

Idéographix au pays des albums



IDÉOGRAPHIX invite à aller regarder du côté d'une construction lexicale chez l'apprenti qui renvoie à de la morphologie et à de la syntaxe puisqu'il ne s'agit pas d'une accumulation de mots et de formes dérivées mais de l'entrée dans un système. Derrière ce genre d'investigation, des questions concrètes cherchent des réponses pratiques quant au rôle des textes dans l'apprentissage de l'écrit et son enseignement aux cycles 1 et 2. S'intéresser, comme nous allons le faire ici aux deux questions suivantes n'impliquera en aucune manière qu'elles soient les seules, les premières ou les plus importantes :

- quel est le lexique qui se retrouve d'un texte à l'autre et comment s'introduit le vocabulaire nouveau ?
- à partir de quel moment ce qui est connu (mots, dérivation, syntaxe, expérience du texte) donne des solutions par rapport à ce qui ne l'est pas encore ?

L'articulation de ces deux questions suggère déjà que l'accumulation primitive ne saurait se faire, comme autrefois dans l'enseignement des langues étrangères, par la mémorisation de listes de vocabulaire mais à l'intérieur et par l'expérience du fonctionnement même des textes rencontrés en lecture.

Nous nous sommes mis dans la situation d'observer et de comparer 13 albums que les enseignants de la recherche INRP utilisent volontiers aux cycles 1 et 2. Dans une première partie, nous considérerons un texte global regroupant les 13 albums afin de découvrir quel matériau lexicale les

élèves auront rencontré à travers eux. Dans une seconde partie, nous essaierons de comparer ces textes entre eux.

1. LES 13 TEXTES PRIS ENSEMBLE

Les 13 textes mis bout à bout équivalent à un seul texte de 5 527 mots. En moyenne ces mots sont répétés 4,6 fois mais avec des disparités importantes : ainsi 'et' est répété 163 fois ; à l'inverse 673 mots n'apparaissent qu'une fois dans le texte global. On dira qu'il y a 163 occurrences du vocable [et] dans le texte et on pourrait comparer les vocables selon leur fréquence : quels sont ceux qui sont le plus répétés ? pourquoi ? IDÉOGRAPHIX ne procédant pas à une analyse grammaticale n'est pas en mesure de réunir les vocables qui appartiennent à la même souche. Il recense comme des vocables différents toutes les formes que peut revêtir le même « mot » : [chien] et [chiens] sont recensés comme 2 vocables ; [aimes], [aimeraient] et [aimez] comme 3 vocables. Dans un dictionnaire, ces 5 formes correspondraient seulement à 2 entrées¹ : *être, aimer*.

Pour aider à s'y retrouver dans ces conventions d'appellation, convenons qu'un mot dans le texte est l'occurrence d'un vocable dans le lexique, lequel vocable peut être une forme fléchée d'une souche, d'une entrée dans un dictionnaire. Le vocable [et] se confond avec le mot souche *et* ; les vocables [chien], [chiens], [chiennne], [chiennes] activent le mot souche *chien*. En bref, ces 13 albums contiennent 5 527 mots qui sont des occurrences de 1 199 vocables,² lesquels relèvent de 828 souches. À l'arrivée, on dira que ces albums sont écrits avec 828 mots (souches) qui sont répétés en moyenne 6,7 fois. Le logiciel permet d'observer la répartition de ces mots.

1.1. Les « mots » de base

Il existe en français 66 mots (souches) dont les différentes formes fléchies (590 vocables) constituent à elles seules environ 50 % de tout texte écrit d'une certaine longueur. Ces 66 mots sont essentiellement liés à la syntaxe (prépositions, pronoms, conjonctions, etc.) mais on y rencontre aussi quelques noms (homme, femme, enfant, etc.), quelques verbes (être, avoir, pouvoir, aller, faire, etc.) et adverbes (pas,...). Toutes leurs formes ne sont pas également fréquentes ([est] vs [fussions]) mais, au total, elles représentent la moitié statistique des textes écrits.

¹ Ces entrées, que nous appellerons ici souches, sont parfois nommées 'lemmes'.

² Est-ce beaucoup ? Le vocabulaire du Cid (1 825 vers) compte pour sa part 1 518 vocables.

Dans la réunion des 13 albums, on recense 3 200 occurrences de ces mots de base, ce qui constitue plus de 50 % de l'ensemble des mots du texte (57,7 %). Ces 3 200 occurrences correspondent à 179 vocables différents (soit 30 % des 590 formes fléchies sous lesquelles peuvent apparaître les mots de base). Chacun de ces vocables est donc répété en moyenne 17,9 fois et ils activent la totalité de la liste des mots de base (66). Chaque souche est utilisée en moyenne 48,5 fois. Cette fréquence (0,9 %) et le fait que tous les mots de base soient activés au moins une fois témoigne de la pertinence empirique de cette catégorisation.

- ♦ C'est la souche **le** (le, la, l', les) qui est ici la plus présente : 408 occurrences, soit pronom soit article, ce qui donnerait une fréquence de 73 820 sur un million de mots alors que la probabilité d'apparition³ en français écrit, d'après la base Frantext, est de 57 887.
- ♦ Vient ensuite **il** (il, elle, ils, elles) : 227 occurrences soit une fréquence réelle de 19 737 contre une probabilité d'apparition de 41 071.
- ♦ C'est **être** qui est le premier verbe : 216 occurrences sous 10 formes différentes, avec une fréquence réelle de 39 458 (probabilité d'apparition de 18 240 sur un million de mots).
- ♦ Ensuite, la souche **de** (de, du, d', des) : 192 fois, article ou préposition soit une fréquence réelle de 34 738 (probabilité d'apparition : 69 466).
- ♦ Ensuite **et** : 163 fois, fréquence 29 492 (contre 18 622)
- ♦ Puis **un** (un, une) : 143 fois soit une fréquence de 25 872 (20 442)
- ♦ Vient alors ce qui se construit autour de **me** (me, m', moi, mes, mon, ma) : 122 occurrences avec une fréquence de 22 073 (probabilité d'apparition : 6 554).
- ♦ Enfin **ce** (ce, c', ces, cet, cette) 119 fois, fréquence 25 530 (9 106 dans Frantext)
- ♦ Les autres souches des mots de base sont activées moins de 100 fois, donc avec une fréquence inférieure à 1,8 %.

Certes, des fréquences calculées sur un texte de 5 500 mots ne peuvent être comparées à celles calculées sur plusieurs millions de mots. Pourtant est-ce une tendance de l'écriture pour les enfants ou l'effet de l'abondance des dialogues ou simplement le hasard si les articles définis et indéfinis, les démonstratifs, le verbe *être*, la coordination *et*, les formes de la première personne sont ici beaucoup plus fréquentes que dans les écrits standards ?

1.2. Tous les autres mots

IDÉOGRAPHIX offre une fonctionnalité liée à la base Frantext : nous avons choisi de répartir tous les vocables rencontrés entre 5 échelons de probabilité d'apparition en français. Il ne s'agit plus ici de souches comme dans le repérage

des 66 mots de base mais bien des formes réelles, des vocables, y compris ceux des mots de base. Par exemple, les vocables [est], [était], [sont] appartiennent au premier échelon, [sera], [sommés] au deuxième, [étions] au troisième, [fûmes] au cinquième.

Le premier échelon comporte les 200 vocables les plus fréquents en français ; le deuxième les 300 suivants ; le troisième les 500 suivants ; le quatrième les 1 000 suivants. Ainsi ces quatre premiers échelons regroupent les 2 000 vocables les plus fréquents dans le français écrit. Le cinquième échelon réunit tous les vocables au delà de ces 2 000 premiers. IDÉOGRAPHIX affecte ainsi environ 220 000 formes pour l'échelon 5. Pour bien répéter qu'il s'agit de la probabilité d'apparition d'une forme écrite précise, deux exemples : [blanc] est de l'échelon 2, [blanche] de l'échelon 3, [blancs] et [blanches] de l'échelon 4 ; [chien] est de l'échelon 4, [chiens] de l'échelon 5.

Ces échelons ne donnent pas une information sur une difficulté orthographique mais seulement un classement selon la probabilité de rencontrer un vocable dans un texte écrit. Ensuite, chaque enseignant fait ce qu'il peut, aussi bien afin que ses élèves soient à l'aise pour lire et écrire les mots qui reviennent le plus souvent et afin que les formes les moins fréquentes s'anticipent plus aisément à partir de principes génératifs : les vocables issus de la souche **chien** et ceux issus de la souche **blanc** obéissent après tout au même principe et [chiennes] n'est, somme toute, pas un vocable nouveau lorsqu'on connaît [chien] et [blanches].

Si on retire des 5 527 mots (que totalisent les 13 albums) les 3 200 occurrences des « mots de base » et 180 noms propres, il reste 2 147 occurrences d'autres mots dont on peut observer la répartition dans nos 5 échelons. Figure également, dans le tableau 1 (page suivante), à la suite des 5 échelons, le rappel des données sur les autres constituants du texte.

Les nombres donnés dans ce tableau appellent d'une certaine manière peu de commentaires puisqu'ils illustrent massivement le principe de fréquences qui est à la base de la constitution des échelons.

- colonnes 1 et 2 et 3 et 4. Quatre manières d'aborder ces effets de fréquences : par échelon, le nombre de mots présents réellement dans le texte (colonne 1) n'est évidemment pas proportionnel au nombre de mots constituant

³ La base Frantext calcule (sur un corpus de textes écrits de tous genres cumulant plusieurs dizaines de millions de mots) une probabilité d'apparition d'un vocable dans un échantillon écrit qui compterait un million de mots. Idéographix calcule la fréquence réelle d'un vocable dans le texte examiné et l'exprime dans le même rapport afin de pouvoir procéder à des comparaisons entre textes ainsi qu'avec la probabilité d'apparition en français écrit.

	1	2	3	4	5	6	7	8
Les « autres mots »	Nombre de mots du texte par échelon	Nombre de vocables représentés	Nombre moyen d'occurrences par vocable	% des vocables de l'échelon présents dans le texte	% de vocables présents 1 seule fois	Longueur moyenne des mots	Nombre de souches activées	Nombre de mots du texte par souches activées
Échelon 1 : les 200 vocables les plus fréquents	380	69	5,51	34,5% de 200	17%	4,30	60	6,3
Échelon 2 : les 300 vocables suivants	277	90	3,08	30% de 300	34%	5,27	74	3,7
Échelon 3 : les 500 vocables suivants	208	94	2,21	18,8% de 500	58%	5,98	78	2,7
Échelon 4 : les 1 000 vocables suivants	250	123	2,03	12,3% de 1 000	57%	5,76	105	2,4
Échelon 5 : vocables au delà (environ 220 000)	1032	583	1,77	0,27% de 220 000	74%	6,42	445	2,3
Total des « autres mots »	2147	959	2,24		63%	5,78	762	2,8
Pour mémoire : les mots de base dans leur ensemble	3200	179	17,9	30,3% de 590	24%	2,74	66	48,5
Les noms propres	180	61	2,95		44%	5,66		
L'ensemble des mots du texte	5527	1199	4,61		57%	4,01	828	6,7

tableau 1

chaque échelon (200, 300, 500, 1 000, 220 000) ; pas davantage le nombre de vocables représentés (colonne 2), ce que confirment la diminution du taux de répétition de chaque vocable (colonne 3) et la diminution du pourcentage de vocables de chaque échelon représentés dans le texte (colonne 4).

- *Colonne 5* : l'approche lexicométrique s'intéresse à ces vocables qui ne sont représentés qu'une fois (hapax) et les compte souvent à part. Leur proportion dans un texte croît de manière assez régulière avec leur rareté dans l'écrit : les échelons successifs contiennent des vocables qui sont de moins fréquents et qui sont de moins en moins répétés.

- *Colonne 6* : elle illustre une caractéristique liée à cet effet de fréquence. Plus les mots sont rares, plus en moyenne ils sont longs. C'est vrai pour les 5 échelons ; c'est encore plus net lorsqu'on regarde la longueur moyenne des mots de base (2,74 lettres) et celle de tous les autres mots (5,78) : plus du double.

- *Colonnes 7 et 8* : nous avons dit que les vocables sont des activations de souches (les vocables [aimerai] et [aimé] sont deux formes de la souche *aimer*). Mais *amour* et *aimer* sont deux souches différentes : les mots souches ne se confondent pas avec l'étymologie et les familles de mots. Entre souches et vocables existent principalement des « variations » liées au nombre, au genre, à la personne et au temps verbal. La colonne 8 est le quotient entre la colonne 1 et la colonne 7. On y retrouve qu'une souche

de mots de base est représentée en moyenne 48 fois alors qu'une souche des « autres mots » l'est 2,8 fois.

Ce tableau permet donc de confirmer quelques constantes lexicométriques à propos de l'observation des textes cumulés de 13 albums de littérature jeunesse. Un peu comme le spécialiste cherche à savoir combien de mots aurait utilisé Corneille dans l'ensemble de son œuvre, il peut être intéressant pour le pédagogue de savoir sur quel fonds lexical roulent des ouvrages rencontrés par les écoliers entre le cycle 1 et le CP. Est-il totalement sans importance que 80 % des mots utilisés dans ce texte (4 315 sur 5 527) appartiennent aux 2 000 vocables les plus fréquents en français et activent seulement 383 souches ?⁴

1.3. Les mots comme système

L'apprentissage de l'écrit ne saurait toutefois se concevoir comme une accumulation indéfinie d'unités. La rencontre de celle-ci n'a d'intérêt que parce qu'elle fait découvrir un système. Dans la pédagogie traditionnelle, le système qui

⁴ Cette question rejoint d'une certaine manière certaines préoccupations de l'enseignement de l'écrit au Japon. On y recense un corpus précisément défini de 1 945 caractères kanji dont l'apprentissage est réparti sur plusieurs années du primaire et à travers lequel l'apprenti construit des règles de la combinatoire et de la déclinaison de ces éléments.

permet de dépasser l'accumulation et d'accéder à l'autonomie se construit grâce au recours à de l'information apportée par l'oral, donc extérieure à l'écrit. Dans nos recherches sur la voie directe, le système tente de se découvrir à l'intérieur de l'écrit lui-même. Il est donc important de savoir si, en prenant l'exemple du corpus réuni par ces 13 textes, on embrasse suffisamment d'événements linguistiques pour accéder au cœur du système graphique et parvenir à étendre la connaissance de l'écrit de l'intérieur.

Les différentes options du dictionnaire d'IDÉOGRAPHIX et la recherche d'occurrences permettent de reclasser les 1 199 vocables ou les 5 527 mots pour découvrir des principes de généralisation et des manquements à « la règle ». Au hasard et pour ne pas lasser : toutes les personnes et au moins 8 temps de la conjugaison (4 modes) permettent d'anticiper une multitude de formes verbales à partir de celles qui existent ; une soixantaine de verbes à la forme négative, plus de 70 phrases interrogatives ; 30 mots qui commencent par [e, n] ou [e, m] ; 58 mots qui se terminent par [e, n, t] dont 19 par [m, e, n, t] dont 14 sont des adverbes et 19 par [a, i, e, n, t] ; des séries comme [chasse chasseur] [voyage voyageur] ; des ensembles comme [chiot caneton chevreau porcelet poussin] ; la majorité des cas de passage au féminin et au pluriel pour les noms et les adjectifs ; etc.. Lorsqu'un texte existe comme produit d'un travail d'écriture, on ne saurait s'étonner d'y rencontrer la langue écrite dans sa spécificité et sa richesse. Ainsi, l'observation des opérations

qui vont d'un mot souche aux différents vocables qui l'activent (et inversement) vont permettre de décontextualiser et de généraliser les savoirs sur l'écrit, voire sur la langue, et donner accès à la connaissance de mots encore jamais vus mais dont on pourra dire néanmoins « quelque chose » en y retrouvant les marques des principes qui les génèrent.

2. LES 13 TEXTES PRIS SÉPAREMENT

Examinons le tableau 2 ci-dessous :

- *colonne 1* : l'identification des 13 albums.⁵ Chaque ligne correspond à un album différent. L'ordre des lignes suit approximativement un ordre de grandeur du texte et en aucune manière son intérêt ou sa « difficulté » pour le cycle 1 et le cycle 2.

⁵ **Atchoum**, Sophie Dufeu, Grasset-Jeunesse ; **Une patte, deux pattes...**, Etienne Bruneel, Père Castor / Flammarion ; **De la petite taupe**, Werner Holzwarth / Wolf Erlbruch, Milan ; **Jojo la Mache**, Olivier Douzou, éd. du Rouergue ; **La Grande Panthère**, François Place, Père Castor / Flammarion ; **La promenade de M. Gumpy**, John Burningham/Catherine Deloraine, Père Castor / Flammarion ; **Le beau ver dodu**, Nancy Van Laan, Kaléidoscope ; **Les trois brigands**, Tomi Ungerer, L'école des loisirs ; **L'ours**, Jenny Millot, AFL ; **Ma nounou**, Jean-Claude Baudroux/Eléonora Bartoloni, éd. du Bastberg ; **Plouf**, Philippe Corentin, L'école des loisirs ; **Porculus**, Arnold Lobel, L'école des loisirs.

1	2	3	4	5	6	7	8
<i>Albums</i>	Nombre de mots	Nombre de vocables	Nombre d'occurrences par vocable	Nombre de souches activées	Longueur des phrases	Nombre de vocables nouveaux	Nombre de souches spécifiques
<i>Une patte, 2 pattes</i>	79	52	1,52	25 3,2	3,16	+52	4
<i>Le beau ver dodu</i>	293	78	3,77	41 7,2	36,75	+71	12
<i>Atchoum</i>	278	103	2,70	45 6,2	5,79	+81	22
<i>L'ours</i>	295	106	2,79	57 5,2	6,53	+63	24
<i>Jojo la mache</i>	281	123	2,30	58 4,8	10,96	+78	21
<i>La promenade</i>	319	140	2,29	58 5,5	11,36	+78	24
<i>Ma nounou</i>	455	145	3,14	81 5,6	10,86	+82	37
<i>Les 3 brigands</i>	450	233	2,32	106 4,2	8,88	+140	74
<i>La petite taupe</i>	498	222	2,25	110 4,5	6,27	+124	70
<i>Poussin Noir</i>	513	195	2,64	123 4,2	12,14	+103	56
<i>Plouf</i>	557	193	2,89	129 4,3	7,56	+82	54
<i>La grande panthère</i>	593	227	2,62	137 4,3	15,03	+108	51
<i>Porculus</i>	916	326	2,81	197 4,6	9,45	+139	91

tableau 2

- *Colonne 2* : nombre total de mots du texte. Un texte court, le premier, 79 mots ; un texte long, le dernier, 916 mots, un rapport de 1 à 11. Entre ces extrêmes, 2 blocs : 5 textes compris entre 278 et 319 mots ; 6 textes compris entre 450 et 593 mots.

- *Colonne 3* : le nombre de vocables : rappelons que [chien] et [chiens] sont recensés comme 2 vocables ; [aimes] et [aimeraient] également. On observe que l'éventail des vocables est nettement plus resserré que celui des mots : de 1 à 6 entre les extrêmes (de 52 à 326).

- *Colonne 4* : c'est le quotient de la colonne 2 par la colonne 3, le nombre moyen de fois qu'un vocable apparaît (occurrence) dans le texte. La variation est assez forte puisqu'elle va de 1,52 à 3,77 avec une moyenne de 2,62. Cette information illustre sans doute des différences entre des types de textes, par exemple ceux qui fonctionnent sur des structures répétitives appliquées à des personnages et des contextes successifs et ceux qui déroulent d'un bout à l'autre une seule « histoire ».

- *Colonne 5* : à gauche, le nombre de souches auxquelles se rapportent les différents vocables ; à droite le nombre moyen de mots du texte qui se rapportent à la même souche. On constate que, dès que les textes atteignent une longueur suffisante, la valeur moyenne des occurrences par souche se stabilise entre 4 et 5.

- *Colonne 6* : on constate la grande diversité de longueur des phrases et le fait qu'elles ne sont pas spécialement courtes. Pour les 13 textes réunis en un seul, la longueur moyenne des phrases est de 9 mots,⁶ ce qui permet à beaucoup d'entre elles de contenir des subordonnées (dont 124 sont introduites par **qui** ou **que**, 14 par **quand**, mais 0 par **dont** !), des infinitives introduites par **pour** (21), etc.

- *Colonne 7* : cette colonne introduit une information tout à fait arbitraire. Si les textes étaient introduits successivement dans cet ordre, chacun apporterait évidemment son lot de vocables inédits jusque là : la totalité des vocables (52) du premier texte, 71 sur 78 pour le second (soit 91 %), 81 sur 103 (78 %) pour le troisième, 63 sur 106 (59 %) pour le quatrième, 42 % pour le 13^{ème}. Cette colonne donne donc la manière dont les 1 199 vocables de l'ensemble de ces 13 textes seraient progressivement abordés. Il est évident que si l'ordre d'introduction était différent, par exemple l'inverse de celui du tableau ci-dessus, les valeurs seraient différentes : 326 pour *Porculus*, 140 pour *La Grande Panthère*, 105 pour *Plouf*, 113 pour *Poussin noir* et, à la fin, 47 pour *Atchoum*, 34 pour *Le beau ver dodu* et 9 pour *Une patte deux pattes*. Ce qui est évident, dans un cas comme dans l'autre, c'est qu'à partir d'un certain rang, sans doute ici au-delà du 10^{ème}, la proportion de vocables nouveaux introduits devient inférieure à 20 % du nombre de mots du texte.

- *Colonne 8* : le même questionnement est ici abordé autrement et sans doute de façon plus pertinente : d'une part, il ne s'agit pas des vocables mais des mots souches que ces vocables activent, d'autre part chaque texte est comparé à l'ensemble des 12 autres comme s'il était le 13^{ème} introduit. Ainsi tous sont traités de la même manière. On peut observer que, hormis 3 textes (*Les trois brigands*, *La petite taupe* et *Poussin noir*), le nombre de souches introduites spécifiquement par un texte est inférieur à 10 % des mots qui le constituent.

3. ET ALORS ?

Ces deux regards sur les mots des albums, l'un global et l'autre séquentiel, suggèrent quelques pistes pour une réflexion pédagogique. En cycle 1, il est rare que la fréquentation des albums, au delà du fait de les lire et de les montrer, fasse exister les textes (ou des extraits) de manière durable comme des objets qui, dans un format d'affiche sur les murs, permettent de vivre avec, de s'y repérer, de garder des traces des parcours, d'en réutiliser des éléments, de les relier et de les comparer, etc. En cycle 2, la GS se soucie rarement de capitaliser toutes les rencontres passées en une sorte de territoire accessible à travers des parcours multiples dont les itinéraires s'ouvrent sur d'autres livres encore. Pourtant l'apprentissage de la lecture trouverait naturellement sa place dans le terreau des émotions littéraires de la petite enfance pour peu que l'école sache les enraciner collectivement dans la matérialité des textes. Il ne suffit pas que l'adulte lise les livres en y mettant le ton ; il faut encore que cette lecture conduise à la fréquentation des textes eux-mêmes parce qu'ils seront relus, manipulés, connus et reliés ensemble, qu'ils existeront parce qu'on s'y référera en permanence, que leur forme sera aussi familière que la lecture qui en est faite.

Il y a extension naturelle de la préoccupation des textes à celle de la langue écrite par laquelle ils ont été produits. Mais



ill : E. Bruneel, Père Castor Flammarion

⁶ par comparaison, dans cet article, les phrases ont en moyenne 10 mots.

le glissement n'opère que s'il y a continuité, c'est-à-dire si, dès le début, la langue écrite est présente dans la classe avec la spécificité d'un matériau à la fois incontournable et que rien ne peut remplacer et si, chemin faisant, son questionnement trouve origine et réinvestissement dans la rencontre et l'approfondissement de ce que les textes produisent. Rien sans les textes, rien sans l'écrit. La question pédagogique revient alors à savoir à quelles conditions un corpus de textes va permettre la rencontre du système de l'écrit. Le double regard que nous avons posé sur ces 13 albums peut contribuer à répondre.

L'entrée, proposée dans la seconde partie, qui prend en compte successivement les albums s'inscrit dans les démarches spontanées de l'école : rencontrer un album (ou tout autre texte), acquérir à travers son étude des connaissances de plusieurs types ; passer alors au suivant, réinvestir ces connaissances dans son exploration, en acquérir de nouvelles ; passer au suivant, et ainsi de suite. On retrouve le rôle des progressions qui tentent d'aller du simple au plus complexe sur laquelle repose, quelle qu'en soit la logique, l'organisation des manuels d'enseignement de la lecture. Aussi semble-t-il normal de commencer par le plus court, celui qui présuppose le moins de préalables. Dans cette logique (tableau 2), on voit que le premier texte compte malgré tout 52 formes différentes dont l'élève est censé ne rien savoir. Combien de temps faudra-t-il pour qu'il les connaisse et comment s'y prendra-t-on ? En supposant qu'il y parvienne et que le texte ait encore quelque saveur, il restera, en abordant l'album suivant, 71 vocables nouveaux pour lesquels il faudra bien faire la même chose. En combien de jours ? Encore 81 pour le troisième... Courage ! L'examen de ce tableau (140 vocables nouveaux pour le 8^{ème}, 124 pour le 9^{ème}, 139 pour le 13^{ème}, sans doute 8 ou 9 mois après avoir abordé le premier) a de quoi décourager.

La possibilité d'une capitalisation progressive est-elle compatible avec les modalités d'un apprentissage linguistique ? Le bébé n'apprend pas sa langue maternelle à partir d'un premier message dont il tirerait un savoir, avant de passer à un second un peu plus complexe, et ainsi de suite. Il est immédiatement plongé dans un océan linguistique (pour éviter le mot 'bain' !) fait de l'ensemble des messages d'un environnement qui l'intègre ; et c'est seulement à l'intérieur de cette globalité (risquons cette fois le mot), à cause de cette globalité où les messages puisent et se recourent, que des éléments se singularisent. L'écrit, comme toute langue, est sans unités de départ, sans données premières qu'il faudrait capitaliser pour construire un système signifiant.⁷ C'est bien de l'inverse qu'il s'agit, et notamment lors de sa rencontre. Apprendre une langue, c'est inventer progressivement les moyens d'isoler de plus en plus d'éléments dont aucun d'eux

ne signifie d'abord par lui-même. Tous ne deviennent porteur de sens que par opposition et par différence ; leur efficacité et leur précision sont proportionnelles à la richesse et à la complexité dans laquelle ils ont dû et pu se spécifier. Ce n'est pas du côté des éléments du code que se joue l'accumulation mais du côté de la simultanéité et de l'entrecroisement des messages. Nous voici plutôt invités à réfléchir à la pédagogie que suggère cette fois notre premier tableau.

Une assurance. Dès le début, l'écrit doit exister en même temps que les textes. Dans la classe, ce qui la distingue de la BCD ou de la famille, dès la Moyenne Section, on ne saurait se contenter de lire des albums en les retournant à intervalles réguliers pour en faire observer les images. Pour certains d'entre eux, comme pour des textes d'autres natures, des pages seront agrandies, affichées, accessibles sur des feuilles où il sera possible d'entourer, de surligner, d'annoter, de relier. Loin d'avoir été seulement entendus, même plusieurs fois, ces textes se mettent à exister matériellement dans un espace visuel, comme des objets qui garantissent la permanence et la disponibilité d'un matériau. C'est sur ces textes qu'on relit, c'est à ces textes qu'on emprunte, c'est dans ces textes qu'on découvre similitudes et différences. Le projet est bien ici de s'assurer de cette double familiarité avec le territoire des rencontres du groupe.

Au cycle des apprentissages, les investigations et les recouplements se systématisent. D'une part, il s'agit d'organiser des activités qui permettent à chacun d'approfondir et d'étendre sa connaissance des textes du patrimoine que le groupe s'est donné et qui est déjà collectivement familier ; d'autre part, il s'agit de puiser transversalement dans l'ensemble de ce qui est ainsi fréquenté et connu pour décontextualiser des éléments, les sérier et les étudier en tant que faits d'écrit et plus seulement comme faits (effets) de textes ; il s'agit enfin, dans des leçons de lecture, d'observer comment ces découvertes au niveau du code s'investissent en retour dans le traitement de nouveaux messages.

L'inquiétude ici devrait être que le territoire des textes exploré soit trop restreint, trop pauvre et rassemble des événements trop monotones et trop peu représentatifs de la richesse de la langue, si bien que le système que l'apprenti en ferait émerger resterait sans pouvoir sur de nouveaux objets ou n'en permettrait qu'un traitement qui échouera ensuite devant ce que la « raison » graphique a de spécifique. Dès lors comment s'assurer que le corpus fréquenté ne trahit

⁷ Cf. l'édition chez Gallimard des manuscrits inédits de Ferdinand de Saussure sous le titre : *Écrits de linguistique générale*, et la présentation qui en est faite par Roger-Pol Droit dans *Le Monde des Livres* du 8 février.

pas la complexité du système de l'écrit et de l'écriture ? En se limitant à la perspective choisie ici, celle de l'observation des « mots », l'introduction d'un quatorzième album peut apporter quelque clarté.

Pour ne pas alourdir la démarche, nous choisissons *Marie Louise* sur lequel Yvonne Chenouf a travaillé parallèlement dans ce numéro. Outre, les 3 noms des personnages, voici les vocables nouveaux introduits par ce texte : *amour ; chair ; convenait ; croisèrent ; défaut ; désunir ; doigts ; duettistes ; embûches ; équilibre ; facile ; inséparable ; intentionnés ; jumelles ; lâcherait ; laisserait ; liées ; mal ; objets ; pourtant ; semée ; séparer ; siamoises ; sœurs ; vie.*

- il saute aux yeux, à 2 ou 3 exceptions, que ces mots seraient nécessaires pour écrire une présentation de l'album ; tous les autres mots appartiennent à un fonds commun, ceux-là, à eux seuls, « racontent » l'histoire. Au fur et à mesure que le fonds commun s'enrichit, les éléments inconnus vont jouer un rôle de plus en plus significatif par rapport à un nouveau texte. Au niveau des albums, il est en outre fort probable que la relation image/texte accompagne cette spécificité.

- D'une bonne partie de ces mots inconnus, on peut dire quelque chose à partir de ce qu'on a déjà observé dans le corpus ; pour ne prendre qu'un exemple *croisèrent* ne fait que s'ajouter à *tombèrent, pétaradèrent, décidèrent, regardèrent, nagèrent, chamaillèrent, dégringolèrent, dressèrent, achetèrent, hésitèrent, emportèrent, naviguèrent, attaquèrent* déjà rencontrés...

- Ces éléments nouveaux représentent moins de 10 % des mots du texte. Bien que déterminants, ils ne sont pas assez nombreux pour rendre le texte incompréhensible et peuvent souvent en retour s'éclairer par le texte lui-même.

Ainsi se dessinent des processus (le savoir lire) qui vont permettre à l'apprentissage de devenir autonome et à la connaissance de l'écrit de s'alimenter de l'intérieur.

Les opérations qu'impliquent, pour l'enseignant et les élèves, la manipulation d'un référentiel commun de textes nombreux et l'organisation de son environnement pédagogique constituent le cahier des charges d'IDÉOGRAPHIX.

Jean FOUCAMBERT