

## HECTOR. PROBLÈMES LEXICAUX

### Résumé

Dans les limites techniques imposées par la téléthèse de communication Hector, et en tenant compte de la spécificité du public visé, nous avons tenté de constituer une base lexicale à la fois économique et performante, destinée à entrer dans la mémoire de la machine.

Au cours de ce travail, nous avons abordé les étapes suivantes:

1. Elaboration et constitution d'une base de données représentatives d'une certaine population (classe d'âge, niveau de langue, situations vécues).
2. Critères de sélection scientifiquement acceptables.
3. Création d'un lexique de 500 mots environ, à partir de la base initiale et en tenant compte d'autres facteurs; description du corpus.
4. Présentation graphique du lexique.
5. Accès au lexique et mise en oeuvre.
6. Autres problèmes à résoudre.

Le choix a été opéré en tenant compte du critère de répartition, complété par les apports des chercheurs du Centre de linguistique appliquée. La présentation des mots retenus (par ordre alphabétique et en fonction des catégories grammaticales traditionnelles) n'est que partiellement satisfaisante. Des progrès sont néanmoins envisageables; certains sont déjà réalisés ou en voie de réalisation dans la nouvelle version d'Hector.

\* \* \*

Im Rahmen der technischen Grenzen der Kommunikationstelethese "Hector" und unter Berücksichtigung der Spezifität des anvisierten Zielpublikums haben wir uns vorgenommen, eine für den Speicher der Telethese bestimmte Wortschatzbasis zu erstellen, die zugleich ökonomisch und leistungsfähig ist.

Im Verlauf dieser Arbeit haben wir uns mit den folgenden Bereichen auseinandergesetzt:

1. Erstellung einer für ein bestimmtes Zielpublikum (Alter, Sprachvermögen, Erfahrungshorizont) repräsentativen Datenvorlage.
2. Aufstellung von wissenschaftlich akzeptablen Selektionskriterien.
3. Auswahl einer aus dem Korpus sortierten und zusätzliche Kriterien mitberücksichtigenden Lexik von etwa 500 Einheiten; Beschreibung des Korpus.
4. Graphische Darstellung der Lexik.
5. Zugang und Inbetriebnahme der Lexik.
6. Weitere lösungsbedürftige Probleme.

Die Auswahl der Lexik wurde unter Berücksichtigung des Kriteriums der Verteilung - ergänzt durch Beiträge des Forscherteams des CLA - vorgenommen. Die vorliegende Auflistung der selektierten Lexik nach alphabetischer Reihenfolge und nach den herkömmlichen grammatischen Kategorien ist nur teilweise befriedigend. Verbesserungen stehen jedoch in Aussicht, von denen einige in der letzten Auflage Hectors bereits verwirklicht worden sind.

\* \* \*

### Introduction

Rappelons, avant toute chose, que la téléthèse Hector, dans sa version la plus répandue à l'heure actuelle, et qui nous intéresse seule ici, dispose de trois niveaux de communication:

Le premier niveau est constitué de 140 cellules pouvant contenir chacune un mot ou un message accessibles à l'aide d'une touche.

Le deuxième niveau, d'un accès moins rapide, offre une capacité plus grande: 900 mots que l'utilisateur atteint en recourant à un code de 3 couleurs ou de 3 chiffres.

De fait, les mémoires liées à ces deux niveaux sont théoriquement vides au moment où un Hector sort de fabrication; un lexique standard existe néanmoins (cf. infra).

Le troisième niveau est comparable à une machine à écrire parlante. Son emploi exige une certaine habileté manuelle (dactylographie). Le choix du lexique et la composition des messages sont donc entièrement laissés à l'initiative de l'utilisateur.

En ce qui concerne les utilisateurs, il convient de préciser qu'Hector, instrument de substitution à la parole, est destiné en priorité à des enfants handicapés d'âges, de formation, de niveau mental et de capacités physiques ou intellectuelles très différents. Pour chaque cas individuel, il faudrait procéder à une analyse des besoins, et constituer une téléthèse "ad personam", en tenant compte également de la motivation plus ou moins forte des sujets handicapés à s'exprimer. Une utilisation par des adultes devenus aphasiques, par exemple, poserait d'autres problèmes tout aussi redoutables, mais que nous n'avons pas abordés ici.

### Mandat

Les premiers Hector mis à disposition par la Fondation suisse pour les téléthèses ont été fournis avec un vocabulaire "standard" de l'ordre de 350 mots, sélectionnés de façon quelque peu empirique. Il a donc paru intéressant au Centre de linguis-

tique appliquée, par mandat de J.-Cl. Gabus, créateur d'Hector, de reprendre cette question et de créer un lexique de base limité à quelques centaines de mots suffisamment "utiles" pour répondre, grâce à leurs capacités de combinaisons, aux besoins des utilisateurs dans un aussi grand nombre de situations que possible.

#### 1. Elaboration et constitution d'une base de données

Déterminer ce contenu idéal se heurte à une première contrainte, purement matérielle, due à la relativement petite taille de la mémoire d'Hector, qui limite considérablement sa capacité. Par rapport à un dictionnaire comme "Le Petit Robert", les 900 mots admis au niveau 2 (auquel nous nous sommes attachés en priorité), paraissent peu de chose. Pour notre part, nous avons prévu de nous en tenir à 500, de façon à laisser aux usagers la possibilité d'enrichir cette banque de mots par des adjonctions ultérieures.

Quoi qu'il en soit, tout choix lexical présente des difficultés réelles, et les auteurs de méthodes de langues ou de dictionnaires s'y trouvent fatalement confrontés.

Mais le problème, dans le cas d'Hector, se pose en termes différents. En effet, il ne s'agit pas de donner des définitions de mots, mais de procurer aux handicapés (que l'on peut qualifier, jusqu'à un certain point, d'apprenants), un répertoire lexical correspondant à celui qu'ils utiliseraient s'ils en avaient les capacités physiques. Le choix doit donc être centré sur des usagers qui, contrairement à une personne étudiant une langue étrangère, connaissent le maniement de la langue et le sens des mots, mais avec de fortes variations individuelles, puisque certains de ces enfants ne savent même pas lire. La nécessité n'en existe pas moins de donner à ce public les matériaux nécessaires à assurer une communication orale

minimale d'abord, mais perfectible, pour l'aider à sortir de l'isolement où le maintient son handicap.

Devant la masse lexicale énorme que représente l'ensemble des mots d'une langue, l'un des premiers critères paraissant à même d'opérer une sélection est celui de l'utilité. D'où l'idée que les mots utiles sont nécessairement fréquents dans l'usage de la langue, et que les mots fréquents présentent à coup sûr un caractère d'utilité. Dans leur enquête sur le français fondamental, Gougenheim, Rivenc, Michéa et Sauvageot ont recouru à ce critère de fréquence pour établir leur liste.<sup>1</sup> 163 conversations enregistrées à cette fin ont été transcrites et dépouillées intégralement. 312.000 mots ont été recueillis, dont 8000 différents. Une liste de fréquences a été établie, et les auteurs ont retenu les mots dont la fréquence était égale ou supérieure à 29.

Les résultats de cette sélection sont connus: le verbe "être" apparaît à 14083 reprises dans les 163 conversations enregistrées, puis le verbe "avoir", puis des mots grammaticaux (de, je, il, ce, ...etc.). Le premier substantif ne figuré qu'au 82ème rang. C'est dire que les mots jugés utiles dans la langue (les mots concrets), n'ont qu'une fréquence inférieure à celle des mots grammaticaux, et beaucoup plus instable. Ils n'apparaissent que dans des circonstances particulières. Pour obtenir ces mots concrets, les chercheurs français ont recouru au système des "centres d'intérêt", à propos desquels les personnes interrogées devaient donner les termes qui leur venaient à l'esprit. On les considère comme disponibles "parce que, même si nous n'avons pas souvent l'occasion de les prononcer, nous les avons toujours à disposition."<sup>2</sup> Ainsi donc, la liste du FF1 comporte deux éléments: les mots fréquents (grammaticaux), qui correspondent aux automatismes du langage, et les mots disponibles, liés à l'intérêt que présentent les notions

exprimées par ces mots.

Dans le cas d'Hector, nous nous sommes heurtés aux mêmes difficultés: comment échapper à l'arbitraire et sélectionner, sur des bases acceptables, quelque 500 mots comprenant un dosage satisfaisant de substantifs, de verbes, d'adjectifs, d'adverbes, de pronoms, de prépositions, de conjonctions et d'interjections sans perdre de vue le public cible et ses spécificités ?

Nous nous sommes partiellement inspirés de l'enquête sur le français fondamental, qui reste exemplaire à plus d'un titre, mais n'est pas exempte de défauts. En effet, les mots qui constituent cette liste apparaissent hors contexte, et la méthode des "centres d'intérêt" permet de multiplier le nombre de mots disponibles, sans se soucier des besoins langagiers des personnes appelées à les utiliser. Les limites d'Hector nous ont contraints à une sélection sévère. Nous avons donc tenté de prendre en compte les mots réellement employés par des enfants francophones, et notamment des enfants handicapés déjà utilisateurs d'Hector, dans des situations de communication authentiques.<sup>3</sup>

Notre base initiale de données se compose de 19 listes de mots représentatifs d'un niveau de langue "standard":

- 350 mots tirés du Français fondamental 1er degré
- le dépouillement d'un livre de lecture composé par des enfants de 6 à 7 ans à l'école primaire<sup>4</sup>: 1234 mots différents dont 689 substantifs, 327 verbes et 128 adjectifs
- la liste résultant d'une enquête sur l'acquisition du langage menée auprès d'enfants de 3 à 4 ans (2 garçonnets et 2 fillettes)<sup>5</sup> soit 514 substantifs, 289 verbes et 79 adjectifs différents

- le vocabulaire standard d'Hector (353 mots au total)
- le relevé de 12 listes extraites de cellules d'Hector utilisées dans divers centres (mots ajoutés à la liste standard).

Soit au total, quelque 1400 mots différents (fréquences non cumulées) qui ont été introduits dans une base de données informatisée, en fonction des catégories grammaticales traditionnelles: substantifs, verbes, adjectifs, ...etc., à l'exclusion des noms propres.

## 2. Critères de sélection

Nous avons tenu compte, à titre principal, du critère de répartition: tout mot apparaissant dans 5 listes au moins sur les 19 constituant notre base de données a été retenu. En effet, mieux que la fréquence (cf. supra), le critère de répartition nous éclaire sur les besoins et les intérêts réels des enfants, grâce à leur caractère de généralité. Le fait qu'un même mot apparaisse très souvent dans une situation donnée (par exemple le mot "dent" lors d'une leçon d'anatomie buccale), ne signifie pas que ce terme soit utile dans d'autres circonstances.

Sur cette base, nous avons pu élaborer un noyau lexical de 239 mots<sup>6</sup>, auxquels nous avons appliqué deux critères supplémentaires:

### l'efficacité et la productivité

1. Efficacité: à égalité d'occurrence, un mot ayant de nombreux dérivés et composés paraît plus important à retenir qu'un mot n'ayant ni dérivé, ni composé (par ex. la suite: changer, changement, change, échanger, est plus riche que le verbe "oindre"). Dans cette perspective, l'idée surgit immédiatement de recourir à la préfixation et à la suffixation pour obtenir des composés ou des dérivés (ainsi le suffixe -ment

permettrait de générer un nombre important d'adverbes de manière, généralement à partir du féminin de l'adjectif).

2. Productivité: comme les éléments chimiques, certains mots peuvent se lier à de nombreux autres pour former des pré-énoncés, alors que d'autres ne présentent qu'un petit nombre d'enchaînements possibles. R. Galisson<sup>7</sup> a introduit les notions de collocatif limité (ainsi le verbe "fumer" ne peut guère fonctionner qu'avec pipe, cigarette, cigare ou cheminée), et de collocatif illimité (avoir, être, acheter, vendre...). A partir de l'inventaire de Galisson, nous avons essayé d'élaborer les syntagmes possibles en recherchant les cooccurrents à même de se greffer sur les termes-noyaux de notre liste, dans le but de créer des amorces d'actes de langage.

Les résultats de ce travail se sont révélés décevants: d'une part, beaucoup de mots utilisés par Galisson dans son inventaire (à partir du Français fondamental 1er degré), font défaut dans notre propre liste (il y a dans le FFl 4 fois plus de substantifs, 3 fois plus de verbes et 2 fois plus d'adjectifs).

D'autre part, 32 mots retenus par nous n'existent pas dans le FFl. Il s'agit de termes spécifiques liés à l'enfant, à l'institution, à la thérapie, etc.

Notre base, enfin, présente des lacunes internes dues en partie à notre mode de sélection. Certaines séries, par exemple, sont incomplètes jours de la semaine, mois de l'année, ...etc.). Plusieurs antonymes font défaut: "homme" apparaît, mais pas "femme" (moins de 5 occurrences; "gauche" a été retenu, et pas "droite" ...etc. Nous avons donc complété notre liste de base (239 mots sélectionnés par le critère de répartition), et nous y avons ajouté les éléments jugés indispensables dans la perspective d'un emploi par des enfants IMC. Nous avons abouti à un corpus de 585 mots et expressions différents, totalisant 700

entrées environ (l'adjectif "vieux", par exemple, est considéré comme une seule unité lexicale, mais il figure également sous la forme "vieille", soit au total 2 entrées).

3. Description du corpus<sup>8</sup>

1. VERBES (au nombre de 62)

Outre l'infinitif, qui figure dans le lexique primitif d'Hector, nous avons introduit une, deux ou trois formes essentielles, de manière à offrir une plus grande souplesse d'emploi (critère de productivité), en fonction du groupe auquel appartiennent ces verbes.

A) Verbes réguliers en -er

une forme complémentaire est ajoutée à l'infinitif:

chercher - cherche  
donner - donne

Ces deux formes permettent, sur le plan de la langue orale, les combinaisons suivantes:

- Indicatif présent: je/tu/il/elle/on cherche  
ils/elles cherche(nt)
- Impératif: cherche !
- Subjonctif présent: (il faut que) je/tu/il/elle/on/ils/elles  
cherche(nt)
- Infinitif: chercher; il faut chercher; je voudrais  
chercher...
- Indicatif présent: vous cherchez
- Impératif: cherchez !
- Passé composé: j'ai, tu as, il a cherché ...

B) Verbes réguliers en -ir, -oir, -re

une, deux ou trois formes sont ajoutées à l'infinitif:

finir - fini -  
recevoir - reçois - reçu  
attendre - attends - attendu

C) Verbes irréguliers

avoir - ai - a - avez - eu  
être - suis - est - êtes - été

Soulignons que les critères orthographiques ne peuvent jouer aucun rôle ici.

2. SUBSTANTIFS (au nombre de 239)

Les substantifs ont été regroupés en champs sémantiques, pour éviter une dispersion trop grande. Les catégories suivantes ont été introduites: les repas, les meubles, la maison, le lieu d'habitation, les vêtements, le corps humain, la famille, les relations sociales, l'hygiène, les moyens de transport, l'école, les animaux, la nature, les loisirs, la vie en institution. Au sein de ces groupes, nous avons été amenés à ajouter des termes écartés au moment du dépouillement, ou inexistantes mais cependant jugés indispensables (vocabulaire de disponibilité).

Il est intéressant de noter le caractère enfantin de certains substantifs figurant dans le corpus: le loup, le hamster, le lapin, les légos, la poupée...

La série "vie en institution" est fortement marquée par les utilisateurs d'Hector: le centre IMC, la chaise roulante, l'ergo-thérapeute, la licorne, la mentonnière, etc.

C) ADJECTIFS (au nombre de 46)

Lorsque le féminin d'un des adjectifs retenus présente une prononciation nettement différente du masculin, on l'a donné, de manière à rendre le message aussi acceptable que possible:

blanc	-	blanche
plein	-	pleine
		...etc.

De plus, les antonymes ont été pris en compte et présentés en face l'un de l'autre:

chaud/e	-	froid/e
pauvre	-	riche
		...etc.

D) ADVERBES (au nombre de 68)

Les adverbes ont été sélectionnés par catégorie: affirmation, doute, lieu, manière, négation, qualité, temps.

E) CONJONCTIONS ET PREPOSITIONS (39) - MOTS-OUTILS (46)

Comme dans les catégories précédentes, le critère de répartition a introduit des lacunes que nous avons cherché à combler. Il en est ainsi pour les adjectifs démonstratifs ou possessifs qui n'apparaissaient dans notre premier dépouillement qu'au masculin ou au pluriel, ou pour les pronoms personnels, regroupés avec les articles sous le terme de "mots-outils", qui n'ont du reste été que très partiellement relevés dans nos 19 corpus. Certaines conjonctions ou prépositions ont été ajoutées à la liste, en raison de leur importance dans la communication.

F) CHIFFRES (au nombre de 23)

Les chiffres sont traités comme des substantifs, et sont accessibles selon le même code. Les adjectifs numériques cardinaux ont été donnés de 1 à 20, ainsi que 50, 100 et 1000.

G) DEBUTS DE PHRASES (36, plus 6 formes négatives)

Pour faciliter une expression plus aisée et surtout plus rapide, nous avons introduit un certain nombre de "début de phrases" qui, employés avec les formes verbales figurant dans la liste, devraient encourager l'utilisateur à composer des phrases relativement élaborées sans trop d'efforts de manipulation. Ainsi l'expression:

j'ai envie de + l'infinitif "manger" = j'ai envie de manger  
 " + le substantif "gâteau" = j'ai envie de gâteau  
 " + infinitif + substantif = j'ai envie de manger du gâteau.

H) ACTES DE LANGAGE (36, plus 6 formes négatives)

On a vu que les cellules du niveau 1 d'Hector peuvent contenir des messages plus ou moins complexes. Il nous a paru judicieux de tenir compte de ces éléments authentiques utilisés par les Hectoriens soit pour entrer en communication:

Bonjour !  
 Comment tu t'appelles ?

soit pour exprimer des opinions:

Ce n'est pas difficile !  
 D'accord !

soit pour exprimer des demandes d'aide:

Il faut arranger ma position !  
 Les toilettes s'il vous plaît !

soit pour exprimer des sensations:

J'ai soif ! je ne me sens pas bien !

soit pour exprimer d'autres formules courantes:

Quelle heure est-il ?

Un instant s.v.p. !

La localisation de ces actes de langage au niveau 1 ou 2 d'Hector dépendra du nouveau software à l'étude actuellement.

#### Remarques

Avec ses 585 mots différents, notre liste se veut à la fois économique, suffisante et efficace. Mais ces exigences, qui semblent se compléter, ne sont pas exemptes de contradictions.

Certes, le répertoire proposé est économique (à la fois par le nombre et le caractère "passe-partout"); mais ses limites mêmes impliquent des lacunes. Aucun usager d'Hector ne pourra à long terme se contenter du choix mis à sa disposition. Certes, par rapport au noyau initial de 239 mots sur lesquels nous nous sommes appuyés, le progrès est évident. Mais les aménagements apportés sont en grande partie fondés sur le jugement des auteurs ou sur leur bon sens. D'où un caractère de subjectivité qui remet en cause notre volonté de recourir à des critères "scientifiques" de sélection. Il est évident aussi que, dans des versions d'Hector dotées d'une mémoire plus généreuse au niveau 2 (ce qui est envisageable), la masse lexicale pourrait être considérablement augmentée. Mais une telle mesure entraînerait fatalement des difficultés de consultation (cf. infra). Il nous semble que le principe d'économie adopté est plus efficace qu'un enrichissement excessif, puisque le passage au niveau 3 autorise l'emploi de n'importe quel mot. De plus, nous imaginons qu'un nombre relativement limité d'entrées favorise la mémorisation des codes et, par conséquent, un emploi plus aisé.<sup>9</sup>

Sur le plan de l'efficacité, enfin, l'introduction de quelques formes verbales "conjuguées", de féminin et de pluriels d'adjectifs ou de pronoms devrait faciliter une expression grammaticalement acceptable, donc plus compréhensible par les interlocuteurs, souvent gênés par les caractéristiques de la voix synthétique d'Hector.

#### 4. Présentation graphique du lexique

Le répertoire lexical préparé par les membres du Centre de linguistique appliquée pour Hector se présente sous deux formes: alphabétique et "grammaticale".

##### 1. Liste alphabétique

Cette première liste, fondée sur l'ordre alphabétique, donne pêle-mêle verbes, substantifs, adjectifs, ...etc. Chacun de ces mots porte obligatoirement un code de trois chiffres, allant de 100 à 999. Les formes verbales, (cf. supra) font l'objet d'une entrée séparée, avec rappel de toutes les autres:

111	aller	(vais 953/ va 950)
953	vais	(aller 111/ va 950)
950	va	(aller 111/ vais 953)

L'utilisateur, en fonction de ses besoins d'expression, peut donc retrouver assez facilement la forme qu'il souhaite utiliser, sans passer nécessairement par l'infinitif. Les pronoms qui permettent de faire fonctionner les formes verbales sont données indépendamment. Mais toute une série d'expressions sont offertes sous le pronom "il" qui sert d'entrée: il faudrait; il me faut du temps ...etc.) ou "je": j'ai faim; j'ai mal; je m'appelle; je ne sais pas ...etc.).

Il en va de même pour les adjectifs qualificatifs dont le féminin présente une prononciation différente de celle du masculin:

588 long (longue 590)

590 longue (long 588)

pour les adjectifs démonstratifs:

244 ce (cet 254/ ces 250/ cette 255)...etc.

pour les adjectifs possessifs:

600 ma (mon 642/ mes 631) ...etc.

et pour les articles:

560 l' (la 561/ le 571/ les 577) ...etc.

Les substantifs sont donnés, sauf exception, au singulier, sans article et sans indication de genre, de façon à faciliter la construction de phrases correctes du point de vue grammatical.

Une telle présentation implique que les utilisateurs d'Hector sachent lire, ce qui n'est pas toujours le cas, et qu'ils soient suffisamment habiles pour tourner les pages du dossier lorsqu'ils cherchent des mots ou des expressions. Sur ce point, les difficultés gestuelles que nous avons pu observer chez des enfants IMC nous amènent à penser que ce type de liste n'est probablement pas l'idéal (sous sa forme actuelle, 13 pages dactylographiées; caractères de dimension moyenne).<sup>10</sup>

D'un autre côté, il faut admettre aussi que l'ordre alphabétique est difficilement maîtrisable par les enfants en général. Chercher un mot selon ce critère nécessite une connaissance approfondie de l'orthographe: pourquoi centime s'écrit-il avec c, et sous avec s? Pourquoi le mot août figure-t-il sous a, après année et avant appareil de photo, alors qu'on devrait le trouver sous ou? D'où l'idée d'une entrée phonologique ou syllabique.<sup>11</sup>

Pour faciliter, dans la mesure du possible, le repérage

du vocabulaire, les mots ont été regroupés sous forme de "tranches" réparties de la façon suivante:

les mots commençant par

A	se trouvent entre	100 et 199 (page 1)
B	"	200 et 229 (page 2)
C	"	230 et 299 (pages 2 et 3)
D	"	300 et 349 (pages 3 et 4)
		...etc.

Ce système permet aux usagers de s'y retrouver plus aisément et d'accéder plus rapidement aux termes qu'ils cherchent puisqu'ils peuvent en gros les situer grâce à leur lettre initiale.

On constatera, de plus, que, en réalité, les mots commençant par A, par exemple, n'occupent que les positions 100 à 156. De 157 à 199, il est donc possible d'introduire d'autres termes commençant par A (mais l'ordre alphabétique intégral n'est plus respecté).

## 2. Classement "grammatical"

Nous avons adopté les catégories qui figurent généralement dans les manuels de grammaire traditionnels, non sans prendre certaines libertés: verbes; substantifs (regroupés en "champs sémantiques"); adjectifs; adverbes; conjonctions et prépositions; mots-outils; chiffres (adjectifs numériques cardinaux), catégories auxquelles nous avons ajouté deux rubriques: les débuts de phrases et les actes de langage.

Au sein de chacune de ces rubriques, l'ordre alphabétique est respecté. On postule donc de la part de l'utilisateur à la fois des compétences orthographiques et alphabétiques, et une capacité de restructuration à partir de catégories plus ou moins arbitraires.

Les numéros d'accès sont les mêmes que ceux de la liste alphabétique. Il sera intéressant d'observer à laquelle de ces deux listes les Hectoriens donneront la préférence.

5. Accès au lexique et mise en oeuvre

Pour accéder au lexique du niveau 2, qui nous intéresse ici, on a vu qu'il est indispensable de passer par un code, formé de 3 chiffres ou, à la rigueur, de 3 couleurs. Pour les enfants qui ne savent pas lire, des symboles remplacent les mots écrits (système Bliss). Quoi qu'il en soit, le processus est lent. C'est ainsi que pour exprimer une phrase comme:

"je (518) cherche (265) mon (642) crayon (296)" douze manipulations sont nécessaires. Pour un sujet dont la motricité est réduite, l'effort est considérable, voire excessif.

Des perfectionnements sont à l'étude, tendant à simplifier la procédure d'accès et à accélérer les processus. Outre l'entrée phonologique ou syllabique, dont il a déjà été question, on a pensé à l'emploi de 1, 2, 3 ou X lettres-clés, accessibles à l'aide de 1, 2, 3 ou X touches:

Ainsi	<u>trn</u>	pourrait donner accès à	train
	<u>trm</u>	"	tram
	<u>trv</u>	"	travail
	<u>trvr</u>	"	travailler

Notre collègue Grosjean a imaginé un code restreint composé de groupes de deux lettres dont la combinaison permet de composer des phrases plus ou moins complexes avec une relative facilité:

MA = maman FA = fait RE = le repas DU = du soir

Dans ce cas aussi, dont l'arbitraire n'est pas absent, l'utilisateur devrait disposer d'une liste codée, lui donnant des lettres-clés correspondant non seulement à des mots isolés, mais

à des substantifs accompagnés d'un article, voire d'expressions plus complexes:

BE = j'ai besoin de ES = est-ce que tu (essequetu)  
JV = je ne veux pas (jveupa) VE = la vérité DI = dire  
...etc.

Une autre solution consisterait, en prenant le problème en sens inverse, à faire afficher à la machine (munie d'un écran plus grand et plus lisible) des "menus" d'un accès rapide, au sein desquels l'utilisateur désignerait le mot ou l'expression qui lui convient à l'aide d'un crayon lumineux, d'un système tactile ou de tout autre procédé. La machine prononcerait immédiatement le terme sélectionné, ou le mettrait en mémoire pour l'exprimer au sein d'une phrase entièrement constituée, au gré de l'Hectorien.

Il n'en demeure pas moins que, à l'heure actuelle, notre lexique est comparable à un annuaire téléphonique. Les mots imprimés sont inertes, et il faut impérativement passer par un code pour les actualiser. Mais, par eux-mêmes, ils ne constituent pas des phrases, pas plus que la juxtaposition de cellules mortes ne crée la vie. On offre à l'enfant handicapé un choix de matériaux bruts, mais aucune structure dans laquelle ces mots agencés entre eux deviendraient porteurs d'un sens clair et univoque.

Cela signifie qu'une enquête lexicale, aussi poussée soit-elle, ne suffit pas. Les auteurs du Français fondamental ne s'y sont pas trompés, puisqu'ils ont complété leur liste de mots par un appendice grammatical<sup>12</sup> et quelques conseils sur les éléments à enseigner dans ce domaine. Mais, dans le cas d'Hector, il ne s'agit pas de recenser ou d'enseigner des règles de grammaire, mais de les mettre en oeuvre. Or, dès le départ, l'utilisateur se heurte à des difficultés causées par la forme des mots proposés. On se trouve entre l'emploi d'énoncés tout faits (par ex.

"Comment ça va ?" (niveau 1) et d'énoncés à construire complètement (niveau 3), aussi bien sur le plan lexical que syntaxique.

Dans sa version standard de première génération, Hector permettait de composer, au niveau 2, des phrases du genre:

"Moi (je) vouloir la boîte blanc"

Il serait souhaitable que l'ordinateur équipant Hector soit capable de produire des formes qui tiennent compte de leur environnement, et que la présence du "je" dans l'exemple ci-dessus entraîne la forme "veux" et que l'adjectif "blanc" prenne la marque du féminin, en raison de la présence de l'article "la".

De fait, l'introduction de sous-programmes de redressement syntaxique représenterait un progrès considérable et serait de nature à améliorer le fonctionnement de la communication, sans parler de l'agrément de l'utilisateur, dont la production est actuellement ralentie par le nombre d'opérations syntaxiques plus ou moins complexes qu'il doit prendre en charge.

Par la suite, les progrès de l'intelligence artificielle aidant, il devrait être possible à la machine de générer des phrases syntaxiquement tout à fait correctes, à l'aide des éléments choisis par la personne handicapée dans les "menus" (qui restent à préparer) qui lui seraient proposés.<sup>13</sup>

#### 6. Autres problèmes à résoudre

Hector est, par définition, une téléthèse de communication orale. Mais le fait que les mots à prononcer s'affichent sur un écran rend le système ambigu et d'un maniement délicat, puisque les règles de graphie-phonie de la machine ne correspondent pas entièrement à celles du français parlé et écrit. On a par exemple dû introduire l'adverbe "plus" sous deux formes (plu/plus), de façon qu'Hector établisse la distinction

entre "il n'est plus malade" et "je voudrais plus à boire".

Dans le cas des verbes en -er, notamment, il n'est pas pertinent d'introduire, au niveau de la langue parlée, une distinction entre le singulier et le pluriel. La forme "mange" est polyvalente, et il n'est pas utile qu'apparaissent le -s de la 2e personne du singulier ou le -nt de la 3e du pluriel, au présent de l'indicatif.

Au niveau 2, où nous nous situons, une autre difficulté doit être relevée: celle de la polysémie. Ainsi le participe plu, du verbe pleuvoir, peut être utilisé également comme participe du verbe plaire, ou comme adverbe de quantité ou de négation. Les formes verbales ferme, montre, paye, porte ...etc. sont utilisables comme substantifs. De même, pour exprimer le mot "travail", on recourra à la forme verbale "travaille", qui existe dans la liste.

Ces particularités, gênantes sur le plan de l'écrit, constituent un avantage dans la langue parlée: "...au lieu d'éviter les homonymes, nous nous en servons dans le lexique: près - prêt; sans - sens - s'en - cent. Ce travail élargit encore le vocabulaire et permet de rendre l'enfant attentif aux différentes graphies d'un même son."<sup>14</sup>

Il est du reste intéressant de constater que certains usagers parviennent à former des mots qui ne sont pas dans le lexique avec des mots qui y figurent. Ainsi

chat - tôt = château

chat - ton = chaton

...etc.

Des conventions s'avèrent nécessaires, et on constate que, même au niveau 3, les utilisateurs se font leurs propres règles de fonctionnement et adoptent un système graphique fort éloigné de l'orthographe, mais dont les résultats sont satisfaisants au

niveau de l'expression orale.

Il n'en demeure pas moins que la tentation est grande de focaliser l'attention sur l'écran plus que sur la production orale d'Hector. Il est difficile aux personnes qui dialoguent avec un Hectorien de ne pas "lire" les messages, alors que seule la communication orale devrait entrer en ligne de compte. Sur ce point, la doctrine (s'il en existe une) n'est pas claire, et plusieurs thérapeutes estiment qu'Hector peut favoriser l'apprentissage de la lecture, voire de l'orthographe, en dépit des difficultés engendrées par un tel usage.

Il est enfin permis de se demander si, dans le domaine de l'apprentissage à l'emploi d'Hector<sup>15</sup>, il ne serait pas souhaitable d'inciter les handicapés à se former en partie eux-mêmes. En effet, en l'absence du thérapeute ou de l'interlocuteur nécessaires pour susciter l'échange, on pourrait simuler des situations de communication, à l'aide d'un magnétophone tout d'abord, qui poserait des questions auxquelles l'utilisateur répondrait à l'aide d'Hector. Le système pourrait être complété par des diapositives ou par des images animées (vidéo), voire d'exercices "à trous", à compléter alors par écrit sur un écran de bonnes dimensions (par ex. ordinateur).

Il nous semble que, d'une part, les enfants, stimulés par ce type de travail, acquerraient probablement une habileté plus grande dans la manipulation de la téléthèse, et, d'autre part, enrichiraient leur vocabulaire, grâce à des exercices ad hoc, comme leurs camarades plus privilégiés le font en classe.

Université de Neuchâtel  
Centre de linguistique appliquée  
CH 2000 Neuchâtel

René Jeanneret

#### Notes

1. G. Gougenheim, P. Rivenc, R. Michéa et A. Sauvageot (1964): L'élaboration du Français fondamental, Paris, Didier.
2. Le Français fondamental 1er degré (1954), Paris, Institut pédagogique national, p. 13.
3. Sur l'importance des documents authentiques, cf. R. Galisson (1983): Des mots pour communiquer, Paris, CLE International.
4. Livre de lecture (1969-1970); classe de Mlle Décosterd, Collège de La Maladière, Neuchâtel.
5. F. Redard (1978): La négation chez quatre enfants francophones à l'âge de trois ans, Bref 14, p. 51-59.  
F. Redard (1976): Etudes des formes interrogatives en français chez les enfants de trois ans", Etudes de linguistique appliquée, 21, p. 93-110.
6. A titre indicatif, voici les mots apparaissant dans un maximum de corpus (sur 19): boire (14); aller (13); bonjour (12); maison (11); content (11); avec (11); comment (6).
7. R. Gallisson (1971): Inventaire thématique et syntagmatique du français fondamental coll. Le Français dans le Monde - BELC, Paris, Hachette-Larousse.
8. Centre de linguistique appliquée (1987): Lexique Hector, Université de Neuchâtel.
9. J. Bailod (1986): Résultats d'application d'Hector à La Chaux-de-Fonds, présentés au Congrès de la FST, Neuchâtel, 1986: "Elle a mémorisé 40 à 50 cases du niveau 1 et une dizaine de codes chiffrés" (p. 5).
10. Idem p. 3.: "Ce qui lui est le plus favorable pour trouver des mots dans le lexique c'est d'utiliser le grand panneau affiché dans la classe. Quand il n'y a pas accès, il doit utiliser un cahier aux pages plastifiées et alourdies pour qu'il puisse les tourner; c'est moins facile et moins rapide".
11. cf. F. Redard: Remarques à propos de la relation graphie-phonie dans le message par téléthèse vocale (p. 83 de ce bulletin).
12. cf. note 2.
13. cf. F. Grosjean: Lorsque Hector rencontre un psycholinguiste expérimentaliste... Que se disent-ils ? (p. 13 de ce bulletin).
14. cf. J. Bailod, note 9, p. 3.
15. G. Monod (1986): Pierre - Hector, rapport présenté au Congrès de la FST, Neuchâtel, 1986.