 /4  
 énoncé, N1 /4  
 énoncer, V4 /4  
 enorgueillir, V18 /4  
 énorme, A31 /4  
 énormément, ADV /4  
 énorment, N21 /4  
 enquêter, V20 /4  
 enquête, N21 /4  
 enquêter, V3 /4  
 enquêteur, N36.N35.A35 /4  
 enquiérent, A32 /4  
 enquiérement, N1 /4  
 enquiéner, V3 /4  
 enquiéneur, N35.A35 /4  
 enracinement, N1 /4  
 enraciner, V3 /4  
 enragé, N32.A32 /4  
 enrager, V5 /4  
 enrayer, V14 /4  
 enregistrement, N1 /4  
 enregistrer, V3 /4  
 enrhumé, A32 /4  
 enrhumé, V3 /4  
 enrichi, A32 /4  
 enrichir, V18 /4  
 enrichissant, A32 /4  
 enrichissement, N1 /4  
 enrober, V3 /4  
 enrôlé, N32 /4  
 enrôlement, N1 /4  
 enrôler, V3 /4  
 enrouer, V3 /4  
 enruler, V3 /4  
 ensablement, N1 /4  
 ensanglantement, N1 /4  
 ensanglanter, V3 /4  
 enseignant, N32.A32 /4  
 enseigne, N1.N21 /4  
 enseigné, N32.A32 /4  
 enseignement, N1 /4  
 enseigner, V3 /4  
 ensemble, N1.ADV /4  
 enserrer, V3 /4  
 ensevelir, V18 /4  
 ensevelissement, N1 /4  
 ensoleillé, A32 /4  
 ensoleillement, N1 /4  
 ensoleiller, V3 /4  
 ensommeillé, A32 /4  
 ensommeillement, N1 /4  
 ensommeiller, V3 /4  
 ensorcelant, A32 /4  
 ensorceler, V9 /4  
 ensorceleur, N35.A35 /4

perception,,N21 /4  
 percer,,V4 /4  
 perceuse,,N21 /4  
 percevable,,A31 /4  
 percevoir,,V46 /4  
 perché,,N1.A32 /4  
 percher,,V3 /4  
 perchoir,,N1 /4  
 perclus,,A61 /4  
 percolateur,,N1 /4  
 percussion,,N21 /4  
 percussionniste,,N31 /4  
 percutant,,A32 /4  
 percuter,,V3 /4  
 perdant,,N32.A32 /4  
 perte,,N21 /4  
 perdre,,V67 /4  
 perdreau,,N3 /4  
 perdrix,,N22 /4  
 perdu,,N1.A32 /4  
 perdurable,,A31 /4  
 perdurer,,V3U /4  
 père,,N1 /4  
 pérégrination,,N21 /4  
 péremptoire,,A31 /4  
 péremptoirement,,ADV /4  
 pérennisation,,N21 /4  
 pérenniser,,V3 /4  
 péréquation,,N21 /4  
 perestroïka,,N22S /7  
 perfectible,,A31 /4  
 perfection,,N21 /4  
 perfectionné,,A32 /4  
 perfectionnement,,N1 /4  
 perfectionner,,V3 /4  
 perfectionnisme,,N1 /4  
 perfectionniste,,N31.A31 /4  
 perfide,,N31.A31 /4  
 perforation,,N21 /4  
 perforé,,A32 /4  
 perforer,,V3 /4  
 performance,,N21 /4  
 performant,,A32 /4  
 perfuser,,V3 /4  
 perfusion,,N21 /4  
 péri,,N21+.N1.A32.(PFX) /4  
 péricliter,,V3U /4  
 péril,,N1 /4  
 périlleusement,,ADV /4  
 périlleux,,A63 /4  
 périmé,,A32 /4  
 périmé,,V3 /4  
 périmètre,,N1 /4  
 période,,N21.N1 /4  
 périodicité,,N21 /4  
 périodique,,A31 /6  
 périodique,,N1.A31 /4  
 périodiquement,,ADV /4  
 péripatéticienne,,N21 /4  
 péripétie,,N21 /4

périphérie,,N21 /4  
 périphérique,,N1.A31 /4  
 périphériquement,,ADV /4  
 périphrase,,N21 /4  
 périple,,N1 /4  
 périr,,V18 /4  
 périscope,,N1 /4  
 périssable,,A31 /4  
 péristyle,,N1.A31 /4  
 perle,,N21 /4  
 perlé,,A32 /4  
 perler,,V3 /4  
 perlimpinpin,,N1 /4  
 permanence,,N21 /4  
 permanent,,N32.A32 /4  
 permanente,,N21 /4  
 perméabilité,,N21 /4  
 perméable,,A31 /4  
 permettre,,V71 /4  
 permis,,N2 /4  
 permissif,,A38 /4  
 permission,,N21 /4  
 permissionnaire,,N1 /4  
 permissivité,,N21 /4  
 permutable,,A31 /4  
 permutation,,N21 /4  
 permuter,,V3 /4  
 pernicieusement,,ADV /4  
 pernicieux,,A63 /4  
 péroration,,N21 /4  
 pérorer,,V3U /4  
 perpendiculaire,,N21.A31 /4  
 perpendiculairement,,ADV /4  
 perpète,,XINC /4  
 perpétrer,,V7 /4  
 perpette,,XINC /4  
 perpétuation,,N21 /4  
 perpétuel,,A40 /4  
 perpétuellement,,ADV /4  
 perpétuer,,V3 /4  
 perpétuité,,N21 /4  
 perplexe,,A31 /4  
 perplexité,,N21 /4  
 perquisition,,N21 /4  
 perquisitionner,,V3 /4  
 perron,,N1 /4  
 perroquet,,N1 /4  
 perruche,,N21 /4  
 perruque,,N21 /4  
 persécuté,,N32.A32 /4  
 persécuter,,V3 /4  
 persécuteur,,N36.A36 /4  
 persécution,,N21 /4  
 persévérance,,N21 /4  
 persévérant,,N32.A32 /4  
 persévérer,,V7U /4  
 persienne,,N21 /4  
 persiflage,,N1 /4  
 persifler,,V3 /4  
 persifleur,,N35.A35 /4

persil,,N1 /4  
 persillade,,N21 /4  
 persillé,,N1.A32 /4  
 persiller,,V3 /4  
 persique,,A31 /4  
 persistance,,N21 /4  
 persistant,,A32 /4  
 persister,,V3U /4  
 personnage,,N1 /4  
 personnalisation,,N21 /4  
 personnaliser,,V3 /4  
 personnalité,,N21 /4  
 personne,,N21.PRON /4  
 personnel,,N1.A40 /4  
 personnellement,,ADV /4  
 personnification,,N21 /4  
 personnifié,,A32 /4  
 personnifier,,V3 /4  
 perspective,,N21 /4  
 perspicace,,A31 /4  
 perspicacement,,ADV /4  
 perspicacité,,N21 /4  
 persuader,,V3 /4  
 persuasif,,A38 /4  
 persuasion,,N21 /4  
 perte,,N21 /4  
 pertinence,,ADV /4  
 pertinence,,N21 /4  
 pertinent,,A32 /4  
 perturbant,,A32 /5  
 perturbateur,,N36.A36 /4  
 perturbation,,N21 /4  
 perturber,,V3 /4  
 péruvien,,N41.A41 /4  
 pervers,,N61.A61 /4  
 perversion,,N21 /4  
 perversité,,N21 /4  
 perverti,,N32.A32 /4  
 pervertir,,V18 /4  
 pervertissement,,N1 /4  
 pesamment,,ADV /4  
 pesant,,N1.A32 /4  
 pesanteur,,N21 /4  
 peser,,V6 /4  
 peseta,,N21 /4  
 pesette,,N21 /4  
 peso,,N1 /4  
 pessimisme,,N1 /4  
 pessimiste,,N31.A31 /4  
 peste,,N21 /4  
 pester,,V3U /4  
 pesticide,,N1.A31 /4  
 pestiféré,,N32.A32 /4  
 pestilence,,N21 /4  
 pestilentiel,,A40 /4  
 pet,,N1 /4  
 pétainisme,,N1 /4  
 pétainiste,,N31.A31 /5  
 pétale,,N1 /4  
 pétanque,,N21 /4

Couche C2 : échantillons prélevés sur les lettres A, E, P

alourdissable,,A31 /4	ambien,,N1.A41 /4	andantino,,N1.ADV /4
aloyau,,N3 /4	amidon,,N1 /4	androgynie,,N31.A31 /4
alpaga,,N1 /4	amidonnage,,N1 /4	andropause,,N21 /4
alpestre,,A31 /4	amincir,,V18 /4	anémiant,,A32 /5
alpha,,N2.A80(NA) /4	amirauté,,N21 /4	anesthésiable,,A31 /5
alphabétiquement,,ADV /4	ammoniac,,N1.A46 /4	anesthésique,,N1.A31 /4
alphabétisme,,N1 /4	amunoniacal,,A76 /4	anévrisme,,N1 /4
alpin,,A32 /4	ammoniaque,,N21 /4	anévrisme,,N1 /4
altérabilité,,N21 /4	ammoniaqué,,A32 /5	anfractuosité,,N21 /4
altérant,,N1.A32 /4	amniocentèse,,N21 /4	angevin,,N32.A32 /4
altérité,,N21 /4	amniotique,,A31 /4	angineux,,N63.A63 /4
alterateur,,N1 /4	amochage,,N1 /4	angiome,,N1 /4
alternation,,N21 /4	amochir,,V18 /4	anglicisant,,N32.A32 /4
alternativité,,N21 /5	amoindrissant,,A32 /5	anglomane,,N31.A31 /4
altièrement,,ADV /4	amoindrissement,,N1 /4	anglomanie,,N21 /4
altimètre,,N1 /4	amollir,,V18 /4	anglophilie,,N21 /4
altocumulus,,N2 /4	amollissant,,A32 /4	anglophobie,,N21 /4
altostratus,,N2 /4	amollissement,,N1 /4	angolais,,N61.A61 /4
altruïstement,,ADV /4	amoralement,,ADV /4	angstroem,,N1 /4
altuglas,,N2 /4	amoralisme,,N1 /4	angström,,N1 /4
alunage,,N1 /4	amoraliste,,N31.A31 /5	anguilliforme,,N1.A31 /5
aluner,,V3 /4	amortie,,N21 /4	angulairement,,ADV /4
alunir,,V18 /4	amourettes,,N22P /4	angulation,,N21 /4
alunissage,,N1 /4	ampérage,,N1 /4	anhydride,,N1 /4
alvéolaire,,A31 /4	ampère,,N1 /4	anhydrite,,N21 /4
alvéolé,,A32 /4	amphétamine,,N21 /4	anima,,N2S /5
amadou,,N1 /4	amphétaminique,,A31 /4	animalcule,,N1 /4
amandine,,N21 /4	amphi,,N1.(PFX) /4	animalerie,,N21 /4
amanite,,N21 /4	amphibie,,N1.A31 /4	animalier,,N42.A42 /4
amarrage,,N1 /4	amphibien,,N1 /4	animaliser,,V3 /4
amaryllis,,N2.N22 /4	amphigouri,,N1 /4	animisme,,N1 /4
ambigument,,ADV /4	amphigourique,,A31 /4	animiste,,N31.A31 /4
ambitionner,,V3 /4	amplitude,,N21 /4	ankylosant,,A32 /5
amble,,N1 /4	amuïr,,V38 /4	ankylose,,N21 /4
amblyope,,N31.A31 /4	amuïssement,,N1 /4	ankyloser,,V3 /4
amblyopie,,N21 /4	amulette,,N21 /4	annal,,A76 /4
ambler,,V3 /4	amusable,,A31 /5	annalité,,N21 /4
ambroisie,,N21 /4	amusette,,N21 /4	annexionnisme,,N1 /4
ambulatoire,,A31 /4	anabaptisme,,N1 /4	annonciation,,N21 /4
amendable,,A31 /4	anabaptiste,,N31 /4	annualiser,,V3 /4
amendement,,N1 /4	anabolique,,A31 /4	annulabilité,,N21 /4
amender,,V3 /4	anabolisant,,N1.A32 /4	anoblissant,,A32 /5
aménorrhée,,N21 /4	anachorète,,N31 /4	anodisation,,N21 /4
amenuisement,,N1 /4	anagogie,,N21 /4	anodiser,,V3 /4
américanisme,,N1 /4	anagogique,,A31 /4	ânonnement,,N1 /4
américaniste,,N31.A31 /4	analgésiant,,N1.A32 /5	anovulation,,N21 /5
américanophile,,A31 /5	analgésique,,N1.A31 /4	antagonique,,A31 /4
américanophobe,,A31 /5	analité,,N21 /4	antécédemment,,ADV /4
amérindien,,N41.A41 /4	analogiquement,,ADV /4	antéchrist,,N1 /4
amerlo,,N31 /4	analoguement,,ADV /4	antéposé,,A32 /4
amerloque,,N31 /4	anaphore,,N21 /4	antéposer,,V3 /4
amerlot,,N1 /4	anaphorique,,N1.A31 /4	antéposition,,N21 /4
amerrier,,V18 /4	anarchisant,,N32.A32 /4	antériorisation,,N21 /4
amerissage,,N1 /4	anarcho,,N31.XINC /5	antérioriser,,V3 /4
améthyste,,N1.N21.A80 /4	anatolien,,N1.A41 /7	antéropostérieur,,A32 /5
ameubler,,V18 /4	anatomiquement,,ADV /4	antérosupérieur,,A32 /5
ameublisement,,N1 /4	anatomiste,,N31 /4	anthologique,,A31 /4
amiablement,,ADV /4	anchoyade,,N21 /5	anthropo,,(PFX) /7
amiante,,N1 /4	ancistrodon,,N1 /5	anthropologiste,,N31 /4
amibe,,N21 /4	andante,,N1.ADV /4	anthropomorphe,,A31 /4

ensacher, V3 /4  
 enseignable, A31 /4  
 ensemblier, N42 /4  
 ensemencement, N1 /4  
 ensemençer, V4 /4  
 ensilage, N1 /4  
 entame, N21 /4  
 entarrage, N1 /4  
 entartrement, N1 /4  
 entarter, V3 /4  
 entérinement, N1 /4  
 entériner, V3 /4  
 entérite, N21 /4  
 enterrage, N1 /4  
 enthousiastement, ADV /4  
 entichement, N1 /4  
 entoilage, N1 /4  
 entoiler, V3 /4  
 entomologie, N21 /4  
 entomologique, A31 /4  
 entomologiste, N31 /4  
 entonnement, N1 /4  
 entortillage, N1 /4  
 entortillement, N1 /4  
 entraînable, A31 /4  
 entrapparaître, V69 /5  
 entravé, A32 /4  
 entrechat, N1 /4  
 entrechoquement, N1 /4  
 entrecouplement, N1 /4  
 entrecroisement, N1 /4  
 entrecroiser, V3 /4  
 entrecuisse, N1 /4  
 entrefermer, V3 /4  
 entregent, N1 /4  
 entrejambe, N1 /4  
 entrelacs, N2 /4  
 entrelardé, A32 /4  
 entrelarder, V3 /4  
 entremarier, V3 /5  
 entremetteur, N35 /4  
 entremettre, V71 /4  
 entrepénétrer, V7 /4  
 entreposage, N1 /4  
 entreposeur, N35 /4  
 entretailer, V3 /4  
 entretoise, N21 /4  
 entrevoie, N21 /4  
 entropie, N21 /4  
 enturbanné, A32 /4  
 enturbanner, V3 /5  
 énucléation, N21 /4  
 énucléer, V3 /4  
 énumérable, A31 /4  
 énumératif, A38 /4  
 énurésie, N21 /4  
 envaser, V3 /4  
 envenimement, N1 /4  
 environnemental, A76 /4  
 envoûteur, N35 /4

envoyeur, N35 /4  
 enzyme, N21+.N1 /4  
 éolien, N1.N41.A41 /4  
 éolienne, N21 /4  
 éosine, N21 /4  
 épaisseur, ADV /4  
 épaisseur, N1.A35 /4  
 épandage, N1 /4  
 épandre, V67 /4  
 épatamment, ADV /4  
 épate, N21 /4  
 épatement, N1 /4  
 épépinage, N1 /4  
 épépiner, V3 /4  
 éperlan, N1 /4  
 éperon, N1 /4  
 éperonnement, N1 /4  
 éperonner, V3 /4  
 épervier, N1 /4  
 éphémèrement, ADV /4  
 éphéméride, N21 /4  
 épiage, N1 /4  
 épicea, N1 /4  
 épicurienement, ADV /4  
 épidémiologie, N21 /4  
 épidémiologique, A31 /4  
 épidémiquement, ADV /4  
 épiderme, N1 /4  
 épierrage, N1 /4  
 épierrement, N1 /4  
 épierrer, V3 /4  
 épieu, N3 /4  
 épigastre, N1 /4  
 épiglotte, N21 /4  
 épigone, N1 /4  
 épigramme, N21 /4  
 épigraphe, N21 /4  
 épilage, N1 /4  
 épilatoire, A31 /4  
 épinette, N21 /4  
 épinglage, N1 /4  
 épiphyse, N21 /4  
 épiquement, ADV /4  
 épiscopalement, ADV /4  
 épiscopalien, N41.A41 /4  
 épiscopalisme, N1 /4  
 épisiotomie, N21 /4  
 épistème, N21 /4  
 épistémique, A31 /5  
 épistémologie, N21 /4  
 épistémologique, A31 /4  
 épistémologiquement, ADV /4  
 épistémologiste, N31 /4  
 épistémologue, N31 /4  
 épistolaiement, ADV /4  
 épistolier, N42 /4  
 épithalame, N1 /4  
 épithélium, N1 /4  
 épithétique, A31 /4  
 épithétisation, N21 /4

épithétiser, V3 /5  
 épizone, N21 /5  
 épitage, N1 /4  
 épitement, N1 /4  
 épinter, V3 /4  
 épongeage, N1 /4  
 éponyme, N1.A31 /4  
 épouillage, N1 /4  
 épouiller, V3 /4  
 épousailles, N22P /4  
 épouseur, N35 /4  
 epsilon, N2 /4  
 épucer, V4 /4  
 épuisable, A31 /4  
 épuisette, N21 /4  
 épuration, N1 /4  
 épurateur, N1.N35.A35 /4  
 épuratif, A38 /4  
 épure, N21 /4  
 épurement, N1 /4  
 équarrissage, N1 /4  
 équarrissement, N1 /4  
 équarrisseur, N35 /4  
 équarri-soir, N1 /4  
 équationnel, A40 /5  
 équeutage, N1 /4  
 équeuter, V3 /4  
 équilatère, A31 /4  
 équilibrer, N1.A35 /4  
 équille, N21 /4  
 équipollent, N1.A32 /4  
 éradication, N21 /4  
 éradiquer, V3 /4  
 éraflage, N1 /4  
 éraffement, N1 /4  
 érailler, V3 /4  
 érecteur, N1.A36 /4  
 érectile, A31 /4  
 érectilité, N21 /4  
 erectus, A2S /5  
 ergonome, N31 /4  
 ergonomie, N21 /4  
 ergonomique, A31 /4  
 ergonomiquement, ADV /4  
 ergot, N1 /4  
 ergothérapie, N21 /4  
 érodabilité, N21 /5  
 érodable, A31 /4  
 érosif, A38 /4  
 érosivité, N21 /5  
 érotisant, A32 /5  
 érotisation, N21 /4  
 érotiser, V3 /4  
 érotogène, A31 /4  
 érotologie, N21 /4  
 érotomane, N31.A31 /4  
 érotomanie, N21 /4  
 erratique, A31 /4  
 erratiquement, ADV /4  
 erse, N1.N21.A31 /4

périssabilité, N21 /5  
 périltoine, N1 /4  
 perlingual, A76 /4  
 perliouse, N21 /4  
 perlouze, N21 /4  
 permanencier, N42 /5  
 permanent, V3 /4  
 permanganate, N1 /4  
 perméabiliser, V3 /4  
 permutable, A31 /4  
 permisivement, ADV /4  
 permittivité, N21 /4  
 permutableté, N21 /4  
 péroné, N1 /4  
 péronisme, N1 /4  
 péroniste, N31.A31 /4  
 péronnelle, N21 /4  
 péroreur, N35.A35 /4  
 perpendiculaireté, N21 /5  
 perpétration, N21 /4  
 perplexement, ADV /4  
 perquisiteur, A36 /4  
 perquisitionneur, N35 /4  
 perruquier, N42 /4  
 pers, A61 /4  
 persan, N1.N32.A32 /4  
 perse, N1.N21.N31.A31 /4  
 persécuteur, A38 /5  
 persona, XINC /5  
 personnalisme, N1 /4  
 personnaliste, N31.A31 /4  
 perspectif, A38 /4  
 perspectivisme, N1 /4  
 perspectiviste, N31 /5  
 persuasivement, ADV /4  
 pertuis, N2 /4  
 pervenche, N21.A80 /4  
 perversement, ADV /4  
 pervertisseur, N35.A35 /4  
 pesage, N1 /4  
 pesée, N21 /4  
 pessimistement, ADV /4  
 pesteux, A63 /4  
 pétinisme, N1 /4  
 pétiniste, N31.A31 /5  
 pétiole, N1 /4  
 pétitionnaire, N31 /4  
 pétitionnement, N1 /4  
 pétitionner, V3 /4  
 pétiochard, N32.A32 /7  
 pétioire, N21 /4  
 peton, N1 /4  
 pétoncle, N1 /4  
 pétrarquiser, V3 /4  
 pétrarquisme, N1 /4  
 pétrarquiste, N31.A31 /4  
 pétrissable, A31 /4  
 pétrochimie, N21 /4  
 pétrochimique, A31 /4  
 pétrolette, N21 /4  
 peupleraie, N21 /4  
 phaéton, N1 /4  
 phagocyte, N1 /4  
 phagocytose, N21 /4  
 phalangette, N21 /4  
 phalangien, A41 /4  
 phalangiste, N31.A31 /4  
 phalanstère, N1 /4  
 phalène, N21+.N1 /4  
 phallisme, N1 /4  
 phallocratie, N21 /4  
 phallocratique, A31 /4  
 phalocratie, N1 /4  
 phantasme, N1 /4  
 pharamineux, A63 /4  
 pharaonien, A41 /4  
 pharaonique, A31 /4  
 pharisaïsme, N1 /4  
 pharisien, N41 /4  
 pharmacologiquement, ADV /4  
 pharmacologiste, N31 /4  
 pharmacologue, N31 /4  
 pharmacopée, N21 /4  
 pharyngal, A76 /4  
 pharyngé, A32 /4  
 pharyngien, A41 /4  
 phasé, A32 /5  
 phatique, A31 /4  
 phénicien, N1.N41.A41 /4  
 phénobarbital, N1 /4  
 phénoménologie, N21 /4  
 phénoménologique, A31 /4  
 phénoménologiquement, ADV /4  
 phénotype, N1 /4  
 phénotypique, A31 /4  
 philanthropement, ADV /4  
 philanthropiquement, ADV /4  
 philanthropisme, N1 /4  
 philatélique, A31 /4  
 philatélie, N1 /4  
 philharmonie, N21 /4  
 philippine, N21 /4  
 philistin, N1.A32 /4  
 philodendron, N1 /4  
 philologie, N21 /4  
 philologique, A31 /4  
 philologiquement, ADV /4  
 philologue, N31 /4  
 philosophale, A21 /4  
 phlébite, N21 /4  
 phlébologue, N31 /4  
 phlegmon, N1 /4  
 phlox, N2 /4  
 phobiquement, ADV /4  
 phobogène, A31 /5  
 phocéen, N41.A41 /4  
 phoenix, N2 /4  
 phonème, N1 /4  
 phonéticien, N41 /4  
 phonétisation, N21 /4  
 phonétiser, V3 /4  
 phonie, N21 /4  
 phonique, A31 /4  
 phoniquement, ADV /4  
 phono, N1.(PFX) /4  
 phonogramme, N1 /4  
 phonographe, N1 /4  
 phonothèque, N21 /4  
 phosphatage, N1 /4  
 phosphate, N1 /4  
 phosphaté, A32 /4  
 phosphater, V3 /4  
 phosphore, N1 /4  
 phosphoré, A32 /4  
 phosphorescence, N21 /4  
 photocellule, N21 /4  
 photocomposer, V3 /5  
 photocomposeuse, N21 /4  
 photocomposition, N21 /4  
 photocopiste, N31 /5  
 photograver, V3 /4<gravure  
 photographeur, N35.A35 /4  
 photogravure, N21 /4  
 photojournalisme, N1 /4  
 photojournaliste, N31 /5  
 photomonteur, V3 /4  
 photon, N1 /4  
 photosensibilisant, N1.A32 /5  
 photosensibilisation, N21 /4  
 photosensibiliser, V3 /4  
 photosensibilité, N21 /4  
 photosensible, A31 /4  
 photostyle, N1 /4  
 photosynthèse, N21 /4  
 photothèque, N21 /4  
 photothérapie, N21 /4  
 photothérapie, A31 /4  
 phraséologie, N21 /4  
 phraséologique, A31 /4  
 phraser, V3U /4  
 phraseur, N35 /4  
 phréatique, A31 /4  
 phrygien, N1.N41.A41 /4  
 phlétis, N21 /4  
 phlétisique, N31.A31 /4  
 phylloxera, N1 /4  
 phylloxéra, N1 /4  
 physiocrate, N31.A31 /4  
 physiocratie, N21 /4  
 physiocratique, A31 /4  
 physiologisme, N1 /5  
 physiologiste, N31 /4  
 physiothérapie, N21 /4  
 physiothérapique, A31 /4  
 phyto, (PFX) /4  
 pi, N2 /4  
 piaf, N1 /4  
 piaffeur, A35 /4  
 piaillerie, N21 /4  
 piailleur, N35.A35 /4

## Couche C3 : échantillons prélevés sur les lettres A, E, P

alogie,.N21 /4	alumine,.N21 /4	amaril,.A32 /4
alogique,.A31 /4	aluminer,.V3 /4	amarillose,.N21 /5
aloine,.N21 /5	aluminerie,.N21 /5	amarinage,.N1 /4
alomancie,.N21 /7	alumineux,.A63 /4	amariner,.V3 /4
alopécie,.N21 /4	aluminium,.N1 /4	amarinier,.N1 /5
alopécique,.A31 /4	aluminier,.V3 /5	amarnien,.A41 /4
alose,.N21 /4	aluminifère,.A31 /5	amaryllidacée,.N21 /4
aloter,.V3 /5	aluminique,.A31 /4	amassette,.N21 /4
alouate,.N1 /4	aluminisation,.N21 /4	amateloter,.V3 /4
alouchier,.N1 /5	aluminiser,.V3 /4	amatir,.V18 /4
alouvir,.V18 /5	aluminite,.N21 /5	amatoxine,.N21 /5
aloyer,.V13 /4	alumino,(PFX) /7	amaurobius,.N2 /5
alpaguer,.V3 /4	aluminochlorure,.N1 /5	amaurose,.N21 /4
alpax,.N2 /4	aluminofluorure,.N1 /5	amaurotique,.A31 /4
alpe,.N21 /4	aluminose,.N21 /5	amazonite,.N21 /4
alpenstock,.N1 /4	aluminosilicate,.N1 /4	ambérique,.N21 /4
alper,.V3 /5	aluminothermie,.N21 /4	ambidextérité,.N21 /5
alphabète,.N31.A31 /4	aluminure,.N21 /5	ambidextrie,.N21 /5
alphabético,(PFX) /7	alumnat,.N1 /5	ambiéqual,.N76.A76 /5
alphabétiseur,.N35 /5	alumnite,.N21 /5	ambilatéralité,.N21 /5
alphabloquant,.N1.A32 /5	alun,.N1 /4	ambiophonie,.N21 /4
alphaglobuline,.N21 /5	alunation,.N21 /5	ambiophonique,.A31 /5
alphalytique,.A31 /5	aluné,.A1 /5	ambisexué,.A32 /4
alphanimétique,.A31 /5	alunerie,.N21 /5	ambix,.N2 /5
alphanumérique,.A31 /4	alunifère,.A31 /5	ambler,.V3 /4
alpharécepteur,.N1.A36 /5	alunite,.N21 /4	ambleur,.A35 /4
alphastimulant,.N1.A32 /5	alunogène,.N1 /5	amblyopode,.N1 /5
alphathérapie,.N21 /5	alurgite,.N21 /5	amblyopsidé,.N1 /5
alphatron,.N1 /5	alvéographe,.N1 /5	amblyopyge,.N1 /5
alphetomancie,.N21 /7	alvéolectomie,.N21 /5	amblyornis,.N2 /5
alpien,.A41 /5	alvéoline,.N21 /5	amblyoscope,.N1 /4
alpiste,.N1 /4	alvéolisation,.N21 /5	amblyostomidé,.N1 /5
alquifoux,.N2 /4	alvéoliser,.V3 /5	amblystome,.N1 /4
altaïque,.N1.A31 /4	alvéolite,.N21 /4	ambon,.N1 /4
altaïte,.N21 /5	alvéolo,(PFX) /5	ambréine,.N21 /4
altazimutal,.A76 /5	alvéolyse,.N21 /5	ambrette,.N21 /4
altéragène,.A31 /4	alveus,.N2 /5	ambroisien,.A41 /5
altératif,.A38 /5	alvier,.N1 /5	ambrosia,.N21 /5
alternat,.N1 /4	alvin,.A32 /4	ambrosiaque,.A31 /4
alternatrice,.N21 /4	alyde,.N1 /5	ambrosien,.A41 /4
alterme,.A31 /4	alymphocytose,.N21 /5	ambulacraire,.A31 /4
altermomoteur,.N1 /5	alysie,.N21 /5	ambulacral,.A76 /5
althaea,.N21+.N1 /4	alysse,.N21 /4	ambulacre,.N1 /4
althée,.N21+.N1 /5	alysson,.N1 /4	ambuler,.V3 /4
altimétrie,.N21 /4	alyssum,.N1 /5	ameive,.N1 /5
altimétrique,.A31 /4	alyte,.N1 /4	amélanche,.N21 /5
altimétriquement,.ADV /5	amabile,.ADV /5	amélanchier,.N1 /5
altiplanation,.N21 /5	amabiliser,.V3 /5	amélie,.N21 /5
altiport,.N1 /4	amadine,.N21 /5	amenage,.N1 /5
altise,.N21 /4	amadouvier,.N1 /4	aménagement,.N35 /4
altiste,.N31 /4	amalgamation,.N21 /4	aménagiste,.N31 /4
altisurface,.N21 /5	aman,.N1 /4	amenée,.N21 /4
altitudinal,.A76 /5	amandaie,.N21 /4	amènement,.ADV /4
alucite,.N21 /4	amanderaie,.N21 /5	ameneur,.N1 /5
alucité,.A32 /5	amandon,.N1 /5	amensal,.A76 /4
aluette,.N21 /4	amanitine,.N21 /5	amensalisme,.N1 /4
alumifère,.A31 /4	amarantacée,.N21 /4	amentale,.N21 /4
aluminage,.N1 /4	amaranthacée,.N21 /5	amentifère,.N1.A31 /4
alumine,.N21 /4	amarantite,.N21 /5	américium,.N1 /4
alumineux,.A63 /4	amareyeur,.N35 /4	amérindianiste,.N31 /5
aluminier,.V3 /5		
aluminifère,.A31 /5		
aluminique,.A31 /4		
aluminisation,.N21 /4		
aluminiser,.V3 /4		
aluminite,.N21 /5		
alumino,(PFX) /7		
aluminochlorure,.N1 /5		
aluminofluorure,.N1 /5		
aluminose,.N21 /5		
aluminosilicate,.N1 /4		
aluminothermie,.N21 /4		
aluminure,.N21 /5		
alumnat,.N1 /5		
alumnite,.N21 /5		
alun,.N1 /4		
alunation,.N21 /5		
aluné,.A1 /5		
alunerie,.N21 /5		
alunifère,.A31 /5		
alunite,.N21 /4		
alunogène,.N1 /5		
alurgite,.N21 /5		
alvéographe,.N1 /5		
alvéolectomie,.N21 /5		
alvéoline,.N21 /5		
alvéolisation,.N21 /5		
alvéoliser,.V3 /5		
alvéolite,.N21 /4		
alvéolo,(PFX) /5		
alvéolyse,.N21 /5		
alveus,.N2 /5		
alvier,.N1 /5		
alvin,.A32 /4		
alyde,.N1 /5		
alymphocytose,.N21 /5		
alysie,.N21 /5		
alysse,.N21 /4		
alysson,.N1 /4		
alyssum,.N1 /5		
alyte,.N1 /4		
amabile,.ADV /5		
amabiliser,.V3 /5		
amadine,.N21 /5		
amadouvier,.N1 /4		
amalgamation,.N21 /4		
aman,.N1 /4		
amandaie,.N21 /4		
amanderaie,.N21 /5		
amandon,.N1 /5		
amanitine,.N21 /5		
amarantacée,.N21 /4		
amaranthacée,.N21 /5		
amarantite,.N21 /5		
amareyeur,.N35 /4		

endothélialiser, V3 /5  
 endothéliochorial, A76 /5  
 endothélium, N1 /4  
 endothérapie, A31 /4  
 endothermique, A31 /4  
 endotoxine, N21 /4  
 endouzainer, V3 /4  
 endrailler, V3 /5  
 induction, N21 /4  
 enduiseur, N35 /4  
 enduro, N1 /4  
 endymion, N1 /4  
 énéma, N1 /4  
 énéo, (PFX) /7  
 énéolithique, N1.A31 /4  
 énergétique, N41 /4  
 énergétisme, N1 /4  
 énergétiste, N31.A31 /5  
 énervation, N21 /4  
 éneyer, V3 /4  
 enfaçonner, V3 /4  
 enfaîteau, N3 /4  
 enfaîtement, N1 /4  
 enfâter, V3 /4  
 enferrage, N1 /4  
 enfeu, N1 /4  
 enficeler, V9 /4  
 enfichisme, N1(XN) /7  
 enfichiste, N31.A31(XA) /7  
 enfieller, V3 /4  
 enfileur, N35 /4  
 enflaquer, V3 /4  
 enflécher, V7 /4  
 enfléchure, N21 /4  
 enfléurage, N1 /4  
 enfleurer, V3 /4  
 enfonceur, N35 /4  
 enfonçure, N21 /4  
 enforcer, V18 /4  
 enformer, V3 /4  
 enfossage, N1 /4  
 enfosser, V3 /5  
 enfourchement, N1 /4  
 enfourchure, N21 /4  
 enfourmage, N1 /4  
 enfournement, N1 /4  
 enfourneur, N35 /4  
 enfourmeuse, N21 /5  
 enfrichement, N1 /4  
 enfricher, V3 /5  
 enfumoir, N1 /5  
 enfûtage, N1 /4  
 enfutailler, V3 /4  
 enfûter, V3 /4  
 enfûteur, N35 /5  
 engageante, N21 /5  
 engageant, A32 /4  
 engainement, N1 /4  
 engainer, V3 /4  
 engaller, V3 /4

engamer, V3 /4  
 engane, N21 /4  
 enganter, V3 /4  
 engargoussage, N1 /4  
 engargousser, V3 /4  
 engaver, V3 /4  
 engeancer, V4 /4  
 enseigner, V3 /4  
 enger, V5 /5  
 engerbage, N1 /4  
 engerbement, N1 /4  
 engerber, V3 /4  
 englaçage, N1 /4  
 englâcer, V4 /4  
 englaçonnement, N1 /4  
 englaçonner, V3 /5  
 englanté, A32 /5  
 engobage, N1 /4  
 engobe, N1 /4  
 engober, V3 /4  
 engommage, N1 /4  
 engommer, V3 /4  
 engorgeur, N35 /5  
 engoulé, A32 /5  
 engouler, V3 /4  
 engoulevent, N1 /4  
 engrain, N1 /4  
 engraisseur, N35 /4  
 engramme, N1 /4  
 engravé, N1.A32 /5  
 engraver, V3 /4  
 engravure, N21 /4  
 engrêlé, A32 /4  
 engrêler, V3 /4  
 engrêlure, N21 /4  
 engrenage, N1 /4  
 engrenant, A32 /5  
 engrénation, N21 /4  
 engrènement, N1 /4  
 engrener, V6 /4  
 engreneur, N1.N35 /4  
 engreneuse, N21 /4  
 engrenure, N21 /4  
 engrisailler, V3 /4  
 engrois, N2 /4  
 engrumeler, V9 /4  
 engueniller, V3 /4  
 engueuser, V3 /4  
 enguiché, A32 /5  
 enguichure, N21 /4  
 enhardé, A32 /5  
 enharmonie, N21 /4  
 enharmonique, A31 /4  
 enharnacher, V3 /4  
 enhayer, V14 /4  
 enhayer, N35 /5  
 enherbement, N1 /4  
 enherber, V3 /4  
 éniellage, N1 /5  
 énieller, V3 /4

enjabler, V3 /4  
 enjaveler, V9 /4  
 enjonçage, N1 /4  
 enjoncer, V4 /4  
 enjouguer, V3 /5  
 enjuguer, V3 /4  
 enjuponnage, N1 /4  
 enjuponner, V3 /4  
 enképhaline, N21 /4  
 enkikinant, A32 /5  
 enkikiner, V3 /4  
 enlaçure, N21 /4  
 enlarmer, V3 /4  
 enlevage, N1 /4  
 enlève, N21 /5  
 enlevure, N21 /4  
 enliasser, V3 /4  
 enlier, V3 /4  
 enlignement, N1 /4  
 enligner, V3 /4  
 enlinceuler, V3 /4  
 ennéade, N21 /4  
 ennéagonal, A76 /4  
 ennéagone, N1.A31 /4  
 enneigeur, N1 /5  
 enniaiser, V3 /5  
 ennoisement, N1 /4  
 ennoyage, N1 /4  
 ennoyer, V13 /4  
 ennuagement, N1 /4  
 ennuager, V5 /4  
 énolique, A31 /4  
 énoisible, A31 /4  
 énoisation, N21 /4  
 énoiser, V3 /4  
 énoper, V3 /4  
 énoptalmie, N21 /4  
 énostose, N21 /4  
 énouage, N1 /4  
 énouer, V3 /4  
 énoueur, N35 /5  
 énoyautage, N1 /4  
 énoyauter, V3 /4  
 énoyateur, N1 /5  
 enquerre, V0 /4  
 enquêté, N32 /5  
 enquiller, V3 /5  
 enquinauder, V3 /4  
 enraidir, V18 /4  
 enraiment, N1 /4  
 enrailler, V3 /4  
 enrayage, N1 /4  
 enrayeur, N1 /5  
 enrayoir, N1 /4  
 enrayure, N21 /4  
 enrènement, N1 /4  
 enrêner, V3 /4  
 enrésinement, N1 /4  
 enrésiner, V3 /5  
 enrhumable, A31 /5

perlitique,,A31 /4  
 perlocution,,N21 /5  
 perlocutoire,,A31 /5  
 perloir,,N1 /5  
 perlon,,N2S /4  
 perlot,,N1 /4  
 permafrost,,N1 /4  
 permalloy,,N1 /4  
 permanentiste,,N31 /5  
 permaner,,V3 /4  
 permanganique,,A31 /4  
 permanoir,,V0 /5  
 perme,,N21 /4  
 perméation,,N21 /5  
 permien,,N1.A41 /4  
 permo,,XINC /7  
 permsélectif,,A38 /5  
 permutant,,N32.A32 /4  
 perniciosité,,N21 /4  
 péronéo,,XINC /7  
 péronier,,N1.A42 /4  
 péronosporacée,,N21 /4  
 péronosporale,,N21 /4  
 peropérateur,,A31 /5  
 pérot,,N1 /4  
 peroxydase,,N21 /4  
 peroxydasique,,A31 /4  
 peroxydation,,N21 /4  
 peroxyde,,N1 /4  
 peroxyder,,V3 /4  
 peroxysomal,,A76 /5  
 peroxysome,,N1 /5  
 perphosphate,,N1 /5  
 perphosphoré,,A32 /5  
 perrayeur,,N35 /5  
 perré,,N1 /4  
 perreyer,,V3 /4  
 perrière,,N21 /4  
 perruquage,,N1 /4  
 perruquer,,V3 /4  
 perruqueur,,N35 /5  
 persécutoire,,A31 /5  
 persel,,N1 /4  
 persévèrement,,ADV /4  
 persévératif,,A38 /5  
 persévération,,N21 /4  
 persicaire,,N21 /4  
 persicot,,N1 /4  
 persiennage,,N1 /5  
 persillère,,N21 /4  
 personae,,XINC /7  
 personale,,N21 /4  
 personé,,A32 /4  
 personée,,N21 /5  
 perspiration,,N21 /4  
 persulfate,,N1 /4  
 persulfure,,N1 /4  
 persulfuré,,A32 /4  
 persulfurer,,V3 /4  
 persulfurique,,A31 /4  
 perthitique,,A31 /4  
 pertuisane,,N21 /4  
 pertuisanier,,N1 /4  
 pervibrage,,N1 /4  
 perrivateur,,N1 /4  
 perribration,,N21 /4  
 perribrer,,V3 /4  
 pesade,,N21 /4  
 peseur,,N35.A35 /4  
 peson,,N1 /4  
 pessaire,,N1 /4  
 pesse,,N21 /4  
 pessière,,N21 /5  
 pestiférer,,V7 /4  
 pétalisme,,N1 /4  
 pétalodie,,N21 /5  
 pétaloïde,,A31 /4  
 pétardage,,N1 /4  
 pétarder,,V3 /4  
 pétase,,N1 /4  
 pétauriste,,N1.N31 /4  
 pétéchal,,A76 /4  
 pétéchie,,N21 /4  
 pétillères,,N22P /5  
 pétiolaire,,A31 /5  
 pétiolé,,A32 /4  
 pétiolule,,N1 /5  
 péutoire,,N1.A31 /4  
 pétotcher,,V3 /4  
 pétouiller,,V3 /4  
 pétré,,A32 /4  
 pétrel,,N1 /4  
 pétreux,,A63 /4  
 pétricherie,,N21 /5  
 pétricole,,N21 /5  
 pétrinal,,N1 /5  
 pétrissée,,N21 /5  
 pétrisseur,,N1.N35.A35 /4  
 pétrisseuse,,N21 /4  
 pétro,,(PFX) /7  
 pétrochimiquement,,ADV /4  
 pétrochimiste,,N31 /4  
 pétrogale,,N1 /4  
 pétrogenèse,,N21 /5  
 pétrogénétique,,A31 /4  
 pétroglyphe,,N1 /7  
 pétrographie,,N31 /4  
 pétrographie,,N21 /4  
 pétrographique,,A31 /4  
 pétrographiquement,,ADV /4  
 pétroïque,,A31 /4  
 pétroléochimie,,N21 /4  
 pétroléochimique,,A31 /4  
 pétroléochimiquement,,ADV /4  
 pétroléochimiste,,N31 /5  
 pétroler,,V3 /4  
 pétroleur,,N35 /4  
 pétrolifère,,A31 /4  
 pétrolisme,,N1 /4  
 pétrolochimie,,N21 /4  
 pétrolochimique,,A31 /4  
 pétrologie,,N21 /4  
 pétrologique,,A31 /4  
 pétrologiste,,N31 /5  
 pétromastoïdien,,A41 /5  
 pétrosélinique,,A31 /4  
 pétun,,N1 /4  
 pétuner,,V3U /4  
 pétunsé,,N1 /4  
 peucédan,,N1 /4  
 peul,,N1.N32.A32 /4  
 peulven,,N1 /4  
 pexer,,V3 /5  
 peyotl,,N1 /4  
 pézize,,N21 /4  
 pfennig,,N1 /4  
 ph,,(NN) /7  
 phacochère,,N1 /4  
 phacoémulsification,,N21 /5  
 phacomatose,,N21 /5  
 phacome,,N1 /5  
 phacomètre,,N1 /4  
 phacophagie,,N21 /5  
 phage,,N1 /7  
 phagédénique,,A31 /4  
 phagédénisme,,N1 /4  
 phagocytaire,,A31 /4  
 phagotrophie,,N1.A31 /5  
 phagotrophie,,N21 /5  
 phalanger,,N1 /4  
 phalange,,N21 /4  
 phalangisation,,N21 /5  
 phalangiser,,V3 /5  
 phalanstérien,,N41.A41 /4  
 phalarope,,N1 /4  
 phalère,,N21 /4  
 phalline,,N21 /4  
 phallo,,(PFX) /7  
 phallocentrique,,A31 /4  
 phallocentrisme,,N1 /4  
 phalloïde,,A31 /4  
 phalloïdien,,A41 /5  
 phanatron,,N1 /4  
 phanère,,N1 /4  
 phanéritique,,A31 /4  
 phanérogame,,N21.A31 /4  
 phanérogamie,,N21 /5  
 phanérogamiste,,N31 /5  
 phanéroptère,,N21 /5  
 phanie,,N21 /4  
 phaonie,,N21 /5  
 pharillon,,N1 /4  
 pharisaïque,,A31 /4  
 pharisaïquement,,ADV /4  
 pharmaco,,(PFX) /7  
 pharmacocinétiq,,N21 /4  
 pharmacodépendance,,N21 /4  
 pharmacodépendant,,A32 /5  
 pharmacodynamie,,N21 /4  
 pharmacodynamique,,N21.A31 /4



## Annexe 9

### Echantillon du corpus de textes

#### Extrait des dépêches d'une journée de l'AFP (Juin 1990)

- [ 9036] la fouille de toutes les maisons de vacances vides des environs où l'enfant  
[ 9037] aurait pu se réfugier.  
[ 9038]  
[ 9039] 6Hallyday mariage,prev  
[ 9040] Soirée Disney pour le **repas de noces** des Hallyday  
[ 9041] Par Claude LAVAL  
[ 9042] RAMATUELLE, 10 juil (AFP) - Johnny et Adeline Hallyday ont fait vivre, pour leur  
[ 9043] **repas de mariage**, lundi soir, à plus de cinq cents invités une nuit à la Disney,  
[ 9044] sur le thème de "Autant en emporte le vent" dans un décor de stuc planté au  
[ 9045] cœur du château de la Messardière à Saint-Tropez.  
[ 9046] Après la mise en scène qui le matin avait fait de la cérémonie de mariage un  
[ 9047] show aussi au point que les spectacles de Johnny Hallyday, la soirée a semblé  
[ 9048] plus terne aux invités.  
[ 9049] Pour la plupart, les participants n'avaient pas pu louer les indispensables  
[ 9050] costumes de l'époque de Scarlett O'Hara et c'est en smocking et robe longue  
[ 9051] qu'ils ont parcouru le long **tapis rouge** qui conduisait aux marches du château,  
[ 9052] de style anglo-mauresque du XIX-ème, juché au sommet d'une colline au-dessus de  
[ 9053] la baie de Saint-Tropez.  
[ 9054] Sur les marches, des figurantes à l'allure d'automates, aux visages plâtrés,  
[ 9055] s'animaient dans leurs **robes à crinoline** au passage des convives. Des éclairages  
[ 9056] violents, soulignaient de fausses colonnes, alors qu'en haut des marches,  
[ 9057] apparaissait la silhouette du château flanqué de tourelles dans le plus pur  
[ 9058] style **carton-pâte**.  
[ 9059] Le palace, récemment inauguré, présentait encore l'allure d'un immense  
[ 9060] chantier, avec des **engins de terrassement** qui voisinaient avec les Harley  
[ 9061] Davidson et les Rolls rutilantes des convives, vite maculées au travers des  
[ 9062] chemins boueux qui conduisaient au lieu de la fête.  
[ 9063] Ils étaient tous là...  
[ 9064] Mais, les amis de Johnny étaient tous là : Eddy Mitchell, Eddy Barclay, Line  
[ 9065] Renaud, Jean-Pierre Aumont, ainsi que de nombreux tropéziens.  
[ 9066] Ce n'est que vers 23H30, alors que les invités avaient été conviés à 21H00,  
[ 9067] que le couple est arrivé à bord d'une calèche tirée par quatre chevaux. Adeline  
[ 9068] a surgi d'un **nuage de poussière** dans une robe corolle d'organdi, ornée de perles  
[ 9069] de chez Nina Ricci, alors que Johnny avait revêtu d'un élégant smocking signé  
[ 9070] Christian Dior.  
[ 9071] Ils ont gravi les marches au son du quatrième mouvement de la cinquième  
[ 9072] symphonie de Beethoven, alors que des projecteurs balayaient le ciel et que  
[ 9073] quelques colombes apeurées tentaient de prendre leur envol.  
[ 9074] Les orchestres country ont succédé à une **formation classique**, puis des  
[ 9075] mariachis ont pris le relai avant que des attractions de Las Vegas ne soient  
[ 9076] proposées au invités.  
[ 9077] Ce n'est qu'au **petit matin**, alors que le soleil commençait à illuminer le  
[ 9078] golfe de Saint-Tropez, que les convives se sont dispersés par petits groupes.  
[ 9079] Johnny et sa jeune épouse ayant gagné la suite royale du palace. Le nouveau  
[ 9080] couple devait entamer dès mardi une croisière vers une destination inconnue.  
[ 9081] 3Mongolie politique / Le PC suspend l'activité de ses cellules au sein de  
[ 9082] l'armée et de la police  
[ 9083] MOSCOU, 10 juil (AFP) - La direction du parti révolutionnaire du peuple de  
[ 9084] Mongolie (communiste), au pouvoir à Oulan-Bator, a décidé de suspendre  
[ 9085] temporairement à partir du 16 juillet l'activité **politique** de ses **organisations**  
[ 9086] **de base** au sein de l'armée, de la police et des **forces de sécurité**, annonce  
[ 9087] mardi l'agence TASS citant le quotidien mongol Unen.  
[ 9088] Selon l'agence officielle soviétique, la suspension sera valable jusqu'à la  
[ 9089] fin des élections aux Hourals du peuple (Assemblées) à tous les niveaux, et  
[ 9090] notamment la formation du grand Houral et du petit Houral d'Etat, les deux  
[ 9091] chambres législatives, prévue pour le 29 juillet.  
[ 9092] La décision du présidium du **Comité Central** du parti a été prise sur  
[ 9093] proposition de certains délégués lors du **Congrès extraordinaire** du parti en  
[ 9094] avril dernier, selon le quotidien Unen toujours cité par TASS. Plusieurs partis  
[ 9095] réclamaient également cette mesure, a ajouté TASS.  
[ 9096] 5Kénya violences / Au moins six morts lundi au Kénya, selon la presse  
[ 9097] NAIROBI, 10 juil (AFP) - Au moins six personnes ont été tuées au cours des  
[ 9098] affrontements entre émeutiers et **forces de l'ordre** lundi au Kénya, rapportent  
[ 9099] mardi les deux quotidiens non-gouvernementaux, The Nation et The Standard.  
[ 9100] The Standard publie la photo du cadavre d'un écolier tué par la police à

[ 9101] Ndenderu, à une vingtaine de kilomètres au nord de Nairobi. Deux autres  
[ 9102] personnes auraient été tuées près de la poste de Kiambu, importante  
[ 9103] agglomération, située également au nord de la capitale.  
[ 9104] Stephen Ndungu et M. Njoroge (pas de prénom), âgés tous deux de 20 ans, ont  
[ 9105] été victimes des balles de la police au moment où il s'apprêtaient à mettre le  
[ 9106] feu à un camion transportant de l'essence, près de Kikuyu, à une quinzaine de km  
[ 9107] de Nairobi, selon le journal.  
[ 9108] Enfin, un manifestant a trouvé la mort dans le quartier de Kawangware, dans  
[ 9109] l'ouest de la capitale, indique The Nation, confirmant une information donnée  
[ 9110] lundi par une **source diplomatique**.  
[ 9111] Trois autres personnes avaient été victimes des tirs de la police dimanche  
[ 9112] soir, selon le plus **grand quotidien** kényan qui tire à quelque 200.000  
[ 9113] exemplaires.  
[ 9114] Le quotidien gouvernemental, The Kenya Times, qui tire à 60.000 exemplaires  
[ 9115] selon ses responsables, fait sa "une" sur la "**guerre totale** contre les bandits"  
[ 9116] décrétée par le gouvernement dans un communiqué publié lundi.  
[ 9117] Le communiqué précise que le **commissaire de police** principal "a été autorisé  
[ 9118] à employer la force pour protéger les citoyens innocents qui pourraient être  
[ 9119] victimes des hooligans dont le but est le vol et la destruction".  
[ 9120] A Nairobi, le travail a repris au ralenti dans le **centre ville** car les  
[ 9121] autobus circulent au compte-goutte et les matatus, les taxis-collectifs, sont en  
[ 9122] grève.  
[ 9123] Plusieurs sources ont fait état d'affrontements dans la nuit de lundi à mardi  
[ 9124] dans plusieurs bidonvilles, notamment celui de Mathare, le plus peuplé d'entre  
[ 9125] eux.  
[ 9126] La tension était cependant retombée en début de matinée, ont indiqué  
[ 9127] plusieurs témoins.  
[ 9128] 4URSS Congrès / Gorbatchev réplique aux critiques devant une salle frondeuse  
[ 9129] MOSCOU, 10 juil (AFP) - Mikhaïl Gorbatchev, dans un discours d'une heure  
[ 9130] mardi matin au 28ème Congrès du **parti communiste** soviétique, a répliqué aux  
[ 9131] multiples critiques de sa politique, s'en prenant notamment à ceux qui seraient  
[ 9132] tentés par une "dictature".  
[ 9133] Il a fait son intervention devant une salle frondeuse qui a marqué sa  
[ 9134] désapprobation par un brouhaha quasi permanent.  
[ 9135] Il a reconnu que "le parti perd de son pouvoir", mais "aucune dictature ne  
[ 9136] résoudra rien, nous sommes sur le bon chemin, il faut continuer", a-t-il  
[ 9137] aussitôt ajouté.  
[ 9138] "Si nous continuons ainsi, je vous le dis franchement, nous conduisons le  
[ 9139] pays à la banqueroute", a poursuivi M. Gorbatchev à propos du refus de  
[ 9140] l'**économie de marché** exprimé au Congrès.  
[ 9141] "Les délégués qui s'adressent à moi n'ont que deux questions à la bouche:  
[ 9142] quand y aura-t-il de l'ordre, que va-t-on faire des spéculateurs", a dit M.  
[ 9143] Gorbatchev.  
[ 9144] "Le risque c'est que le peuple se détourne d'un choix socialiste discrédité",  
[ 9145] a encore affirmé dans son plaidoyer le **numéro un** soviétique.  
[ 9146] Il s'est élevé contre les critiques de sa **politique étrangère**, notamment à  
[ 9147] propos de l'Europe de l'Est: "Et quoi, faut-il encore envoyer les chars?, allons  
[ 9148] nous encore leur apprendre à vivre?".  
[ 9149] Le discours de M. Gorbatchev a été troublé au bout d'une demi-heure par un  
[ 9150] brouhaha qui n'a plus cessé, au point que le président s'est adressé plusieurs  
[ 9151] fois aux délégués pour savoir s'il pouvait terminer. "Dois-je continuer, dois-je  
[ 9152] continuer", a-t-il demandé. "Ce n'est pas facile: je parle seul et vous êtes  
[ 9153] cinq mille en face de moi", a-t-il constaté.  
[ 9154] 3Justice Médecine- / Inculpation d'une anesthésiste après la mort d'un patient au  
[ 9155] cours d'une opération  
[ 9156] SAINT-DIE (Vosges), 10 juil (AFP) - Une anesthésiste de l'hôpital de Gérardmer  
[ 9157] (Vosges) vient d'être inculpée d'**homicide involontaire** par un **juge d'instruction**  
[ 9158] de Saint-Dié, après le décès, en février 1989, d'un patient au cours d'une  
[ 9159] opération pour une fracture de la jambe, a-t-on appris mardi de source  
[ 9160] judiciaire.  
[ 9161] Victime d'une fracture de la jambe, au cours d'une bagarre, M. Jean-Paul  
[ 9162] Sasso, 44 ans, était décédé le 23 février sur la **table d'opération**, apparemment  
[ 9163] privé d'oxygène. L'appareil qui, à l'époque, assurait l'arrivée en oxygène du  
[ 9164] malade sous anesthésie, était dépourvu de **signal sonore**, et les médecins qui  
[ 9165] entouraient M. Sasso, se sont aperçus très tard qu'il ne fonctionnait pas

## Annexe 10

### Echantillons des sous-dictionnaires C1, C2 et C3 tirés du corpus de textes du LADL

C1	allure	améliorable	amuser
	allusion	améliorant	amuseur
	aloi	amélioration	anachronique
	alors	améliorer	anachronisme
	alouette	amen	anagramme
	alourdir	aménageable	analogie
	alourdissement	aménagement	analogique
	alphabet	aménager	analogue
	alphabétique	amende	analphabète
	alphabétisation	amène	analphabétisme
	alpiniste	amenée	analysable
	alsacien	amener	analysant
	altérable	aménité	analyse
	altération	amenuiser	analysé
	altercation	amer	analyser
	altéré	amèrement	analyseur
	altérer	américain	analyste
	alternance	américaine	analytique
	alternant	amertume	anarchie
	alternatif	ameublement	anarchique
	alternative	ami	anatomie
	alternativement	amiable	anatomique
	alterne	amical	ancestral
	alterné	amicale	ancêtre
	alterner	amie	anchois
	altier	amincissement	ancien
	altitude	amitié	anciennement
	alto	amnésie	ancienneté
	altruisme	amnistie	ancrage
	altruiste	amnistier	ancre
	aluminium	amoindrir	ancrer
	amadouer	amoncellement	âne
	amaigrissement	amont	anéantir
	amalgame	amorce	anéantissement
	amalgamer	amorcer	anecdote
	amande	amorphe	anecdotique
	amant	amorti	anémie
	amarre	amortie	anémier
	amarrer	amortir	anesthésie
	amas	amortissement	anesthésier
	amasser	amour	anesthésiste
	amateur	amoureux	ange
	amateurisme	amoureux	angélique
	amazonien	amovible	anglais
	ambages	amphithéâtre	anglaise
	ambassade	ample	angle
	ambassadeur	ablement	anglican
	ambiance	ampleur	angliciser
	ambiant	amplifiant	anglicisme
	ambigu	amplificateur	anglophone
	ambiguïté	amplification	angoissant
	ambitieux	amplifié	angoisse
	ambition	amplifier	angoissé
	ambivalence	ampoule	angoisser
	ambivalent	amputation	animal
	ambulance	amputé	animalité
	ambulant	amputer	animateur
	âme	amusant	animation

empathie	épistémologie	expérimentateur
emplir	épistémologique	explicitation
empourprer	épistémologue	exponentiel
empreindre	épistolairement	exponentielle
emprunteur	épithélium	expressionniste
émuler	équarrissage	extensif
encapsuler	équipollent	extériorité
enceindre	éradication	extinctif
encéphalite	éradiquer	extralinguistique
enchâsser	ergonomie	extranéité
enclaver	ergonomique	extrinsèque
enclore	érosif	faber
encollage	érotiser	façonnage
encrier	erratique	factoriel
encyclopédisme	escarboucle	factorielle
encyclopédiste	escarcelle	factoring
endémie	escarre	fantasia
endémique	essaimage	fantomal
endogène	estampage	fariner
endormissement	esthétisant	faucheur
endoscopie	estimatif	faucille
enfermement	étalager	fauteur
enfièvrement	étatisme	féal
engineering	éteignoir	fécal
enjoindre	ethnocentrisme	fédéré
enluminer	ethnolinguistique	fédérer
énonciateur	éthologie	fellah
énonciatif	éthylène	féméniser
énonciation	étiage	féodalisme
enrobage	étiologie	fermoir
enroulement	étiquetage	ferrailier
ensabler	étirage	ferry
ensemblier	étoiler	feutrer
ensemencement	étudiable	fibreur
ensemencer	eurodevise	fibreur
entériner	eurodollar	fichtrement
entomologiste	évangéliste	fidéliser
entrechat	éventrer	fiducial
entrecroisement	évidement	figaro
entrecroiser	évider	filage
entrelacs	évolutif	filmographie
entremettre	évolutionniste	filtrage
entreposage	exaction	finaliser
entropie	examinable	finitude
énumérable	excavateur	fioul
énumératif	excaver	fiscalement
environnemental	excellément	fiscalisation
enzyme	excepter	fission
éolien	exécutoire	fixing
éolienne	exégèse	flandrin
épannage	exégète	flécher
éphéméride	exemption	flexibilité
épidémiologique	exigibilité	flexion
épiderme	exorbiter	florès
		florin

inexécuté	invertase	laïcisme
inexécution	invétérer	lainer
inexister	irréfragablement	lamellé
infinitude	isabelle	lameller
inhabile	isoélectrique	lamer
inintelligent	isolateur	laminaire
inintentionnel	isolationniste	lamper
initialer	isomérase	lan\on
injecteur	isonomie	landier
inlandsis	isostatique	langres
inopportunité	isothermique	langue
inopposabilité	isotope	langué
inopposable	isotopie	langueter
inorganique	isotopique	languetter
insaisissabilité	isotrope	lanthane
insatisfaisant	isotropie	lao
insoler	itération	lapiner
institutes	ithyphallique	larmer
insusceptible	ive	latérite
intension	ivraie	latéritique
intensionnel	ivrogner	laurer
intercellulaire	jacent	laye
intercensitaire	jachérer	layer
interchangeabilité	jack	lectine
interclubs	jacot	lectorat
interdunaire	jacqueline	lenter
interfacial	jacquot	léonard
interférogramme	jamber	lépidoptère
interférométrie	jan	lésionnaire
interférométrique	japonaiser	lette
interféron	jasper	lettrier
interfluve	jointer	leucine
interglaciaire	joliet	leucopénie
intergranulaire	joseph	levurer
intérieurisation	jubé	li
interjecter	julien	libérien
interjeter	juridictionnel	libero
intermétallique	jurisconsulte	licitation
internalisation	jurisprudentiel	liciter
internaliser	k	liéger
interneurone	ka	ligase
interpellatif	kan	lignager
interphase	khan	ligner
interpréteur	kip	ligneur
intertropical	Kodak	liliacée
intestat	krill	limonine
intracellulaire	ksi	linéament
intranasal	labelliser	linger
intransigeance	lactase	linkage
intransmissibilité	lactique	lipase
intransmissible	lactose	liposome
intransmissible	lactoser	liser
intransmissible	laguiole	lisp
intransmissible	laguner	lithographier

## Exemples de sélection de noms composés (type Nadj)

	groupe molaire	le/ms; - -
	groupe morphologique	un/N1/A31/ms; - +
	groupe moteur	un/N1/A36/ms; - +
	groupe motopropulseur	un/N1/A1/ms; - +
	groupe multimédia	un/N1/A31/ms; - +
	groupe multiplicatif	un/N1/A38/ms; - +
	groupe musculaire	un/N1/A31/ms; - +
	groupe natal	un/N1/A32/ms; - +
	groupe nominal	un/N1/A76/ms; - +
	groupe orthocentrique	un/N1/A31/ms; - +
	groupe orthogonal	un/N1/A76/ms; - +
	groupe paramilitaire	un/N1/A31/ms; - +
	groupe parlementaire	un/N1/A31/ms; - +
CA	groupe pétrolier	un/N1/A42/ms; - +
CA	groupe pharmaceutique	un/N1/A31/ms; - +
	groupe phonétique	un/N1/A31/ms; - +
CA	groupe politique	un/N1/A31/ms; - +
CA	groupe pop	un/N1/A80/ms; - +
	groupe prépositionnel	un/N1/A40/ms; - +
	groupe primaire	un(1/N1/A31/ms; - +
	groupe primaire	un(2/N1/A31/ms; - +
CA	groupe privé	un/N1/A32/ms; - +
	groupe projectif	un/N1/A38/ms; - +
	groupe public	un/N1/A46/ms; - +
CA	groupe punk	un/N1/A80/ms; - +
	groupe quaternaire	un/N1/A31/ms; - +
	groupe réfrigérant	un/N1/A32/ms; - +
	groupe reproducteur	un/N1/A36/ms; - +
	groupe réversible	un/N1/A31/ms; - +
CA	groupe rival	un/N1/A76/ms; - +
	groupe rythmique	un/N1/A31/ms; - +
CA	groupe sanguin	un/N1/A32/ms; - +
CA	groupe scolaire	un/N1/A31/ms; - +
	groupe sculpté	un/N1/PP/ms; - +
	groupe sculptural	un/N1/A76/ms; - +
	groupe secondaire	un(1/N1/A31/ms; - +
	groupe secondaire	un(2/N1/A31/ms; - +
	groupe sémantique	un/N1/A31/ms; - +
	groupe sénatorial	un/N1/A76/ms; - +
	groupe sidérurgique	un/N1/A31/ms; - +
	groupe simple	un/N1/A31/ms; - +
CA	groupe social	un/N1/A76/ms; - +
CA	groupe sportif	un/N1/A38/ms; - +
	groupe statutaire	un/N1/A31/ms; - +
	groupe succursaliste	un/N1/A31/ms; - +
	groupe sucrier	un/N1/A42/ms; - +
	groupe symétrique	un/N1/A31/ms; - +
	groupe symplectique	un/N1/A31/ms; - +
	groupe syntagmatique	un/N1/A31/ms; - +
	groupe syntaxique	un/N1/A31/ms; - +
	groupe technique	un/N1/A31/ms; - +
CA	groupe terroriste	un/N1/A31/ms; - +
	groupe tertiaire	un/N1/A31/ms; - +
	groupe textile	un/N1/A31/ms; - +
	groupe thermoélectrique	un/N1/A31/ms; - +
	groupe tissulaire	un/N1/A31/ms; - +
	groupe topologique	un/N1/A31/ms; - +
	groupe touristique	un/N1/A31/ms; - +
	groupe tribal	un/N1/A76/ms; - +
	groupe unitaire	un/N1/A31/ms; - +
	groupe verbal	un/N1/A76/ms; - +
	groupe vocal	un/N1/A76/ms; - +
	groupement adapté	un/N1/PP/ms; - +
	groupement aérien	un/N1/A41/ms; - +
	groupement aéroporté	un/N1/A32/ms; - +
	groupement agricole	un/N1/A31/ms; - +
	groupement blindé	un/N1/A32/ms; - +
	groupement bovin	un/N1/A32/ms; - +
	groupement chimique	un/N1/A31/ms; - +
	groupement communautaire	un/N1/A31/ms; - +
	groupement corporatif	un/N1/A38/ms; - +
	groupement fonctionnel	un/N1/A40/ms; - +
	groupement forestier	un/N1/A42/ms; - +
	groupement gestionnaire	un/N1/A31/ms; - +
	groupement hydrophobe	un/N1/A31/ms; - +

CA	handicap moteur	un/N1/A36/ms; - +
	handicap neuropsychosensoriel	un/N1/A40/ms; - +
CA	handicap physique	un/N1/A31/ms; - +
CA	handicap scolaire	un/N1/A31/ms; - +
	handicapé cardiaque	un/N32/A31/ms; + +
	handicapé coronarien	un/N32/A41/ms; + +
	handicapé isolé	un/N32/A32/ms; + +
	handicapé lourd	un/N32/A32/ms; + +
CA	handicapé mental	un/N32/A76/ms; + +
CA	handicapé moteur	un/N32/A36/ms; + +
	handicapé patronymique	un/N32/A31/ms; + +
CA	handicapé physique	un/N32/A31/ms; + +
	handicapé psychique	un/N32/A31/ms; + +
CA	handicapé scolaire	un/N32/A31/ms; + +
	handicapé sensoriel	un/N32/A40/ms; + +
	hangar agricole	un/N1/A31/ms; - +
	Hanse germanique	la/fs; - -
	Hanse teutonique	la/fs; - -
	harangue politique	une/N21/A31/fs; - +
	haras national	un/N2/A76/ms; - +
CA	harcèlement sexuel	un/N1/A40/ms; - +
	harem royal	un/N1/A76/ms; - +
	hareng guai	un/N1/A1/ms; - +
	hareng pec	un/N1/A1/ms; - +
	hareng saur	un/N1/A1/ms; - +
CA	harengs marinés	des/mp; - -
CA	haricot blanc	un/N1/A47/ms; - +
CA	haricot rouge	un/N1/A31/ms; - +
CA	haricot vert	un/N1/A32/ms; - +
CA	haricots blancs	des/mp; - -
CA	haricots jaunes	des/mp; - -
CA	haricots noirs	des/mp; - -
CA	haricots panachés	des/mp; - -
CA	haricots rouges	des/mp; - -
CA	haricots secs	des/mp; - -
CA	haricots verts	des/mp; - -
	harle huppé	un/N1/A32/ms; - +
	harmonica chimique	un/N1/A31/ms; - +
	harmonica chromatique	un/N1/A31/ms; - +
	harmonica diatonique	un/N1/A31/ms; - +
	harmonica thermique	un/N1/A31/ms; - +
	harmonie classique	la/fs; - -
	harmonie consonante	la/fs; - -
	harmonie dissonante	la/fs; - -
	harmonie imitative	la/fs; - -
	harmonie municipale	une/N21/A76/fs; - +
	harmonie préétablie	la/fs; - -
	harmonie tonale	la/fs; - -
	harmonie vocalique	la/fs; - -
	harmoniques inférieures	des/fp; - -
	harmoniques naturels	les/mp; - -
	harmonisation fiscale	une/N21/A76/fs; - +
	harmonisation vocalique	la/fs; - -
	harnois blanc	un/N2/A47/ms; - +
	harpe celtique	une/N21/A31/fs; - +
	harpe éolienne	une/N21/A41/fs; - +
	harpe galloise	une/N21/A61/fs; - +
	harpe indienne	une/N21/A41/fs; - +
	harpe irlandaise	une/N21/A61/fs; - +
CA	hausse démographique	une/N21/A31/fs; - +
	hausse indiciaire	une/N21/A31/fs; - +
	hausse nominale	une/N21/A76/fs; - +
CA	hausse pétrolière	une/N21/A42/fs; - +
CA	hausse salariale	une/N21/A76/fs; - +
	hauteur absolue	la/fs; - -
	hauteur angulaire	une/N21/A31/fs; - +
CA	hauteur apparente	une/N21/A32/fs; - +
	hauteur barométrique	une/N21/A31/fs; - +
	hauteur métacentrique	une/N21/A31/fs; - +
CA	hauteur relative	la/fs; - -
	hauteur tonale	la/fs; - -
	hauteur vraie	une/N21/A32/fs; - +
	haveuse intégrale	une/N21/A76/fs; - +
	haveuse universelle	une/N21/A40/fs; - +
	hayon élévateur	un/N1/A36/ms; - +

## **Annexe 12**

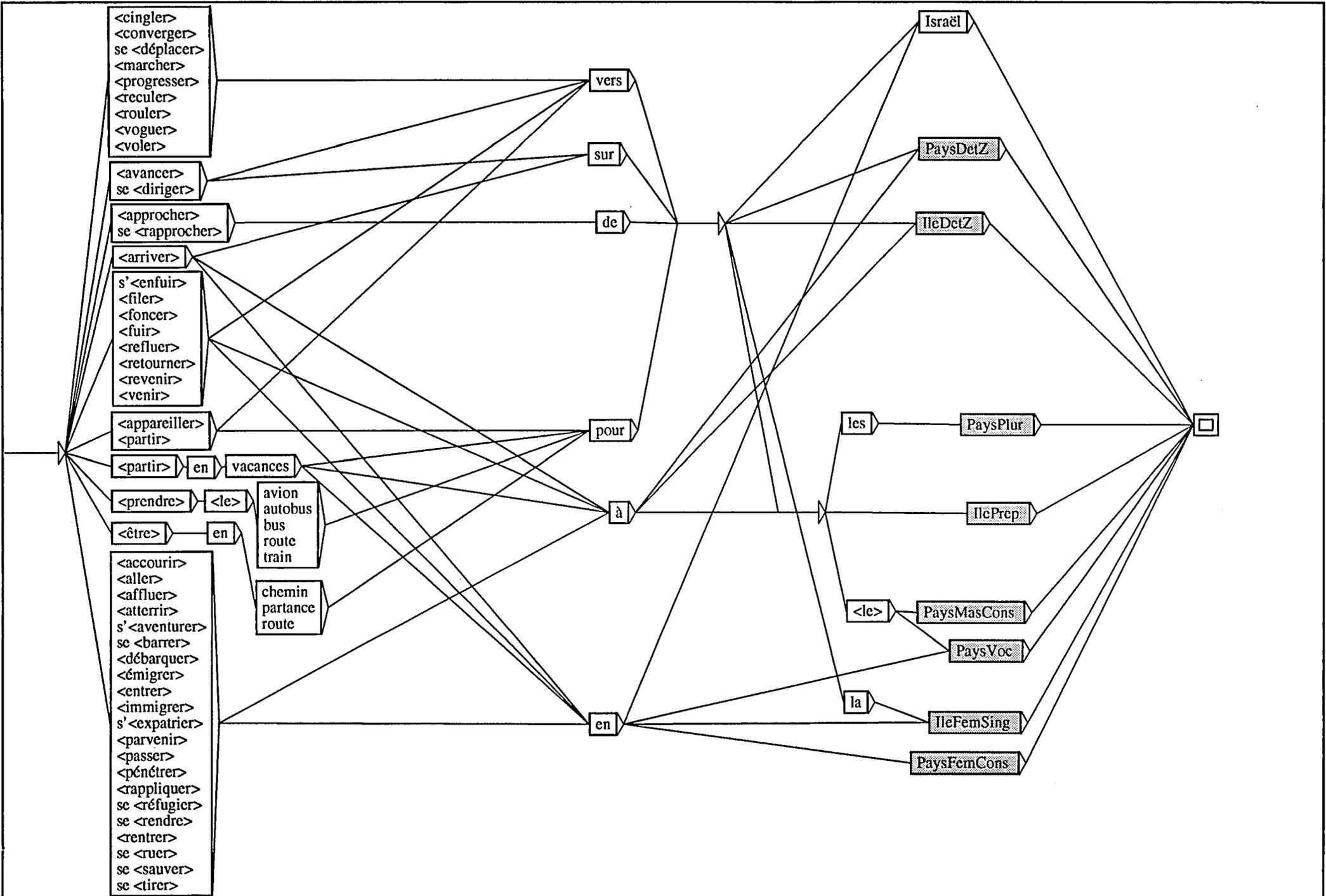
**Prédicat sémantique "Mouvement de déplacement d'un sujet humain par rapport à un lieu désigné"**

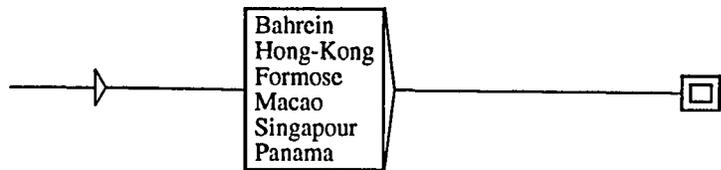
**Champ sémantique : Pays, îles**

**Orientation : ALLER**

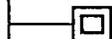
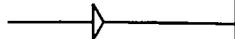
**- Automate N V PrépN**

**- Automate N V N**

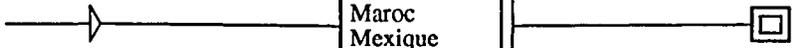


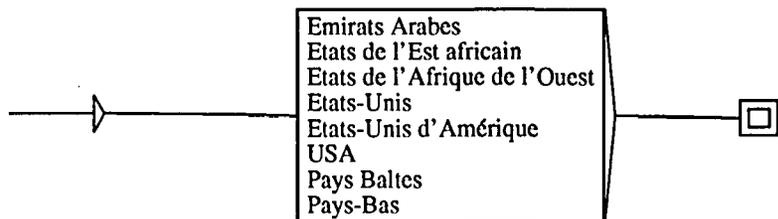


Belgique  
Birmanie  
Bolivie  
Bulgarie  
République Centrafricaine  
Chine  
Colombie  
Communauté des Etats Indépendants  
CEI  
Communauté Economique Européenne  
CEE  
Corée  
Côte d'Ivoire  
République Dominicaine  
Estonie  
Finlande  
France  
Gambie  
Grande-Bretagne  
Grèce  
Guinée  
Hollande  
Hongrie  
Jordanie  
Lettonie  
Libye  
Lituanie  
Malaisie  
Mauritanie  
Mongolie  
Namibie  
Norvège  
Nouvelle Calédonie  
Nouvelle Zélande  
Papouasie  
Pologne  
Polynésie Française  
République Fédérale Allemande  
RFA  
Roumanie  
Russie  
Sierra Leone  
Somalie  
Suède  
Suisse  
Syrie  
Tanzanie  
Tchécoslovaquie  
Terre Adélie  
Tunisie  
Turquie  
Yougoslavie  
Zambie



Bangladesh  
Bénin  
Brésil  
Burkina Faso  
Brundi  
Cameroun  
Canada  
Chili  
Congo  
Costa Rica  
Danemark  
Japon  
Gabon  
Ghana  
Guatemala  
Honduras  
Groenland  
Kenya  
Koweït  
Laos  
Liban  
Libéria  
Luxembourg  
Mali  
Maroc  
Mexique  
Mozambique  
Népal  
Nicaragua  
Niger  
Nigeria  
Paraguay  
Pérou  
Portugal  
Qatar  
Royaume-Uni  
Rwanda  
Salvador  
Sénégal  
Soudan  
Sri-Lanka  
Surinam  
Swaziland  
Tchad  
Togo  
Venezuela  
Vietnam  
Yémen  
Zaïre  
Zimbabwe





Afghanistan  
Afrique du Sud  
Albanie  
Algérie  
Allemagne  
Angleterre  
Angola  
Arabie Saoudite  
Argentine  
Australie  
Autriche  
Ecosse  
Egypte  
Equateur  
Espagne  
Ethiopie  
Inde  
Indonésie  
Irak  
Iran  
Irlande  
Islande  
Italie  
Ouganda  
URSS  
Uruguay



