

# Acquisition des Langues Assistée par Ordinateur (ALAO)

Thierry Chanier

▶ **To cite this version:**

Thierry Chanier. Acquisition des Langues Assistée par Ordinateur (ALAO). Linguistique. Université Blaise Pascal - Clermont-Ferrand II, 1995. tel-00180340

**HAL Id: tel-00180340**
**https://theses.hal.science/tel-00180340**

Submitted on 18 Oct 2007

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**UNIVERSITÉ BLAISE PASCAL (CLERMONT 2)**
**LABORATOIRE DE RECHERCHE SUR LE LANGAGE**

# DOSSIER DE TRAVAUX

PRÉSENTÉ POUR L'OBTENTION D'UNE

## HABILITATION À DIRIGER DES RECHERCHES

SPÉCIALITÉS : INFORMATIQUE ET LINGUISTIQUE

PAR

**Thierry Chanier**

HABILITATION SOUTENUE LE 13 JANVIER 1995 DEVANT LA COMMISSION D'EXAMEN :

| | |
|---|---|
| **Mr Robert Bouchard** | |
| **Mr Michel Chambreuil** | Rapporteur |
| **Mme Françoise Demaizière** | |
| **Mme Monique Grandbastien** | Présidente |
| **Mr Daniel Kayser** | Rapporteur |
| **Mr Jacques Rouault** | |
| **Mr Daniel Véronique** | Rapporteur |

# ABRÉVIATIONS

- ALAO : Acquisition des Langues Assistée par Ordinateur
- ALS : Acquisition d'une Langue Seconde
- EAO : Enseignement Assisté par Ordinateur
- EIAO : Environnements Interactifs d'Apprentissage avec Ordinateur
- ESF : Expression semi-figée
- FLE : Français Langue Étrangère
- GU : Grammaire Universelle
- IA : Intelligence Artificielle
- IA-ED : Intelligence Artificielle et Éducation
- ITS : Intelligent Tutoring Systems
- L1 : La langue source, ou langue maternelle
- L2 : La langue cible
- LOS : apprentissage ou usage d'une Langue pour des Objectifs Spécifiques
- LRL : Laboratoire de Recherche sur le Langage
- NTF : Nouvelles Technologies d'aide à la Formation
- TAG : Tree Adjoining Grammar
- TALN : Traitement Automatique du Langage Naturel

# TABLE DES MATIÈRES


**1. INTRODUCTION ..... 4**

1.1. UNE PREMIÈRE DÉFINITION DE L'ALAO ..... 4
1.2. DISCUSSION ..... 5
1.3. RECHERCHES ANTÉRIEURES ..... 7

**2. ELEMENTS CONSTITUTIFS DE L'ESPACE DE RECHERCHE . 9**

2.1. ACQUISITION DES LANGUES ..... 10
2.2. ALAO : TRADITIONS ET OBJECTIFS ..... 13

**3. POINTS D'INVESTIGATION DANS MA RECHERCHE ..... 15**

3.1. OBJECTIFS, RECHERCHE D'EXPERTISES ET CRÉATION DE MATÉRIAUX ..... 15
3.1.1. Analyse des besoins, besoin des apprenants ..... 16
3.1.2. Quelle norme pour la langue cible ? ..... 23
3.1.3. Création des matériaux, études linguistiques, procédures de TALN .... 24
3.2. MODÉLISATION DE L'APPRENANT, DIAGNOSTIC ..... 31
3.2.1. ELEONORE ..... 32
3.2.2. ALEXIA ..... 36
3.2.3. "Travailler en France": suivi de l'apprenant ..... 38
3.3. PÉDAGOGIE: CONTENU ET STRATÉGIES ..... 40
3.3.1. ELEONORE et le dialogue collaboratif ..... 40
3.3.2. CAMILLE et l'auto-apprentissage ..... 42
3.4. EVALUATION, INTÉGRATION DANS LE CURRICULUM ..... 44

**4. ORGANISATION ET DIFFUSION DE LA RECHERCHE EN ALAO ..... 46**

4.1. FORMATION DE CHERCHEURS, ORGANISATION D'ÉQUIPES DE RECHERCHE ..... 46
4.2. CONTRIBUTION À LA DIFFUSION ET À L'ORGANISATION DE LA RECHERCHE .... 47

**5. BIBLIOGRAPHIE ..... 49**

# 1. INTRODUCTION

Ce document a pour but de définir le domaine de "l'Acquisition des Langues Assistée par Ordinateur" (ALAO), et d'y décrire mes travaux. Ce premier chapitre esquisse une définition de l'ALAO. Les liens entre ce thème et le contenu de ma thèse sont discutés. Dans le deuxième chapitre, j'examine les éléments constitutifs de l'espace de recherche : l'acquisition d'une langue seconde d'une part, le milieu de l'ALAO d'autre part, avec ses orientations passées et ses nouveaux principes de recherche. Le chapitre 3 développe mes travaux suivant un plan en partie identique aux étapes de déroulement d'un projet de recherche ALAO, en mettant particulièrement l'accent sur l'expertise du domaine, la modélisation de l'apprenant, et les stratégies pédagogiques adoptées. Dans le dernier chapitre, j'évoque ma contribution à l'organisation et à la diffusion de la recherche en ALAO.

## 1.1. Une première définition de l'ALAO

L'ALAO est un domaine de recherche et développement qui intéresse plusieurs disciplines intervenant dans le champ des sciences cognitives :

- la linguistique (dans son acceptation large comprenant la linguistique théorique, appliquée, la didactique des langues),
- la linguistique-informatique,
- l'informatique (en particulier l'intelligence artificielle (IA)),
- la psycholinguistique.

Certaines disciplines sont centrées sur la description des connaissances linguistiques des apprenants, sur la description de l'évolution de ces connaissances, sur celle des stratégies d'apprentissage ou sur l'élaboration d'expertises pédagogiques propres au domaine.

D'autres voient dans l'étude de l'acquisition des langues l'occasion d'appliquer des méthodes et techniques générales destinées à prendre en charge des connaissances, diagnostiquer des états de connaissances, ou l'occasion d'introduire des outils et des formalismes issus du traitement automatique du langage naturel; leurs travaux sont essentiellement orientés vers la construction d'environnements interactifs d'aide à l'apprentissage.

Les environnements informatiques orientés vers l'acquisition des langues recouvrent deux axes de recherche et développement distincts, mais complémentaires:

- les environnements permettant d'effectuer des recherches fondamentales sur l'acquisition des langues au travers d'une interaction avec l'ordinateur.
- les environnements d'apprentissage finalisés pour l'acquisition de compétences linguistiques (tant grammaticales/lexicales, que sociolinguistiques ou communicatives), en écriture, lecture, oral, conçus à partir d'approches pédagogiques multiples.

Les premiers reposent sur des modèles informatisés du traitement (en reconnaissance ou en production) ou de l'acquisition des langues; modèles destinés à valider, en les programmant, des théories générales ou bien des aspects ponctuels de l'acquisition en langue

maternelle ou seconde. S'ils ne s'intéressent pas directement à l'enseignement, ils dévoilent les mécanismes fondamentaux de l'apprentissage/acquisition.

Les seconds visent à construire les modules fondamentaux d'un système d'aide à l'apprentissage, qui, si l'on reprend la présentation usuelle en IA et Éducation, seraient : un module de diagnostic de l'apprenant, un module expert ou module d'usage des compétences linguistiques visées, un module traduisant les expertises pédagogiques appropriées, voire un module prenant en charge la gestion de l'interface de communication. Une partie importante de la recherche consiste à élaborer des théories assez formalisées pour être implémentées dans des logiciels sans que la formalisation n'apporte trop de réduction.

Dans ces deux types d'environnements, l'expérimentation constitue un apport essentiel à la méthodologie de recherche, soit pour apporter des données fondamentales à la construction des modèles de simulation, soit pour valider la pertinence des modèles d'aide à l'apprentissage.

## 1.2. Discussion

Quelques commentaires s'imposent après cette présentation que nous avions adoptée pour le colloque SCIAL'93 (publication 28).

Nous utilisons à la fois les termes "acquisition" et "apprentissage". L'acquisition est quelquefois différenciée de l'apprentissage lorsqu'on souhaite discriminer les processus cognitifs qui les sous-tendent. Le terme "acquisition" est alors utilisé pour signifier qu'une langue est assimilée souvent inconsciemment dans certaines conditions d'exposition (ou encore d'immersion), alors que le terme "apprentissage" renvoie à l'étude consciente d'une langue. Des usages encore plus spécialisés et non concordants sont adoptés dans les théories chomskiennes (White, 1989) et dans celle du Contrôleur de Krashen (1988). Nous ne tenons pas à ouvrir une discussion sur les positions partisanes de ces deux approches. Nous utiliserons, sauf mention explicite, de façon relativement interchangeable les deux termes, sans préjuger du caractère conscient ou inconscient des processus sous-jacents. La seule nuance recherchée est celle qui tient à la distinction des situations d'exposition à une langue. Nous parlerons plus spécifiquement d'apprentissage lorsqu'un individu fait le choix conscient de maîtriser des compétences nouvelles, dans une langue, dans une situation où intervient un intermédiaire d'aide à sa formation (soit un non-natif, un enseignant, un autre apprenant, ou un système informatique). Le terme acquisition,lui, sera souvent utilisé de façon plus large, englobant non seulement les situations d'apprentissages institutionnels mais également celles, dites naturelles, rencontrées par un individu vivant à l'étranger. Dans toutes les situations, les processus d'acquisition/apprentissage peuvent être conscients ou inconscients. Il est bien connu aujourd'hui que même lors d'un enseignement explicite de règles d'une langue, celles-ci ne peuvent être maîtrisées par l'apprenant qu'après avoir été intériorisées de façon partiellement inconsciente. Notre interprétation générique du terme "acquisition" correspond à l'usage anglo-saxon courant en linguistique : les appellations *first language acquisition research* (Ingram, 1989) et *second language acquisition research* (Ellis, 1985) recouvrent tous les travaux de recherche en langue maternelle et seconde, en situations institutionnelles ou naturelles.

Si nous nous référons fréquemment aux travaux anglo-saxons en acquisition des langues, c'est parce que ce que l'on appelle "la didactique des langues" ne recouvre que partiellement

nos domaines d'intérêts. En France, l'usage veut que l'apprentissage des mathématiques, de l'informatique, des langues, etc. soient essentiellement l'affaire des didacticiens. L'étude de l'enseignement (élaboration de programmes, de curriculums) et du rôle de l'enseignant est un domaine de recherche digne d'intérêt. Il trouve son application dans les recherches sur les environnements informatiques d'apprentissage lorsqu'il s'agit, par exemple, de développer certaines stratégies pédagogiques, de planifier l'intégration d'un logiciel dans le cadre d'une formation institutionnelle. Mais le problème primordial est celui de la compréhension de l'acquisition, donc de la focalisation sur l'apprenant (Tarone, 1989 Nunan, 1988). Placer la détermination des connaissances et des stratégies de l'apprenant, et de l'évolution de ces connaissances au centre de nos préoccupations établit un trait d'union avec la linguistique théorique, la psycholinguistique et l'intelligence artificielle. La recherche en Intelligence Artificielle et Éducation (IA-ED) cherche d'abord à modéliser ces connaissances de façon à individualiser l'interaction entre l'apprenant et le système. De plus les besoins de formation en langues obligent à concevoir la formation dans des cadres non scolaires, cadres d'auto-formation où l'apprenant choisit ses méthodes, ses rythmes de travail. La conception didactique traditionnelle répond mal à ces besoins (Porcher, 1990). Ce cadre d'apprentissage est au contraire celui naturellement (ce qui ne veut pas dire sans problèmes) offert par les environnements informatiques.

En ce qui concerne les environnements informatiques d'aide à l'apprentissage, les objectifs des chercheurs en EAO (Enseignement assisté par Ordinateur) ne correspondent que partiellement aux miens, et cela pour deux raisons. Tout d'abord, si une partie de mes travaux a bien pour visée d'aider à apprendre, elle ne cherche pas à le faire uniquement en faisant jouer au système informatique le rôle d'un enseignant. La relation maître-élève[1] n'est qu'une des relations qu'il est possible d'établir entre l'usager et le système. Suivant les remarques du paragraphe précédent, il s'agit d'abord de placer l'apprenant au centre d'investigations, dans lesquelles sont explorées les situations multiples d'apprentissage que peuvent créer spécifiquement les technologies informatiques : exerciseurs, bien sûr, mais aussi apprentissage guidé par la découverte, apprentissage collaboratif, simulation, etc. Cette extension du champ initial de l'EAO est partagée par de nombreux chercheurs et est symbolisé par la transformation de certains sigles : le sigle EAO est peu à peu supplanté par celui de NTF (Nouvelles Technologies d'aide à la Formation) (Demaizière & Dubuisson, 1992) ; de même EIAO, en partie pour les raisons évoquées ci-dessus, n'est plus décliné "Enseignement Intelligemment Assisté par Ordinateur", mais "Environnement Interactifs d'Apprentissage par Ordinateur" ; le terme "Intelligent Tutoring Systems " (ITS), appellation favorite des systèmes IA-ED dans les années 80 (Wenger, 1987), est aujourd'hui peu à peu remplacé par celui de "Intelligent Learning Environnements".

La seconde raison qui explique l'écart avec les objectifs traditionnels de l'EAO est la volonté d'inscrire la recherche en ALAO dans une perspective de confrontation des différentes théories sur l'acquisition des langues. En effet, une méthodologie de recherche appropriée place l'ALAO dans une position privilégiée en acquisition des langues : un domaine d'acquisition étant défini, il est possible d'en dresser une expertise linguistique fine, de trouver sa correspondance en linguistique-informatique, de s'intéresser à la façon dont

[1] On devrait d'ailleurs plutôt parler de relation tuteur-élève puisqu'il s'agit d'échange 1-1 par opposition à celui 1-n, caractéristique de la classe.

l'être humain comprend et produit des énoncés dans ce domaine, à la façon dont il construit ses connaissances, aux stratégies qu'il y déploie et enfin d'examiner les approches pédagogiques adéquates. Alors peut se construire un espace de recherche continu, fondé sur les travaux de plusieurs disciplines, disciplines dont les recherches ne se recouvrent pas vraiment, du fait de leur objectifs spécifiques.

Dans cette synthèse de mes travaux de recherche, je ne chercherai pas à faire nettement la part des choses entre ceux qui relèveraient de l'informatique ou bien des sciences du langage. En effet, si, en termes informatiques, je m'inscris dans le courant IA-ED (publication 27), celui-ci a pour principe de ne pas ignorer les spécificités du domaine d'apprentissage étudié, comme le souligne cet extrait d'un livre à paraître (Intelligent Learning Environments (Self, Brna, Parkes & Twidale, 1995)) : "[The book] tries to give a balanced picture of how the subtleties of a specific learning-teaching situation should be taken account of in designing an appropriate combination of educational philosophy and computational technique".

## 1.3. Recherches antérieures

Mes premiers travaux dans le domaine de l'ALAO datent de la fin 1989. Mais c'est dès le début de mon travail de thèse, en 1986, que l'intelligence artificielle et le langage naturel ont été au centre de mes préoccupations. Plus précisément deux thèmes m'ont particulièrement intéressé et ont été développés en parallèle: celui du traitement automatique du langage naturel (TALN) (travail de thèse), et celui des environnements informatiques d'aide à des apprentissages (travail du petit groupe de recherche à l'origine du Laboratoire de Recherche sur le Langage (LRL)).

Ce deuxième axe de recherche se situait au point de convergence des centres d'intérêts des membres du groupe, venant de disciplines différentes, en représentation des connaissances et en sciences cognitives. En bref, le problème posé était le suivant : de quelles façons les modèles, souvent formels, utilisés pour représenter les connaissances, et les modèles de développement cognitif permettaient-ils de rendre compte de l'état, et de l'évolution des connaissances d'un apprenant dans un domaine donné ? Et de quelles façons des représentations du suivi des connaissances d'un apprenant nous permettaient-elles d'intervenir automatiquement dans le processus d'apprentissage ? Nous avions choisi pour domaine d'apprentissage celui de la lecture (publication 11), car il correspondait à une expertise double de membres de notre groupe : expertise théorique, sur les travaux récents en apprentissage de la lecture, et expertise de terrain, provenant d'instituteurs habitués à collaborer avec nous et ouverts aux nouvelles technologies. Le glissement, lors de mon année post-doctorale, de l'étude de l'apprentissage de la lecture à celui concernant l'apprentissage d'une langue seconde n'a pas modifié la problématique évoquée ci-dessus. Les travaux que nous exposerons tentent bien de répondre aux mêmes questions. Et la démarche adoptée est la même : l'étude de l'apprentissage ne peut se construire qu'à partir d'une expertise partagée par une équipe de recherche. Dans mon cas, il s'agit du français pour des apprenants non francophones. Ce choix s'est naturellement imposé du fait de la décision des partenaires européens du projet CAMILLE, auquel je participe, de ne travailler que dans leur langue maternelle, de façon à suivre au mieux l'évolution du parler des natifs (base de notre expertise) et à garantir la pertinence des aspects socioculturels de chaque langue

étudiée. De plus l'équipe du laboratoire comprend des enseignants-chercheurs de FLE (Français Langue Étrangère).

Fixer ainsi un domaine d'apprentissage n'implique nullement une mise à l'écart des recherches en IA-ED. Ce champ de recherches s'est certes ingénié, à ses débuts à la fin années 1970, à définir des principes généraux coordinateurs, tels la décomposition des systèmes en modules modélisant les partenaires potentiels du processus d'apprentissage: l'apprenant, l'expert, le pédagogue, et par extension l'interface du système (Polson & Richardson, 1988 ; Wenger, 1987). Mais la stagnation des travaux sur les ITS à la fin des années 1980 a montré que les multiples tentatives de définitions d'architecture de systèmes indépendants de type d'apprentissage particuliers étaient encore prématurées. Suivant un phénomène souvent observé lors de l'inauguration d'une nouvelle voie de recherche (cf. le cas de la traduction automatique), les travaux se sont ensuite diversifiés et circonscrits à des problèmes plus pointus. Dans ce document nous aurons l'occasion de montrer que cette circonscription à un domaine d'apprentissage particulier ne nuit pas à l'approfondissement de questions générales en IA-ED.

Mon deuxième axe de recherche en 1986 comportait deux volets distincts en traitement informatique du langage naturel. Le premier volet, plus linguistique informatique, avait pour objectif d'intégrer dans un même formalisme certaines caractéristiques propres aux grammaires d'unification et une théorie sémantique sur la représentation du discours, la Théorie de la Représentation du Discours. Le second volet concernait l'application de méthodes de raisonnements formels issues de l'intelligence artificielle à l'extraction automatique de connaissances à partir de textes de spécifications, écrits dans un sous-ensemble du langage naturel, et à la vérification de leur cohérence. Nous n'évoquerons pas ici le projet ACTES dans lequel j'ai pu inclure ces deux volets (un résumé est fourni dans le document complémentaire listant mes activités de recherche ; le lecteur pourra également se reporter aux publications 18 et 8).

Revenons au premier volet. J'ai tenté d'intégrer mes premiers outils de développement de grammaires, utilisant l'unification et la théorie du discours, dans un système de compréhension de textes. L'intérêt était de confronter les prétentions généralistes de ces formalismes syntaxiques et sémantiques à des problèmes bien concrets, liés à un corpus défini et cohérent. La confrontation m' a conduit à modifier ces outils, les étendre, voire les "tordre" pour qu'ils puissent prendre en compte certains phénomènes non prévus initialement. En retour, cela m'a permis de discuter les limitations des théories dont ils s'inspiraient, ainsi que le type d'architecture appropriée à la collaboration de plusieurs théories partielles.

Cette démarche intégrative est toujours fréquemment adoptée en TALN. Aucune approche venant de la linguistique théorique et cherchant à intégrer aspects syntaxiques et sémantiques n'est assez mûre pour faire de la compréhension automatique de textes, de tailles significatives. Ces dernières années, habitude a été prise de travailler sur de larges corpus électroniques, pour effectuer des traitements de profondeurs variables sur les textes. On a vu alors ressurgir, pour connaître de nouveaux développements, les anciennes approches statistiques et probabilistes (Charniak, 1993; Lebart & Salem, 1994 ; Computational Linguistics, 1993), assez souvent couplées à des approches plus traditionnelles en linguistique informatique. Même dans le champs de la syntaxe formelle, où des progrès significatifs ont été accomplis par des théories de puissances assez comparables (Abeillé,

1993 ; Borsley, 1991), on a vu certains auteurs associer aux règles de grammaire d'unification des facteurs probabilistes.

Qu'en est-il de l'impact des travaux en TALN et linguistique-informatique sur la recherche en ALAO ? Ils ont été, et sont encore, importants pour la communauté de chercheurs de ce domaine, non par leur efficience, mais surtout parce qu'ils ont fixé une certaine image de ce que l'intelligence artificielle pouvait apporter aux environnements d'aide à l'apprentissage. Les techniques de TALN pouvaient offrir de nouvelles capacités de traitements qui allaient permettre de le système de connaissances explicites et calculables sur le domaine d'apprentissage, et par là-même lui permettre d'interagir avec l'apprenant. Les formalismes linguistiques, grâce à leur puissance déclarative et leurs capacités d'analyse, allaient pouvoir servir de référence pour l'enseignement.

Ainsi, il était devenu presque coutumier, à la fin des années 1980, de voir dans les conférences de linguistique-informatique des concepteurs de systèmes de TALN ou de formalismes linguistiques proposer des applications de leur travail à l'enseignement des langues. Certains auteurs ont également imaginé des architectures de systèmes d'EIAO très ambitieuses, dont chaque module était spécifié à partir d'interprétations trop optimistes d'articles de recherches représentant l'état de l'art en linguistique-informatique ((Farghaly, 1989) par exemple).

Il ne s'agit pas ici de nier la valeur de certaines de ces propositions, mais de remarquer qu'elles n'ont pas connu de suite. Leurs auteurs se seraient alors heurtés à l'écueil inévitable de leur manque de fondement pédagogique, et/ou de leur ignorance de la problématique spécifique de l'acquisition des langues. Depuis, les choses ont évolué. Aujourd'hui des chercheurs, qui ont des compétences en linguistique-informatique et ont décidé de fixer leur attention sur le domaine de l'ALAO, apportent réellement une contribution nouvelle. Nous reviendrons largement sur ces questions importantes, en précisant notre contribution, étayée par notre expérience de recherche en linguistique-informatique.

## 2. ELEMENTS CONSTITUTIFS DE L'ESPACE DE RECHERCHE

Dans la partie précédente, j'ai donné une brève définition de l'ALAO, indiqué dans quelles sous-disciplines s'inscrivaient mes recherches actuelles, et enfin quel était le fil conducteur avec mes recherches antérieures. Avant de présenter ces recherches dans la troisième partie, il m'apparaît nécessaire d'indiquer brièvement quelles sont les données de départ.

D'abord qu'en est-il de la spécificité de l'acquisition des langues ? J'en explique ici la nature par touches contrastives avec l'apprentissage en sciences exactes. Cela me permet d'introduire une terminologie minimale, de donner des indications sur ce qui nous intéresse en acquisition des langues, et plus particulièrement des langues secondes. Ce n'est pas le lieu de référencer une bibliographie détaillée. Les références bibliographiques ont été réduites au minimum, pour ne pas surcharger l'exposé. Le lecteur désireux d'avoir une vue d'ensemble de la recherche en acquisition des langues secondes pourra consulter les deux ouvrages de base, non scolastiques de Ellis (1985) et Gaonac'h (1987).

Ensuite, je présenterai le milieu de l'ALAO. Je rappellerai les axes dominants de recherche jusqu'à ces dernières années, et listerai les nouveau principes de travail, qui me serviront de référence par la suite.

## 2.1. Acquisition des langues

Alors que, en recherche IA-ED, la programmation informatique et les mathématiques sont des domaines habituels d'études, celui des langues semble avoir reçu beaucoup moins d'attention. Par exemple, lors de la dernière conférence internationale IA-ED en 1993, 2 articles seulement sur 70 s'intéressaient à l'apprentissage des langues. En ce qui concerne les mathématiques, deux raisons principales sont avancées pour expliquer son rôle prédominant (Nwana , 1993 : p 5) :

1. Mathematics is highly structured and its algorithms are well defined, making it easier to concentrate on the features of the Intelligent Tutoring System itself, rather than those of the domain.
2. Mathematics is an important educational domain because it underpins most of science and engineering.

L'importance de la perspective éducationnelle ne peut sûrement pas expliquer la différence de situation entre mathématiques et langues. Apprendre une langue est peut être l'un des premiers défis éducationnels de nos sociétés. Des premières années à l'âge adulte, la maîtrise d'une langue et du langage est un problème crucial, qu'il s'agisse de la langue première, i.e. la langue maternelle (il suffit de penser à l'extension de l'illettrisme), ou d'une langue seconde. Les barrières linguistiques ne posent pas que des problèmes d'ordre éducationnel, mais également d'ordre sociologique et politique.

C'est bien plutôt la première des raisons évoquées qui explique le fossé important qui sépare mathématiques et langues. Si les mathématiques scolaires ont une structure bien définie, on ne peut en dire autant des langues. Il est aujourd'hui bien établi parmi les chercheurs et praticiens en acquisition des langues que parler couramment une langue ne peut se réduire à l'acquisition d'une compétence grammaticale. Parce que langage et communication sont indissociables, c'est bien une *compétence communicative* qu'il s'agit d'acquérir. La compétence communicative comprend (suivant la présentation de Bérard (1991), d'après (Hymes 1971 ; Canal & Swain 1980 ; Brumfit & Johnson 1979 ; Moirand 1982)):

- la compétence formelle/grammaticale, se manifestant par une capacité à produire et comprendre des énoncés corrects sur les plans syntaxique, lexical et phonologique. Cette compétence est souvent appelée "compétence linguistique", lorsque l'on donne une interprétation très étroite au champ couvert par la linguistique.
- la compétence sociolinguistique ou socioculturelle : capacité à faire usage d'une langue de façon appropriée au contexte socioculturel ; connaissance de l'histoire culturelle et des relations entre les différents acteurs sociaux.
- la compétence discursive : capacité à comprendre et produire des discours construits de façons différentes suivant les paramètres de la situation de communication (discours oral ou écrit, lettre officielle ou simple correspondance, etc.).
- la compétence référentielle : connaissance des domaines d'expérience quotidienne, professionnelle, de leurs objets et relations ; capacité à trouver l'information correspondante.
- la compétence stratégique : capacité à transmettre efficacement de l'information à un allocutaire, incluant la capacité à mettre en oeuvre des stratégies communicatives

(Bialystok, 1990) de façon à résoudre les problèmes qui peuvent survenir lors de l'échange.

Nombre de chercheurs et praticiens s'accordent à penser qu'une langue est utilisée pour faire des choses, et que, par conséquent, une grande partie des situations d'apprentissage doivent relier son usage à des tâches de résolution de problèmes. Les défenseurs de l'approche communicative en enseignement des langues ont toujours insisté sur ce point, comme en témoigne cet extrait de Johnson (1979, p 200), qui a fait oeuvre de pionnier en la matière :

> << [...] Fluency in communicative process can only develop within a 'task-oriented teaching' - one which provides 'actual meaning' by focusing on tasks to be mediated through language, and where success or failure is seen to be judged in terms of whether or not these tasks are performed >>

La situation est bien différente en mathématiques ou physique, où abstraction et pratique intensive de manipulations symboliques ont, pour ainsi dire, oblitéré toute relation avec la réalité qu'elles représentent. Des expériences fameuses ont montré que cette absence de connexion n'est pas sans conséquence sur l'éducation : par exemple, des étudiants scientifiques peuvent maîtriser la manipulation de formules et être incapables de faire des prédictions correctes dans des expériences simples de dynamique.

Une telle variété d'objectifs d'apprentissage condamne, sans doute, toute tentative de définition d'un type idéal d'environnement d'apprentissage. Au contraire, on peut s'attendre à voir surgir des types différents suivant les compétences pratiquées.

Le listage de cet ensemble de compétences montre que, lorsque l'on utilise une langue, une grande variété d'aptitudes cognitives et de connaissances sont en jeu. Par exemple, les connaissances linguistiques sont traditionnellement décomposées en phonologie, morphologie, syntaxe, sémantique et pragmatique. Un tel éventail de connaissances et savoir-faire associé a donné naissance à un grand nombre d'écoles linguistiques et de théories. L'absence de consensus a eu l'inconvénient en ALAO de surtout privilégier la représentation des connaissances du domaine (au contraire de la situation rencontrée en mathématiques, comme l'a remarqué Nwana dans son premier argument), et par là-même de négliger les stratégies d'acquisition et les aspects pédagogiques. Pour modéliser les connaissances de l'expert et de l'apprenant, les chercheurs ont souvent adopté les points de vue d'une théorie particulière, utilisé son formalisme abstrait pour écrire une grammaire computationnelle[2] et bien sûr utilisé sa terminologie propre. Cela a produit des clivages en linguistique avec les personnes d'abord concernées par l'aspect acquisition des langues, et pas nécessairement par la linguistique formelle. Il existe une position intermédiaire entre engagement complet et rejet total de toute théorie linguistique généraliste. Ainsi en TALN, les chercheurs ont compris que les contraintes imposées par le champ d'application sont de toute première importance. Dans cette optique, chacun est alors libre de décider de fonder son approche en ALAO sur des bases théoriques ou plus "pragmatiques". Une théorie est intéressante lorsqu'elle nous apporte un ensemble de réponses et d'explications suffisamment cohérent et consistant pour nous permettre de décrire et de traiter automatiquement un ensemble, assez

[2] C'est ainsi que j'appellerai toute grammaire d'un sous-ensemble du langage naturel ayant été soit implémentée, soit décrite dans un formalisme présentant des propriétés de calculabilité attestées.

étendu, de phénomènes linguistiques. Elle peut être encore plus attrayante lorsqu'elle s'intéresse à la façon dont s'acquièrent le langage et les langues.

Si l'on considère maintenant l'aspect acquisition des langues, la multiplicité des situations d'acquisition/apprentissage contraste fortement avec le cas des mathématiques. La première langue est acquise en environnements partiellement naturels, partiellement institutionnels. Une deuxième langue peut être acquise comme étant soit une *langue étrangère*, i.e. une langue enseignée à l'école mais qui n'est pas utilisée comme medium d'instruction à l'école, ni comme langue de communication dans le pays (cas du français en Angleterre ou au Japon) ; soit une langue seconde, i.e. une langue qui n'est pas la langue native d'un pays mais est largement utilisée comme medium de communication, aux côtés d'une autre langue (cas du français en Algérie). Suivant la tradition actuelle en linguistique appliquée anglo-saxonne, j'utiliserai le terme "*langue seconde*" de façon générique, recouvrant donc les aspects langue étrangère et langue seconde proprement dite[3].

Lorsqu'une personne apprend une deuxième langue, elle maîtrise déjà (mais pas toujours, lorsqu'il s'agit d'un enfant) une première langue (L1), aussi appelée langue native ou *langue source*, et cherche à acquérir une *langue cible* (L2). Se met alors en place, chez l'apprenant un système intermédiaire, appelé *l'interlangue*. Des analyses d'erreurs des apprenants ont permis de mettre en évidence ce système d'interlangue. Les processus cognitifs, causes d'erreurs, les plus fréquents sont :

- l'emprunt de structures et schémas d'organisation à la langue maternelle (processus de transfert),
- l'extension de structures de la langue cible, par analogie par exemple (processus de surgénéralisation)
- l'expression de nouvelles significations à partir de mots et connaissances partielles déjà acquises (stratégies de communication).

L'existence de deux systèmes (sans compter l'interlangue) en L2, et par conséquent de processus de transfert, n'a pas d'équivalent en mathématiques. Pour s'en convaincre, il suffit d'imaginer les difficultés que rencontrerait un professeur européen de mathématiques s'il devait enseigner notre approche de l'algèbre en Inde ... à des jeunes du pays qui auraient appris l'algèbre suivant la tradition indienne des siècles précédents !

Mais il ne faut pas surestimer le processus de transfert (Odlin, 1989). Les chercheurs en Acquisition d'une Langue Seconde (ALS) essayent d'abord de comparer, pour une langue donnée, les étapes d'acquisition que traversent les enfants dans cette langue avec celles des apprenants pour qui cette langue est la L2. Et ils ne pensent pas que le transfert à partir de la L1 des apprenants soit la cause première des problèmes acquisitionnels, au moins lorsque ce ne sont plus des vrais débutants. Apprendre une seconde langue n'est pas plus difficile (ou est aussi difficile) qu'en apprendre une première. Le terme "naturel", souvent utilisé pour qualifier la situation d'apprentissage en L1, n'implique pas que la maîtrise de sa langue maternelle soit une tâche aisée. Les premières années, les enfants consacrent une grande partie de leurs facultés cognitives à assimiler ce qu'est le langage et à acquérir leur langue maternelle, parce que c'est la seule façon pour eux d'être compris et d'avoir accès au monde contrôlé par les adultes. Des expériences ont montré que, sur certaines tâches linguistiques,

[3] Sur ces problèmes terminologiques, il n'existe en français, à ma connaissance que le dictionnaire de Galisson et Coste (1976) qui reflète mal la réalité des recherches actuelles en acquisition.

les apprenants adultes en L2 sont plus efficaces que les enfants natifs. Cependant le temps limité d'immersion dans la langue cible est sans doute la cause principale des difficultés d'acquisition. C'est une des raisons qui rendent les environnements informatiques d'apprentissage aussi attractifs.

Une telle variété de situations d'apprentissage en L2 a tout naturellement motivé le développement d'un large éventail de théories sur l'acquisition, théories contradictoires ou complémentaires. Ce n'est pas le lieu de les décrire ici (MacLaughlin, 1987), mais nous ferons allusion à plusieurs d'entre elles dans la partie suivante.

## 2.2. ALAO : traditions et objectifs

L'usage des environnements informatiques d'ALAO s'étend aujourd'hui rapidement. Depuis plus de 10 ans, une importante minorité des secteurs éducatifs supérieurs de plusieurs pays, et plus récemment d'institutions de formation, exploitent les possibilités offertes par l'informatique pour améliorer ou soutenir l'apprentissage des langues (Thomson & Chesters, 1992). Le milieu ALAO est bien organisé, en particulier grâce à :

- ses journaux spécialisés tels Computer Assisted Language Learning (CALL), System, CALICO (Computer Assisted Language Learning and Instruction Consortium) et ReCALL ; sans compter, bien sûr, les nombreux journaux spécialisés en enseignement des langues et qui accordent, aujourd'hui, une place non négligeable à l'ALAO.
- ses associations et conférences annuelles telles CALICO, EUROCALL (European Association for CALL) ou la conférence d'Exeter en Angleterre,
- des centres de ressources, comme le CTI, Centre for Modern Languages de l'université de Hull en Angleterre, qui offre informations, formations, évaluations de logiciels à tous les partenaires institutionnels du secteur éducatif britannique.
- des programmes européens (DELTA, LINGUA, COMETT) et nationaux (comme le consortium TELL - Technology Enhanced Language Learning - en Grande-Bretagne) qui concentre et unifie les développements d'environnements informatiques sur une trentaine de sites universitaires.

Une telle situation[4] représente pour l'approche IA-ED en ALAO à la fois une opportunité et un défi. Du fait que l'ALAO n'est plus circonscrite à un groupe restreint d'enthousiastes mais qu'elle a gagné un plus large public, et du fait qu'elle est aujourd'hui un enjeu commercial, la nécessité de garantir un espace pour des recherches avancées est ouvertement légitimée. Corollairement, cette situation pose un réel défi aux chercheurs, qui doivent démontrer qu'une partie de leurs travaux peuvent être applicables, directement ou indirectement, et que leurs objectifs pédagogiques sont concordants avec les centres d'intérêts mentionnés dans la section précédente.

A ses début, dans les années 1980, l'approche IA-ED en ALAO n'a pas été très convaincante parce que ses objectifs pédagogiques n'étaient pas innovants et parce que l'éventail des méthodes ou techniques extraites du TALN ou de l'IA était trop étroit. De plus

[4] On se contentera de remarquer, pour l'instant, que : 1) le milieu de la recherche française est pratiquement complètement absent des publications et manifestations scientifiques en ALAO ; 2) ministères et institutions publiques français s'intéressent encore trop peu aux environnements informatiques d'aide à l'apprentissage en général, aux niveaux scientifiques ou applicatifs. Je reviendrai sur ce constat en dernière partie.

les méthodes utilisées induisaient des calculs de trop grande complexité. La plupart des systèmes IA-ED en apprentissage des langues étaient des *systèmes de calcul automatique d'erreurs*. De tels systèmes ont pour objectifs de permettre la pratique et l'évaluation de la compétence grammaticale. Ils mettent l'accent sur le développement du module expert, organisé autour d'une grammaire computationnelle censée refléter les connaissances expertes du linguiste. Des procédures , directement inspirées de l'approche "bug/mal-rules" en IA-ED, sont ajoutées à la grammaire computationnelle de façon à gérer le diagnostic d'erreurs. Ces procédures mettent à jour le modèle de l'apprenant et sont supposées informer le système sur l'état des connaissances et des compétences de l'apprenant. La plupart du temps, le système interagit avec l'utilisateur au travers d'énoncés qui ont été détachés de toute situation réelle de communication (une présentation détaillée de ces systèmes est donnée dans (publication 3)). Les points faibles de cette approche sont : son absence de prise en compte des données acquisitionnelles réelles, son manque de motivation cognitive, sa présentation déformée des langues, qui ignore les autres compétences communicatives, enfin sa surestimation de l'importance des techniques de traitement d'erreurs en TALN.

Bien sûr, ces critiques des systèmes de calculs automatiques d'erreurs ne cherchent pas à condamner tout effort dans ce domaine. Elle visent seulement à ne pas confiner les travaux dans cette seule perpective ; tout comme la recherche en IA-ED, en général, ne saurait se limiter à l'approche bug/mal-rule, quelle que puisse avoir été son importance historique.

A partir des années 1990, de nombreux travaux présentés dans des colloques internationaux spécialisés[5] prouvent que les centres d'intérêts et les objectifs ont changé dans les recherches IA-ED appliquées à l'ALAO. Je citerai, ci-dessous, neuf desiderata clés pour l'ALAO, extraits de (Oxford, 1993). Ils me serviront de référence dans la suite de ma présentation. Si dans cette section 2.2, j'ai distingué, comme le voulait la tradition, l'ALAO traditionnelle des approches spécifiques IA-ED du domaine, je n'opérerai plus cette distinction, à partir de maintenant, suivant en cela la définition ouverte de l'ALAO que j'ai donnée en partie 1. Ces desiderata pour l'ALAO sont :

1. La compétence communicative doit être la pierre angulaire de l'ALAO.
2. L'ALAO doit fournir une assistance linguistique appropriée aux besoins courants de l'apprenant. Le concept d'assistance en situation d'apprentissage renvoie à ce que Vygotsky appelait la "zone proximale de développement" (publication 4).
3. L'ALAO doit offrir un "Input" linguistique riche et authentique.
4. Le modèle de l'apprenant en ALAO doit être construit en intégrant des styles différents d'apprentissage.
5. Le contenu linguistique offert par un environnement ALAO est plus facilement assimilé s'il est présenté sous forme d'associations, elles-mêmes organisées autour de thèmes intéressants et pertinents, et autour de tâches proches de celles rencontrées en situation réelle de communication en L2.
6. L'ALAO doit impliquer des types différents d'interaction, ne se limitant pas à celles de type tuteur-élève.

[5] Comme le colloque de l'OTAN tenu à Washington en 1990 (Swartz & Yazdani, 1992), celui de UMIST en Angleterre en 1991 (Thompson & Zähner, 1992), ou le colloque ARI (Holland, Kaplan, & Sams, sous presse), tenu à Alexandria, Etats-Unis, en 1993.

7. L'ALAO doit fournir un système de correction d'erreurs utile et pertinent en concordance avec les besoins changeants de l'apprenant.
8. L'ALAO doit impliquer toutes les aptitudes langagières appropriées (écriture, lecture, parler, écoute) et étudier comment la mobilisation de certaines capacités peut aider le développement d'autres.
9. L'ALAO doit former les utilisateurs à devenir progressivement des apprenants autonomes en leur enseignant explicitement les stratégies d'apprentissage en langues appropriées (O'Malley & Chamot, 1990).

# 3. POINTS D'INVESTIGATION DANS MA RECHERCHE

La présentation de mes travaux suivra un plan en partie identique aux étapes de déroulement d'un projet de recherche ALAO, en mettant particulièrement l'accent sur l'expertise du domaine, la modélisation de l'apprenant, et les stratégies pédagogiques adoptées. Seront discutés les points suivants :

- Objectifs, recherche d'expertises et création de matériaux,
  - -- Analyse des besoins, besoin des apprenants
  - -- Quelle norme pour la langue cible ?
  - -- Création des matériaux, études linguistiques, procédures de TALN
- Modélisation de l'apprenant, diagnostic
- Pédagogie: contenu et stratégies
- Evaluation, intégration dans le curriculum

Chaque thème est décliné par rapport à la contribution respective de chaque projet de recherche. Chaque projet n'est pas nécessairement représenté dans chaque thème. En effet, certains avaient des objectifs limités, ou bien n'ont pas encore atteint un stade de développement suffisamment avancé. Chaque point discuté renvoie aux publications correspondantes fournies en annexe, sauf dans le cas du projet ALEXIA, qui est ici développé pour la première fois (en effet, la seule publication donnée en annexe sous ce label est antérieure à la spécification précise qu'il a aujourd'hui atteinte). Les objectifs de chacun de mes projets de recherche sont explicités dans la première sous-partie, "Analyse des besoins".

## 3.1. Objectifs, recherche d'expertises et création de matériaux

Où l'on illustre les principes :

1. La compétence communicative doit être la pierre angulaire de l'ALAO.
3. L'ALAO doit offrir un "input" linguistique riche et authentique.
5. Le contenu linguistique offert par un environnement ALAO est plus facilement assimilé s'il est présenté sous forme d'associations, elles-mêmes organisées autour de thèmes intéressants et pertinents, et autour de tâches proches de celles rencontrées en situation réelle de communication en L2.
8. L'ALAO doit impliquer toutes les aptitudes langagières appropriées (écriture, lecture, parler, écoute) et étudier comment la mobilisation de certaines capacités peut aider le développement d'autres.

Adopter ces principes de travail déplace considérablement le point de départ du travail des concepteurs de systèmes d'ALAO. En effet dans les systèmes à calcul automatique d'erreurs, les besoins des apprenants sont définis à priori et le travail commence directement par l'écriture d'une grammaire computationnelle qui doit analyser un corpus d'énoncés fabriqués artificiellement. Ici, au contraire tout commence par la définition du besoin des apprenants par rapport au registre de compétence choisi. Il s'agit ensuite de rechercher l'expertise correspondante, qui déterminera la nature de l'Input proposé aux apprenants.

Ensuite seulement peuvent être créés les matériaux langagiers et implémentés les systèmes de TALN, si besoin est.

### 3.1.1. ANALYSE DES BESOINS, BESOIN DES APPRENANTS

La définition des besoins est loin d'être une question simple si l'on décide de s'intéresser à la façon dont les sujets apprennent, sans se référer à des normes linguistiques définies à priori. Nous verrons que théories linguistiques et besoins des apprenants peuvent coïncider ou, au contraire diverger.

Le projet BELLOC nous donne l'occasion de discuter des besoins en apprentissage grammatical. J'explique pourquoi un formalisme utilisé en linguistique-informatique ne peut servir de grammaire pédagogique. Le projet ELEONORE nous offre l'opportunité de sortir du cadre trop étroit de la syntaxe prise isolément, et examine ses relations avec la pragmatique. La notion de grammaire pédagogique est alors étendue aux aspects notionnels-fonctionnels. ELEONORE concerne donc l'acquisition de compétence grammaticale et sociolinguistique. Avec le projet ALEXIA, nous délaissons la grammaire pour nous tourner vers l'acquisition lexicale. Les mots et expressions lexicales sont aujourd'hui vus comme des entités linguistiques complexes, pouvant imposer des contraintes sur la structure de tout ou partie d'un énoncé. De plus ce sont des entités interconnectées de multiples façons. Variétés linguistiques et associations doivent servir d'hypothèses de base de travail pour aider l'apprenant à accroître son vocabulaire, dont la limitation peut constituer un handicap majeur. Enfin avec le projet CAMILLE "Travailler en France" nous abordons la question de l'apprentissage pour des objectifs spécifiques. Les besoins des apprenants délimitent des objectifs d'utilisation de la langue. La formation est alors construite à partir des tâches à accomplir. Le contenu linguistique est ensuite adapté à ces buts.

#### 3.1.1.1. BELLOC et l'apprentissage grammatical

Le projet BELLOC, auquel j'ai participé en Angleterre après mon travail de thèse s'est centré autour du problème du diagnostic des connaissances grammaticales des apprenants en L2 (publications 2 et 3). Le problème de l'enseignement ou non de la grammaire a fait couler beaucoup d'encre (Ellis, 1990 ; Besse, 1986). Toute une série d'expérimentations, ainsi que l'examen des résultats de pédagogie communicative outrancière (refusant tout enseignement de grammaire) dans les pays anglo-saxons ont montré que l'acquisition d'une compétence grammaticale passait par un apprentissage explicite de la grammaire. Toute la question est alors de savoir quel type de grammaire enseigner et comment le faire.

Ayant juste terminé mes travaux de thèse, pendant lesquels j'avais développé une grammaire d'unification, j'ai alors examiné quel était l'applicabilité de ces grammaires computationnelles. Beaucoup de chercheurs ont essayé d'utiliser directement de telles grammaires, expliquant qu'elles permettaient à l'apprenant de taper librement des énoncés qui seraient ensuite diagnostiqués automatiquement. Pour étayer leur argumentation les utilisateurs de ces formalismes ont coutume de rappeler leurs qualités (Shieber, 1986): déclarativité, calculabilité, adéquation linguistique, i.e. naturalité de description des phénomènes linguistiques pour les linguistes ou spécialistes en charge du développement de la grammaire. En fait, il y a deux interprétations très différentes de la naturalité suivant que l'on adopte le point de vue de la linguistique informatique (ou théorique) et celui de la linguistique appliquée. Dans le premier cas, l'objectif est de décrire et expliquer les

phénomènes linguistiques. Les grammaires produites sont alors de nature descriptive. En linguistique appliquée, au contraire, c'est un point de vue prescriptif qui prévaut.

Prenons l'exemple des travaux de thèse de Patricia Manning (1995) sur l'acquisition du genre en français L2 par des anglophones. Le traitement du genre (attribution de traits morphologiques aux items lexicaux, règles d'accord) ne posent plus de problèmes majeurs en linguistique-informatique, sauf peut-être dans certains cas de dépendances lointaines et de coordination. Pour des apprenants anglophones c'est un problème important au début et qui perdure jusqu'à des stades très avancés. Ils rencontrent des difficultés pour appliquer la plupart des règles où le genre est en cause : accord déterminant-nom, adjectif-nom, élision et contraction, combinaison de toutes ces règles. Les apprenants éprouvent donc des difficultés sérieuses pour tirer parti du contexte syntaxique pour déterminer le genre des noms, sans parler de celles rencontrées en situation de production. Quant à leurs stratégies de devinement du genre des mots, sans contexte ou en contexte restreint, elles s'appuient de façon inappropriée sur des indices orthographiques, phonétiques ou sémantiques. Ces stratégies de devinement sont d'ailleurs bien différentes de celles mises en oeuvre par les natifs. Des théories sur l'acquisition des langues, comme le Modèle de Compétition (MacWhinney, 1989) ou la Grammaire Universelle (GU) chomskienne (White, 1989), appuyées par de nombreuses expérimentations, offrent des explications à ces difficultés acquisitionnelles. Brièvement, le genre en anglais est un épiphénomène de nature purement sémantique, et dans cette langue l'ordre des mots joue un rôle très important, au contraire de la morphologie ; en français, la morphologie joue un rôle de tout premier plan (donc les natifs savent bien utiliser tous les indices de cette nature), le système du genre est très important et est extériorisé par des variations d'ordre morphologique. A partir de ces explications (et également d'études sur les stades de développement de l'acquisition en L1 et L2), il est possible de construire une intervention pédagogique en *ordonnant les règles* et en choisissant les *contextes appropriés*, permettant de favoriser la mise en place de certaines stratégies et de minimiser les phénomènes de transfert négatif. Il est alors possible de développer dans l'environnement informatique des activités permettant à l'apprenant de travailler de façon explicite, implicite ou exploratoire sur ces règles ; puisque l'on sait (Ellis, 1990), et Patricia Manning l'a encore vérifié, qu'il est nécessaire de varier ces stratégies pédagogiques en fonction des niveaux des apprenants et de leurs styles cognitifs propres.

Cette évocation du besoin des apprenants concernant l'acquisition du genre montre bien qu'une grammaire computationnelle n'apporte aucune réponse en soi, ni sur la définition de ces besoins, ni sur leur traitement. Cependant, Matthews (1994) établit une distinction, pour l'apprentissage grammatical, entre des grammaires computationnelles construites avec, ou sans, référence à des théories linguistiques, particulièrement des théories comme la Grammaire Universelle qui ont un point de vue explicatif sur certains problèmes d'acquisition des langues. Il condamne l'utilisation des grammaires DCG, comme étant sans principe. A mon avis, l'auteur introduit une certaine confusion entre théories linguistiques et formalismes linguistiques. En effet, des formalismes comme DCG, PATR, etc. (Shieber, 1986) permettent d'implémenter des théories linguistiques variées. Des implémentations de GU ont d'ailleurs eu lieu en DCG. Mais le problème n'est pas là. Le système SCRIPSI (Catt & Hirst, 1990), par exemple, utilise une grammaire computationnelle construite en référence à GU. Si le diagnostic d'erreurs grammaticales géré par SCRIPSI est sans doute l'un des plus fins décrits jusqu'à présent, rien n'est dit de la façon dont cette grammaire permet de répondre aux

besoins des apprenants. Si la théorie chomskienne et (surtout) les expérimentations qu'elle a suscitée apportent des informations déterminantes sur l'acquisition, il reste cependant à prouver comment une grammaire computationnelle inspirée du modèle chomskien répond à la détermination du contenu et de la forme de ce qui doit être enseigné.

Dans mes articles sur BELLOC, j'ai expliqué, comme d'autres chercheurs, la nécessité de mettre au point une *grammaire pédagogique* et proposer une structure de "règles applicables", qui peuvent servir de cadre à la description de cette grammaire. Le choix du métalangage à introduire progressivement, de l'ordonnancement des règles en fonction des difficultés d'apprentissage et du savoir des apprenants, de leur description consistante, des exemples significatifs associés est un problème ouvert. Les grammaires pédagogiques présentées dans les manuels ne répondent pas à ces problèmes (ni même sur le choix de la norme de langue, comme nous le verrons plus loin). Elles ont été fortement critiquées par les chercheurs de linguistique appliquée, mais ceux-ci n'ont à ce jour proposé que des descriptions partielles, sans rechercher la rigueur formelle nécessaire à leurs implémentations. La traduction d'une grammaire pédagogique en grammaire computationnelle est en effet la seule garantie de conhérence des descriptions, et elle est la condition nécessaire à tout diagnostic automatique.

L'arrêt de mes travaux dans ce domaine, et le manque d'informations actuelles sur la portée de travaux très intéressants entrepris autour de grammaires à caractère pédagogique comme IPG (Pijls, Daelemans & Kempen, 1987), m'incitent à croire qu'il est plus pertinent de travailler en ALAO sur des problèmes grammaticaux limités, suivant la méthodologie utilisée en partie par Manning, méthodologie dont nous explorons les différents aspects dans ce chapitre 3.

#### 3.1.1.2. ELEONORE et l'apprentissage syntactico-pragmatique

D'après Valdman (1990 : p. 88):

> "Une grammaire pédagogique compatible avec un enseignement de langue étrangère orienté vers la communication doit être solidement ancrée dans la sociopragmatique ; c'est-à-dire qu'elle doit refléter un usage fonctionnel de la langue imbriqué dans des situations de communications. Autrement dit, elle doit être notionnelle-fonctionnelle en nature, et doit se centrer sur le sens et la fonction des marques linguistiques plus que sur leur forme de surface".

Valdman souligne l'imbrication des connaissances et usages de natures syntaxiques, sémantiques et pragmatiques. Après avoir travaillé dans BELLOC sur des problèmes purement grammaticaux à contexte très réduit, j'ai donc décidé de choisir le domaine plus ouvert des interrogatives. ELEONORE est un environnement d'apprentissage des interrogatives directes du français par des anglophones de niveau intermédiaire à avancé(publications 1, 4, 9 et (Renié, 1995)).

L'exemple 1 montre la variété des formes interrogatives du français par rapport à l'anglais. Cette variété est le reflet des aspects multiples, sociolinguistique, sémantique et conversationnel de l'interrogative en français, le reflet des relations entre interrogative et interrogation.

(EX 1)
Where is the restaurant ?

Où se trouve le restaurant ?
Où le restaurant se trouve-t-il ?
Le restaurant se trouve où ?
Où est-ce que le restaurant se trouve ?

Nous avons retenu cinq paramètres pragmatiques et un sémantique pouvant influencer le choix du type syntaxique de l'énoncé lorsque l'on pose une question en français. Ces paramètres sont : le canal de communication, la force illocutoire, la fonction dialogique de la question dans un échange conversationnel, le contexte social et la position sociale réciproque des interlocuteurs/correspondants. Le locuteur doit savoir comment, quand et pourquoi se formule tel énoncé interrogatif. Donc une combinaison de critères pragmatiques contraignent la construction syntaxique. De plus, travailler sur l'interrogative permet de plonger naturellement l'apprenant dans un cadre conversationnel, du fait de l'importance de ces formes linguistiques dans l'interaction langagière.

Tous les natifs adultes savent utiliser (et utilisent) ces différentes formes, mais de façon différenciée suivant le contexte, en particulier le contexte social. Il est donc intéressant pédagogiquement de proposer aux apprenants une vision plus fine de la langue, c'est-à-dire une vue qui prenne en compte une variété de niveaux, afin que ceux-ci

> "viennent à comprendre comment [certains aspects de la compétence communicative] sont intégrés à une culture cible et à une culture sociale entière, et que naisse une tolérance envers des particularités que l'on ne retrouve pas dans le discours de la communauté d'origine des apprenants." (Valdman, op. cit. : p.84)

Qu'en est-il de l'enseignement institutionnel des formes interrogatives en FLE ? Une approche "traditionnelle", aujourd'hui quelque peu dépassée en France, classerait ces formes suivant des critères purement syntaxiques, en les faisant apprendre toutes simultanément, ou éventuellement en introduisant un ordre sur des critères de fréquence et simplicité (les formes CONSERV et ESK arrivant en tête). En réaction, une stratégie évidente d'apprentissage conduit les apprenants à n'utiliser qu'un registre restreint de formes, puisque rien ne justifie vraiment dans cette perspective leur variété. Quant aux méthodes de FLE basées sur la compétence communicative, étudiées dans (Renié 1995), elles sont passées d'une approche dans laquelle les différentes structures syntaxiques étaient étiquetées de façon rigide selon le niveau de langue auquel elles étaient censées appartenir, à une tendance plus souple et plus fidèle à la langue parlée. Cependant l'approche récente, rejetant toute description rigide, s'interdit souvent d'expliquer tout phénomène linguistique. De plus la variété des paramètres pragmatiques qu'elles retiennent est trop limitée. Enfin une question est toujours laissée en suspens : comment la pragmatique est-elle acquise en L2 ? Comment la faire apprendre ?

Nous reviendrons sur ces questions qui doivent être reliées à celles concernant l'acquisition de la pragmatique en L1, et la transposition ou non de cette compétence en langue maternelle à la L2.

#### 3.1.1.3. ALEXIA et l'apprentissage lexical

Après avoir été considérée pendant près de 20 ans comme un problème tout à fait auxiliaire dans l'acquisition d'une langue seconde, la compétence lexicale retrouve aujourd'hui une

place de choix dans les préoccupations des chercheurs sur l'acquisition des langues (Carter & McCarthy, 1988 ; Helmy, 1989) . Le fait qu'un apprenant ne maîtrise pas un niveau seuil de vocabulaire est considéré comme un facteur fortement préjudiciable à la communication. Cette constatation n'entraîne pas pour autant un retour aux méthodes d'apprentissage de listes de mots, répandues dans les années 60. Les études en linguistique théorique et appliquée, ainsi qu'en psycholinguistique montrent que le mot est une unité linguistique complexe porteuse de connaissances morphologiques, syntaxiques, sémantiques et même pragmatiques. Le choix d'un mot dans la production d'un énoncé peut imposer une structure sur tout l'énoncé. Certains mots (nombreux et d'usage fréquent par les natifs) ont des fonctions discursives essentielles: expression d'actes de parole, enchaînement de parties du discours.

Les spécialistes de l'apprentissage et de l'enseignement du lexique définissent aujourd'hui le problème de la mesure de la connaissance d'un mot par un apprenant en des termes proches de ceux utilisés en linguistique lexicale. Par exemple, connaître un mot c'est (Richards, 1985):

- savoir dans quel contexte il est utilisé à l'oral ou à l'écrit (fréquence d'utilisation, collocations associées)
- appréhender les limitations de son usage suivant les variations de fonctions ou de situations,
- connaître ses comportements morphologiques, syntaxiques, sémantiques,
- pouvoir l'intégrer dans son propre lexique mental et établir les relations nécessaires avec les autres mots dans le réseau associé.

Le taux de l'acquisition du vocabulaire en L2 est étonnamment faible. Certains chercheurs estiment, par exemple, que l'acquisition de 2000 mots en 5 ans est au-delà de la capacité de beaucoup d'apprenants. Ceci peut être opposé au taux d'apprentissage estimé en L1: 3000 mots par an durant les années d'école. On considère que la plus grande part de l'apprentissage en L1 s'effectue au travers de l'oral, d'une manière moindre de la lecture et de la télévision, plutôt qu'au travers d'un environnement institutionnel spécifiquement dédié à l'instruction du vocabulaire. Les apprenants en L2 sont donc dans la pire des situations: il faut des années d'immersion aux natifs pour acquérir une compétence lexicale minimum, or l'apprentissage en L2 est généralement beaucoup plus court (puisqu'il n'est pas intensif) et se passe surtout en situation institutionnelle, celle où le rendement d'acquisition est le plus bas. Enfin lorsque le taux d'acquisition lexical est trop bas, les apprenants ne peuvent transférer certaines stratégies efficaces qu'ils adoptaient en L1, dans la L2[6]. Il y a donc là un défi intéressant pour l'ALAO.

Un certain nombre de systèmes d'ALAO se sont orientés vers la mise à disposition des apprenants de ressources lexicales "traditionnelles" (dictionnaire électronique organisé alphabétiquement, concordanceurs et bases de textes plus ou moins homogènes) autour desquelles il construisent des exercices lexicaux et surtout des activités où l'apprenant peut collecter et organiser ses informations lexicales. Pour notre part, compte tenu de ce que l'on connaît de l'organisation du lexique mental (Aitchison, 1987 ; Levelt, 1989) et la définition que l'on donne de l'acquisition d'un mot (cf. ci-dessus), il nous est apparu important d'organiser ces activités lexicales autour d'une base de données "active" contenant toutes les

[6] Un bon lecteur en L1 doit connaître 3000 mots en L2 pour être capable de transférer ses compétences de lecture dans la langue cible

informations nécessaires (morphologiques, syntaxiques, sémantiques et pragmatiques). Par "active", nous entendons le fait que le système doit savoir mettre en oeuvre ses connaissances pour:

- dans une situation d'aide à la compréhension, offrir des accès multiples à l'apprenant, visualiser sous formes de réseaux multiples les associations entre mots,
- dans une situation d'aide à la production , diagnostiquer les énoncés de l'apprenant de façon à savoir non seulement corriger certaines erreurs, mais également offrir des paraphrases.

Enfin, nous avons également décidé de ne pas restreindre le lexique aux mots simples, mais d'y inclure les collocations et expressions semi-figées (dont les expressions idiomatiques). Ces expressions sont très peu considérées en ALAO. Pourtant ces unités linguistiques sont numériquement beaucoup plus importantes dans les dictionnaires électroniques de langues telles que le français ou l'anglais que les entrées lexicales simples[7]. Elles sont d'un usage fréquent par les natifs, mais sont difficiles à acquérir par les apprenants non-natifs. Aux yeux des didacticiens en L2, les justifications pour les enseigner sont multiples: ces expressions fournissent un matériel de base pour l'analyse et la segmentation du discours. Leur maîtrise doit permettre à l'apprenant de ne pas violer certaines restrictions lexicales, lui éviter de commettre des erreurs de registre dans la production de son discours. Elles peuvent faciliter ses productions écrites ou orales, puisqu'elles lui évitent de porter une attention soutenue à chaque mot pris individuellement. Elles autorisent l'apprenant à porter son attention sur des structures plus larges du discours et sur les aspects sociaux de l'interaction.

#### 3.1.1.4. CAMILLE "Travailler en France" et l'apprentissage à objectifs spécifiques

Le projet européen CAMILLE a pour but de fournir, à partir d'une même plate-forme informatique hypermédia des cours de formation dans quatre langues (anglais, espagnol, français et hollandais), correspondant soit à des niveaux de formation générale pour débutants (en espagnol, français et hollandais), soit à des formation sur objectifs spécifiques pour des apprenants de niveaux intermédiaire à avancé (en anglais et français) (publication 5). C'est, à notre connaissance, la première fois en ALAO qu'une telle expérience de développement de formations aussi larges à partir d'une même plate-forme informatique, est tentée. Une approche différente est explorée par le projet britannique TELL (Thompson, 1992), où plusieurs dizaines d'universités cherchent à coordonner spécifications, développements et évaluations de nombreux logiciels, couvrant chacun des besoins plus limités, dans de plus nombreuses langues.

Au laboratoire LRL, à Clermont, notre équipe s'est intéressée au français de spécialité (publication 7). Dans cette publication, nous discutons de l'intégration d'une approche

[7] Les travaux au LADL et ailleurs ont montré leur forte proportion dans la langue (20000 expressions verbales figées contre 8000 ou 12000 verbes libres ; 6000 adverbes figés contre 2000 libres ; 80 000 noms simples contre 300 000 ou 400 000 noms composés) et leur forte probabilité d'occurrences dans les textes. Quand il y a ambiguïté potentielle, c'est presque toujours l'interprétation idiomatique qui est la bonne, sauf en cas de jeux de mots.

communicative pour des besoins spécifiques en langues (LOS)[8] dans le cadre d'un environnement hypermédia.

Le public visé est celui des employés de PME et des étudiants de sciences économiques et de gestion. L'analyse des besoins se décompose en analyse de la situation cible et des besoins d'apprentissage. La première concerne la détermination des tâches que les apprenants auront à accomplir et des usages langagiers correspondants. La seconde doit révéler comment les apprenants vont apprendre à utiliser la L2 : quelles motivations ont-ils pour se former ? Comment apprennent-ils ? Quelles ressources leur sont accessibles ? Quel est leur profil ? Quand et où la formation aura-t-elle lieu ?

Pour répondre à toutes ces questions nous avons élaboré un questionnaire et mené une enquête dans les entreprises et auprès des responsables de formation de langues. Les résultats de cette enquête recoupés avec ceux des études européennes, nous ont permis de confirmer la place que pouvait occuper l'apprentissage autonome et de déterminer les situations cibles autour desquelles organiser notre développement sur certains aspects de la vie professionnelle en France, à savoir la recherche d'un emploi et la situation de vente. Suivant le point de vue de Johnson (cité en section 2.1) nous avons construit le premier module autour d'une tâche principale "A la Recherche d'un Emploi", puis développé les bases de connaissances et les activités permettant à l'apprenant d'accomplir cette simulation de tâche. Dans cette approche résolution de problèmes, ce sont donc les connaissances et procédures immergées dans un contexte socioculturel défini qui ont déterminé l'architecture du logiciel. Les connaissances linguistiques, présentes dans les ressources associées au logiciel, et le contenu linguistique des activités sont seulement alors déterminés à partir du contexte du problème à résoudre (cf. figure 1). Certaines activités linguistiques peuvent être spécifiquement ajoutées de façon à favoriser une systématisation des compétences visées afin qu'elles puissent être appliquées à d'autres situations.

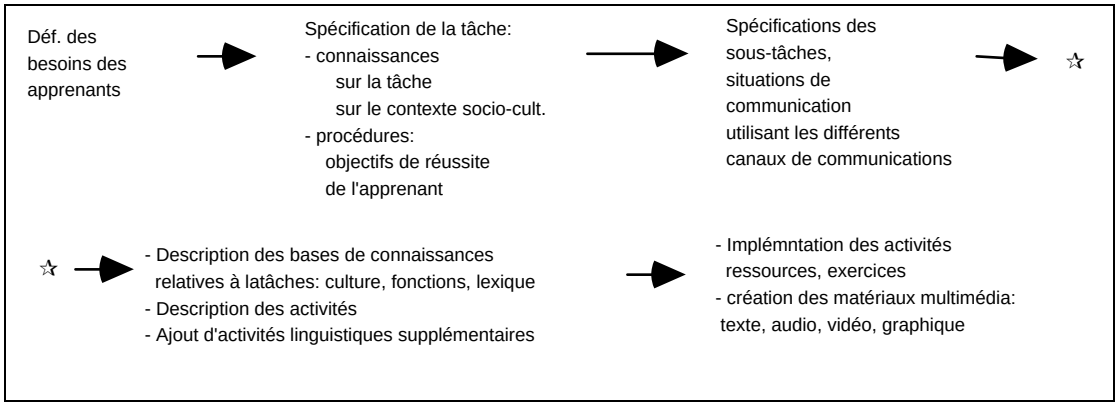



**Figure 1**: Méthodologie de création d'un cours hypermédia LOS.

De façon à éclairer les discussions ultérieures, il peut être utile de rappeler les principes théoriques sous-jacents, principes expliquant l'adéquation entre une approche avec objectifs spécifiques et l'apprentissage d'une L2 en général (Hutchison & Waters, 1987):

[8] Nous utiliserons l'abréviation LOS pour référer à l'apprentissage ou l'usage d'une langue pour des objectifs spécifiques (Lehman, 1993). Cette abréviation est l'équivalent de LSP (Language for Specific Purposes) couramment utilisée en anglais (Hutchison & Waters, 1987).

l'apprentissage d'une langue peut être en partie assimilée à une *procédure de prise de décisions* , et ne se réduit pas seulement à un problème de connaissances linguistiques (l'apprenant a des connaissances de base et un certain savoir-faire dans les tâches abordées) ; l'apprentissage de la langue L2 s'appuie sur *l'expérience de l'apprenant dans sa L1* (il sait se servir de sa langue source pour mener à bien des tâches similaires dans son pays) ; *l'apprentissage d'une langue s'accomplit partiellement de façon incidente* ( une langue peut s'apprendre en résolvant un problème particulier, sans que cela soit présenté comme une tâche de nature linguistique).

### 3.1.2. QUELLE NORME POUR LA LANGUE CIBLE ?

Quelle doit être la norme de référence pour l'apprentissage d'une langue ? En ALAO, l'expert a pratiquement toujours été (et l'est sans doute encore majoritairement aujourd'hui) le linguiste. Examinons les implications d'un tel choix à partir du cas des interrogatives. Le linguiste Le Goffic (1993), dans son étude de la "Grammaire de la phrase française", lorsqu'il examine la forme interrogative Qu-Sujet-Verbe qualifie les énoncés de l'exemple 2 comme étant des "énoncés marqués, à connotation de langage relâché ou vulgaire, et clairement réprouvés par la norme".

(EX 2)
A quoi tu penses ?
Où tu vas ?

Cette caractérisation cherche à indiquer que ces énoncés se distinguent de ceux de l'exemple 3.

(EX 3)
A quoi penses-tu ?
Où vas-tu ?

Cependant, parler de "langage relâché ou vulgaire" - outre le fait que nous ne puissions que difficilement admettre le qualificatif de "vulgaire" ici - ne signifie pas grand chose si le lecteur ne sait pas à quelle réalité l'auteur fait référence. D'autre part, il serait facile de déduire d'une telle définition que des énoncés comme ceux de l'exemple 2 sont marginaux dans la langue par rapport à ceux de l'exemple 3 ... ce qui est le contraire de la réalité à l'oral, toutes classes sociales confondues. La forme interrogative Qu-Sujet-Verbe est donc présentée par rapport à une norme figée qui néglige un ensemble d'aspects importants de la communication et s'appuie sur des critères pragmatiques très limités.

Di Vito (1991) traite du fait qu'un apprenant typique en L2, lorsqu'on lui présente une règle grammaticale, n'a aucune notion de ce qu'est, dans la vie réelle, sa fréquence d'occurrence, son champ fonctionnel, ou usage social. Si le contexte pédagogique ne lui fournit pas de telles informations, l'apprenant n'a d'autres choix que de supposer que les formes grammaticales sont également généralisables, également importantes sur le plan communicatif et également productives en langue cible. Ce qui est, bien sûr, faux. Dans son article, elle établit une comparaison entre la présentation de quatre problèmes grammaticaux importants proposée par des ouvrages de référence en FLE et l'analyse de ces mêmes points grammaticaux à partir d'un corpus de français courant. Elle relève une différence importante

entre le français tel qu'il est réellement parlé ou écrit et tel qu'il est présenté dans les manuels. Par exemple, des structures négatives utilisées par tous les natifs à l'oral, en communication soutenue ou non, sont complètement absentes des manuels.

Cette situation a des conséquences importantes. En effet, outre le fait que l'on n'enseigne pas le français tel qu'il est réellement utilisé, on construit des normes sur des critères subjectifs qui présentent la langue cible à partir de formes trop complexes à maîtriser par les apprenants. Car les formes non décrites, mais couramment utilisées sont en partie le résultat d'applications de stratégies de simplification de la part des natifs. Il est de plus nécessaire de construire une recherche sur l'acquisition sur des données réelles. Ainsi le Modèle de Compétition, qui s'est d'abord intéressé à l'acquisition en L1, fait reposer sa théorie sur une compétition entre une série d'indices, qui expliquerait les stades et les difficultés d'acquisition. Pour qu'un indice soit considéré (par exemple, la position préverbale) il faut d'abord (c'est une condition nécessaire, mais pas suffisante) qu'il soit disponible et fiable. La disponibilité exprime le fait que l'indice soit présent à chaque fois qu'on en a besoin ; la fiabilité, le fait que sa prise en compte conduise à une interprétation correcte. Disponibilité et fiabilité sont traduits quantitativement à partir de calculs sur corpus, donc de données réelles.

### 3.1.3. CRÉATION DES MATÉRIAUX, ÉTUDES LINGUISTIQUES, PROCÉDURES DE TALN

Si donc on admet que l'expert doit être le natif en acquisition d'une L2, alors se pose la question de trouver les données existantes, de créer celles indispensables, de mener une étude linguistique et d'implémenter les programmes de traitement automatique sur la langue cible. Le développement de l'expertise est, comme souvent en IA-ED, un processus long et coûteux. Les données sur la langue cible sont la plupart du temps inexistantes, dispersées ou manquent de précision. Quant aux programmes, soit ils n'existent pas, soit ils ne traitent qu'une partie des informations nécessaires (essentiellement syntaxiques), en utilisant des formalismes dont les implémentations ne sont pas portables, ou sont trop complexes pour être intégrées dans un environnement informatique d'apprentissage.

#### 3.1.3.1. ELEONORE

L'élaboration d'un modèle informatique reflétant les compétences du natif francophone sur les interrogatives (et plus précisément d'un natif métropolitain, car les Québécois n'utilisent pas tout à fait de la même façon les structures syntaxiques pour construire des interrogatives) passe naturellement par une étape de description linguistique articulant aspects syntaxiques et pragmatiques.

Description linguistique

Les travaux sur la syntaxe des interrogatives du français sont nombreux. Ils révèlent une grande disparité terminologique associée à des descriptions parfois contradictoires. La disparité des analyses s'explique par le fait que certains auteurs se sont focalisés sur certains types de constructions, perdant ainsi un point de vue global. Notre étude a consisté à fixer une terminologie propre, justifiée par une mise en correspondance des structures de surface des énoncés et du type de réponses associées à ces énoncés interrogatifs (Renié 1995). Cette correspondance nous a conduit, par exemple, à rejeter certaines terminologies comme question totale versus question partielle (ou question ouverte/fermée, question oui/non) , et à retenir au contraire celle plus consistante de question pleine versus question non-pleine.

Quatre types de constructions des interrogatives directes ont été décrits :

- l'interrogative qui reprend l'ordre des mots de la proposition déclarative, appelée CONSERV,
- l'interrogative incluant la particule est-ce-que et ses composés, appelée ESK,
- l'interrogative par inversion du clitique sujet, appelée INV-CLIT,
- et l'interrogative par inversion du sujet nominal, appelée INV-STYL.

L'exemple 4 illustre chaque type :

(EX 4)
Où se trouve le restaurant ? (inversion stylistique, INVSTYL)
Où le restaurant se trouve-t-il ? (inversion clitique, INVCLIT)
Le restaurant se trouve où ? (conservation de l'ordre de la déclarative, CONSERV)
Où est-ce que le restaurant se trouve ? (introduction de la particule est-ce-que, préservation partielle de l'ordre de la déclarative, ESK)

Ont été écartées lors de notre analyse les interrogatives indirectes et enchâssées, les formes elliptiques, les interrogatives semi-figées et celles comportant un verbe modal. Il ne s'agit évidemment pas de dissimuler la complexité de la langue. Mais puisque notre but est d'établir une correspondance entre formes de surface et usage, la prise en considération de ces derniers cas nous auraient amené à élargir beaucoup trop le champ de cette correspondance, et donc à traiter dans un même modèle des phénomènes linguistiques trop peu connexes.

Nous avons ensuite étudié les types sémantiques de questions associées à ces formes syntaxiques: questions assertives, sur le rôle, les circonstances et l'actance. Cela nous a permis de dresser un premier tableau de contraintes entre types sémantiques et syntaxiques, car l'orientation sémantique d'un énoncé interrogatif limite l'éventail des choix syntaxiques possibles. Mais ce premier niveau de contraintes laisse toujours ouvert un choix beaucoup trop grand de formes syntaxiques.

Afin d'essayer de lever cette indétermination, nous avons examiné les travaux des génératives chomskiens. Ces travaux offrent une description assez exhaustive, sur le plan syntaxique des interrogative directes, mais ils se focalisent uniquement sur les manières d'opérer, à partir d'une forme déclarative, les transformations nécessaires pour obtenir les différentes constructions syntaxiques. Leur souci majeur est de ne pas surgénérer, ni sous-générer, d'une part, et de réduire le nombre des transformations à appliquer, d'autre part. Leur modèle devient beaucoup plus flou lorsqu'ils tentent d'expliquer les facteurs qui peuvent rendre compte des transformations obligatoires ou non, fréquentes ou rares. Le modèle chomskien réduisant, par principe, le contexte linguistique ne permet pas d'extraire les facteurs déterminants.

Sur ce point l'approche psycholinguistique est beaucoup plus constructive. Par psycholinguistique nous entendons, non pas la branche chomskienne, mais celle que l'on appelle aussi "psychologie du langage" et qui se situe dans le champ d'investigation de la psychologie cognitive. Elle est représentée en France par des chercheurs comme Caron ou Gaonac'h. Pour eux, "le but de la psycholinguistique est de construire et valider un modèle du locuteur humain" (Caron, 1989 : p. 52). La pragmatique est alors amenée à prendre en compte le discours, et avec lui apparaît la notion de règles conversationnelles. Il existe selon Caron deux catégories de moyens pour renvoyer à l'acte d'énonciation : les moyens lexicaux et syntaxiques, et les marques linguistiques ; les moyens lexicaux et syntaxiques spécifiant la fonction du contenu propositionnel dans l'acte de communication particulier. Nous voilà bien au coeur du problème où doit être étudié le rapport entre la syntaxe et l'acte d'énonciation.

Cela dit, les interrogatives n'ont pas été étudiées par ce courant, à l'exception des formes semi-figées qui jouent un rôle particulier dans l'acquisition d'une langue.

Pour déterminer les paramètres pragmatiques, nous nous sommes inspirés des études linguistiques sur l'échange verbal. Cinq paramètres ont été retenus et leurs domaines de valeurs fixés : le canal de communication (oral/écrit), la force illocutoire (demande d'information, suggestion, demande d'action), la fonction dialogique de la question dans un échange conversationnel (énoncé d'ouverture, emphatique, enchaîné), le contexte social (travail, maison, restaurant) et la position sociale réciproque des interlocuteurs/correspondants (égale, supérieure, inférieure). Ces paramètres étant fixés, nous pouvions alors réinterpréter les résultats des travaux sur l'usage des interrogatives par les natifs des chercheurs français et anglo-saxons (Anscombre, Coste, Gadet, Coveney, etc.); Il s'agit bien de réinterprétation puisque chaque investigation ne mettait en jeu qu'un sous-ensemble des paramètres cités ci-dessus. Le produit de ce travail est un ensemble de tableaux de contraintes établissant une correspondance entre formes syntaxiques et contexte d'énonciation, indiquant, non pas des choix impératifs, mais préférentiels.

Arrivé à ce stade, il était possible de construire un modèle du locuteur natif.

Implémentation du modèle expert

Le modèle doit avoir, pour un contexte d'énonciation et un contenu sémantique fixés, la même probabilité qu'un locuteur natif de produire tel ou tel énoncé. Donc si, pour un ensemble de paramètres sémantiques et pragmatiques, il existe plusieurs interrogatives possibles de structures syntaxiques différentes, nous voulons pouvoir produire ces énoncés dans l'ordre des préférences d'un natif. Le module de génération intègre deux programmes principaux, une table de contraintes sémantico-pragmatiques et le générateur de langage naturel proprement dit.

L'implémentation du modèle est justifiée par plusieurs raisons. D'abord, par principe méthodologique (qui différencie les approches IA-ED de l'ALAO traditionnelle), il est nécessaire de valider la description opérée manuellement. L'informatique oblige à une description fine et cohérente et permet d'observer si le modèle se comporte selon les spécifications initiales. Ensuite, dans ELEONORE l'apprenant peut jouer avec les différents critères sémantico-pragmatiques et observer les productions opérées par le système. La combinatoire est tellement élevée qu'il aurait fallu, en l'absence de programme informatique, pourvoir des centaines d'énoncés pour couvrir toutes les situations. Enfin, sans générateur, tout diagnostic des productions de l'apprenant aurait été impossible. A ce propos, l'approche souvent adoptée par des concepteurs de systèmes d'ALAO consiste à implémenter un analyseur, plutôt qu'un générateur. Le système doit alors faire face à une situation inutilement complexe, puisqu'il lui incombe la tâche supplémentaire de deviner le contenu sémantique du message de l'utilisateur. Il est plus simple, et raisonnable pédagogiquement, de fixer de concert avec l'apprenant le contenu sémantique, à partir duquel peut être produit l'énoncé.

#### 3.1.3.2. ALEXIA

Explorer les associations entre items lexicaux nécessite de fixer un champ sémantique. De façon à ne pas disperser nos recherches, compte tenu des objectifs visés dans CAMILLE, nous avons choisi le champ "travail, emploi et chômage". Puisque l'expertise doit se référer à la façon dont la langue est réellement utilisée, j'ai décidé de construire un corpus

électronique de textes de français courant, à partir duquel ont ensuite été menées des études linguistiques, qui serviront à la création de la base de données lexicales active.

Corpus

Nous avons commencé à construire un corpus électronique. Il compte aujourd'hui 250 000 mots, du français courant. Pour ce faire, nous avons sélectionné des articles de journaux, revues, magazines de toutes natures : presse économique ou sociale, quotidiens "populaires" ou non, périodiques édités par les associations de chômeurs, etc.. Ce corpus, composé de textes écrit, contient néanmoins des interviews.

Les raisons de la constitution de ce corpus sont multiples, il doit : comme nous l'avons dit servir à construire un dictionnaire des termes du domaine ; fournir des exemples d'utilisation de ces termes en contexte, en associant item lexical et phrases dans lesquelles il apparaît ; être une base de textes pour la constitution d'activités lexicales d'aide à l'acquisition lexicale en L2 pour les environnements informatiques ALEXIA et CAMILLE, puisque les champs sémantiques sont les mêmes, et pour une expérimentation autour du dictionnaire DICOLOGIQUE, expérimentation dont les résultats intéressent directement le projet ALEXIA (cf. section 3.2).

Extraction des mots et expressions du domaine

Nous avons ensuite élaboré une procédure d'extraction automatique des items lexicaux (mots simples, collocations et expressions terminologiques) caractéristiques du domaine (Rabefitia, 1994), en ayant recours à des méthodes statistiques et probabilistes. Pour extraire les mots de base nous avons calculé l'ordre lexicométrique en relevant la liste de chaque forme du corpus associée à sa fréquence, par ordre décroissant. Les premières formes de la liste correspondaient, bien entendu, aux mots grammaticaux, mais on y trouvait également des mots lexicaux, dont l'utilisation exceptionnellement fréquente était conditionnée par la thématique du corpus (pour les substantifs "travail", "embauche", "emploi", etc.). Ces mots lexicaux, associés à quelques verbes supports ("faire", "mettre", etc.) souvent employés dans les expressions, ont constitué les mots de base. La complétion de cette liste s'est faite manuellement.

Cette liste des mots de base a servi de référence pour l'extraction des expressions, dont un des constituants devaient appartenir à cette liste. Cette façon de procéder élimine certaines expressions, comme les expressions à caractère métaphorique (puisqu'il se peut qu'aucun des mots d'une telle expression n'appartienne au champ lexical étudié), mais présente l'avantage de réduire le bruit (les structures ne caractérisant pas le domaine, mais la langue française en général). Ces cooccurrences ont été construites à partir des lemmes et non des formes fléchies, de façon à observer les variations syntaxiques, le degré de figement, de ces expressions . Pour une forme pôle donnée, nous avons décidé de sélectionner un ensemble de formes qui ont tendance à se trouver souvent dans un voisinage à l'intérieur duquel on considère que deux formes sont cooccurrentes. Nous avons retenu pour limite de voisinage la phrase. Le voisinage a été évalué en calculant la distance entre les deux mots de la cooccurrence au sein de la même phrase.

A partir de cette liste de cooccurrences et de l'ordre lexicométrique du corpus, nous avons pu appliquer un modèle probabiliste. Ce modèle s'inspire de celui de Lafon (1984), qui vise à rechercher les associations privilégiées dans l'ensemble de toutes les associations. Contrairement à Lafon, qui travaille sur toutes les formes fléchies, nous nous sommes centrés

sur les lemmes des mots de base et avons éliminé les mots grammaticaux présents dans les associations. Ayant obtenu les probabilités pour chaque collocation et fixé des seuils d'acceptabilité, nous sommes intervenus une seconde fois manuellement pour éliminer (10 à 15% des collocations) les collocations qui ne présentent pas d'intérêt du point de vue de leur utilisation dans le champ lexical de la recherche d'emploi. Ces deux interventions manuelles permettent de réduire la récolte trop large résultant d'approches purement statistiques et probabilistes (cf. Jacquemin, 1994). Ensuite chaque expression retenue a été indexée avec les phrases dans lesquelles elles apparaissaient, de façon à en étudier les variations syntaxiques et à servir d'illustration pour les apprenants.

Études linguistiques

La procédure d'extraction automatique a fourni une première liste que nous avons croisée avec celle obtenue en explorant les lexèmes reliés aux thèmes du travail et du chômage. L'exploration des différentes relations sémantiques lexicales a permis de sélectionner d'autres mots, collocations et ESF. 150 unités lexicales ont été finalement retenues et décrites, une à une, à partir d'un patron très proche de celui utilisé dans la théorie sens-texte de Mel'cuk (1992). Le dictionnaire explicatif et combinatoire offre sans doute la description (partielle) la plus fine du français, description en partie opératoire et orientée vers la génération et la paraphrase.

BD lexicale, réseaux et TAG

Nous abordons à l'heure actuelle la phase d'organisation de la base de données lexicales, base de données dans laquelle seront décrites les informations relevées lors des études linguistiques mentionnées précédemment. Une entrée lexicale est un objet complexe rassemblant : des informations de base ( définitions des sens, exemples d'utilisation en contexte, registres de langue) ; des informations permettant de construire automatiquement des réseaux d'items lexicaux reliés soit par des liens sémantiques, soit par leur appartenance à des expressions ou ayant des relations de synonymie avec ces expressions ; des informations morpho-syntaxiques. Une partie des liens sémantiques correspondent à des fonctions lexicales du dictionnaire explicatif et combinatoire. Il sera nécessaire ensuite d'écrire des algorithmes de construction automatique de réseaux. Les informations syntaxiques pour une entrée lexicale correspondent à la structure syntaxique de cette entrée, définie sous forme d'un arbre au sens des grammaires d'Arbre Adjoints (TAG) (Abeillé 1991), et la définition des transformations morpho-syntaxiques que peut subir cette entrée lexicale. En ce qui concerne les ESF, il nous paraît utile d'indiquer à l'apprenant et de pouvoir calculer automatiquement, pour une aide à la production, les transformations telles que la passivation, la pronominalisation, l'introduction de modificateurs, les variations lexicales, le passage de la forme déclarative aux formes interrogatives. Les informations correspondant à ces transformations ont été décrites lors des études linguistiques. Le formalisme TAG, couplé à l'analyseur TAG, permettent d'opérer automatiquement ces transformations. De plus, nos partenaires du LIPN de l'université Paris 13 développent un nouvel algorithme pour l'analyseur permettant de corriger les productions de l'apprenant, en lui proposant des paraphrases de son énoncé (Issac, 1994). Cette intégration d'informations lexicales de différents niveaux n'est pas sans rappeler l'approche "oecuménique" de Sérasset (1994). Dans son cas, il s'agit de définir un langage de description des structures linguistiques de façon à proposer des solutions génériques pour la constitution de bases lexicales multilingues. Dans

notre cas, il s'agit de constituer une base lexicale monolingue dont les fonctionnalités correspondent à l'aide à l'acquisition lexicale en L2.

Les obstacles que nous avons à surmonter sont, bien sûr, la détermination du nombre minimum de liens de nature syntaxique ou sémantique et la contrainte de "reproductibilité". En effet, au contraire des travaux courants en lexicographie, portant sur la construction automatique de dictionnaires, nous nous limitons à un seul champ lexical. Nous pouvons donc travailler en profondeur et finement sur la langue, avec la contrainte que pratiquement toutes les informations associées à une entrée lexicale soient opérationnels, c'est-à-dire puissent servir d'arguments à des opérations qui permettent, elles, d'accéder à d'autres items lexicaux. De plus, une partie des compositions d'opérations doit correspondre à des opérations motivées sur le plan cognitif (par rapport à des aides en compréhension, production, ou apprentissage). La contrainte de reproductibilité, correspond au fait que cette démarche perdrait sans doute beaucoup de son intérêt s'il elle n'était pas reproductible à d'autres champs lexicaux.

#### 3.1.3.3. "Travailler en France"

Hutchinson et Waters (1987: p.106) ont dit que la création de matériaux était, en pratique, l'un des traits caractéristiques du développement d'une formation de langue à objectifs spécifiques. Cela vaut pour des formations conçues traditionnellement ou autour d'un environnement informatique. Si pour les environnements orientés LOS le slogan "expert = natif" tient toujours, ce n'est pas seulement le savoir linguistique du natif qui nous intéresse, c'est aussi son savoir (et savoir-faire) par rapport à l'objectif spécifique visé. Rappelons que la résolution de la tâche (rechercher un emploi, savoir négocier une vente, qui est le sujet du deuxième module, dont le développement sera terminé fin 1994) définit le contenu de la formation et que le contenu linguistique n'est introduit que dans un deuxième temps. Il est donc nécessaire de trouver et intégrer cette expertise du natif. C'est une différence importante avec les projets habituels en ALAO, même ceux intégrant l'approche IA-ED, où le contexte des échanges langagiers (lorsqu'il existe) entre l'apprenant et le système est celui limité à quelques aspects de la vie quotidienne.

L'autre différence importante, concerne l'éventail des contenus linguistiques. Si dans tous les projets dont nous parlons dans ce document, la pierre angulaire est bien l'approche communicative, dans les environnements LOS toutes les compétences qui la composent doivent être intégrées ; particulièrement les aspects socioculturels ne sauraient être omis. D'autre part, comme la vidéo fait partie des ressources technologiques offertes par les environnements multimédia, il devient même possible d'aborder certains aspects de la communication non verbale. En effet, il est bien connu (Cosnier & Brossard, 1984) que certains gestes conventionnels sont étroitement liés au discours oral, soit pour répéter l'information communiquée verbalement, soit pour permettre d'interpréter un énoncé qui, autrement, serait ambigu, soit pour se substituer complètement à un énoncé. Cette liaison avec la communication verbale explique que ces gestes et comportement soient souvent enseignés à des apprenants qui doivent entretenir des relations régulières avec des personnes

de la communauté linguistique cible. L'ALAO ne va donc pouvoir l'ignorer[9]. Dans "Travailler en France" nous avons commencé "timidement" par l'introduction, dans le module 2 sur l'acte de vente, d'activités d'éveil (awareness) à l'importance de la gestuelle qui peut se substituer à certains messages verbaux idiomatiques.

Bien sûr, l'éventail de l'expertise à prendre en compte dans "Travailler en France", a concentré le travail sur la création de matériaux authentiques et de ressources associées, et non sur les descriptions linguistiques ou les procédures de TALN, au contraire des autres projets ALEXIA et ELEONORE. Il y a complémentarité et non opposition. Car, insistons bien sur le fait que la qualité et la richesse de l'Input sont déterminantes dans l'aide à l'acquisition d'une langue, comme cela a été prouvé expérimentalement ; même si nous ne sommes pas en mesure d'analyser, de décrire, ou de modéliser aussi finement que nous le désirions les contenus discursifs ; si nous ne savons pas encore définir avec précision et certitude les conditions garantissant la transformation de l'Input en Intake (appropriation).

Je rappelle maintenant quelques points caractéristiques de notre travail concernant la création de documents authentiques et des ressources, points abordés aux sections 4.1, 5 et 6.3 de la publication 7).

Les matériaux authentiques

Les raisons généralement avancés pour défendre la nécessité de matériaux authentiques sont de trois sortes :

- la motivation de l'apprenant. On connaît l'importance de ce facteur dans l'apprentissage des langues. Or l'apprentissage de façon partiellement incidente de la langue (cf. les principes de la LOS en section 3.1) ne peut avoir lieu que si matériaux et contexte revêtent un minimum de réalité.
- le développement de l'apprentissage autonome. Les stratégies déployées par l'apprenant pour traiter des documents authentiques peuvent être réinvesties dans la vie professionnelle.
- la qualité de l'Input.

Mais le développement de matériaux authentiques pose plusieurs problèmes: quel type d'authenticité recherché ? comment produire et rassembler ces matériaux ? est-il nécessaire d'avoir des compétences particulières dans le domaine d'expertise ?

Dans "A la Recherche d'un Emploi" nous utilisons deux types de matériaux authentiques. Le premier type, qui correspond à une part minoritaire de nos données linguistiques, est constitué de documents textuels véridiques, collectés, puis disséminés dans chaque unité du logiciel et dans les ressources, pour offrir des activités complémentaires, fournir des exemples alternatifs dans les exercices et apporter de l'information complémentaire d'ordre socioculturel. Le second type de matériaux a été créé par nous-mêmes. Il compose la totalité des documents vidéo et audio et partiellement des documents textuels. C'est sur ces matériaux qu'est véritablement construit le logiciel. Avant de créer ces données textuelles, audio et vidéo, il a fallu rassembler des informations précises sur des profils de chercheurs

[9] Même si une résistance de certains linguistes est à craindre. Il est d'ailleurs assez surprenant de constater que dans le Dictionnaire Encyclopédique de Pragmatique (Moeschler & Reboul, 1994), par ailleurs excellent, ces aspects non verbaux de la communication soient complètement ignorés. Les auteurs sont très soucieux de définir la nature des relations entre pragmatique et linguistique, mais excluent définitivement, brièvement et sans explication, toute relation entre pragmatique et sémiotique (p. 503) !

d'emploi, étudier la littérature destinée aux chercheurs d'emploi et celle destinée aux recruteurs, etc.. Pour le module sur l'acte de vente, nous avons rencontré et filmé des professionnels, assisté à des simulations de vente, etc. Une fois les données collectées, il faut les analyser (gestuelle comprise), faire ressortir les critères déterminants qui seront l'objet de calculs dans les simulations, puis écrire les scénarios, les faire jouer par des acteurs. Mais garantir la véracité de l'Input ne suffit pas, comme nous l'avons dit. De nombreuses études en acquisition ont montré l'importance de la simplification et de la redondance de la communication entre parents et enfant ("mother talk") et entre natifs et apprenants ("foreigner talk"). Dans un environnement d'ALAO, la simplification ne consiste pas à réduire la richesse des dialogues, mais à décomposer par étapes successives l'imbrication des données vidéo, audio et textuelles. La redondance est assurée par la possibilité qu'a l'apprenant de rejouer tout ou partie des séquences qui lui sont proposées et à lui laisser le temps de réfléchir pour apprécier les conséquences de ces choix, ou pour s'enregistrer.

Les ressources

Les ressources sont un ensemble de livres hypertextuels accessibles à tout instant de la session de travail soit globalement (accès par "la table des matières"), soit ponctuellement, lorsque des liens ont été définis entre certaines de leurs informations et des activités. Il existe quatre livres différents, couvrant le spectre des ressources linguistiques et extra-linguistiques nécessaires dans une approche communicative : culture, actes de langages, grammaire et lexique. Le lexique est un peu particulier puisqu'il regroupe un dictionnaire spécifique et un ensemble de réseaux lexicaux. Ces réseaux ont été construits à partir de schémas simplifiés de nos travaux dans ALEXIA. Dans l'article (publication 7), j'explique pourquoi ces ressources ne sont pas accessoires, mais indispensables. Outre l'aide en ligne qu'elles fournissent, elles peuvent faire l'objet d'un travail spécifique. Ainsi nous comptons utiliser les ressources culturelles, comme un véritable "dossier de civilisation", qui fait habituellement l'objet d'un travail particulier en classe. Comme l'apprenant doit travailler de façon autonome, il est nécessaire de lui fournir ces informations (accompagnées d'activités supplémentaires) en ligne. Et, pour l'instant, l'accès à des ressources électroniques extérieures à CAMILLE est impossible pratiquement et se révélerait inapproprié.

## 3.2. Modélisation de l'apprenant, diagnostic

Où l'on illustre les principes :

4. Le modèle de l'apprenant en ALAO doit être en partie construit en intégrant des styles différents d'apprentissage.
7. L'ALAO doit fournir un système de correction d'erreurs utile et pertinent en concordance avec les besoins changeants de l'apprenant.

Étudier la façon dont un groupe de sujets développent leur interlangue sur un thème déterminé, étudier les difficultés qu'ils rencontrent, et les stratégies qu'ils déploient est la première étape à franchir pour tenter d'individualiser l'apprentissage. A partir de ce recueil d'expertises s'élaborent les stratégies pédagogiques et le modèle de l'apprenant. Self (1988) a identifié 6 types de fonctions du modèle de l'apprenant, en examinant les principaux ITS de l'époque :

- corrective : pour aider à éradiquer les connaissances erronées.
- élaborative : pour aider à étendre ses connaissances.
- stratégique : pour guider le système dans ses choix de stratégies pédagogiques.

- diagnostique : pour aider à lever des indéterminations sur ce que l'apprenant maîtrise réellement.
- prédictive : pour aider le système à prédire les réactions de l'apprenant aux interventions pédagogiques.
- évaluative : pour aider à juger les progrès accomplis par l'apprenant ou bien le système.

Dans ELEONORE, que nous examinons en premier, le recueil d'expertises a été constitué à partir des résultats des investigations menées en linguistique appliquée sur l'acquisition des interrogatives en L1 et L2. Mais comme ces travaux ne couvraient que les aspects syntaxiques, et pas les aspects pragmatiques, nous avons dû les compléter par une expérimentation. La fonction corrective du modèle de l'apprenant est gérée par le module de diagnostic d'erreurs qui fait appel à un catalogue de règles déviantes et d'autres procédures permettant de discriminer les problèmes grammaticaux des problèmes pragmatiques. La fonction élaborative est assurée par le module de remédiation qui, prenant la relève du module de diagnostic d'erreurs, engage un dialogue collaboratif pour amener l'apprenant à réviser ses points de vue. La fonction diagnostic est discutée (mais non encore implémentée). Elle devrait identifier si une erreur est l'expression d'une règle de l'interlangue de l'apprenant, ou seulement une erreur de performance. Les autres fonctions sont discutées dans les sections 3.3 et 3.4.

ALEXIA n'a pas encore atteint un stade de développement suffisamment avancé pour que les questions précédentes soient évoquées. Mais, au préalable, il nous faut déterminer les stratégies cognitives d'accès lexical déployées par les apprenants disposant d'un dictionnaire électronique, structuré en réseau. J'explicite les grandes lignes de l'expérimentation que nous allons bientôt réaliser et dont les résultats nous serviront à fixer précisément les modes d'aide à la compréhension et à la production dans ALEXIA. J'explique ensuite l'intérêt de la constitution automatique d'un dictionnaire personnalisé de l'apprenant, qui remplira une partie des fonctions du modèle de l'apprenant.

Dans CAMILLE, il n'est pas question de modéliser l'apprenant. Mais je détaille le module de suivi de l'apprenant qui a été implémenté, et explique pourquoi il ne peut remplir la fonction d'évaluation des connaissances de l'apprenant. Il est cependant indispensable dans un contexte d'auto-formation, indispensable pour le formateur, en cas de suivi à distance, et pour l'utilisateur afin de l'aider à faire le point, se repérer et soutenir sa motivation. Je discute enfin la possibilité d'offrir à l'apprenant de s'évaluer à l'aide de documents accompagnant le logiciel.

### 3.2.1. ELEONORE

Suivant une présentation similaire à celle adoptée pour la modélisation du natif, nous examinons les travaux sur l'acquisition des interrogatives afin d'en dégager les grands traits d'un modèle d'apprenant. Nous verrons qu'il a été nécessaire de monter des expériences avec des apprenants afin de trouver des réponses à des questions laissées de côté par la linguistique appliquée (Renié 1995). Puis nous discutons de la façon dont nous avons conçu le diagnostic des connaissances de l'apprenant dans ELEONORE (publication 1). C'est une démarche assez similaire qu'a adoptée Manning (1995) dans sa thèse sur l'acquisition du genre en français L2. A la différence près que, les résultats dans ce domaine étant plus homogènes, elle a pu s'appuyer plus précisément sur des modèles existants. Autre différence notable, elle

a utilisé son environnement d'apprentissage ITSICALL de façon essentiellement expérimentale. ITSICALL permet de pratiquer certaines règles de grammaire contenant des contraintes linguistiques sur le genre suivant des modes exploratoire, implicite ou explicite. L'environnement lui a servi à collecter automatiquement des données sur les apprenants, qu'elle a ensuite interprétées manuellement de façon à tirer des conclusions sur les mérites respectifs de chacun des modes d'apprentissage. Elle n'a donc pas cherché à implémenter un modèle d'apprenant, ni à fortiori à mettre en place un diagnostic automatique.

**Acquisition des interrogatives**

Les chercheurs en linguistique appliquée ALS concilient deux approches :

- une approche dans laquelle l'apprenant est pris comme locuteur de la langue seconde,
- une approche dans laquelle l'apprenant est pris comme sujet participant à un apprentissage.

L'hypothèse sous-jacente est que les stratégies d'acquisition et de production mises en oeuvre par l'apprenant l'amèneront à produire telle ou telle forme, mais que la situation d'apprentissage/acquisition, qui constitue son environnement linguistique peut influencer également son rapport à la langue (la compréhension qu'il en a, la production qui en résulte). Ainsi, sur le domaine de l'acquisition des interrogatives en L2, on trouve un éventail d'études portant sur des production d'enfants ou d'adultes, à partir desquelles les chercheurs ont tenté :

- d'identifier l'ensemble des stratégies d'acquisition, de production et de communication mises en oeuvre pour produire un énoncé interrogatif, stratégies caractérisant le système linguistique qu'est l'interlangue de l'apprenant,
- d'inférer des observations générales sur l'acquisition d'une L2 (identité L1/L2, existence d'un mode d'acquisition naturel, etc.).

Les stratégies ne peuvent être identifiées qu'en explorant le contexte d'acquisition/apprentissage. C'est pourquoi l'interaction en classe est un sujet important de recherche. L'étude de la langue des enseignants (teacher talk) et celle des natifs s'adressant à des non-natifs (foreigner talk) dans des classes d'anglais langue seconde a montré une différence importante entre ces langues et l'anglais tel qu'il est parlé par les natifs hors cadre institutionnel, différences concernant aussi bien les formes d'interrogatives, que les types d'interrogation. L'Input en situation institutionnelle est donc particulier.

A partir de nombreux travaux comparant, en anglais, l'acquisition des interrogatives en L1 et en L2, il est possible d'identifier (sur le plan syntaxique) certains stades, qui seraient de plus en partie identiques en L1 et L2. C'est-à-dire qu'un enfant anglophone qui acquiert les structures interrogatives de sa langue passera par les mêmes stades qu'un adulte non anglophone. Il s'agit cependant là d'un schéma approximatif, car des différences concernant des constructions intermédiaires ont été notées. D'autre part, il ne faut pas négliger les variations individuelles. Chaque individu a, en effet, ses préférences et un rapport particulier à la langue qui se traduisent par des écarts avec la "norme" d'acquisition. Cette norme d'acquisition correspond donc à un stéréotype d'apprentissage d'un individu abstrait.

Mais les investigations évoquées précédemment se rapportaient à l'anglais. Aucune recherche comparable n'ayant eu lieu sur le français, nous avons procédé par transposition pour construire, à partir de résultats fragmentaires un stéréotype d'apprentissage en français. Les étapes d'acquisition dégagées ont ensuite servi à pondérer chaque forme d'interrogative comme étant plus ou moins facile à acquérir.

Nous avons ensuite travaillé sur le diagnostic grammatical des productions des apprenants. Rappelons que les stratégies utilisées par les apprenants (surgénéralisation, transfert, fossilisation etc.) sont la cause d'erreurs de production. Une analyse d'erreurs (procédure bien connue en linguistique appliquée) permet d'établir une correspondance entre certaines formes déviantes produites par les apprenants et l'état de leurs connaissances. Nous avons donc monté une expérience avec des apprenants, puis formalisé notre diagnostic de façon à obtenir un répertoire de *règles grammaticales déviantes* (bug catalogue).

Toute cette discussion sur l'acquisition ne concernait que les aspects grammaticaux. Qu'en est-il de l'acquisition de la pragmatique en L2 ? En particulier :

- les aspects pragmatiques et les aspects linguistiques (morpho-syntaxiques et lexicaux) s'acquièrent-ils simultanément ?
- les connaissances pragmatiques déjà acquises en L1 interviennent-elles dans l'acquisition de connaissances pragmatiques en L2 ?

Il existe très peu de travaux sur l'acquisition de la pragmatique en L1 et pratiquement aucun en L2. En nous appuyant sur les études de Bates (1976) et de Coste (1970), il nous a été possible d'élaborer un certain nombre d'hypothèses sur la non similitude d'acquisition de la pragmatique en L1 et L2. Hypothèses que nous avons vérifiées dans le deuxième volet de l'expérimentation mentionnée précédemment (expérimentation longitudinale avec des apprenants FLE, résidant en France depuis 3 à 6 mois). Ce volet a permis de confirmer les choix pédagogiques fixés dans ELEONORE, à savoir que : les apprenants L2 acquièrent en priorité une compétence grammaticale, plutôt que pragmatique (donc il y a place pour un travail spécifique sur ce point avec des apprenants de niveau intermédiaire à avancé) ; l'apprenant acquiert des connaissances pragmatiques et un jugement de grammaticalité plus par l'interaction avec un natif qu'en contexte institutionnel (cf. le rôle du compagnon et de l'expert natif dans notre système).

**Modélisation de l'apprenant**

Comme nous l'avons dit ELEONORE s'adresse à des apprenants de niveau intermédiaire à avancé, donc ayant déjà une connaissance (complète ou partielle) des structures syntaxiques des interrogatives. Ce qui nous intéresse n'est pas d'établir le diagnostic d'une compétence abstraite (contrairement à de nombreux systèmes IA-ED) mais d'estimer la capacité d'un apprenant à mobiliser ses compétences dans des tâches de difficultés cognitives variables. Or l'on sait qu'une règle peut faire partie de la connaissance passive d'un apprenant sans que pour autant il soit capable de l'appliquer dans des situations concrètes, particulièrement lorsque la charge cognitive requise pour accomplir des tâches spécifiques s'accroît. Notre idée a donc été de concevoir plusieurs types d'activités, de façon à pouvoir contrôler les niveaux de difficultés.

Dans la publication 1 (section 3.2), nous montrons pourquoi des modèles d'apprenant, utilisés classiquement en IA-ED (Wenger, 1987), conçus sur le principes du recouvrement entre les connaissances de l'expert et celles de l'apprenant sont inapplicables dans notre cas. Nous discutons ensuite de la technique faisant appel à un catalogue de règles déviantes. Ce catalogue, nous l'avons construit à partir de l'expérimentation évoquée dans la partie précédente. Il peut être utilisé pour diagnostiquer une partie de l'interlangue de l'apprenant constituée de connaissances déviantes. Une critique souvent adressée à ce type de modélisation (Ohlsson, 1992) est que, dans les domaines procéduraux où elle a été appliquée

(soustraction, calcul algébrique, etc.), les catalogues construits pour une population d'apprenants ne pouvaient être transposés à d'autres populations. Mais je pense qu'en ce qui concerne les langues, les procédures d'analyse d'erreurs communément utilisées en ALS ont prouvé l'extensibilité des résultats obtenus. En fait le problème est surtout de savoir ce qui a été la cause d'une erreur : si c'est une stratégie d'apprentissage alors la règle déviante correspondante peut être considérée comme faisant partie de l'interlangue de l'apprenant ; si c'est une stratégie de communication, il n'en est pas de même. Deux types d'informations peuvent nous permettre de trancher : le niveau de difficultés de la tâche où l'erreur est apparue (comme nous l'avons dit) et une utilisation consistante de cette règle.

Examinons maintenant la procédure de diagnostic (publication 1, section 3.3). Dans ELEONORE c'est un module particulier qui en assure la gestion, le compagnon natif. C'est également le compagnon qui possède l'expertise sur le français évoquée en section 3.1 et qui peut se comporter comme un collaborateur selon les besoins.

Dans plusieurs activités, deux types d'erreurs peuvent survenir simultanément : une erreur de type pragmatique (mauvaise sélection du type de l'interrogative en fonction du contexte) ou erreur de type grammatical. De façon à fournir rapidement un diagnostic, et à le faire dans presque tous les cas, le programme a été conçu sous forme d'un tableau noir simple, où différents sous-modules gèrent des diagnostic partiels à base de repérage d'indice, de relaxation de contraintes et d'appel du catalogue de règles déviantes.

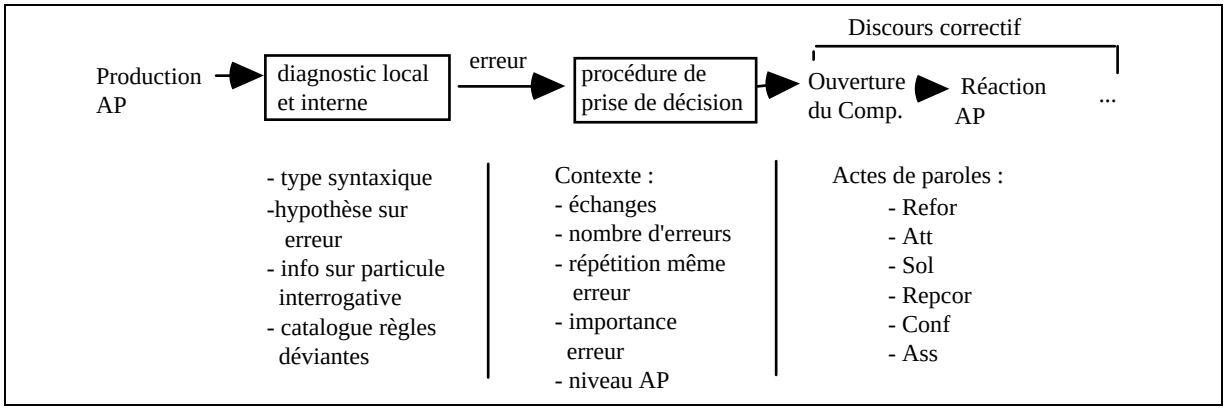



**Figure 2 :** Procédure de diagnostic et remédiation. C est le compagnon et L l'apprenant..

Il est important de remarquer que ce diagnostic n'est qu'une *procédure interne et locale* : interne, parce qu'à ce stade le compagnon garde pour lui le résultat du diagnostic ; locale car cette analyse ne s'est faite qu'en examinant la production de l'apprenant (et donc pas le modèle de l'apprenant, ni l'historique des interactions). Beaucoup de systèmes IA-ED s'arrêtent à ce stade et indiquent l'erreur à l'apprenant. En fait, en ALS le diagnostic interne n'est que la première étape d'une série de trois, dont les deux autres, *procédure de prise décision* et *discours correctif* , sont toutes aussi importantes. En effet, pédagogiquement parlant, l'objectif ultime n'est pas l'affectation d'une règle déviante à l'interlangue de l'apprenant. Il s'agit bien plutôt de lui permettre de modifier son interlangue progressivement, sans le décourager, en lui fournissant l'appui nécessaire pour l'amener à réviser ses points de vue. C'est pourquoi, dans ELEONORE, le compagnon ouvre un dialogue et initialise le premier tour de parole en sélectionnant parmi un ensemble d'actes de paroles la réponse la plus appropriée à partir du contexte global d'occurrence de l'erreur. Nous reviendrons dans la section 3.3 sur le type de dialogue qu'est capable de gérer le compagnon et qui lui permet

d'adopter une attitude collaborative. La figure 2 représente les trois étapes du processus de remédiation.

### 3.2.2. ALEXIA

Dans ALEXIA, avant de songer à modéliser l'apprenant, il nous faut savoir comment il peut utiliser un dictionnaire organisé sous forme de réseau lexical. C'est l'objectif d'une expérience prévue prochainement. Ensuite je parlerai de la nécessité de la construction d'un dictionnaire personnalisé. Comme cette deuxième partie n'est encore qu'en l'état de spécifications, des considérations pédagogiques seront introduites dans la discussion et ne seront plus reprises dans la section suivante 3.3.

**Stratégies cognitives d'accès lexical**

Compte tenu des objectifs de l'environnement ALEXIA, nous avons besoin d'évaluer:

- les stratégies adoptées par les apprenants dans des tâches de compréhension d'un texte, celles adoptées dans des tâches de production,
- les moyens que les apprenants utilisent pour construire leur lexique personnel,
- dans quelle mesure les différents types d'accès lexical et les ressources lexicales disponibles aident à la compréhension, à la production et à la rétention de vocabulaire.

Les travaux en psycholinguistique ou en linguistique appliquée n'apportent pas de réponses à ces questions pour les raisons suivantes.

Les expérimentations sur le lexique menées en psycholinguistique (cf. par exemple (Lecoq & Segui, 1989 ; Taylor, 1990) consistent généralement à inférer la structure du lexique mental à partir de tâches de reconnaissance, conduites en temps très limité, de mots ou segments de phrases dans des contextes très restreints. Les réponses des sujets sont limitées à des sélections de choix prédéfinis. Même si nous prenons en considération les résultats de ces expériences, la méthodologie expérimentale utilisée ne permet pratiquement jamais de mesurer les performances des sujets en situation de production lexicale. Le contexte trop restreint ignore la combinatoire linguistique rencontrée dans les textes et oblitère donc les choix de hauts niveaux que doivent accomplir ces sujets. La vitesse de réponse des sujets aux stimuli ne donne aucune information sur les stratégies de choix mises en oeuvre sur des durées plus longues, c'est à dire lorsque les sujets ont à intégrer un ensemble d'informations provenant du contexte textuel et à opérer une sélection sur ces informations. Enfin, l'essentiel des travaux en psycholinguistique se concentre sur la construction de modèles statiques du lexique mental et ignore donc la dynamique de l'acquisition lexicale.

Quant aux expériences, en nombre limité, menées par les chercheurs de linguistique appliquée s'intéressant au lexique, elles ont pour support unique les dictionnaires papiers (Bogaards, 1991 ; Laufer, 1993) et s'intéressent à des tâches de traduction. Les résultats de ces expériences sont difficilement transposables puisque les tâches données aux sujets ne sont pas celles généralement rencontrées en situation de communication en langue seconde. De plus, ces expériences ignorent les environnements informatiques, donc toutes les possibilités et limitations spécifiques de ces supports, ainsi que toutes les stratégies spécifiques induites par ces environnements sur les apprenants.

Nous préparons donc une expérimentation qui constituerait une évaluation a priori des hypothèses sous-jacentes dans ALEXIA et pourrait ainsi nous amener à réorienter le

développement de notre environnement informatique si nécessaire. Souvent, pour les environnements informatiques, ce type d'évaluation est réalisé a posteriori (après le développement du logiciel). Il sert alors généralement plus à justifier le travail réalisé qu'à le modifier. L'apport de l'expérimentation pour ALEXIA serait d'autant plus significatif qu'elle utiliserait le corpus de textes établi.

L'environnement ALEXIA étant basé sur un réseau lexical, il est nécessaire de pouvoir mener une expérimentation sur ce type de support. Il existe déjà deux dictionnaires sur support informatique qui offrent une recherche lexicale fondée sur les liens sémantiques entre items lexicaux: l'un en anglais, WORDNET, et l'autre en français, DICOLOGIQUE (Dutoit, 1992). A notre connaissance il n'y a pas eu d'expériences, répondant à nos objectifs, menées sur ces deux environnements. Notre système est orienté vers l'apprentissage lexical du français, justifiant le choix de DICOLOGIQUE comme cadre d'expérimentation. Mener une expérimentation nous permettra d'observer comment les sujets exploitent ce type de réseau en situation de résolutions de tâches et de répondre aux question suivantes:

- Comment les apprenants procèdent-ils pour accéder à un item lexical? (Accès):
- Quelles stratégies emploient-ils pour comprendre le sens d'un mot nouveau ? (Compréhension)
- Quelles sont les stratégies employées lors de l'utilisation d'un nouveau mot ? (Production)
- Quelles sont les méthodes employées par les apprenants pour retenir le vocabulaire, et est-ce que l'utilisation d'un environnement informatique basé sur un réseau lexical favorise cette rétention? (Apprentissage)
- Un réseau lexical permet-il une meilleure approche des tâches de production et de compréhension a effectuer, quels sont les avantages et les désavantages de la représentation adoptée: est-elle mieux adaptée aux stratégies des apprenants et correspond-elle mieux à leurs représentations mentales ? (Apports d'informations pour la construction de notre réseau lexical).
- Quelles sont les informations les plus demandées dans une recherche lexicale (synonymies, antonymie, structures syntaxiques, exemples, définitions...) ?

L'expérimentation se déroulera, début 1995, auprès d'un public d'apprenants de français langue étrangère de niveaux différents, en formation à l'institut du CAVILAM (cf. expérimentation sur CAMILLE en 3.4). Les sujets seront donc en phase d'acquisition lexicale soutenue, et auront sur le plan des champs sémantiques les mêmes connaissances qu'un sujet natif du français. Il s'agira de constituer deux groupes au sein desquels seront formés des binômes et des monômes : un groupe travaillera avec un dictionnaire sur support informatique, l'autre avec un dictionnaire sur support papier. Le fait de prendre des sujets de niveaux différents nous permettra de comparer leurs stratégies utilisées, de façon à pouvoir ensuite, dans ALEXIA, introduire un niveau réflexif permettant de dialoguer avec chaque apprenant sur sa façon de travailler.

**Lexique personnel et activités lexicales**

Le premier principe pédagogique défendu par Goodfellow (1994) est l'assistance à la construction d'un dictionnaire propre à l'apprenant. Trois critères, interdépendants, justifient l'appartenance d'un item lexical à ce dictionnaire personnalisé : une expression pourra avoir été abordée en compréhension, en production, ou introduite dans des regroupements

d'expressions opérés directement par l'apprenant. Tout comme la présentation d'informations lexicales sous forme de réseaux favorise l'acquisition, Goodfellow (1993) a montré qu'un environnement informatique devait permettre à l'utilisateur d'effectuer ses regroupements personnels. La question reste ouverte de savoir si les types de liens entre les items lexicaux doivent être laissés au libre arbitre de l'utilisateur ou, au contraire, être choisis parmi ceux effectivement utilisés par la base lexicale. Le statut d'un item lexical du dictionnaire peut donc être différent suivant le critère qui a présidé à son mode d'introduction, et suivant qu'il a, ou non, été validé par des activités lexicales appropriées.

MacWhinney (1992) suggère, à partir de ces travaux sur le Modèle de Compétition, quelques principes pédagogiques dépendant des stades d'acquisition lexicale d'un apprenant (nous omettons le stade de restructuration phonologique et de transfert initial):

- l'apprentissage par coeur d'items lexicaux joue un rôle important au début, mais devra ensuite s'effacer.
- lors de l'acquisition d'un nouvel item lexical, il est nécessaire d'expliciter sa structure syntaxique et casuelle et de le relier explicitement à d'autres items lexicaux. A ce stade, des erreurs dues à des transferts depuis la L1 devraient être corrigées explicitement.
- même si l'application de simple procédures de transfert continue à être source d'erreurs, il vaut mieux ne pas chercher à détecter, ni corriger de telles erreurs en production. La compréhension, essentielle à ce stade, doit être travaillée à partir de matériaux riches et difficiles.
- à un stade plus avancé, il est nécessaire de détecter et corriger les erreurs de façon à éviter l'apparition de phénomènes de fossilisation et d'aider à la restructuration fonctionnelle des connaissances de l'apprenant.

A partir de cette description générale, suivant le niveau de l'apprenant, il est possible de construire une série d'activités lexicales traditionnelles en compréhension et production sur des mots simples, des collocations ou des ESF : reconstitution d'une partie de réseau, groupements de mots suivant différents critères, exercices à trous, paraphrase d'une expression, choix entre différentes paraphrases en contexte, détermination du sens d'un mot ou d'une expression à partir d'un texte dans lequel, il (elle) est introduite de façon redondante, retrouver le sens d'un mot ou d'une expression à partir à partir d'informations fragmentaires (concept du "mot sur le bout de la langue"), etc. Ces activités nécessitent pour une grande part l'utilisation du corpus de textes, de la base de données lexicales et de l'analyseur TAG, lorsqu'il sera nécessaire de diagnostiquer les productions de l'apprenant.

Au travail linguistique proprement dit, il convient d'ajouter la possibilité d'un dialogue avec l'apprenant sur la tâche à accomplir, c'est-à-dire de pouvoir discuter explicitement des stratégies qu'il emploie, en particulier pour l'accès lexical.

Bien sûr, tout ceci n'est encore que pure spéculation, mais cette construction du dictionnaire personnalisé et l'introduction d'activités acquisitionnelles correspondant aux recommandations des chercheurs en psycholinguistique et linguistique appliquée ne peut avoir lieu sans la mise en place de ressources linguistiques importantes, organisées en fonction des besoins des apprenants. Les ressources que nous avons mentionnées dans la section précédente font, encore aujourd'hui, défaut aux systèmes lexicaux d'ALAO. C'est donc prioritairement sur ce point que portent nos efforts actuels.

### 3.2.3. "TRAVAILLER EN FRANCE": SUIVI DE L'APPRENANT

"Travailler en France" n'a ni les connaissances, ni le savoir-faire (en termes de capacités à résoudre les problèmes posés à l'utilisateur, à discuter les stratégies possibles) pour se livrer à un diagnostic de ce que fait ou de ce que connaît l'apprenant. Il n'a pas été conçu pour cela. Rappelons que son but est de mettre l'apprenant au contact de la L2, de créer un certain nombre de situations où il doit interagir dans la L2 à travers différents canaux de communication. En particulier, les questions posées à l'apprenant dans les activités ne servent pas à mesurer ses connaissances, mais à entretenir l'interaction, c'est-à-dire à motiver, orienter son exploration des matériaux offerts.

Lorsque le système garde la trace de ce qu'a fait l'apprenant (réponses aux questions, cheminement dans les simulations, enregistrement), il construit un suivi et non une évaluation. Cette distinction est très importante, mais est souvent ignorée en ALAO. Ce suivi ne fournit pas une évaluation pour les raisons suivantes (cf. section 4.3 de la publication 7):

- compte tenu d'un nombre d'activités offerte dans le logiciel, il n'est pas dit que les apprenants les accomplissent plusieurs fois, ce qui empêche toute mesure différentielle systématique,
- les réponses fournies aux questions mesurent un produit du travail, pas la façon dont il a été accompli (à l'exception des simulations, où le cheminement est mémorisé et certaines fois évalué pas à pas),
- dans certaines activités, il peut n'exister aucun système de mesure des réponses/performances de l'apprenant (productions orales , notamment),
- enfin, et surtout, dans une approche communicative, l'apprentissage ne peut se mesurer par seule référence aux compétences linguistiques, mais doit également mesurer les capacités à communiquer dans les situations spécifiques.

Concernant le dernier point, il paraît difficile de développer un logiciel spécifique permettant de tester ces capacités. C'est pourquoi, nous proposons de construire un test sans support informatique, établi sur le modèle du DELF ou DALF (diplômes internationaux de français langue seconde où sont validées connaissances et capacités à communiquer). Nous reviendrons sur ce point en section 3.4.

Revenons à la question du suivi. C'est une fonction importante car elle soutient la motivation de l'apprenant et certifie le travail qu'il a accompli ; particulièrement en situation d'autonomie, où le système doit fournir à l'utilisateur les moyens de faire le point sur ce qu'il a fait et ce qui reste à faire. C'est pourquoi nous avons conçu un programme qui donne une vue d'ensemble des activités réalisées, permet d'accéder aux données de l'utilisateur pour chacune d'entre elles. De plus, lorsque cet apprentissage autonome est contrôlé à distance par un formateur, ce suivi est la seule interface entre les deux partenaires.

Cependant, si le suivi est la base nécessaire aux prises de décisions de l'apprenant (éventuellement de concert avec le formateur), son coût de développement est élevé. En effet des informations doivent être prélevées à presque tous les points de passage de l'utilisateur ; les données collectées dans les activités ne se limitent souvent pas à des scores ou notes, mais permettent de reconstituer les productions de l'apprenant ; ces données multimédia doivent pouvoir être rejouées lorsqu'il s'agit d'enregistrements et de simulations ; le suivi doit être indépendant de la machine sur laquelle les sessions ont été réalisées (nous n'avons que

partiellement satisfait cette condition, du fait de l'espace de stockage requis par les données audio-vidéo). Le développement du suivi a représenté une fraction significative du temps total d'implémentation du logiciel. Minimiser ce coût reviendrait-il à adopter une meilleure méthodologie de développement en termes de génie logiciel, ou à introduire plus explicitement dans l'architecture initiale un modèle d'apprenant simplifié ? La question reste ouverte.

## 3.3. Pédagogie: contenu et stratégies

Où l'on illustre les principes :

2. L'ALAO doit fournir une assistance linguistique appropriée aux besoins courants de l'apprenant. Le concept d'assistance en situation d'apprentissage réfère à ce que Vygotsky appelait la "zone proximale de développement".
6. L'ALAO doit impliquer des types différents d'interaction, ne se limitant pas à celles de type tuteur-élève.
9. L'ALAO doit former les utilisateurs à devenir progressivement des apprenants autonomes en leur enseignant explicitement les stratégies d'apprentissage en langues appropriées.

Avec ELEONORE nous discutons de la nécessité d'ordonner les activités d'apprentissage par difficultés cognitives croissantes. Le système propose une variété de modes d'apprentissage en les reliant aux activités linguistiques les plus appropriées. Dans les activités les plus difficiles, le système peut adopter une attitude collaborative. Le module du compagnon natif ouvre alors un dialogue avec l'apprenant. Mais parler de collaboration n'est pas une panacée. Un examen de la littérature en ALS, sciences de l'éducation et IA-ED souligne que le problème est de définir les conditions pour qu'une interaction entre deux personnes devienne collaborative, et les conditions pour qu'une situation collaborative conduise à des interactions acquisitionnelles.

Avec CAMILLE nous abordons la question de l'auto-apprentissage. C'est une situation de fait que connaissent déjà beaucoup d'apprenants, sans avoir jamais eu recours aux NTF. Nous montrons comment le système peut aider l'utilisateur à maîtriser son propre apprentissage aux niveauX global (place de cette "tranche" de formation offerte par le logiciel dans le cursus général de l'intéressé) et local (guidage et orientation dans l'ensemble des activités proposées).

### 3.3.1. ELEONORE ET LE DIALOGUE COLLABORATIF

L'approche pédagogique dans ELEONORE revêt un caractère multi-formes. Des activités de conception assez traditionnelle en côtoient d'autres fondées sur des principes plus communicatifs. L'apprenant travaille d'abord sur un mode inductif les contraintes pragmatiques gouvernant les interrogatives. Des exercices peuvent l'aider à réviser ses connaissances grammaticales. Dans un second temps des activités collaboratives lui sont proposées, activités dans lesquelles l'objectif est l'accomplissement d'un but spécifique appartenant au contexte de la vie quotidienne. Donc, suivant les activités offertes par le système, le langage est soit le but, soit le moyen (voire les deux à la fois).

**Diversité des modes d'apprentissage**

D'un point de vue pédagogique, l'apprentissage doit être ordonné, débutant par des activités faciles pour finir sur des plus complexes. De plus le système doit offrir des modes d'interaction variés correspondant à des stratégies d'apprentissage complémentaires. Des

activités qui nécessitent seulement la compréhension/analyse d'énoncés ou de paramètres linguistiques sont bien plus aisées que celles requérant la production/génération d'énoncés. La psycholinguistique nous a même appris que les processus gouvernant compréhension et production sont différents (Levelt, 1989). Dans ces conditions, demander simultanément à un apprenant d'analyser les composantes d'une situation de communication et de produire des énoncés représente une surcharge cognitive, seulement acceptable si les sous-processus en jeu ont été relativement bien maîtrisés. L'un des grands avantages des NTF est justement de pouvoir briser, selon les besoins, cette simultanéité toujours rencontrés dans les dialogues de la vie réelle.

Notre système offre à l'apprenant trois séries ordonnées d'activités d'apprentissage :

- 1) "Choix d'une interrogative" ou apprentissage des paramètres pragmatiques. A ce stade compréhension et sélection de paramètres sont seuls mis en jeu.
2) "Production d'une interrogative" ou activation de la compétence grammaticale pour produire des interrogatives en contexte.
3) "Choisir et produire une interrogative". Mode mixte où l'apprenant doit sélectionner les paramètres et produire les questions appropriées.

Le système propose une variété de modes d'apprentissage (découverte, exercice, explicatif, collaboratif) en les reliant aux activités linguistiques les plus appropriées (compréhension seule, production seule, ou combinaison des deux) suivant l'ordre :

- étape 1 : l'apprenant découvre par lui-même (modes découvertes et explicatif),
- étape 2 : il teste ce qu'il a découvert (mode exercice),
- étape 3 : il utilise en contexte les connaissances acquises (mode exercice et collaboratif).

**Le dialogue collaboratif**

La collaboration n'est pas une panacée. Avant de l'introduire dans notre système, nous en avons exploré les fondements à partir des points de vue de l'ALS, des sciences de l'éducation et de l'IA-ED en reliant trois notions principales : collaboration, interaction et acquisition (publication 4).

En sciences de l'éducation, l'apprentissage collaboratif est une situation étudiée depuis longtemps, comparée aux situations d'apprentissage compétitif ou individualiste. Il a reçu des critiques positives et négatives (Slavin, 1983). Le problème est de définir les conditions pour qu'une interaction entre deux personnes devienne collaborative, et les conditions pour qu'une situation collaborative conduise à des interactions acquisitionnelles.

En ALS, la collaboration a été étudiée dans le cadre des recherches sur *l'interaction exolingue* entre natif et non-natif (Encrages, 1993). De façon plus générale, les aspects positifs de l'interaction mis en avant en ALS sont :

- l'interaction peut ajouter une dimension socioculturel à l'apprentissage des langues,
- elle permet à l'apprenant d'utiliser ses connaissances tout en facilitant leur révision, si nécessaire,
- apprendre en interagissant prépare l'apprenant aux situations réelles de communication.

Le premier aspect est en rapport avec notre d'étude de l'acquisition à partir de problématiques langagières qui ont une dimension linguistique (au niveau lexical et morpho-syntaxe) et une dimension (socioculturelle) pragmatique. Le second aspect est à mettre en

relation avec les procédures de diagnostic et remédiation évoquées précédemment. Dans l'interaction natif - non-natif, le premier peut aider le second en cas de misconception, en instaurant un dialogue qui porte à la fois sur la forme et le fond, en prodiguant une série de rétroactions (feedback) correctives de façon à aider l'apprenant à trouver l'énoncé satisfaisant.

Considérons maintenant les séquences interactives où il peut y avoir collaboration et acquisition. Il existe un certain consensus entre les points de vue de l'ALS et ceux des sciences de l'éducation sur les conditions nécessaires à la transformation d'une séquence interactive entre natif et non-natif en "séquence potentiellement acquisitionnelle" : de la part de l'apprenant il faut qu'il y ait conscience des difficultés, désir d'auto-comparaison de ses productions avec celles du natif, que les règles en question fassent partie de son interlangue ; de la part du natif, il faut qu'il y ait désir de collaborer et qu'il en attende quelque chose ; la situation doit être une situation de résolution de problème, donnant lieu à l'établissement d'un contrat didactique et à une bi-focalisation (sur le fond et la forme).

Dans ELEONORE, le module compagnon représentant un natif peut avoir une attitude collaborative dans deux types de circonstances (publication 1, section 4) : lors des procédures de remédiation dans les activités où l'apprenant produit des énoncés, et dans la dernière activité où l'apprenant doit s'adresser à un interlocuteur imaginaire pour accomplir une tâche (réserver une table de restaurant, par exemple). Pour se comporter de façon collaborative, le module compagnon doit avoir les compétences simulées d'un natif sur le domaine des interrogatives ; il doit savoir ce que fait l'apprenant ; et il doit adopter une attitude coopérative en étant attentif, toujours conscient des buts que doit atteindre l'utilisateur et en sachant communiquer avec lui.

Utilisateur du système et module compagnon ont chacun la possibilité d'engager un dialogue. Il y a alors interaction dans le cadre d'un apprentissage de la langue et collaboration dans la perspective de résoudre à deux une tâche non linguistique. Compagnon et apprenant dialoguent au travers de séquences potentiellement acquisitionnelles, composées d'échanges. Un échange est un tour de parole composé de deux actes de langage (speech acts). Une grammaire computationnelle spécifique gère ce dialogue. Les actes de parole sont de différente nature :

- certains peuvent être produits par l'apprenant uniquement, par le compagnon uniquement, ou par les deux.
- certains focalisent seulement sur la forme, d'autres seulement sur le contenu, mais la plupart permettent d'entretenir une bi-focalisation.
- certains actes de parole du compagnon apparaissent seulement après la production d'une interrogative par l'apprenant, ou seulement en réponse à un acte de parole de l'utilisateur, et un troisième groupe, dans chacun des deux cas.

### 3.3.2. CAMILLE ET L'AUTO-APPRENTISSAGE

"Travailler en France" a deux caractéristiques fortes (publications 7 et 24): il permet de pratiquer un auto-apprentissage, et cet auto-apprentissage porte sur une langue utilisée pour des objectifs spécifiques. L'auto-apprentissage ne saurait avoir exactement la même forme suivant le profil du public auquel il s'adresse. Le public visé est ici clairement un public non captif, adulte, professionnel ou en passe de le devenir, et qui ne dispose pas d'un temps infini pour son apprentissage d'une deuxième langue, celui-ci n'étant pas un but en soi mais un

moyen. Les apprenants sont ici typiquement des clients, des professionnels aux demandes précises et pour lesquels deux aspects sont essentiels : la souplesse et la rapidité.

La rapidité va sans dire, Porcher (1992) l'a bien mise en évidence ; la souplesse concerne les moments possibles d'apprentissage en même temps que les modalités de travail. Les nouvelles technologies ont un avantage important : elles sont à la disposition de l'apprenant quand il le désire, sur son lieu de travail ou à la maison ; elles offrent suffisamment de choix pour que l'apprenant ait l'impression de suivre son propre chemin ; elles peuvent fournir sur un seul support tous les outils et informations nécessaires pour entreprendre un niveau spécifique de formation. Cette dernière caractéristique permet d'identifier la niche particulière que peut occuper un seul logiciel du type "Travailler en France", par rapport à un ensemble de ressources informatiques. Développons ce point précis.

Il existe une tradition importante de recherches en France sur l'auto-apprentissage (qui donc examine la position de l'apprenant face à un ensemble de ressources disponibles). Elle a signalé que le risque d'un tel apprentissage intempérant est de favoriser les favorisés, de conforter les apprenants ayant déjà un capital culturel au détriment de ceux qui n'en ont pas ou beaucoup moins. Holec (1991) signale trois domaines que l'apprenant doit maîtriser pour mener à bien son propre apprentissage : l'aspect méthodologique (savoir se définir des objectifs, trouver le matériel correspondant et s'auto-évaluer), linguistique et la culture d'apprentissage.

Face à des logiciels comme "Travailler en France ", et plus généralement comme ceux de CAMILLE, l'apprenant ne se trouve pas dans la situation où il doit mettre bout à bout des morceaux de logiciels pour se construire un partie de sa formation. Notre ambition est de fournir une "tranche" de formation à l'apprenant (correspondant à plus de 30 heures de travail interactif suivant les résultats de nos premières évaluations). Le système doit donc aider l'intéressé à maîtriser son propre apprentissage aux niveaux global et local.

D'abord globalement, du fait de l'orientation LOS, des objectifs spécifiques d'utilisation de la langue ont été clairement définis ; pour ce qui concerne son positionnement par rapport à une formation globale de FLE, nous avons cherché à établir une correspondance par rapport à certaines unités du diplôme DELF ; pour ce qui concerne son mode d'utilisation, il faut distinguer séparation complète de tout cursus de formation et intégration dans un cursus où le formateur n'a qu'un contact épisodique et/ou distant avec l'apprenant. Le second cas est en cours d'expérimentation dans des instituts de formation (nous y reviendrons dans la partie évaluation). Le premier cas présente bien évidemment des inconvénients, puisque l'apprenant a seul la charge de l'organisation de sa formation. Mais cette situation correspond à un réel besoin car nombre de formations très traditionnelles se vendent déjà beaucoup aujourd'hui sur supports livres, K7, voire CD-ROM. Je pense qu'il est possible d'aider l'apprenant à se repérer, à situer les limites de tels logiciels en lui fournissant une documentation papier précisant les points globaux que nous venons de mentionner, indiquant des voies d'explorations privilégiées, des travaux supplémentaires à partir de matériaux disponibles dans le logiciel (cas du dossier culture), des moyens d'auto-évaluation. Le document que nous devons élaborer pour les besoins des évaluations sommatives pourrait sans doute remplir cette fonction.

En ce qui concerne les implications locales d'une utilisation autonome d'un tel environnement informatique, plusieurs aspects sont à prendre en considération : le guidage de

l'exploration dont l'apprenant a la maîtrise ; l'organisation et le séquencement pédagogiques des contenus ; la facilité d'accès aux informations nécessaires pour accomplir les activités ; l'offre d'un éventail de façons différentes de travailler répondant aux styles d'apprentissages variés des apprenants. Nous répondons à ces différents points (dont certains recoupent les problèmes d'utilisation d'environnements basés sur des hypertextes) dans la publication 7. Nous ne les développerons pas ici. Notons simplement que leur traitement nécessite une intégration des points de vue informatique (génie logiciel, interface, environnement informatique construit sur un hypertexte (McAleese & Green, 1990 ; Baron, Baudé & La Passardière, 1993)) et pédagogique sur l'acquisition des langues.

## 3.4. Evaluation, intégration dans le curriculum

La littérature sur les environnements informatiques d'aide à des apprentissages insiste fréquemment sur le fait que l'évaluation doit être partie intégrante des projets de recherche et/ou développements. Bien que cela soit assez peu mis en pratique, il est néanmoins vrai que les hypothèses de travail, qui ont été assertées, doivent être (au moins partiellement) vérifiées ; et cela pour deux raisons essentielles. Premièrement, une expérimentation, réalisée avant la fin du cycle de développement du logiciel, peut aider à vérifier que les apprenants utilisent réellement l'environnement dans l'esprit où il a été initialement conçu, qu'ils l'utilisent complètement, et que l'interface ne présente pas de lacunes importantes ou n'introduit pas d'interférences négatives. Deuxièmement, à la fin du processus de développement, d'autres expériences peuvent servir à mesurer la réalité d'un apprentissage au regard des objectifs pédagogiques initiaux. Cette distinction correspond aux deux types d'évaluation, généralement appelées *formative* et *sommative*.

Cette contrainte pèse-t-elle de la même façon sur les deux types de travaux de recherche que nous avons distingués en ALAO, à savoir ceux ayant pour objectif premier la modélisation informatique de l'acquisition, et ceux plaçant au premier plan des objectifs pédagogiques finalisés et n'incorporant pas de modèle explicite d'acquisition ?

L'évaluation formative devrait être, à mon avis, un impératif pour tout projet de recherche qui a atteint le stade du développement d'un prototype. Il peut être par contre sans objet, voire même pénalisant, d'ériger en principe l'évaluation sommative pour le premier type de travaux de recherche. Les chercheurs en IA-ED consacrent presque inéluctablement dans leurs articles ou livres (Polson, 1988) un chapitre final à la question de l'évaluation de type sommatif. Si l'on comprend bien les raisons pratiques qui les poussent à trouver une argumentation pour convaincre les sponsors de la rentabilité des investissements dans la recherche IA-ED, il n'est pas sûr que cela soit la bonne façon d'y répondre[10], il est même possible que cela soit une erreur méthodologique. A quoi cela sert-il de mettre sur pied une expérimentation lourde pour des versions provisoires de prototypes qui correspondent à des élaborations successives de modèles d'acquisition ?

Quel type de validation de ces modèles est-il alors possible d'élaborer ? Self (1989) propose de faire fonctionner l'environnement informatique comme un simulateur. On part de l'hypothèse que l'on dispose d'une théorie (même partielle) d'apprentissage L, qui à partir d'un

[10] Il est possible en effet d'effectuer un transfert des résultats de ces recherches sur des projets à finalité plus pratiques. C'est du moins dans cet esprit que j'essaye de travailler.

modèle S d'apprenant fait correspondre à toute action pédagogique A un nouveau modèle S'. Disposer d'une théorie d'un modèle pédagogique P, permet de sélectionner la séquence d'actions pédagogiques qui maximise E(Sf), où Sf est l'état du modèle de l'apprenant en fin de session, et E une fonction d'évaluation. A partir d'un S déterminé, il s'agit alors de constituer des séquences d'actions de l'apprenant, en faisant varier les paramètres interactionnels à sa disposition (lorsque des productions langagières sont nécessaires, mieux vaut les extraire d'une expérimentation limitée de type formatif). Puisque les modèles et théories partielles utilisées ne sont pas strictement corrects, il est possible d'affiner le modèle de l'apprenant en introduisant une analyse probabiliste par exemple.

Cette méthodologie a peu été expérimentée (VanLehn & al, 1994) et demanderait, bien sûr à être précisée. Elle semble avoir l'avantage de se construire de façon montante à partir des paramètres du modèle d'acquisition dont le chercheur dispose, de façon à en vérifier sa cohérence interne et à l'affiner. L'expérimentation avec des apprenants, toujours nécessaire in fine, n'interviendrait que lorsque ce modèle serait stabilisé. Concrètement, pour ce qui nous concerne directement, une expérimentation formative est prévue pour ALEXIA début 1995. Je compte ensuite reprendre cette idée de simulation.

Revenons maintenant aux évaluations concernant les environnements à finalité plus pratique. Nous venons de terminer l'évaluation formative du premier module "A la Recherche d'un Emploi". Le dépouillement des données est en cours et sera achevé à temps pour que certains résultats puissent être pris en compte dans le développement du deuxième module. L'évaluation sommative du premier module est prévue au premier semestre 1995. C'est un processus compliqué, nécessitant une lourd travail de préparation : une coordination européenne, une préparation de documents de travail pour les pré et post-tests permettant d'aborder les différents aspects de la compétence communicative, une coordination avec l'équipe de formateurs, la création de scénarios d'utilisation pour les groupes de sujets, etc.

L'évaluation de l'apprentissage est un problème complexe que le système éducatif a su parfaitement évacuer. Il n'est peut-être pas inutile de rappeler qu'en ce qui concerne les systèmes de formations traditionnelles (Porcher, 1990: p. 27):

> "On évalue beaucoup plus l'enseignement que l'apprentissage, la conformité que la compétence, l'académique que le savoir-faire. C'est moins ce que l'élève est capable de faire, qui est contrôlé, que la manière dont il est capable de reproduire le cours suivi. Il s'agit d'une évaluation spéculaire et non pas d'un contrôle de capacité."

Les contraintes qui pèsent sur l'ALAO sont donc assez singulières ! D'autant que souvent les dés sont "pipés". En effet, l'évaluation d'un logiciel n'est qu'un aspect du problème. L'étude de l'intégration des logiciels dans les cursus de formation est souvent laissée de côté. Or elle est déterminante. S'il s'agit d'un logiciel remplissant des fonctions limitées, la coordination de son utilisation avec les programmes des formateurs est essentielle ; et, dans tous les cas (donc, en particulier pour les systèmes visant un apprentissage autonome), l'accès aux ressources technologiques est lui aussi essentiel. Les institutions qui ont su penser et planifier l'introduction des NTF savent bien, qu'à partir des ressources déjà existantes aujourd'hui, il est possible de déléguer une fraction importante des charges de formation (jusqu'à 50 % du temps de formation chez certains de nos partenaires dans le projet CAMILLE). Le processus d'évaluation dans ces cas là est simple. Il ne fait pas de différence

avec l'enseignement traditionnel : les tests mis au point pour l'évaluation des enseignements traditionnels sont appliqués sans changement, lorsqu'à un cours est substituée une formation uniquement basée sur les NTF. L'amélioration des résultats est d'ailleurs souvent spectaculaire. Bien sûr, ce type d'opérations n'est possible que pour l'acquisition de certaines compétences. C'est l'objectif de la recherche appliquée que d'ouvrir cet éventail de possibilités.

Cela dit, l'élaboration d'une évaluation sommative pour "Travailler en France" dans le cadre du projet CAMILLE m'intéresse à plusieurs titres. Nous manquons de critères de référence pour les environnements de types hypermédia de large échelle. La détermination de certains critères permettra de relancer des recherches plus ponctuelles. Enfin, nous espérons pouvoir extraire des documents de test un guide d'auto-apprentissage qui accompagnera le logiciel.

# 4. ORGANISATION ET DIFFUSION DE LA RECHERCHE EN ALAO

Dans cette partie je présente ma contribution à l'organisation de la recherche en ALAO. Des compléments d'informations sont développés de façon dispersée dans le document de CV étendu fourni en annexe.

## 4.1. Formation de chercheurs, organisation d'équipes de recherche

Un axe de recherche ne peut vivre sans l'engagement de jeunes chercheurs. C'est pourquoi, à partir de la structure existante du DEA de linguistique-informatique de Clermont 2, nous avons ouvert une nouvelle option sur le thème du langage naturel et des environnements informatiques d'aide à l'apprentissage. Lors de sa création en 1991, j'ai cherché à faire participer des enseignants-chercheurs de différentes disciplines : informatique et intelligence artificielle, linguistique computationnelle et linguistique appliquée (ou plutôt didactique des langues), et psycholinguistique. Cette option est commune aux projets de recherche sur l'acquisition de la lecture et sur l'acquisition d'une langue seconde. Nous recrutons des étudiants d'origine littéraire ou scientifique. Notre ambition n'est pas, bien entendu, de transformer des informaticiens en linguistes, ni des étudiants de cursus littéraire en informaticiens, mais de les sensibiliser, autour de la problématique de l'acquisition d'une langue ou du langage, aux questions que chaque discipline juge primordiale et aux méthodologies de recherche respectives. Suivant leur formation antérieure, les étudiants seront orientés vers des projets de DEA où ils auront le choix de travailler, soit sur des problèmes de linguistique computationnelle, des problèmes IA-ED, à construire des expérimentations avec des apprenants, ou à décrire des phénomènes linguistiques particuliers. Si le stage de DEA ne requiert qu'un type particulier de savoir-faire, il n'en est pas de même du travail de thèse. D'où cette sensiblisation initiale.

En ce qui concerne cette pédagogie de la recherche, une discipline est particulièrement mal représentée, non seulement localement mais également au niveau national, il s'agit de la linguistique appliquée (au sens anglo-saxon). S'il existe une communauté relativement forte de didactique des langues en France (représentée aussi localement), il n'en existe aucune qui explore l'éventail des thèmes de recherche en ALS repris internationalement. Cela se traduit

par l'absence d'ouvrages de référence, en langue française, sur cette méthodologie de recherche (du type de ceux proposés par Brown (1988) et Nunan (1992)), sans parler d'ouvrages sur les théories d'acquisition en L2 (à l'exception de ceux, notables, proposés par les psycholinguistes). Pire même, les livres et journaux en langue anglaise du domaine sont absents des bibliothèques universitaires nationales, ce qui témoigne bien de l'orientation très particulière de la linguistique française. C'est ce déséquilibre que constate Lehman (1993) dans son ouvrage sur l'apprentissage à objectifs spécifiques

Parlons maintenant de l'organisation de la recherche. Le domaine de l'ALAO est par nature un domaine pluri-disciplinaire. Le montage et la participation à des projets rassemblant des équipes de spécialités différentes localisées en France ou à l'étranger est bien sûr nécessaire pour l'échange d'idées et le transfert de savoir-faire (cf. le projet ALEXIA et les discussions concernant CAMILLE dans les publications 5 et 7). J'ai constaté cependant une réelle difficulté à intégrer dans des travaux collaboratifs des chercheurs de formation psycholinguistique au niveau français, malgré nos multiples démarches dans ce sens.

Mais cette collaboration autour de projets ne saurait se substituer à la création d'une ou plusieurs équipes au sein d'un même laboratoire, que ce soit pour mener des recherches personnelles ou collaboratives. Il existe un seuil critique minimum en dessous duquel toute recherche en ALAO risque d'être compromise. Pour pouvoir correctement travailler une équipe doit, bien évidemment, disposer d'un potentiel de ressources informatiques dont la gestion nécessite un investissement en temps non négligeable. L'équipe doit être pluri-disciplinaire et rassembler dans notre cas des compétences en :

- informatique aussi bien au niveau génie logiciel, qu'au niveau micro-informatique (connaissance détaillée des matériels et des couches "basses" des logiciels et systèmes d'exploitation),
- intelligence artificielle (aspects théoriques et techniques, connaissances des différents langages de programmation couramment utilisées dans cette spécialité, capacités à faire communiquer des applications écrites dans différents langages),
- linguistique computationnelle,
- linguistique appliquée, avec un minimum de compétences en didactique des langues.

Pour qu'un réel travail collaboratif puisse s'installer, il faut non seulement que chaque participant soit ouvert aux méthodologies de recherche et de travail des autres, mais également qu'existe une pratique commune intensive, dont la durée minimum doit souvent être supérieure à une année pour devenir fertile.

Enfin, puisque l'expérimentation avec des apprenants intervient systématiquement à des stades différents du développement d'un projet d'ALAO, l'équipe doit avoir accès localement à un institut spécialisé dans la formation en langue. Un institut à caractère professionnel est bien plus approprié qu'un centre de formation universitaire pour étudiants étrangers, car, dans le second cas, la durée des séjours des étrangers est souvent trop limitée et leurs motivations faibles.

## 4.2. Contribution à la diffusion et à l'organisation de la recherche

Pour que la recherche en ALAO puisse se développer, il faut non seulement qu'une collaboration pluri-disciplinaire s'installe au sein d'un laboratoire, autour de projets de

recherche communs à plusieurs laboratoires, mais également au niveau des communautés de chercheurs des disciplines concernées.

Pour atteindre ce dernier objectif, j'ai orienté mes actions dans deux directions complémentaires. D'abord, de concert avec différents collègues, nous avons introduit le thème de "la modélisation informatique de l'acquisition des langues secondes" au sein de la communauté nationale des sciences cognitives (projet ALEXIA, accepté dans le GDR "Sciences cognitives de Paris"), au sein de la communauté nationale de TALN (organisation sous notre responsabilité d'une journée ATALA), au sein de la communauté française EIAO (participation au comité de programme des journées EIAO de Cachan) et de la communauté internationale IA-ED (édition d'un numéro spécial du journal "Journal of Artificial Intelligence in Education" (publication 27), apparition du thème de l'apprentissage des langues pour la première fois en 1995 au congrès Artificial Intelligence in Education). Ces différents milieux opposaient au départ une certaine résistance (surtout au niveau français) du fait de l'assimilation entre les thèmes acquisition des langues et didactique des langues.

Le deuxième axe d'interventions concerne les colloques et congrès scientifiques. Certains ne rassemblent qu'un nombre trop limité de disciplines (uniquement des enseignants de langue, ou des linguistes et psycholinguistes sans chercheurs en acquisition sur des thèmes pourtant pertinents pour l'ALAO, des chercheurs en ALS sans chercheurs en linguistique computationnelle ni linguistique théorique). Mais il ne suffit pas de rassembler des communautés différentes pour que de réels échanges aient lieu (publication 31 bis). Les comités de programme doivent installer des procédures de sélection sérieuses en fixant certaines contraintes aux auteurs, comme celles de situer des approches de type linguistiques-informatiques en ALAO par rapport à l'état de l'art en ALS. Notre contribution dans ce domaine a consisté à organiser le colloque SCIAL (publication 28) en 1993 qui a rassemblé chercheurs et praticiens autour du thème "sciences cognitives, informatique et acquisition des langues".

Enfin, l'ALAO, de par ses objectifs, ouvre sur des thématiques de recherche à caractère plutôt fondamental, ou, au contraire, à caractère appliqué. J'espère avoir montré dans ce document que ces deux thématiques étaient conciliables et que les échanges entre elles pouvaient être bi-directionnels.

# 5. BIBLIOGRAPHIE


Abeillé A. (1991): "Quand l'arbre ne cache pas la forêt : analyse du français à l'aide d'une grammaire d'Arbre Adjoints", *TA Informations*, vol. 32, 2, Klincksieck, pp 51-70.

Abeillé A. (1993): *Les Nouvelles Syntaxes : Grammaires d'unification et analyse du français*. Armand Colin : Paris.

Aitchison J. (1987): *Words in the mind*. Blackwell.

Baron G.L., Baudé J. & de La Passardière B. (1993): *Hypermédias et Apprentissages*. Actes des deuxièmes journées. INRP.

Bates E.(1976): *Language and context: the acquisition of pragmatics.* New york: Academic Press.

Bérard E. (1991) : *L'approche communicative : Théories et pratiques.* CLE International : Paris.

Besse H. (1986): "Enseignement/apprentissage des langues étrangères et connaissances grammaticales et linguistiques. *Les langues Modernes: faire réfléchir sur la langue*, n°2.

Bialystok E. (1990): *Communication strategies: a psychological analysis of second-language use*. Basil Blackwell: Oxford.

Bogaards P. (1991): "A propos de l'usage du dictionnaire de langue étrangère". *Cahier de Lexicologie*, N°52, Vol 1, pp 131-152.

Borsley R.D. (1991): *Syntactic Theory: A Unified Approach*. Edward Arnold.

Brown J.D. (1988) : *Understanding Researh in Second Language Learning.* Cambridge University Press.

Brumfit C. & Johnson K. (1979) (eds): *The Communicative Approach to Language Teaching.* Oxford University Press.

Canale M. & Swain M. (1980): "Theoretical bases of communicative approaches to second language teaching and testing". *Applied Linguistics*, vol 1 (1), p 28.

Caron J. (1989): *Précis de psycholinguistique*. PUF.

Carter R., McCarthy M. (1988): *Vocabulary and Language Teaching*. Longman: London.

Charniak E. (1993): *Statistical Language Learning*. MIT Press.

Computational Linguistics (1993): Special Issue on Using Large Corpora, *Computational Linguistics*, vol 19, 1/2.

Cosnier J. & Brossard A. (1984): *La communication non verbale.* Delachaux et Niestlé : Neuchâtel.

Coste, D.(1970)."Les procédés d'interrogation directe.", *Etude de conversations d'enfants de neuf ans."*, CREDIF, ENS St-Cloud.

Carter R., McCarthy M. (eds) (1988): *Vocabulary and Language Teaching*. Longman.

Catt M. & Hirst G. (1990): "An Intelligent CALI System for grammatical Error Diangosis. *Computer Assisted Language Learning (CALL)*, vol 3, pp 3-26.

Cosnier J. & Brossard A. (1984): *La communication non verbale*. Delachaux et Niestlé : Lausanne.

Demaizière F. & Dubuisson C. (1992): *De l'EAO aux NTF. Utiliser l'ordinateur pour la formation.* Ophrys : Paris.

Di Vito N. O'Connor (1991) : "Incorporating Native Speaker Norms in Second Language Materials". *Applied Linguistics*, vol 12, 4, pp. 383-396.

Dutoit D. (1992): "A set-theoretic approach to lexical semantics". *Proc. of COLING 92*, Nantes, Aug. 23-28, pp 983-987.

Ellis R.(1985): *Understanding Second Language Acquisition*, Oxford: Oxford University Press.

Ellis R.(1990) :*Instructed second language acquisition*. Oxford: D.Crystal & K.Johnson.

Encrages (1993): *Interaction et acquisition d'une langue étrangère*. Aile, revue de l'association Encrages, n°2.

Farghali A. (1989): "A Model for Intelligent Computer Assisted Language Instruction (MICALI)" . *Computer and the Humanities*, vol. 23, 1, pp 235-250.

Galisson R. & Coste D. (1976): *Dictionnaire de didactique des langues*. Hachette : Paris.

Gaonac'h D.(1987): *Théories d'apprentissage et acquisition d'une langue étrangère*, Paris: Hatier-Credif.

Goodfellow R. (1993): "Call for vocabulary : requirements, theory and design", *Computer Assisted Language Learning*, Volume 6, Part 2.

Goodfellow R. (1994): "Design principles for computer-aided vocabulary learning". *Computers and Education*, vol 23, 1/2, pp 53-62.

Helmy. Ibrahim Amr H. (ed) (1989): *Lexiques.* Hachette, collection Recherches et Applications.

Holland M., Kaplan J., & Sams M. (eds.) (in press): *Intelligent language tutors: balancing theory and technology*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Holec H. (1991): "Autonomie et apprentissage auto-dirigé, quelques sujets de réflexion". *Les auto-apprentissages, actes des 6èmes rencontres de l'ASDIFLE*.

Hutchinson T., Waters A. (1987): *English for Specific Purposes. A learning-centred approach*. Cambridge University Press.

Hymes D. H. (1971) : "On communicative competence" in (Brumfit & Johnson 79), pp 5-26. Excerpts from a 1971 publication.

Johnson K. (1979): "Communicative Approaches and Communicative Processes" in (Brumfit & Johson 1979), pp 192-205.

Ingram D.(1989): *First Language Acquisition: method, description and explanation*. Cambridge: Cambridge University Press.

Issac F. (1994).: “Un algorithme d’analyse pour les grammaires d’arbres adjoints”, *Colloque international sur les Grammaires d’Arbres Adjoints*. Paris.

Jacquemin C. (1994): "Les entrées lexicales complexes: à l'interface entre le lexique et la syntaxe. *Actes des journées TALN*, Marseille, avril. pp 99-108.

Johnson K. (1979): "Communicative Approaches and Communicative Processes" in Brumfit C. & Johnson K. (eds.): *The Communicative Approach to Language Teaching*. Oxford University Press, pp 192-205.

Krashen S.D. (1988): *Second Language Acquistion and Second Language Learning*. Prentice-Hall.

Lafon P. (1984): *Dépouillements et statistiques en lexicométrie*. Genève: Slatkine-Champion.

Last R.W. (1989): *Artificial Intelligence Techniques in Language Learning*. Ellis Horwood.

Laufer B. (1993): "The effect of dictionnary definitions and examples on the use and comprehension of new L2 words". *Cahiers de Lexicologie*, N°63, 2, pp 131-142.

Lebart L. & Salem A. (1994): *Statistique Textuelle*. Dunod : Paris.

Lecocq P, Segui J.(eds) (1989): *L'accès lexical*. Revue Lexique. Presse Universitaire de Lille.

Le Goffic P. (1993): *Grammaire de la phrase française.* Hachette.

Lehmann D. (1993) : *Objectifs spécifiques en langue étrangère : les programmes en question*. Hachette : Paris.

Levelt W.J.M. (1989): *Speaking: From Intention to Articulation*. MIT Press.

McLaughlin B. (1987): *Theories of Second Language Acquisition*. London: Edward Arnold.

MacWhinney B., Bates E. (1989): *The Cross-linguistic Study of Sentence Processing*. Cambridge University Press: New York.

MacWhinney B. (1992): "The Competition Model and foreign language acquisition". In *Intelligent Tutoring Systems for Second-Language Learning*, Swartz, M. & Yazdani, M. (eds). New York: Springer-Verlag, pp. 39-50.

Manning P. (1995): *Investigating Teaching Strategies in Computer Assisted Language learning*. Thèse de l'Open University, Milton Keynes, GB.

Matthews C. (1994): "intelligent Computer Assisted Language Learning as a Cognitive Science : The choice of syntactic frameworks for language tutoring". Special Issue on Language Learning, Chanier T. (ed). *Journal of Artificial Intelligence in Education (JAIE)*, à paraître.

McAleese R. & Green C. (1990): *Hypertext : state of the art*. Intellect : Oxford.

Mel'cuk I. (1992): *Dictionnaire explicatif et combinatoire du français contemporain, Vol. III.* Montréal: Les Presses de l'université de Montréal.

Moeschler J. & Reboul A. (1994): *Dictionnaire Encyclopédique de Pragmatique*. Editions du Seuil : Paris.

Moirand S. (1982): *Enseigner à communiquer en langue étrangère*. Hachette: Paris.

Nwana H., S. (ed.) (1993): *Mathematical Intelligent Learning Environments*. Oxford: Intellect.

Nunan D. (1988) : *The Learner-Centred Curriculum. : A study in second language teaching.* Cambridge University Press.

Nunan D. (1992) : *Research Methods in Language Learning.* Cambridge University Press.

Odlin T. (1989): *Language transfer : Cross-linguistic influence in language learning.* Cambridge University Press.

Ohlsson S. (1992): "Constraint-Based Student Modelling". *Journal of Artificial Intelligence in Education*, vol. 3 (4), pp 429-447.

O'Malley J.M, Chamot A.U.(1990): *Learning Strategies in second language acquisition*. Cambridge University Press.

Oxford R.L. (1993): "Intelligent computers for learning languages: The view for Language Acquisition and Instructional Methodology", *CALL journal*, vol. 6 (2), pp 173-188.

Pijls F., Daelemans W. & Kempen G. (1987): "Artificial Intelligence tools for grammar and spelling instruction". *Instructional Science*, vol 16, pp 319-336.

Polson M.C., Richardson J.J. (eds) (1988): *Foundations of Intelligent Tutoring Systems.* Lawrence Erlbaum.

Porcher L. (1990) : " L'évaluation des apprentissages en langue étrangère". *Etudes de Linguistique Appliquée* , n° 80. pp.5-36.

Porcher L. (ed.) (1992): *Les auto-apprentissages*, numéro spécial du Français dans le Monde, Hachette : Paris.

Rabefitia M. (1994): *Extraction automatique des collocations et des expressions terminologiques.* Mémoire de DEA MIASH, Université Paris 4 et Paris 5.

Renié D. (1995): *Acquisition des interrogatives directes du français langue seconde assistée par ordinateur : aspects syntaxiques et pragmatiques*. Thèse de l'université Clermont 2 : Clermont-Ferrand.

Richards J.C. (1985): "Lexical knowledge and the teaching of vocabulary", in *The Context of Language Teaching*, Richards J.C. (ed), Cambridge University Press.

Schwind C. (1990): Critique du livre "Artificial Intelligence Techniques in Language Learning", *Computational Linguistics*, vol 16,4, décembre. pp 242-244.

Self J. (1988): "Student Models : what use are they ?". In *Artificial Intelligence Tools in Education*, Ercoli P. & Lewis R. (eds), Elsevier Science Publishers.

Self J. (1989) : "Theoretical Foundations for Intelligent Tutoring Systems. Invited paper, *4th International Conference on AI and Education.* Amsterdam, May.

Self J., Brna P., Parkes A. & Twidale M. (eds) (1995): *Intelligent Learning Environments.* Lawrence Erlbaum Associates.

Sérasset G. (1994): "Approche oecuménique au problème du codage des structures linguistiques". *Actes des journées TALN*, Marseille, avril. pp 109-118.

Shieber S. M. (1986): *An Introduction to Unification-Based Approaches to Grammar*. Univeristy of Chicago Press : Chicago.

Slavin R.(1983). *Cooperative Learning.* Longman: New York.

Swartz, M. & Yazdani, M. (eds.) (1992): *Intelligent Tutoring Systems for Second-Language Learning*, . New York: Springer-Verlag.

Tarone E. & Yule G. (1989) : *Focus on the Language Learner*. Oxford University Press.

Taylor I. (1990): *Psycholinguistics:Learning and Using Language*. Prentice Hall.

Thompson J. & Chesters G. (1992): "CALL for All". *Computers and Education*, vol. 19 (1).

Thompson J. & Zähner C. (eds.) (1992): *Proceedings of the ICALL Workshop, UMIST, September 1991*. CTI for Modern Language, University of Hull, UK.

Valdman A.(1990): "Anthenticity, variation, and communication in the foreign language classroom." in *Text and context: cross-disciplinary perspectives on language study*. Lexington, Mass: Heath, pp.79-97.

VanLehn K., Ohlsson S. & Nason R. (1994) : "Applications of Simulated Students: An Exploration". *Journal of Artificial Intelligence in Education*, vol. 5, 2. pp 135-176.

Wenger E. (1987): *Artificial Intelligence and Tutoring Systems*. Morgan Kaufmann.

White L. (1989) : *Universal Grammar and Second Language Acquisition.* (vol.1), John Benjamins : Amsterdam.