

Ressources
pour l'école primaire

Le vocabulaire et son enseignement

Évaluation de la difficulté des textes
La part du lexique

Jean Mesnager
Professeur d'IUFM honoraire

Novembre 2011

Application à la réflexion pédagogique

On retiendra de ce texte la présentation d'un outil d'aide à l'évaluation de la difficulté des textes, rapportée à la présence d'un vocabulaire plus ou moins connu des élèves.

L'auteur rappelle à la fois l'importance du vocabulaire lors de la mise en œuvre de la compréhension de l'écrit mais également l'attention qu'on doit porter aux mots lors du choix des textes, supports aux apprentissages et à la lecture.

La reconnaissance des mots écrits est l'acte le plus élémentaire du lecteur qui lit pour comprendre. S'il y a trop peu de mots identifiés (plus précisément identifiés de manière dite "orthographique", c'est à dire sans déchiffrement lettre à lettre), la construction du sens devient très difficile, voire impossible.¹

De fait, le pourcentage de mots reconnus à la lecture est dépendant du bagage lexical du lecteur².

Par ailleurs, la pratique régulière, voire intense de la lecture amène le jeune élève vers l'expertise, dans un contexte social où cette expertise est une condition majeure d'insertion.

Dans cette perspective de progrès, bien engagée pour certains, beaucoup moins - voire pas du tout - pour d'autres, un principe devrait être retenu, celui édicté par Michel Fayol dans "Maîtriser la lecture"³ : on veillera *soigneusement à ce que la difficulté augmente régulièrement mais reste toujours dans les possibilités de l'élève et de l'effort qu'il est susceptible de fournir.*

Corollaire : il est nécessaire d'évaluer la difficulté des textes proposés aux élèves, en lecture autonome comme en lecture aidée ou assistée. Or, l'affaire n'est pas simple, pour de nombreuses raisons.

- ▶ Plusieurs facteurs jouent dans ce qui fait qu'un texte est difficile ou facile ; on peut les pointer, mais comment les intégrer pour obtenir un niveau de difficulté ?
- ▶ Les élèves ayant des niveaux variables de compétence, et la lecture étant affaire d'interaction entre un lecteur et un texte, établir un score objectif de difficulté n'aurait pas de sens⁴.

Une démarche rationnelle pour évaluer la difficulté est néanmoins envisageable. Qui plus est, il est possible d'automatiser dans une certaine mesure cette évaluation. C'est ce que nous allons exposer en développant plus précisément la question du lexique. On se limitera à un seul domaine de la lecture parmi d'autres, mais non des moindres, participant au progrès des compétences : *(Evaluer la difficulté) paraît surtout indispensable dans le domaine ... de la Littérature de Jeunesse. Voisinant avec de rares lecteurs boulimiques, bon nombre d'élèves aux capacités peu affirmées ne peuvent faire l'expérience du plaisir de lire qu'avec des ouvrages à leur portée. Et les échecs répétés d'un enfant, occasionnés par des entreprises au-dessus de ses forces génèrent une carrière... de non-lecteur⁵.*

Les facteurs de la difficulté

A la lumière des travaux maintenant bien établis sur les dimensions mises en jeu dans le processus de lecture, on distinguera...

- ▶ la typographie : la lisibilité des caractères et l'aération de la présentation (interlignage, paragraphes, intertitres) ;
- ▶ le lexique et la syntaxe (difficulté de la langue) ;
- ▶ au-delà de la phrase, les éléments de grammaire du texte (inférences, points de vue, système anaphorique) et la clarté de sa progression, la visibilité de sa structure narrative ou expositive.

¹ Alain Lieury a analysé le phénomène, en termes de "saturation", dans "Mémoire et réussite scolaire"

² Quelle que soit la forme "cérébrale" que prend le stockage, mots isolés et/ou réseaux de sens

³ Publication de l'ONL, Odile Jacob

⁴ D'autres objections plus ou moins théoriques sont avancées ; on y répond dans divers articles cités en bibliographie

⁵ "Evaluer la difficulté des textes, Cahiers pédagogiques, Oct 2007

Les perspectives de traitement automatisé que nous développons ici concernent la difficulté de la langue, et plus précisément du lexique.

Vers une évaluation automatisée de la difficulté lexicale

1) Quand on constate qu'un texte contient trop de mots "difficiles", on parle en général des mots « rares ». On apportera plus loin quelques nuances à cette assimilation. Mais il est évident que : "*Caché derrière la haie, le loup surveillait patiemment les abords de la maison*"⁶ se lit mieux que "*Dissimulé dans le hallier, le canidé épiait les parages de l'édifice*"

Les travaux de l'INALF⁷, ont abouti dès les années 70 à un dictionnaire des fréquences classant sur ce critère tous les mots du français. Ainsi, pour évaluer la difficulté lexicale d'un texte, on pourrait a) affecter chaque mot de son coefficient de fréquence et b) calculer la rareté moyenne du lexique. La première des deux opérations serait néanmoins une opération très lourde.

La solution imaginée depuis longtemps⁸ pour évaluer plus aisément consiste à partir d'une liste de base (LB), un vocabulaire "fondamental" de la langue considérée ; on ramènera alors la fréquence des mots du texte au pourcentage de mots qui ne sont pas dans cette liste (MA : Mots Absents de la liste), ce qui est beaucoup plus facile à déterminer. Ainsi, on trouvera MA= 12,4% dans un texte de Colette ou Giono, mais MA =1,4 % dans un extrait du "Petit Nicolas" de René Goscinny. Aujourd'hui, l'opération de mesure se fait en une fraction de seconde avec un outil informatique adapté.

Pour établir une LB, nous nous appuyons ici sur une ressource lexicométrique plus récente : "Lexique 3" (<http://www.lexique.org/>). Cette base, maintenant la plus significative, fournit pour 135 000 formes du français avec leurs représentations orthographiques et phonémiques, la syllabation, la catégorie grammaticale, le genre et le nombre, et surtout les fréquences de ces différentes formes. On en a tiré une LB de 35000 formes Elle est constituée de 8500 mots racines, le reste regroupe les formes fléchies : formes conjuguées les plus fréquentes (par ex *courus*, mais pas *courusse* !) féminins et pluriels ; puis les dérivés et composés.

2) Assimiler difficulté à rareté a cependant ses limites.

D'abord, *un seul* mot inconnu ou ambigu dans un texte peut en perturber la compréhension. Comment comprendre l'histoire d'Ali-Baba sans connaître le mot *caverne* ?

D'autre part, le traitement automatique ou informatique utilisé est défaillant dans les cas suivants :

- ▶ homographie (*tour*, édifice ≠ *tour*, procédé),
- ▶ polysémie ou sens figuré (*solution*, résultat ≠ *solution*, mélange chimique)⁹
- ▶ relativité de l'usage des mots suivant les régions (*cime*, commun pour les enfants du midi, rare pour ceux du nord)
- ▶ surtout, présence d'expressions imagées plus ou moins figées, ou de périphrases, obscures pour un enfant, et pourtant faites de mots simples : *sortir le grand jeu*, *perdre ses moyens*, *essuyer une défaite*.

⁶ Marcel Aymé, Les contes du Chat perché.

⁷ CNRS - Institut National de la Langue Française.

⁸ Depuis 1929, premiers indices de lisibilité américains.

⁹ Il existe des outils informatiques qui prétendent prendre en compte ce cas et celui qui précède, mais ils sont d'un emploi très lourd ; ex Visusyn, CNRS

3) Théoriquement, de tels phénomènes suffiraient à mettre gravement en cause le traitement automatique uniquement fondé sur la fréquence. Mais l'expérience a montré que dans la plupart des textes destinés aux élèves du cycle 3, voire aux premières années de collège, les accidents de ce type ne représentent que peu de cas : moins d'un dixième des obstacles lexicaux présents dans ces textes.

Le traitement automatique de la rareté, en tant qu'indicateur de difficulté lexicale, nous paraît donc très pertinent pour la plupart des textes destinés à cette tranche d'âge. La part d'incertitude est limitée : l'approximation n'est pas l'erreur.

4) On peut néanmoins se poser plusieurs autres questions légitimes sur les résultats obtenus. Comment le niveau (âge, étape du cursus scolaire) est-il déterminé ? Les écarts étant souvent importants entre les enfants, ce jugement général est-il vraiment utile ? Que faire ensuite de cet éclairage sur la difficulté, quelles exploitations utiles sont envisageables ?

Un outil pour évaluer la difficulté de la langue : *Lisi*

La description brève que nous faisons maintenant du logiciel "*Lisi*"¹⁰ répond à ces questions. Cet outil évalue la difficulté *de la langue* des textes qui lui sont soumis.

L'outil prend évidemment en compte, en plus du lexique, la complexité de la syntaxe. Un développement serait nécessaire sur ce dernier point, on le trouvera dans différents articles de la bibliographie.¹¹ On notera simplement ceci : faute de pouvoir mesurer directement la complexité des phrases, on a adopté le critère "nombre de mots par phrase". Statistiquement, plus la moyenne des phrases d'un texte est longue, plus la syntaxe est complexe. Ce phénomène se vérifie en général, pour un texte d'au moins 350 mots.

Lisi ouvre le texte à évaluer et le traite, intégrant syntaxe et lexique. Il produit un résultat chiffré, un indicateur de difficulté situé dans une échelle de 2 à 10, auquel est associée une appréciation exprimée dans ces termes : "La langue de ce texte est accessible à la plupart des élèves de (ex) CE2-CM1". Dans la formulation, l'accent est mis sur deux points : a) seule la langue est en jeu ; b) il faut se résoudre à un certain degré d'approximation.

Le tableau de résultats permet en outre de faire apparaître les mots considérés "difficiles" du texte. Les résultats ne sont pertinents que pour des textes d'au moins 350 mots, destinés aux élèves du CE2 à la 5^e, ou si l'on veut les 8-12 ans.

La mise au point de l'outil a été effectuée à partir de plusieurs centaines de textes.

On l'a ensuite validé de deux manières :

- ▶ en comparant ses préconisations à des épreuves de compréhension proposées à 300 élèves de la tranche d'âge indiquée plus haut ;
- ▶ en l'appliquant aux items de compréhension des textes dans les épreuves nationales de la DEPP proposées aux classes de 6^e : on a comparé les scores de *Lisi* sur ces textes et les performances des élèves sur plusieurs années¹².

¹⁰ "Evaluation de la difficulté des textes", Jean Mesnager, Nathan

¹¹ En particulier "Pour une étude de la difficulté des textes", Le français aujourd'hui, n°137

¹² Compte-rendu dans l'article "Evaluer la difficulté des textes, Cahiers pédagogiques, Oct 2007

Aménagements des scores bruts et exploitation pédagogique

Le maître dispose donc d'un indicateur pour choisir ou rejeter (et même modifier) les textes qu'il proposera aux élèves, en lecture autonome ou assistée. On a rappelé plus haut les vertus de cette lucidité, compte tenu des enjeux. Mais tenir compte ne veut pas dire suivre aveuglément les préconisations brutes d'un outil automatisé comme *Lisi*.

Ce logiciel autorise d'abord un petit aménagement manuel : si l'adulte juge que certains mots pointés comme rares ne le sont pas vraiment pour les élèves de sa classe, il peut les supprimer de la liste des rares et l'évaluation est automatiquement corrigée.

Plus généralement, il est de sa responsabilité d'en tirer des conséquences individualisées. Si tel élève de CE2 est en grosse difficulté, il ne le laissera pas affronter sans aide des textes que le logiciel considère comme de ce niveau ; à l'inverse, les plus avancés pourront se lancer dans des récits étiquetés de niveau CM1, voire CM2. De plus, il n'hésitera pas à utiliser des écrits de niveau supérieur à celui de sa classe pour des ateliers d'entraînement : des textes difficiles dans ce cadre, voire résistants : pourquoi pas ? Mais dans des limites raisonnables.

Un détail pédagogique qui compte : le repérage des "mots difficiles" qu'opère l'outil permet au maître, s'il maintient l'étude de ce texte en atelier de compréhension, d'éclairer leur sens au début de la séance, *avant la lecture*.

Malgré son caractère d'approximation revendiquée, l'indicateur de *Lisi* est très supérieur à l'appréciation pragmatique faite par la plupart des adultes, même experts en matière de littérature de jeunesse. Celle-ci s'opère en effet dans la plupart des cas par un parcours cursif du texte, dans le pire des cas un simple feuilletage, terminé par un jugement global qui tient plus de "l'impression" que du repérage méthodique.

Mais on ne lui demandera pas trop : son domaine d'expertise est limité à la langue, et non à l'ensemble des difficultés de nature plus avancée liées au "fonctionnement" des textes. Et bien sûr, il n'est pas dans son pouvoir d'émettre un avis sur la richesse ou des contenus, ni de pronostiquer un plaisir de lecture.

Bibliographie

DENHIERE G. ET BAUDET S, Lecture et compréhension des textes, Aspects cognitifs, 1988

FAYOL M, Questions de compréhension, in Regards sur la lecture et ses apprentissages, Observatoire National de la lecture, 1996

GIASSON J, La lecture, théorie et pratique, Editions De Boeck

GROMER WEISS, Lire, A Colin, chapitre lisibilité, pp 70 et suivantes.

LABASSE B, La lisibilité rédactionnelle est-elle soluble dans la psycholinguistique ? Article disponible sur Internet : <http://www.univ-lyon1.fr/spr/GEV/READABILITY.html>

MESNAGER J Pour une étude de la difficulté des textes, ou la lisibilité revisitée, Le Français aujourd'hui, n° 137, avril 2002

MESNAGER J Evaluer la difficulté des textes : l'approximation n'est pas l'erreur, Cahiers pédagogiques, Oct 2007

Logiciel Lisi, in Evaluer la difficulté des textes, Nathan, Sept 2008