

LOGICIEL D'AIDE À L'ÉCRITURE
INTÉGRANT DES TECHNIQUES
D'INTELLIGENCE ARTIFICIELLE

RAPPORT DE RECHERCHE

Rédigé par Laurence BEER
sous la direction de
COLETTE DUBUISSON
ET MARIE NADEAU

UQAM

JANVIER 1991

INTRODUCTION

Ce rapport a pour but de présenter l'état d'avancement d'une recherche sur l'élaboration d'un logiciel d'aide à l'écriture destiné aux sourds de niveau collégial et universitaire. Un prototype de ce logiciel permet de se rendre compte du fonctionnement général du système et d'évaluer ce qui a déjà été réalisé. Ce rapport comprend deux parties. Il commence tout d'abord par un bref rappel de la problématique, de la finalité et par une description générale du projet. Il se poursuit ensuite par une description détaillée du travail déjà accompli.

L'équipe qui a participé à cette recherche est constituée par :

Laurence Beer,
Alain Boulanger,
Bruno Dima,
Colette Dubuisson,
Daniel Leiby,
Christopher Miller,
Marie Nadeau
Chantal Payant.

PREMIERE PARTIE : PROBLÉMATIQUE, FINALITÉ ET DESCRIPTION GÉNÉRALE

1.1. PROBLÉMATIQUE

L'accès à la langue chez les personnes déficientes auditives, partiel et tardif, influence de façon évidente la maîtrise de la langue écrite. En effet, les recherches dans ce domaine montrent que les personnes, sourdes de naissance ou devenues telles en bas âge et avant l'acquisition d'une langue orale, ne parviennent que très rarement à bien assimiler le code écrit.

Voici un bref aperçu des études antérieures à ce projet : Fوسفeld (1958) parle de l'immaturité des sourds dans la langue écrite. Myklebust (1964) montre que les habiletés des sourds en langue écrite ne dépassent jamais deux tiers du niveau normal tout au long de leur scolarité. Schulze (1965) parle d'un retard de 4 ans dans le vocabulaire des sourds. Templin (1966) parle lui aussi du niveau lexical inférieur. L'étude de Thomson (1936), confirmée plus tard par Myklebust (1960) et Perry (1968) montre qu'il est possible de regrouper les erreurs des sourds selon les catégories suivantes, par ordre d'importance : omission, substitution, addition et ordre des mots. L'omission des mots "nécessaires" est la plus fréquente. Power (1971) affirme que l'enfant sourd ne peut atteindre la longueur moyenne d'une phrase d'un enfant entendant de huit ans avant l'âge de dix-sept ans. Dans une de leurs recherches, Quigley et Marschall (1970) ont étudié sur une période de neuf ans la production écrite de sujets sourds. Tout comme Heider et Heider (1940), ils démontrent qu'il y a une amélioration au cours des années, mais que malgré une plus grande complexité syntaxique, les personnes sourdes conservent leur retard par rapport aux sujets entendants.

Assurer un enseignement individualisé grâce à des rétroactions adaptées à chaque erreur et suivant le rythme d'apprentissage de chacun est certainement l'avantage le plus reconnu des applications pédagogiques de l'ordinateur. Pour les déficients auditifs, un système qui propose des phrases correctes semble un moyen d'autant plus justifié qu'il rejoint un mode d'apprentissage dont les sourds sont le plus souvent privés. En effet, un enfant sourd a beaucoup moins d'interactions avec son milieu qu'un enfant entendant, à moins d'avoir des parents qui signent, auquel cas il apprendra la langue des signes comme un entendant apprend le français. Dans l'apprentissage de la langue écrite, les rétroactions sont moins abondantes et surtout rarement immédiates. Schlesinger (1988) explique même certaines difficultés des sourds en langue écrite par la pauvreté du langage qu'on leur adresse.

1.2. FINALITÉ

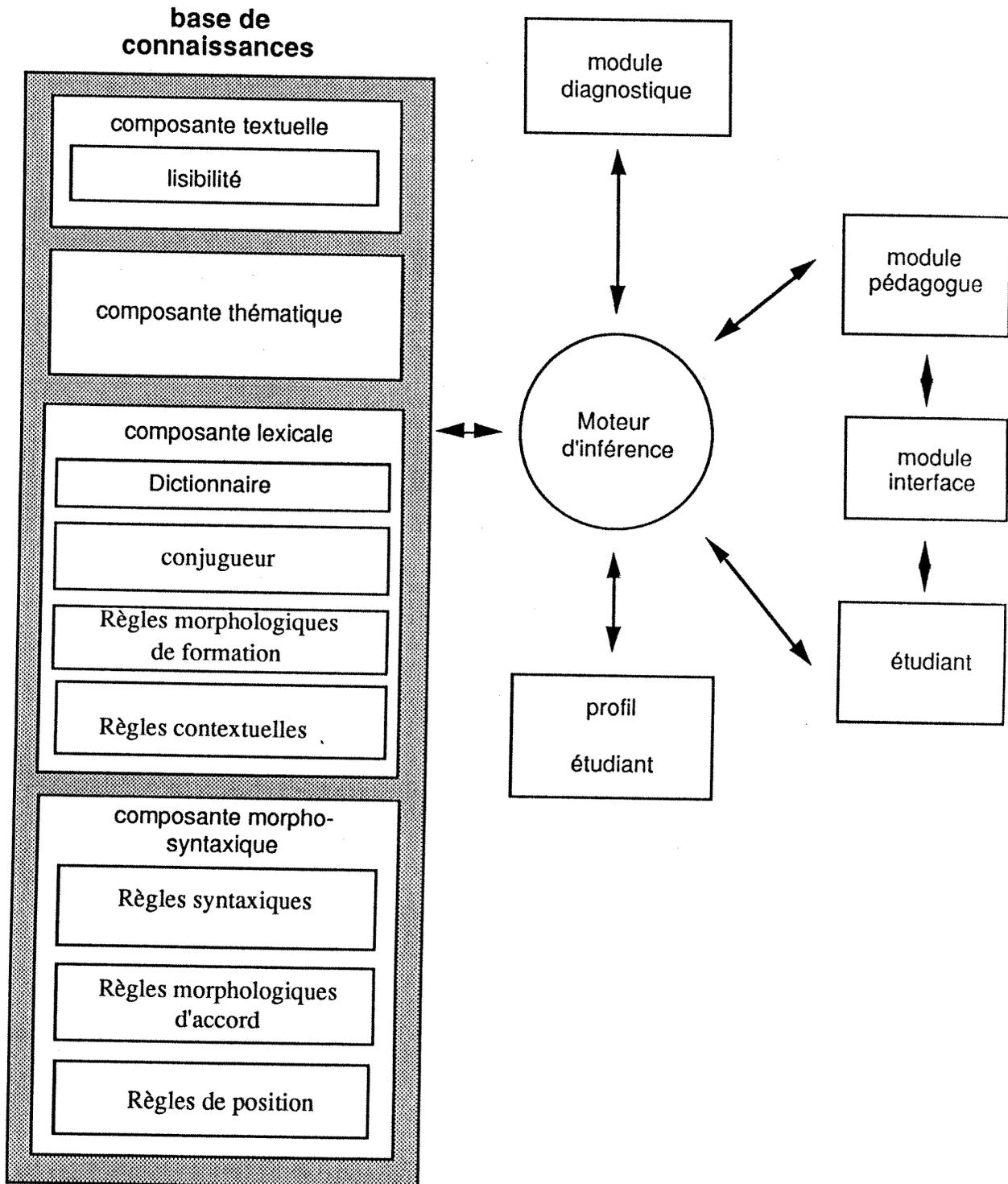
Le logiciel d'aide à l'écriture intégrant des techniques d'intelligence artificielle est donc un tuteur intelligent qui a comme ambition d'apprendre au déficient auditif de niveau collégial ou universitaire à écrire un texte en français correct sur un thème donné.

Ce logiciel fonctionne avec un moteur d'inférence exploitant une base de connaissances provenant d'un domaine d'expertise humaine, en l'occurrence la langue française. Il est important de préciser que l'apprenant compose son texte sans l'intervention de la machine et sans limite de temps.

1.3. DESCRIPTION GÉNÉRALE DU LOGICIEL

Etant donné la difficulté de réaliser un tel projet, il ne peut être question d'élaborer un système qui aide un apprenant sourd à écrire, quelle que soit la situation d'écriture dans laquelle il se trouve. Il faut donc préciser qu'à cette étape de la recherche, un cadre réaliste a dû être défini. Le tuteur intelligent est limité au micro-monde d'une courte bande dessinée (cf. annexe 1). Bien entendu, un enrichissement et un élargissement ultérieurs sont envisagés.

Le tuteur intelligent peut être représenté par le schéma suivant :



1.3.1. Base de connaissances

Elle comprend quatre composantes :

- *composante textuelle* : elle a pour but de délimiter les frontières entre les phrases et de compter les mots;
- *composante thématique* : elle permet de s'assurer d'une compréhension minimale de l'exercice demandé de la part de l'apprenant, en vérifiant la présence des mots-clefs ou de leurs synonymes;
- *composante lexicale* : elle a pour objectif d'assigner une catégorie grammaticale à chacun des mots du texte produit par l'apprenant. Elle comprend un dictionnaire (avec des traits de sous-catégorisation portant sur les verbes et des traits sémantiques pour les noms), un conjugueur, des règles morphologiques de formation et des règles contextuelles;
- *composante morpho-syntaxique* : elle comprend un analyseur syntaxique, des règles morphologiques d'accord et des règles de position.

1.3.2. Module diagnostique

Il est constitué d'une grammaire des fautes permettant de catégoriser les erreurs commises par les apprenants afin de leur proposer une correction adaptée.

1.3.3. Module pédagogique

Il contient la liste des savoir-faire à développer organisés de façon hiérarchique, ainsi que les stratégies pédagogiques qui leur sont associées.

1.3.4. Profil de l'étudiant

Il est constitué des informations sur la performance accomplie par l'apprenant tout au long de l'exercice. Il permettra de réaliser un bilan complet du travail accompli avec le détail des erreurs commises et les diverses réactions aux stratégies pédagogiques mises en oeuvre.

1.3.5. Interface en langue naturelle

Elle comprend les différents messages nécessaires au dialogue de la machine avec l'utilisateur.

1.4. PRÉEXPÉRIMENTATION

Le prototype a été élaboré à partir de l'analyse des données d'une préexpérimentation non informatisée.

Les textes déjà recueillis ont formé un CORPUS DE PRÉEXPÉRIMENTATION que l'on peut diviser en trois parties :

- 10 textes de personnes sourdes, dont on trouvera un exemple en annexe 2;
- 3 textes de personnes mi-sourdes,
- 19 textes de personnes entendantes (corpus témoin).

Ces textes ont été analysés avec le logiciel SATO (Système d'analyse de Textes par Ordinateur), créé par François Daoust à l'UQAM. Ce logiciel permet de calculer le nombre de mots, le nombre de phrases, la longueur moyenne de ces dernières ainsi que le niveau de lisibilité. Il fournit également une liste précise du lexique et donne la fréquence de chaque mot à travers le corpus global. SATO permet enfin d'obtenir une mesure de comparaison entre les différents textes de notre corpus.

L'anonymat de ceux qui ont accepté de se prêter à l'expérience a été respecté.

DEUXIEME PARTIE : DESCRIPTION DÉTAILLÉE DU LOGICIEL

2.1. BASE DE CONNAISSANCES

Cette base comprend quatre parties :

- 2.1.1. composante textuelle
- 2.1.2. composante thématique
- 2.1.3. composante lexicale
- 2.1.4. composante morpho-syntaxique

2.1.1. Composante textuelle

La composante textuelle découpe le texte en phrases afin de vérifier sa lisibilité.

Elle doit, premièrement, délimiter les phrases et, deuxièmement, vérifier que celles-ci ne sont pas trop longues. Il faut donc préalablement définir de façon précise ce qu'est un mot, ce qu'est une phrase et ce qu'est un texte; on peut alors compter les mots et calculer leur nombre par phrase. Voici ces définitions :

- Un mot est une chaîne de caractères entre des délimiteurs (cf. annexe 3);
- Une phrase est un ensemble de mots dont la première lettre du premier élément est une majuscule et qui se termine par un délimiteur de phrase (cf. annexe 3);
- Un texte est un ensemble de phrases qui se termine par le délimiteur "*FIN*" et un retour de chariot;

Nous nous sommes basés sur la norme généralement appliquée dans l'écriture efficace, qui stipule qu'un texte devient plus difficilement lisible si la moyenne de mots par phrase dépasse 13. Cette moyenne ne sera qu'indicative : elle donnera une idée générale de la lisibilité du texte de l'étudiant. Si une phrase est très longue (plus de 20 mots), il est demandé à l'étudiant de vérifier la ponctuation. Bien entendu, si un texte n'est pas ponctué, il est demandé à l'apprenant de le ponctuer.

2.1.2. Composante thématique

Après la lisibilité, il faut vérifier la présence de mots-clefs pour s'assurer d'un minimum de compréhension de l'histoire par l'apprenant. Cela consiste en un ensemble des thèmes jugés indispensables à la narration correcte du récit. Une liste restreinte (cf. annexe 4) de ceux-ci a été constituée; le logiciel vérifie la présence de chacun de ces thèmes (mot-clef ou synonyme) grâce à cette liste.

Si un des termes indispensables (ou un synonyme) est absent, le système questionne l'étudiant. Par exemple, "Dans la bande dessinée, il y a des soucoupes volantes et tu n'en parles pas. Sais-tu ce que veut dire "soucoupe volante"?. Une explication sera donnée si elle est demandée; cela signifie que l'on affichera la définition (simple et concise) du terme inconnu. L'étudiant sera ensuite invité à modifier son texte et l'interface n'interviendra plus avant que le message "*FIN*" n'ait été donné. Cette deuxième analyse nous semble nécessaire car nous nous sommes rendus compte que certains étudiants sourds ne connaissaient pas des mots comme "soucoupe volante" ou "extra-terrestre".

La liste des termes indispensables à une relation minimale de l'histoire a été établie en étroite collaboration avec les autres chercheurs de l'équipe et l'énumération des synonymes a été complétée à l'aide du Petit Robert 1.

2.1.3. Composante lexicale

2.1.3.1. Le dictionnaire

La première partie de ce travail a été la constitution d'un dictionnaire. Celui-ci est formé de 2005 entrées relevées dans le corpus de préexpérimentation. Ces mots se présentent sous une forme minimale (ou neutre : par exemple, tous les adjectifs sont au masculin singulier). Le dictionnaire est complété de façon à contenir tous les éléments lexicaux jugés indispensables à la narration de l'histoire représentée par la bande dessinée.

Les cas d'exception aux règles morphologiques régulières pour la formation du féminin (ajout d'un "e",...), du pluriel (ajout d'un "s",...), et les verbes irréguliers avec leur conjugaison complète entrent également dans le dictionnaire. En outre, on a ajouté certaines entrées lexicales composées : locutions verbales, adjectives, adverbiales, prépositives, conjonctives et interjectives. Au fur et à mesure que le corpus de préexpérimentation s'enrichira, de nouveaux éléments lexicaux viendront s'ajouter au dictionnaire.

Tous les mots entrés dans l'ordinateur sont classés selon le code ASCII (American Standard Code for Information Interchange), ce qui implique que tous les mots comportant des voyelles accentuées soient situés à la fin de cette liste. Ainsi, *échapper* est à la suite de *vouloir*, et de même, *réaction* suit directement *ruser*.

A chaque entrée lexicale, une catégorie grammaticale est assignée selon les ouvrages de référence suivants : le Petit Robert 1, le Multidictionnaire, le Bon Usage et le nouveau Bescherelle. Dans le cas où un élément lexical correspond à plus d'une catégorie lexicale (ex : *devant* : adverbe, préposition, nom masculin singulier et participe présent du verbe *devoir*), on a indiqué autant d'entrées lexicales qu'il y a de catégories grammaticales. L'annexe 5 reproduit le dictionnaire. Dans ce document, les catégories grammaticales réellement rencontrées dans le corpus de préexpérimentation sont en typographie

grasse (exemple : parmi les quatre entrées de *devant*, on trouve “nom masc. sing.”). L'annexe 6 comporte la liste des différentes catégories grammaticales avec les fréquences pour chacune d'entre elles.

2.1.3.2. Modalités d'informatisation du dictionnaire :

Ce programme permettra au linguiste d'encoder de nouvelles entrées lexicales dans le dictionnaire. Cette opération (cf. annexe 7) se fait en cinq étapes :

- on dactylographie au clavier l'entrée lexicale,
- un tableau comportant les 9 grandes catégories grammaticales apparaît alors; une seconde série de tableaux propose les sous-classes des catégories concernées, c'est-à-dire les verbes, adverbess, déterminants, pronoms et locutions. Lorsque la catégorie (et la sous-catégorie) grammaticale est déterminée, un tableau général divisé en 5 parties permet d'apporter à celle-ci des informations complémentaires (le temps, le mode, la personne, le nombre et le genre) afin de définir de façon exhaustive et complète une entrée grammaticale.
- on entre ensuite la référence aux règle(s) morphologique(s) éventuelle(s)
- on entre ensuite la référence à la règle contextuelle éventuelle;
- on signale la référence éventuelle du mot; par exemple, une forme verbale irrégulière nécessite un renvoi à l'infinitif.

2.1.3.3. Règles morphologiques de formation :

Enfin, 13 règles morphologiques (cf. annexe 8) ont permis de définir les diverses formations régulières du féminin des adjectifs qualificatifs et des noms variables, ainsi que la formation régulière du pluriel (6 règles) pour les noms et ces mêmes adjectifs. Pour les règles du féminin, “eux ---> euse” signifie qu'une forme masculine se terminant par “eux” devient “euse” au féminin. Pour les règles du pluriel, “al --> aux” signifie que la forme singulière “al” devient “aux”

au pluriel. Cela permet d'alléger le dictionnaire et de rechercher plus aisément les racines des mots.

2.1.3.4. Règles contextuelles :

On trouvera dans ce dictionnaire la référence aux 33 règles contextuelles (cf. annexe 9); les règles contextuelles ont pour but d'assigner, dans un contexte donné, une seule catégorie grammaticale à des mots qui peuvent en avoir plusieurs. Ces règles ont été constituées à partir des différents contextes du corpus de préexperimentation. Par exemple, "V ---> [v. pron.] / se _____" signifie qu'un verbe est pronominal dans le contexte où il est précédé de "se".

2.1.3.5. Conjugueur

Le programme analysant les formes verbales (cf. annexe 10) a été placé dans la composante lexicale afin que le verbe étudié par le système soit identifié morphologiquement avant d'être analysé du point de vue syntaxique (quatrième composante).

Ce conjugueur se compose d'un fichier des 194 verbes réguliers relevés dans le corpus de préexperimentation. Chaque verbe est suivi de 4 chiffres indépendants, dont voici le détail :

- le premier chiffre réfère au numéro du groupe selon Bescherelle; ainsi, le premier groupe reprend les verbes en *-er*, le deuxième groupe ceux terminés par *-ir* (avec *-issant* au participe présent) et le troisième rassemble tous les autres. Ce premier chiffre ne peut donc être que 1, 2 ou 3.
- le deuxième chiffre renvoie aux verbes types (modèles de conjugaison) selon Bescherelle, qui en comporte 82. Exemple : le verbe *aimer* porte le numéro 6.
- le troisième chiffre rattache le verbe du conjugueur à l'entrée correspondante dans le dictionnaire. Par conséquent, il indique le nombre de lettres à ôter pour retrouver la racine du verbe en

question. Par exemple, pour *fumer*, il faut ôter la terminaison de l'infinitif : *fum*er*.

- le quatrième chiffre établit la correspondance entre la forme verbale employée par l'apprenant et les formes du conjugueur. Par conséquent, il indique le nombre de lettres à ôter pour retrouver la racine modifiée du verbe en question, ceci pour tenir compte des voyelles accentuées et variables dans la conjugaison (exemple : verbes en *-eler/-eter*). En effet, il faut ôter cette voyelle (et tout ce qui la suit) afin de retrouver la racine du verbe dont elle est issue. Donc, pour la forme *lève*, le conjugueur ôte les trois dernières lettres et effectue les recherches à partir de la lettre *l* uniquement.

Un tableau des terminaisons est intégré au conjugueur. Ce tableau à double entrée comporte en lignes (horizontalement) les six personnes de la conjugaison française. En colonnes (verticalement), on trouve les différents temps du verbe dont voici la liste :

- indicatif présent
- indicatif imparfait
- subjonctif présent
- conditionnel présent
- impératif présent (3 personnes), participe présent, participe passé (masculin et féminin) et les deux terminaisons de la troisième personne du passé simple.

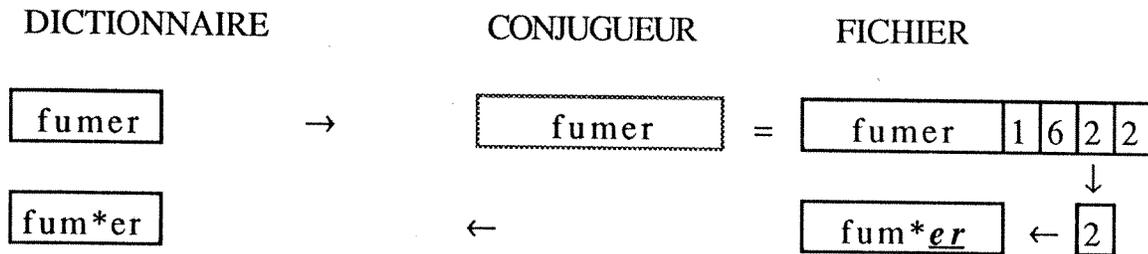
2.1.3.6. Informatisation du conjugueur :

Le conjugueur est un logiciel utilitaire à deux fonctions :

- recherche des racines des verbes;
- recherche de toutes les possibilités d'identification des formes conjuguées que l'apprenant entre au clavier;

- Recherche des racines des verbes

Cette fonction est utilisée lors de la compilation du dictionnaire. En effet, le compilateur trouve les racines de tous les mots, sauf celles des formes verbales. Lorsqu'il rencontre un verbe, le compilateur l'envoie au conjugueur. Ce dernier vérifie la présence du verbe en question dans son propre fichier (cf. annexe 10). Puis, il procède à l'extraction de la racine. Enfin, il renvoie le verbe et la racine au compilateur de dictionnaire. Par exemple :



- Recherche des formes conjuguées

Cette fonction est utilisée avant l'analyse des règles morphologiques de formation. Tous les mots que l'apprenant entre au clavier pouvant être considérés comme des formes verbales, ils sont envoyés au conjugueur afin de vérifier cette possibilité.

Par exemple : "le fermier fume la pipe"

le	fermier	fume	la	pipe
Det	N	V	Det	N

MOT DE L'ÉTUDIANT

DICTIONNAIRE

CATÉGORIE
GRAMMATICALE

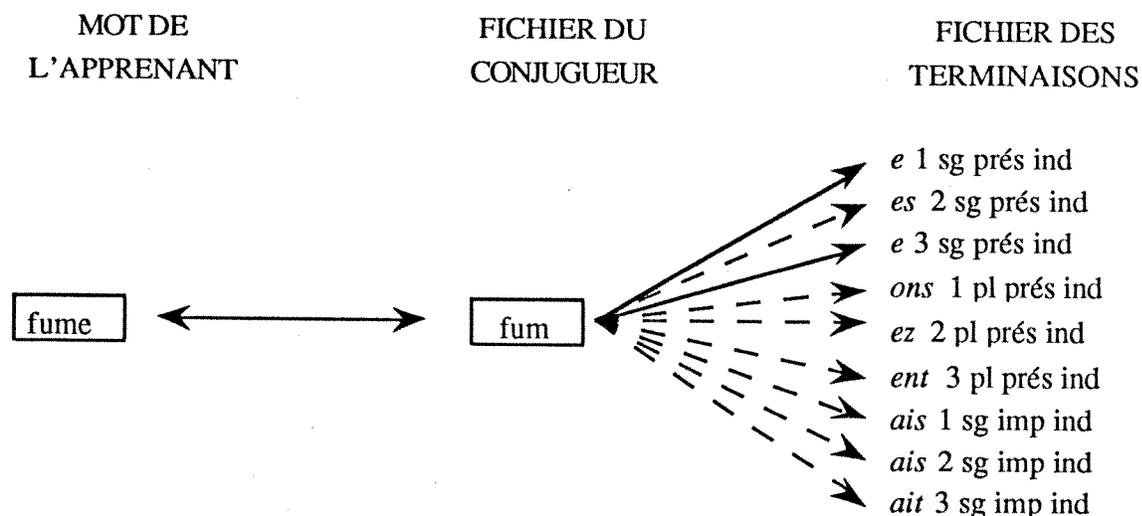
fum*e

↔

fum*er

verbe

On envoie donc le mot au conjugueur.



Les flèches en trait plein indiquent que les terminaisons sont bien identifiées, tandis que le trait pointillé signale que la recherche n'a pas abouti.

Le conjugueur tente d'établir toutes les correspondances avec le verbe de l'apprenant si celui-ci est bien orthographié. Puis, il renvoie la liste des possibilités au module d'analyse des règles morphologiques.

Ainsi dans l'exemple ci-dessus, le logiciel détermine 5 possibilités, soit :

- 1ère personne singulier du présent de l'indicatif;
- 3ème personne singulier du présent de l'indicatif;
- 1ère personne singulier du présent du subjonctif;
- 3ème personne singulier du présent du subjonctif;
- 1ère personne singulier du présent de l'impératif.

2.1.3.7. Traitement informatique de la composante lexicale (cf. annexe 11) :

- Initialisation du système :

Afin de pouvoir analyser une phrase, le prototype initialise tout d'abord des variables de travail. Celles-ci sont des structures logiques qui recevront de l'information provenant du logiciel (la classe des mots, les règles morphologiques attachées à un adjectif ou à un nom quelconque...).

Ensuite, le programme effectue le chargement du dictionnaire. Ce dernier n'est pas sous la forme présentée en annexe 5, car il a été préalablement traduit par un compilateur, ceci pour le rendre plus compact afin de le charger entièrement en mémoire centrale plutôt que d'utiliser un disque physique comme support.

- Recherche dans le dictionnaire :

La période d'initialisation terminée, le logiciel est prêt à travailler. On peut donc entrer une phrase au clavier.

Cette phrase est lue par le logiciel puis chacun des mots sont extraits et emmagasinés dans une structure logique qui servira tout au long du programme.

Exemple :

phrase =

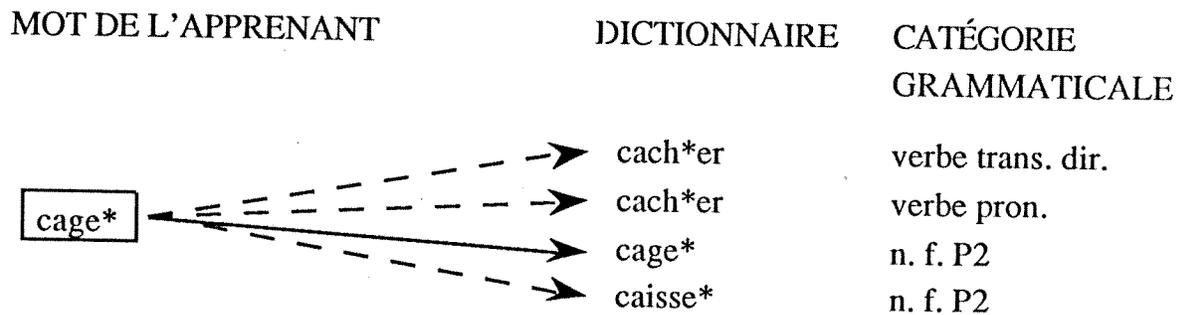
la	petite	cage
----	--------	------

structure =

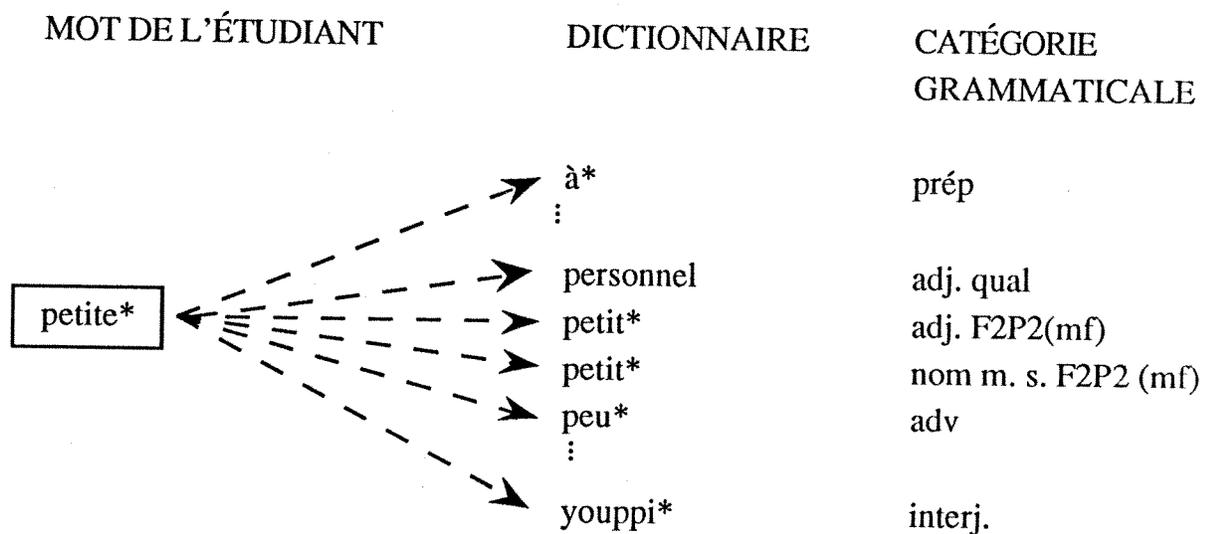
la
petite
cage

Une fois tous les mots emmagasinés dans la structure, chacun est alors décomposé et comparé avec les entrées du dictionnaire. Si ce mot existe déjà tel quel dans le dictionnaire, la recherche s'arrête et l'information est alors stockée en mémoire.

Voici un exemple de fonctionnement :



La comparaison entre le mot de l'apprenant et les mots du dictionnaire se fait lettre par lettre jusque l'astérisque.



Dans ce cas, *petite* est reconstruit et l'orthographe du mot de l'étudiant est valide.

- Règles contextuelles

Quand tous les mots sont validés, il faut désambiguïser les mots (*la* et *petite*) possédant plusieurs catégories.

La est soit déterminant soit pronom. La règle contextuelle n° 16 indique qu'un mot devient pronom s'il se trouve devant un verbe. Or *la* se trouve devant *petite* qui est soit nom soit adjectif. La règle contextuelle permet donc de déterminer que *la* est un déterminant.

A la fin du processus, on trouve donc :

la	dét
petite	adj
cage	nom

2.1.4. Composante morpho-syntaxique

La composante morpho-syntaxique est une mini-grammaire du français (règles morphologiques d'accord, règles syntaxiques et règles de position). Elle doit permettre de rendre compte d'un maximum de phrases grammaticales dans le cadre de l'exercice demandé. De même, les règles devraient pouvoir engendrer un ensemble de phrases correctes proches des énoncés agrammaticaux produits par l'apprenant, en utilisant un maximum de ses propres éléments lexicaux.

2.1.4.1. Règles syntaxiques

Les règles syntaxiques comprennent des règles du type : “une phrase peut être formée d’un groupe nominal suivi d’un groupe verbal.”. L’analyseur syntaxique du prototype identifie les syntagmes nominaux, prépositionnels et adjectivaux correctement structurés dans une phrase simple.

Les règles qui permettent de générer ces syntagmes sont les suivantes :

- Syntagme nominal : SN = (Det) (SAdj) N (SAdj) (SPrép);
SN = Pro;
- Syntagme adjectival : SAdj = (Adv) Adj;
- Syntagme prépositionnel : SPrép = Prép + SN;

2.1.4.2. Informatisation des règles syntaxiques :

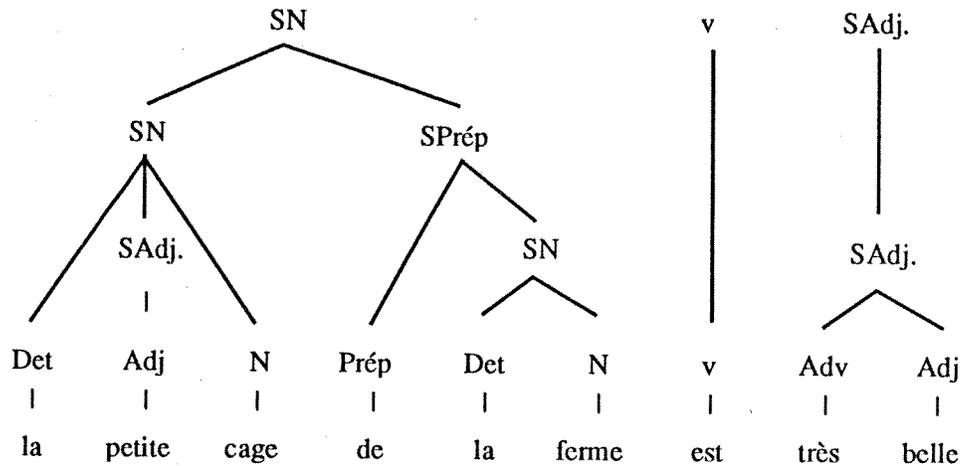
L’analyseur syntaxique (cf. annexe 12) a pour but de déterminer tous les syntagmes dans une phrase et de les analyser afin de détecter les erreurs. L’analyseur du prototype ne traite que les syntagmes simples, de type nominaux, adjectivaux et prépositionnels. Le cas des syntagmes avec conjonction(s) n’est pas encore étudié.

Outre la reconnaissance des trois types de syntagmes, le programme indique s’il y a des erreurs dans la construction. D’autre part, le prototype détecte la présence d’une forme verbale sans toutefois la traiter.

L’analyseur reçoit du module des règles contextuelles un ensemble de mots dont chacun a une étiquette descriptive. Par exemple :

la	petite	cage	de	la	ferme	est	très	belle
Det	Adj	N	Prép	Det	N	V	Adv	Adj

L'analyseur, au début, attend un premier syntagme exclusivement nominal.



Si la phrase ne comporte pas de verbe, le prototype traite quand même l'ensemble des mots, mais il indique un message d'erreur à la fin de l'analyse.

Voici comment le logiciel traite une phrase fautive :

“la chaise de la très maison”

- très → erreur
- la très maison → SN erreur
- de la très maison → SP erreur
- la chaise de la très maison → SN erreur

Voici maintenant comment le logiciel traite la phrase correcte :

“la chaise de la très belle maison”

- très belle → SAdj
- la très belle maison → SN
- de la très belle maison → SP
- la chaise de la très belle maison → SN

2.1.4.3. Règles morphologiques d'accord

Les règles morphologiques comprennent tous les cas d'accord; ainsi, on trouvera dans cette partie des règles du type : "si dans un groupe nominal l'article est au pluriel, alors, le nom doit aussi être au pluriel". Toute cette partie grammaticale reste à faire.

2.1.4.4. Règles de position

On a inséré dans cette partie des règles de position (cf. annexe 13) concernant la place de certaines catégories lexicales : les adjectifs et les pronoms. Ces règles définissent des sous-classes à l'intérieur de ces catégories grammaticales. Ainsi, certains adjectifs se trouvent toujours après le nom (exemples : l'engin spatial, les étoiles brillantes, un homme âgé...) tandis que d'autres doivent impérativement être avant le nom (exemples : un beau tableau, le vieux manteau...).

2.2. MODULE DIAGNOSTIQUE

Dans un contexte d'apprentissage, il est indispensable d'introduire dans le système des règles erronées que l'on sait être courantes chez les apprenants. Le système est ainsi mieux à même de repérer les erreurs de raisonnement. Ce module est donc constitué d'une grammaire des fautes, c'est-à-dire une série de règles rendant compte des erreurs les plus fréquentes que nous avons déjà remarquées dans les corpus de déficients auditifs. Lorsque les erreurs sont identifiées, une explication adaptée est proposée à l'étudiant. Cette partie du tuteur intelligent est une des plus difficiles à réaliser car elle suppose une formalisation des erreurs observées. On trouvera en annexe 14 un texte témoin du corpus de préexpérimentation (cf. annexe 2) dont les fautes ont été identifiées en vue de la catégorisation des erreurs.

2.3. MODULE PÉDAGOGUE

Le module pédagogique interprète les entrées de l'apprenant pour établir son profil pédagogique, pour le solliciter et le guider dans son apprentissage (en lui posant des questions, en lui proposant des objectifs spécifiques à atteindre, en identifiant et en analysant ses erreurs...). Ce module a donc pour mission de contrôler le dialogue avec l'étudiant en utilisant explicitement des règles et stratégies pédagogiques du type suivant :

- si la phrase est grammaticale, on en informe l'étudiant;
- si la phrase est agrammaticale, et que la cause de l'agrammaticalité a été reconnue avec certitude, on propose une correction à l'étudiant en lui expliquant l'erreur commise;
- si la phrase est agrammaticale, et que la cause de l'agrammaticalité n'a pas été reconnue avec certitude, on déclenche une procédure de questionnement (cf. interface).

2.4. PROFIL ÉTUDIANT

Les capacités à juger les connaissances et aptitudes de l'étudiant devront permettre de définir un "profil de l'étudiant", qui servira de paramètre au comportement pédagogique du programme. Ce modèle étudiant sera basé sur une série de savoir-faire (cf. annexe 15) que nous souhaitons développer chez l'apprenant. Au départ, tous les savoir-faire seront à zéro. Ils prendront des valeurs en fonction des actions de l'étudiant au fur et à mesure qu'il utilisera le système. De même, les diverses réactions de l'élève aux stratégies pédagogiques mises en oeuvre seront enregistrées dans ce module. Le profil de l'étudiant se construira donc dynamiquement au cours de la séance de travail.

2.5. DÉROULEMENT DE L'UTILISATION

2.5.1. ÉTUDIANT

Après avoir observé la minibande dessinée, l'étudiant sourd en raconte l'histoire par écrit. L'interface entre en action dès qu'il déclare avoir terminé : message "*FIN*" suivi de RETOUR.

2.5.2. ANALYSE

2.5.2.1. Composante textuelle

Il s'agit dans un premier temps de délimiter les frontières des phrases du texte, c'est-à-dire de voir si ces dernières commencent par une majuscule et finissent par un signe de ponctuation repris dans les délimiteurs de phrase (cf. annexe 3). S'il y a absence complète de ponctuation ou si celle-ci n'est pas cohérente, un message apparaîtra et proposera à l'étudiant de corriger sa version.

2.5.2.2. Composante thématique

Il faut ensuite vérifier la présence des mots-clefs pour s'assurer d'un minimum de compréhension de l'histoire par l'apprenant. Si un de ces termes indispensables (ou un synonyme) est absent, le système questionne l'étudiant pour qu'il modifie son texte.

2.5.2.3. Composante lexicale

La troisième étape est la reconnaissance du vocabulaire utilisé :

- Si un mot n'est pas reconnu, le système en avertit l'apprenant et lui demande de vérifier l'orthographe ou de modifier son texte.
- une ou plusieurs catégories grammaticales sont assignées à chaque mot utilisé par l'apprenant, en fonction des entrées lexicales du dictionnaire, en fonction du conjugueur et des règles morphologiques de formation;
- on applique les règles contextuelles de façon à n'obtenir qu'une seule catégorie grammaticale pour chaque mot.

2.5.2.4. Composante morpho-syntaxique

On passe à l'analyse proprement dite : on reconstruit les différents syntagmes et la phrase de façon à pouvoir juger de sa grammaticalité. On applique ensuite les règles morphologiques d'accord puis les règles de position . Si la phrase est déclarée agrammaticale au niveau de la composante morpho-syntaxique, elle est traitée dans le module diagnostique.

2.6. INTERFACE ÉTUDIANT-MACHINE

Pour effectuer les dialogues qu'il contrôle, le module pédagogue utilise une interface en langue naturelle. Cette interface comprend les messages correspondant à toutes les étapes du déroulement :

- un message d'accueil;
- un ensemble de remarques sur la ponctuation de l'apprenant;
- un ensemble de remarques ayant trait à la la liste restreinte de la composante thématique;
- une banque de commentaires de félicitations (pour les cas où les phrases composées par l'étudiant sont grammaticales);
- une banque de formulations de causes d'agrammaticalités (pour les cas où les phrases comportent une cause d'agrammaticalité prévue par le module diagnostique);
- une procédure de questionnement (pour les cas où la cause d'agrammaticalité n'a pas été reconnue au niveau du module-diagnostic);
- fonction d'aide : tout au long du travail, et déjà pendant la rédaction de l'exercice, une fonction d'aide est aisément accessible à l'apprenant. Pour l'instant, elle comporte un conjugueur, c'est-à-dire les tableaux de conjugaison des verbes réguliers des 3 groupes. Elle comprendra aussi l'accès à la liste de mots du dictionnaire.

BIBLIOGRAPHIE

- Le nouveau Bescherelle, L'art de conjuguer - dictionnaire des huit mille verbes usuels, Hurtubise H. M. H. Ltée, Montréal, 1972.
- DAOUST, F., SATO (Système d'analyse de texte par ordinateur), Manuel de référence, Centre d'analyse de texte par ordinateur, Université du Québec à Montréal, 1986.
- DE VILLERS, M. F., Multidictionnaire des difficultés de la langue française, Éditions Québec/Amérique, Québec, 1988.
- FUSFELD, I.S., "How the deaf communicate - Written Language", American Annals of the deaf, vol 103, n°2, 1958, pp. 255-263.
- GRÉVISSE, M., Le Bon Usage - Grammaire française avec des Remarques sur la langue française d'aujourd'hui, Onzième édition revue, Duculot, Paris-Gembloux, 1980.
- HEIDER, F.K., HEIDER, G.M., "A comparaison of sentence structure of deaf and hearing children", Psychological Monographs, Vol. 52, n°1, 1940, pp. 42-103.
- MARSHALL, W., QUIGLEY, S., Quantitative and qualitative analysis of syntactic structure in the written langage, Institute of research on exceptional children , University of Illinois, 1970.
- MYKLEBUST, H. R., The psychology of deafness, Grune and Statton, N. Y., 1964.
- PERRY, F. R., " The psycholinguistic abilities of deaf children : An exploratory investigation - II, The Australian teacher of the deaf, 1968, pp. 153-160.

POWER, D., Deaf children's acquisition of the passive voice, University of Illinois at Urbana-Champaign, 1971.

ROBERT, P., Le Petit Robert 1 - Dictionnaire alphabétique et analogique de la langue française, Paris-Montréal, 1990.

SCHLESINGER, I. M., "The grammar of sign language and the problems of language universals", dans Morton, John, Biological and Social Factors in psycholinguistics, Logos, Leicester, 1971, pp. 98-121.

SCHULZE, G., "An evaluation of vocabulary development by thirty-two deaf children over a three year period", American Annals of the deaf, vol. 110, pp. 424-435.

TEMPLIN, M. C., "Vocabulary problems of the deaf child", International Audiology, vol. 5, pp. 349-354.

THOMPSON, H., "An analysis of errors in written composition by deaf children", American Annals of the deaf, vol. 81, n°2, 1936, pp. 95-99.

TABLE DES MATIERES

INTRODUCTION

1. PROBLÉMATIQUE, FINALITÉ ET DESCRIPTION GÉNÉRALE

1.1. Problématique	2
1.2. Finalité	3
1.3. Description générale du logiciel	3
1.3.1. Base de connaissances	5
1.3.2. Module diagnostique	5
1.3.3. Module pédagogue	5
1.3.4. Profil de l'étudiant	6
1.3.5. Interface en langue naturelle	6
1.4. Préexpérimentation	6

2. DESCRIPTION DÉTAILLÉE DU LOGICIEL

2.1. Base de connaissances	
2.1.1. Composante textuelle	7
2.1.2. Composante thématique	8
2.1.3. Composante lexicale	
2.1.3.1. Le dictionnaire	9
2.1.3.2. Modalités d'informatisation du dictionnaire	10
2.1.3.3. Règles morphologiques de formation	10
2.1.3.4. Règles contextuelles	11
2.1.3.5. Conjugueur	11
2.1.3.6. Informatisation du conjugueur	12
2.1.3.7. Traitement informatique de la composante lexicale	15

2.1.4. Composante morpho-syntaxique	
2.1.4.1. Règles syntaxiques	19
2.1.4.2. Informatisation des règles syntaxiques	19
2.1.4.3. Règles morphologiques d'accord	21
2.1.4.4. Règles de position	21
2.2. Module diagnostique	22
2.3. Module pédagogique	22
2.4. Profil étudiant	23
2.5. Déroulement de l'utilisation	
2.5.1. Étudiant	23
2.5.2. Analyse	
2.5.2.1. Composante textuelle	23
2.5.2.2. Composante thématique	23
2.5.2.3. Composante lexicale	24
2.5.2.4. Composante morpho-syntaxique	24
2.6. Interface étudiant-machine	25
BIBLIOGRAPHIE	26

