

Évaluer en français

Comment prendre en compte la difficulté des items et des textes

Verónica Sánchez Abchi
Jean-François de Pietro
Murielle Roth



Évaluer en français

Comment prendre en compte la difficulté des items et des textes

Verónica Sánchez Abchi
Jean-François de Pietro
Murielle Roth

IRDP
Faubourg de l'Hôpital 43
Case postale 556
CH-2002 Neuchâtel

Tél. ++41 (0) 32 889 86 18
Fax ++41 (0) 32 889 69 71

E-mail: documentation@irdp.ch
<http://www.irdp.ch>

Remerciements

Nous remercions Juliette Bourquin, Anne-Marie Broi, Martine Wirthner et Nancy Favre Zeiser pour leurs nombreuses remarques pertinentes à propos d'une première version du texte. Nous remercions également Isabelle Deschenaux, Françoise Landry et Doris Penot pour leur travail de vérification, de relecture et de mise en page.

Cette publication de l'IRDP est un document de travail. Il reflète l'état actuel de la situation. La diffusion de ce document est restreinte et toute reproduction, même partielle, ne peut se faire sans l'accord de son(ses) auteur(s).

Afin de faciliter la lecture, seul le genre masculin est employé pour désigner de manière égale les personnes des deux sexes.

Cette publication est également disponible sur le site de l'IRDP www.irdp.ch

Photo de couverture : Maurice Bettex - IRDP

Table des matières

Résumé	5
1. Introduction	7
2. Méthodologie	9
3. La difficulté de l’item	11
3.1 Les facteurs qui caractérisent un item.....	11
3.2 Les facteurs retenus.....	14
3.3 Vers un système d’estimation de la difficulté des items	19
4. Mesurer la difficulté des textes.....	23
4.1 Les formules de lisibilité.....	24
Vers une approche élargie	25
Le vocabulaire	25
La cohésion.....	26
Mise en page et typographie	26
Les genres textuels	27
La structure textuelle	28
La familiarité du contenu	28
Et encore.....	29
4.2 Quels critères peut-on conserver pour l'analyse des textes ?.....	29
4.3 Exemples d’analyse des textes.....	32
4.4 Des propositions pour la pratique	36
Critères à mesurer de manière précise.....	37
Facteurs généraux à estimer de manière qualitative	38
Retour au texte « Recette de cuisine »	40
4.5 Comment traiter les textes discontinus ?	40
5. Les items d’évaluation de la production et du fonctionnement de la langue	43
6. Vers une mesure de la difficulté des items : synthèse	47
7. En guise de conclusion	51
8. Références bibliographiques	53
Annexe 1	57
Annexe 2	61
Annexe 3.....	65

Résumé

Dans le cadre du projet de développement d'une base de données devant servir dans un deuxième temps à la création d'épreuves romandes communes (EpRoCom), ce travail s'intéresse à la manière de prendre en compte la difficulté des « items » – qui constituent les unités centrales de la base de données en cours d'élaboration – et des matériaux langagiers (textes, corpus...) utilisés pour l'évaluation du français. L'accent est mis en particulier sur les items mesurant la compréhension et sur l'estimation de la difficulté des textes. L'objectif est de proposer un outil opératoire qui guide les enseignants et les concepteurs et conceptrices d'épreuves, à quelque niveau que ce soit (tests de référence nationaux, EpRoCom, épreuves cantonales...), dans l'analyse et l'estimation de la difficulté des items et des matériaux langagiers et qui les aide à les élaborer ou les choisir.

Après avoir proposé, à toutes fins utiles, une définition de la notion d'item, nous examinons les multiples facteurs qui, d'après la littérature, permettent de cerner la problématique de la difficulté des items. Ensuite, nous décrivons – dans le cadre des trois composantes qui forment selon nous l'item : son « contenu » (objets langagiers concernés et opérations effectuées¹), son « enveloppe » (consigne, format de questionnement et contexte) et le « matériel langagier » (texte ou corpus) auquel il est attaché – ceux que nous avons décidé de retenir pour la caractérisation de l'item. Nous proposons alors un système d'estimation et/ou de calcul – sur une échelle comportant 3 valeurs : -1 (= facile), 0 (= intermédiaire) et 1 (= difficile) – de la difficulté de chacun de ces facteurs et nous établissons, de manière séparée pour chaque composante (contenu, enveloppe et matériel langagier), une moyenne des valeurs que nous convertissons en un *indice qualitatif de difficulté* pour la composante concernée ($I_{Cont}, I_{Env}, I_{Mat}$).²

Finalement, l'indice de difficulté de l'item est déterminé non par un nouveau calcul qui serait basé sur les indices qualitatifs des trois composantes mais, simplement, en *juxtaposant* ces trois indices. Il nous est en effet apparu qu'il n'était pas pertinent de les fondre en une valeur unique qui masquerait les informations apportées par chacun d'entre eux et qu'il était par conséquent préférable de les conserver, pour la base de données, dans une formule qui les juxtaposent.

En outre, dans la partie finale, une réflexion est menée en vue d'étendre le propos à l'estimation de la difficulté des items d'évaluation dans les domaines de la production, écrite et orale, et du fonctionnement de la langue.

¹ Quelques-uns des termes utilisés dans ce texte sont précisés dans l'Annexe 3.

² Signalons toutefois que le système de calcul est légèrement différent pour la composante « matériel langagier », pour laquelle nous passons par l'établissement d'un indice intermédiaire – calculé selon la procédure présentée ci-avant – qui est ensuite modulé, de manière qualitative, sur la base d'une grille de questions donnant lieu à des bonus et des malus.

1. Introduction

Le Plan d'études Romands (PER), introduit depuis 2011 dans la partie francophone de la Suisse, définit les objectifs d'apprentissage pour l'ensemble de la scolarité obligatoire. Dans une perspective d'évaluation, il importe de pouvoir mesurer si ces attentes sont effectivement satisfaites et, dans ce but, la Conférence intercantonale de l'instruction publique de la Suisse romande et du Tessin (CIIP) envisage l'implémentation d'épreuves romandes communes (EpRoCom). L'Institut de recherche et de documentation pédagogique (IRD) a été mandaté par les autorités éducatives pour développer, dans un premier temps, le dispositif donnant accès à une banque de données validées qui pourrait servir, dans un deuxième temps à la création de telles épreuves. Le travail consiste notamment à mener une réflexion didactique en amont : quels sont les « objets » du PER à prendre en compte pour l'évaluation ? Comment définir des degrés / seuils de réussite pour l'évaluation de ces objets ? Et, à cette fin, comment estimer la difficulté de ce qui est évalué ?... Ce faisant, nous nous efforçons donc de contribuer – avec d'autres – à l'amélioration des « outils » utilisés pour l'évaluation des compétences et des connaissances dans le contexte romand.

Ce travail s'inscrit dans nos réflexions en cours. À l'exemple de l'évaluation de la compréhension en français, il porte sur l'une des questions susmentionnées : la manière de prendre en compte la « difficulté » des *items* et, en premier lieu, des matériaux langagiers (textes, corpus...) utilisés pour l'évaluation.³

Initialement, la focale de ce travail a été mise sur l'estimation de la difficulté des textes utilisés, car le texte s'avère un facteur crucial dans l'évaluation de la compréhension. Il suffit, par exemple, de mal choisir un texte pour observer que les élèves le comprennent très bien, bien, un peu, voire pas du tout... Dans plusieurs contextes, des formules de lisibilité permettant de calculer un score de difficulté des textes ont été utilisées. Ces formules fournissent des indications utiles mais, selon nous, elles ne sont pas suffisamment adaptées au contexte scolaire dans lequel nous travaillons et doivent être complétées et réinterprétées à cette fin.

Cependant, même si le choix du texte est essentiel pour évaluer la compréhension, il n'est qu'un facteur parmi d'autres qui interviennent dans la construction des *items*⁴. Pour estimer la difficulté d'un item, il s'avère également nécessaire d'analyser la manière dont les questions pour évaluer sont formulées, le format des réponses attendues et ce qu'elles demandent comme travail aux élèves (questions à choix multiples, questions ouvertes demandant une activité d'écriture, etc.), les aides fournies pour réussir la tâche (dictionnaires, tableaux de conjugaison, etc.), les opérations qu'ils sont censés mobiliser (repérage, analyse, interprétation, etc.) – c'est-à-dire de prendre en considération tout ce qui rend l'item plus ou moins complexe. L'ensemble de ces facteurs définissent en quelque sorte la manière dont le texte est appréhendé et permettent de mieux saisir les difficultés potentielles que les élèves vont rencontrer : il est possible, par exemple, de poser des questions très simples sur un texte très difficile, comme il est possible de poser des questions très difficiles sur un texte d'apparence simple.

³ Le travail présenté ici fait par conséquent partie d'un ensemble bien plus vaste, comprenant notamment une conceptualisation du processus d'évaluation et des objets à évaluer, une description / classification critériée des items fondée et l'élaboration d'une base de données.

⁴ Nous revenons sur cette notion dans le chapitre 2.

Par ailleurs, il est nécessaire de distinguer entre la difficulté *a priori* d'un item – c'est-à-dire les caractéristiques de l'item qu'une analyse basée sur des critères scientifiquement fondés considère comme susceptibles de faciliter ou compliquer la tâche des élèves – et la difficulté *a posteriori* – celle qui émerge du niveau de réussite obtenu par les élèves pour chaque item : la difficulté d'un item est alors estimée par la proportion d'élèves qui donnent une réponse correcte à cet item (Vrignaud, 2006). Dans ce travail, nous nous intéressons à la difficulté *a priori*, ce qui a pour but de mieux prévoir les possibles obstacles que les items et les textes peuvent présenter à l'élève.

Cette contribution vise donc aussi à rendre attentifs les praticiens sur les facteurs qui peuvent constituer des obstacles pour les élèves dans la réalisation de tâches d'évaluation, en particulier dans le processus de compréhension d'un texte.

Nous tenterons ainsi de mieux comprendre les enjeux de ces deux problématiques, la construction des items et la sélection de textes, en vue de développer, au niveau romand, des modalités d'évaluation aussi solidement fondées que possible. Nos analyses et propositions concerneront essentiellement l'évaluation de la compréhension écrite (et orale). Après avoir parcouru la littérature et décrit les principales approches théoriques et/ou méthodologiques, nous proposerons quelques pistes pour mieux estimer la difficulté des items. Ensuite, nous nous concentrons sur la problématique de la sélection des textes et nous esquisserons une méthode pour estimer le niveau de complexité textuelle, en nous appuyant sur les mesures quantitatives déjà existantes et les évaluations qualitatives réalisées par quelques didacticien-ne-s dans le cadre d'une expérimentation de contrôle. Finalement, nous tenterons de faire des propositions pour intégrer les 3 composantes – contenu, enveloppe et texte – qui, ensemble, définissent la difficulté des items.

Les exemples présentés dans cette contribution concerneront surtout la compréhension écrite. Celle-ci constitue une première entrée dans la problématique et elle nous permet d'explorer et de modéliser le rôle des différents facteurs en jeu. Il est évident, toutefois qu'il faudra mener une réflexion similaire pour les autres domaines du PER – compréhension orale, production écrite et orale, fonctionnement de la langue, etc. Bien qu'un travail d'une telle envergure dépasse l'objectif de cette contribution, nous ferons quelques observations concernant l'évaluation de la production écrite et du fonctionnement de la langue. Nous évoquerons les enjeux des tâches d'évaluation dans ces domaines et relèverons quelques points communs et différences avec le domaine de la compréhension.

2. Méthodologie

Le texte, d'une part, et l'item d'évaluation, d'autre part, constituent donc nos unités de travail. Pour nous, un item est constitué par une ou plusieurs questions dépendant d'une même consigne et liées à un même matériel, visant à mesurer des objets globalement « de même nature »⁵. L'item est autonome, c'est-à-dire qu'il peut être utilisé indépendamment de tout autre item. Il correspond ainsi à une tâche qui permet de donner des indices (traces, données) sur ce que les élèves savent et savent faire (compétences, connaissances, attitudes ou références culturelles).⁶

Le travail s'est déroulé en deux temps. D'abord, nous avons parcouru la littérature concernant la problématique de la difficulté de l'item et de la difficulté du texte et nous en avons retiré un panorama des facteurs d'influence étudiés.

Dans un deuxième temps, nous avons procédé à une sélection des facteurs à retenir pour cette estimation et défini des modalités de calcul.

Enfin, nous avons tenté d'intégrer l'ensemble de nos propositions afin d'établir, dans la perspective de la conception de la base de données, une procédure permettant d'estimer, d'une manière praticable, la difficulté des items.

En ce qui concerne l'analyse de la difficulté des textes, nous avons, à titre d'exemple, appliqué ces critères à trois textes, tirés des épreuves pilotes d'HarmoS pour la 8^e. En parallèle, trois didacticiens, proches de notre projet, ont effectué une analyse « intuitive » de la difficulté de ces textes, en explicitant quels facteurs avaient eu le plus de poids dans leur estimation; et nous avons établi une comparaison entre nos évaluations critériées et ces évaluations intuitives en vue de proposer au final des pistes aussi simples et pertinentes que possible pour définir un indice de difficulté.

⁵ Pour un exemple, voir p. 14

⁶ Cette définition de l'item, issue des travaux en cours, est encore en discussion et devra être précisée davantage (cf. Rapport EpRoCom, Bourquin *et al.* 2015). Ainsi, la formule « globalement de même nature » n'est pas aisée à définir; elle vise notamment à nous permettre de séparer, dans une épreuve existante, des questions – parfois réunies sous une même consigne – qui portent sur la compréhension d'un texte et d'autres qui concernent des analyses grammaticales, pour lesquelles le texte n'est qu'un prétexte. Nous sommes conscients des limites de cette définition qui a avant tout, pour nous, une valeur heuristique et doit nous permettre d'élaborer une base de matériaux opératoires pour une évaluation au niveau romand.

3. La difficulté de l’item

Dans la section 3.1, en nous basant sur la littérature, nous présentons un répertoire de facteurs permettant de caractériser un item. C’est seulement dans un deuxième temps – dans la section 3.2 – que nous procéderons à une sélection parmi ces facteurs et à la formulation d’hypothèses sur la manière dont chacun d’entre eux est relié à un aspect du niveau de difficulté des items et contribuent à la construction d’un *indice de difficulté* global des items.

3.1 Les facteurs qui caractérisent un item

Les enquêtes internationales qui évaluent les compétences en lecture, comme PISA, PIAAC, ALLS ou IALS (OECD, 2012; Goodman *et al.*, 2013; OCDE, 2013), prennent en compte l’ensemble des caractéristiques de la tâche pour estimer sa difficulté. Ces caractéristiques font référence à trois éléments principaux : a) le contexte, b) les processus et/ou stratégies, et c) les textes et les matériels (Kirsch, 2001 a et b, 2003).

a) Le *contexte*. L’étude PISA parle, pour décrire les tâches de compréhension, de *situation*, « c’est-à-dire l’éventail des contextes dans lesquels la lecture intervient ou des usages qui en sont faits », soit la lecture à usage personnel, public, éducatif et professionnel (OCDE, 2013 : 67). Dans le cas de l’enquête PIAAC, la notion de contexte est précisée davantage et elle renvoie à la variété des contextes dans lesquels les tâches de compréhension s’inscrivent et des contenus sur lesquels elles portent. Ces contextes sont clairement définis et regroupés de manière thématique : « maison et famille », « santé et sécurité », « communauté et citoyenneté », « économie », « travail », « plaisir et récréation », etc. Le contexte vise à donner un sens aux tâches de lecture : en effet, dans la vie quotidienne, les activités de lecture ont un but – sont finalisées –, répondent à des besoins qui émergent dans différentes situations – que les enquêtes essaient de reproduire pour rendre les tâches significatives.

Dans le cas des épreuves élaborées par les différents cantons de la Suisse romande, il n’y a pas toujours de « contextes » – dans le sens évoqué par l’OCDE – liés à l’évaluation de la lecture. Tandis que dans les enquêtes internationales ceux-ci visent à donner du sens aux tâches de lecture, dans le cas des épreuves suisses romandes, ce qu’on peut considérer comme le contexte constitue plutôt un rapprochement thématique : les questions et les textes sélectionnés se regroupent souvent autour d’un « fil rouge » thématique qui fournit une sorte de « toile de fond ». Ces « contextes » ne rapprochent par conséquent pas nécessairement les tâches de lecture demandées de situations de la vie quotidienne, mais ils permettent de lier les items à partir d’une même thématique, ce qui facilite notamment l’activation de schémas cognitifs et d’un vocabulaire commun aux diverses questions figurant dans l’épreuve.

Dans la section 3.2 qui concerne *l’enveloppe de l’item*, nous essayerons à notre tour d’analyser le rôle du contexte dans la définition de la complexité de l’item.

b) Les **processus** et/ou **stratégies**⁷, de leur côté, font référence aux processus mis en place par les lecteurs afin de répondre correctement aux questions. Dans le cadre de l'enquête PIAAC (OECD, 2012), trois processus principaux, regroupant les différentes opérations sollicitées dans les items, sont cités : « situer et identifier » (ou « localiser et extraire » dans le cas de PISA; OCDE, 2012), « intégrer et interpréter », « évaluer et réfléchir ».⁸

La localisation et l'identification consistent à repérer des informations dans un texte. Il peut s'agir d'une opération relativement simple – quand l'information recherchée est clairement identifiable dans le texte – mais aussi d'une opération bien plus compliquée, nécessitant par exemple un raisonnement par inférence ou la compréhension de procédés rhétoriques subtils utilisés par l'auteur.

L'intégration et l'interprétation consistent à comprendre les relations entre les différentes parties d'un texte pour construire du sens et tirer des conclusions sur la base du texte considéré dans son intégralité. Ces relations peuvent être explicites ou supposer des inférences, les composantes à mettre en relation peuvent être contigües ou séparées... Des consignes supposant de « comparer » ou de « mettre en relation » constituent des exemples mettant en jeu de tels processus.

L'évaluation et la réflexion exigent que le lecteur relie l'information du texte à d'autres informations, à des connaissances qu'il possède déjà et à son expérience, par exemple pour évaluer la pertinence ou la crédibilité du texte. Selon le contexte et la visée de la lecture, le lecteur peut ainsi être amené à évaluer la pertinence d'un texte et sa crédibilité, sa cohérence argumentative et sa force de persuasion, voire son adéquation linguistique (registres utilisés, structure textuelle, etc.).

Pour l'enquête ALLS (Kirsch, 2003), un quatrième processus est cité : la génération d'informations. Il s'agit de créer une idée ou un résumé en utilisant des connaissances acquises grâce à la lecture, ou d'écrire un texte sur la base d'une lecture. Cette dernière catégorie nous semble un peu plus problématique pour deux raisons : d'un côté parce qu'elle relève de la production textuelle (oral ou écrite) et non plus seulement de la compréhension, et de l'autre parce que cette génération d'information est aussi le produit des autres opérations décrites, telles que l'intégration ou l'évaluation.

Comme nous l'avons vu à propos de l'identification, un processus peut s'avérer plus ou moins complexe selon l'objet auquel il s'applique. Cependant, les différents processus suggèrent tout de même des niveaux de difficulté distincts : *évaluer* est, tendanciellement, plus complexe que *repérer*. Les processus mis en jeu permettent ainsi de supposer *a priori* le niveau de difficulté de l'item et de relier – selon ceux qui sont ciblés dans un item – les résultats à des niveaux de compétence et/ou de connaissances de ceux qui participent à l'enquête. Nous retiendrons par conséquent les processus comme un critère clé pour évaluer la difficulté des items. Toutefois, nous parlerons plutôt d'« opération », car ce terme nous semble plus concrètement décrire l'activité à laquelle il est fait référence ici.

c) Finalement, les enquêtes OCDE citent **les textes et les matériels** pour définir les caractéristiques des tâches.

Liés de manière plus ou moins étroite au texte, divers facteurs sont évoqués pour décrire la tâche : la transparence de l'information présentée dans le texte (informations littérales vs implicites), la

⁷ Selon les auteurs il est parfois question de processus, parfois de stratégies, parfois des deux notions. Pour notre part, nous utiliserons la notion de processus.

⁸ Le cadre théorique de l'enquête PISA (OCDE, 2013) regroupe ces processus sous la catégorie « aspects » de la tâche; dans le cadre d'HarmoS (CDIP, 2011), ils peuvent être rapprochés des « aspects » de compétence.

complexité des inférences à faire, la complexité sémantique et syntaxique des textes (par exemple, les phrases négatives et les questions sont plus complexes que les phrases positives), la quantité des informations nécessaires pour résoudre la tâche, l'emplacement de l'information dans le texte (si l'information requise se trouve dans un endroit proéminent du texte ou dans une note en bas de page, par exemple) et, finalement, la présence ou non d'informations en concurrence – s'il y a plusieurs informations potentiellement utiles pour répondre à une question et qu'il faut sélectionner (OECD, 2012). Certains de ces facteurs, reliant les opérations et les textes, seront retenus pour l'évaluation de la difficulté du texte et, *a fortiori*, de l'item. Nous reviendrons sur ces questions plus en détail dans le chapitre 4 : « mesurer la difficulté des textes ».

En plus de ces trois ensembles de facteurs tirés des cadres théoriques des enquêtes de l'OCDE, il nous semble important de prendre en considération deux autres aspects qui contribuent à la description des items : la formulation de la consigne et le format de l'item.

d) La formulation de la consigne. La formulation des consignes constitue un facteur crucial dans le cadre des évaluations, car elle peut simplifier ou complexifier la tâche (voir synthèse chez Marc & Wirthner, 2013). Selon la littérature à ce propos, pour les consignes écrites – qui s'avèrent généralement plus compliquées que les consignes orales (Henny, 1998) –, différents aspects sont à prendre en considération si on veut estimer le niveau de difficulté de l'item. Par exemple, les phrases complexes (relatives, coordonnées, juxtaposées) et les négations sont des éléments syntaxiques difficiles à comprendre pour les élèves. De même, le fait de présenter une multiplicité d'instructions qui demandent plusieurs choses à faire complique la tâche de l'élève : dans ces consignes dites « multiples », il y a toujours le risque de n'effectuer qu'une partie de la tâche demandée, ce qui rend l'activité sensiblement plus difficile (voir synthèse chez Guay, 2011).

Divers aspects à prendre en considération sont ainsi mentionnés : les relations d'ordre temporel indiquées dans l'énoncé, la longueur de la consigne (en termes de nombre de mots et de nombre de phrases), la mise en page et la présentation de la consigne, le lexique, etc. Ce dernier facteur est particulièrement important. En effet, la compréhension de la consigne est rendue plus difficile lorsque le vocabulaire n'est pas fréquent ou s'il est très abstrait.

En outre, le contexte de la consigne doit aussi être pris en compte. Il s'agit des phrases d'introduction ou des informations, parfois accessoires, qu'on y trouve souvent et qui peuvent perturber la réception ou surcharger la tâche du point de vue cognitif. Le contexte peut jouer soit le rôle d'un facilitateur soit d'une source de difficulté. En effet, les élèves peuvent avoir des problèmes pour établir le lien entre leurs connaissances et les contenus évalués d'un côté et les phrases qui « contextualisent » la consigne de l'autre (voir à ce propos Guay, 2011, et Marc & Wirthner, 2013).

En fait, en tant que texte, le niveau de difficulté de la consigne peut être évalué avec des critères similaires à ceux que l'on utilise, à une plus large échelle, pour l'évaluation de la difficulté des textes (*cf. infra*, chapitre 4).

e) Le format de l'item. Par format de l'item, nous entendons ici le cadre du questionnaire dans lequel il s'inscrit – QCM, question ouverte, etc. – et le type de « réponse » que cela implique. De nombreux articles ont montré l'importance de ce facteur dans l'évaluation de la compréhension écrite (voir synthèse chez Fletcher, 2006).

Pini *et al* (2006) ont répertorié les types de questions les plus communs pour les items utilisés dans les évaluations cantonales genevoises. Les catégories de format relevées, avec les définitions fournies par les auteurs, sont les suivantes :

- *Questions à choix multiples (QCM)* : il y a un début qui présente le problème et plusieurs choix qui représentent les solutions possibles.
- *Questions vrai-faux* : il s'agit d'une seule proposition dont il faut évaluer la véracité. Pour ce faire, les élèves doivent cocher les choix « vrai » ou « faux » (oui/non; d'accord/pas d'accord).
- *Questions d'appariement* : c'est une variante de la question à choix multiple. Pour ce type de question l'élève doit appairer plusieurs amorces à plusieurs réponses. Cela constitue une sorte de complexification de la QCM, car c'est l'élève même qui doit sélectionner les réponses pertinentes pour chaque amorce.
- *Questions de classement* : les élèves doivent classer les éléments d'une liste d'après des critères donnés. Bien que la difficulté soit a priori plus élevée que dans une QCM, tout dépend en fait des critères de classement proposés et de la nature des éléments fournis.
- *Questions d'ordination* : les élèves sont appelés à sérier les éléments d'une liste, selon un critère donné (l'argument le plus fort, les événements d'une histoire d'après l'ordre chronologique, etc.).
- *Questions à réponse brève* : pour ce type de questions, l'élève doit produire « un mot, une phrase, un nombre » qui complète ou répond à la question énoncée. Les questions à réponses brèves se divisent encore en deux formats :
 - les textes lacunaires;
 - les champs de réponses.
- *Questions à réponse narrative*, que nous avons décidé d'appeler *questions à réponse textuelle* pour éviter de possibles confusions avec les regroupements de genres « texte qui raconte » et « texte qui relate ». Dans ce cas, l'élève doit rédiger une réponse plus ou moins longue.
- *Questions demandant une performance* : dans ce cas, l'élève doit effectuer des actions en suivant une consigne.

Par ailleurs, le format du questionnement peut encore varier selon qu'il s'agit d'un test sur papier ou informatisé. Sur ordinateur, par exemple, au-delà de la réponse à un QCM ou de la production d'une réponse écrite plus ou moins étendue, on trouve fréquemment des tâches qui demandent la sélection et/ou le déplacement d'une information requise (cliquer et déplacer une phrase, un mot, etc. dans le texte ou dans le graphique...), la création d'un graphique ou d'un tableau, etc.⁹

3.2 Les facteurs retenus

À la lumière des positions recensées dans la section 3.1, nous avons défini les éléments qu'il nous paraissait important de retenir – pour le projet qui nous concerne, à savoir, rappelons-le, l'élaboration d'outils servant notamment à la construction d'une base données – afin d'analyser le niveau de complexité d'un item d'évaluation. Nous considérons trois grandes catégories qui nous permettent de rassembler les éléments relevés dans la littérature : le **contenu** de l'item, son **enveloppe** et le **texte** ou, plus largement, le *matériel langagier* qui lui est lié. Nous parlons de contenu pour décrire les opérations mobilisées et les objets langagiers sur lesquels elles portent et pour nous référer aux objectifs du plan d'études qui sont concernés. L'enveloppe concerne

⁹ voir Enright *et al.*, 2000, pour une synthèse.

plusieurs facteurs, tels que la consigne et le contexte de tâche qui est attribuée à l'élève, le format de questionnement et le matériel complémentaire mis à disposition pour résoudre la tâche. Le texte, de son côté, sera traité plus en détail dans le chapitre 4.

Voyons maintenant chaque catégorie plus en détail.

Le contenu de l'item

Cette catégorie vise à cerner ce qui est l'objet même de l'évaluation, qu'il s'agisse d'une connaissance ou d'une compétence¹⁰, en quelque sorte indépendamment des moyens utilisés pour le rendre visible.

Pour définir le contenu de l'item, nous reprenons le concept de processus mais nous lui préférons l'expression « opération » pour décrire l'activité attendue de l'élève. Les opérations sont formulées en termes de prédicats qui s'appliquent à des objets langagiers pour former des contenus d'évaluation. De plus, ces contenus correspondent plus ou moins étroitement à des objectifs d'apprentissage du PER, que ceux-ci soient formulés en termes d'objectifs, d'attentes fondamentales ou à l'intérieur des progressions d'apprentissage. Ainsi, afin de nous inscrire dans le cadre défini par le PER, nous essayons dans la mesure du possible d'identifier les contenus en tenant compte de ceux qui y figurent ou qui sont ciblés dans des items présents dans les épreuves cantonales¹¹.

Nous avons constaté que, dans le domaine de la compréhension (et de la production), les contenus sont liés à des opérations. En effet, tant pour l'évaluation, dans les tableaux de spécification des épreuves que nous avons analysées, que pour l'enseignement / apprentissage, dans le PER, la formulation des contenus se fait le plus souvent à l'aide de verbes (ou prédicats) qui décrivent les opérations en jeu et qui sont appliqués à des objets langagiers plus ou moins définis : *repérer [une information], comparer [... et ...], etc.*

En voici quelques exemples. Dans le PER, on observe par exemple le contenu suivant : « repérage des différentes parties du texte (*matériel nécessaire, ingrédients, marche à suivre,...*) » (PER, L1, 21). Ce contenu comporte une opération renvoyant à l'un des processus rencontrés dans les études internationales, « situer et identifier », opération qui s'applique à un matériel langagier donné, les énoncés d'un texte qui expriment des informations d'un certain type; en outre, ce contenu est contextualisé (« dans un texte qui règle des comportements »).

De même, le contenu de la progression « Questionnement du texte, émission et vérification d'hypothèses en cours de lecture » suppose de mettre en place plusieurs opérations d'intégration, d'interprétation et d'évaluation à appliquer aux segments du texte concernés par les hypothèses, voire à l'ensemble du texte. Les contenus « Reformulation d'un passage du texte dans la perspective d'approfondir sa compréhension » ou « Résumé oral d'un texte ou d'une partie de texte pour en dégager le sens » impliquent eux aussi des opérations se rapprochant de l'intégration et de l'interprétation, mais ils impliquent en plus une activité de génération d'information qui s'appuie sur d'autres opérations encore, propres à la production écrite (sélection de l'information, planification, mise en texte, etc.).¹²

¹⁰ Pour cette distinction, voir Roth *et al.* (2014).

¹¹ Les épreuves cantonales ou « communes » ont pour but d'évaluer les acquisitions des élèves à l'échelle d'un canton.

¹² Ce qui non seulement augmente le nombre d'opérations à effectuer mais change fortement la nature même de la tâche à effectuer, qui n'est plus seulement de compréhension mais aussi de production. Or, les caractéristiques du texte produit par l'élève peuvent influencer la perception de l'enseignant et, *a fortiori*, son évaluation de la « compréhension » du texte par l'élève...

Nous constatons également que si un item fait appel à plusieurs opérations et objets en même temps, il va vraisemblablement s'avérer plus complexe.

Par ailleurs, comme on l'a signalé précédemment, le niveau de difficulté des différentes opérations ne peut être établi sans tenir compte des objets langagiers auxquels elles sont appliquées, voire d'autres éléments encore de l'activité, comme le type d'information concerné. Tous ces éléments s'articulent avec les opérations pour constituer les différents niveaux de complexité du contenu. Ainsi, même si les différentes questions d'un test mobilisent des opérations identiques, la difficulté de chaque question peut varier en interaction avec ces éléments – qui nous permettent en quelque sorte de concrétiser (ou actualiser) les opérations et de cerner ainsi les contenus de l'évaluation.

Par exemple, pour « comparer et intégrer », des inférences sont généralement nécessaires; mais il peut s'agir d'inférences très simples ou plus complexes. Ainsi, un item qui évalue l'élément du PER « faire des inférences », et qui mobilise une opération d'intégration des informations, s'avèrera moins complexe si l'inférence se limite par exemple à identifier l'antécédent d'un pronom, et vraisemblablement plus complexe s'il s'agit d'inférer les causes d'une action dans un récit (Graesser, Singer & Trabasso, 1994, cité par Lavigne, 2008). De même, *repérer et identifier une information* peut s'avérer très simple s'il s'agit d'une information littérale. Toutefois, si le texte est complexe, même l'opération de repérage d'une information littérale peut poser des difficultés au lecteur...

Pour estimer le niveau de difficulté du contenu des items, nous avons par conséquent fait le choix de considérer plusieurs facteurs en même temps. Nous pouvons ainsi imaginer un système de classement qui illustre l'ordre de complexité « in crescendo » des opérations retenues d'un côté¹³, des objets langagiers auxquels elles s'appliquent de l'autre. La difficulté estimée du contenu découle alors du croisement de ces deux ensembles. Il importe de souligner que ce classement *a priori* des opérations sur une échelle de difficulté croissante n'est jamais absolu et qu'il doit être modulé selon l'objet auquel elles s'appliquent – objet dont la complexité est essentiellement définie par son « ampleur » et son « statut ». De plus, dans ce système, la complexité de chaque opération – toujours liée à un objet langagier – se combine avec la quantité d'opérations requises pour réussir la tâche. Dès lors, la complexité du contenu d'un item dépend finalement, pour ce qui est des opérations, du nombre d'opérations impliquées et du degré de complexité de l'opération la plus complexe.

L'exemple ci-après illustre la manière dont ces facteurs peuvent être appliqués à un item concret d'évaluation de la compréhension écrite, issu des épreuves pilotes du projet HarmoS. En annexe, nous présentons le texte sur lequel il porte (*Huile de millepertuis – médicament phytothérapeutique – information destinée aux patients*), et dont la difficulté sera analysée dans le chapitre 4.

Qu'utilise-t-on pour fabriquer l'huile de millepertuis ?

Coche la bonne réponse.

- Du millepertuis, de l'huile de tournesol et de l'huile d'olive.
- De l'huile de tournesol, des fleurs et des boutons de millepertuis.
- Différentes huiles, des fleurs et des boutons de millepertuis.
- Des feuilles, des boutons de millepertuis et de l'huile.

¹³ Les opérations citées émergent de nos analyses des épreuves disponibles. Il ne s'agit pas d'une liste fermée et exhaustive. Sa fonction est d'orienter l'estimation de la difficulté des opérations répertoriées.

L'enveloppe de l'item

Comme on l'a indiqué précédemment (cf. section 3.2), « l'enveloppe » de l'item concerne la formulation de la consigne, le format de questionnement et les matériels à disposition. Nous présentons ici les éléments de l'enveloppe que nous avons retenus pour définir la difficulté de l'item.

a) Formulation de la consigne

Pour évaluer le niveau de complexité de la consigne, nous allons retenir deux aspects principaux : la complexité syntaxique et lexicale et le contexte.

1. La complexité syntaxique et lexicale : nous allons estimer si la formulation est claire et si le vocabulaire est accessible pour le lecteur concerné. Ainsi, dès que, dans la formulation de la consigne, il y a un mot ou un élément structurel susceptible de poser problème aux élèves, on peut considérer que la difficulté de la consigne augmente.
2. Le contexte de la consigne : sur la base de nos lectures, nous avons pris la décision de considérer le « contexte » en lien avec la consigne.¹⁴ En ce qui concerne la difficulté de l'item, la présence ou l'absence d'une contextualisation ne constitue pas en soi un facteur ayant une influence décisive sur le niveau de complexité : le contexte n'est pas nécessairement, comme on pourrait le penser, un facilitateur en soi; il peut même, s'il s'agit d'un contexte ambigu ou trop artificiel, compliquer la compréhension de la tâche.

Il s'agit donc de rester prudent. Par rapport à chaque item, nous nous limiterons à estimer, de manière subjective, si les éléments de contextualisation de la consigne semblent plutôt soulager la tâche de l'élève ou si, au contraire, ils en entravent la réalisation.

b) Format de questionnement :

Pour l'estimation de la difficulté des formats de questionnement, nous nous basons sur la typologie fournie par Pini *et al.* (2006). Aux deux sous-formats qu'ils proposent pour les *questions à réponse brève* (textes lacunaires et champ de réponse), nous en avons ajouté un troisième, qui émerge de l'observation de certaines épreuves cantonales et constitue une variation parmi les questions à réponse brève : les questions à réponses visuelles, soit entourer/souligner/colorier la bonne réponse.

Par ailleurs, nous préférons parler de *questions à réponse textuelle* plutôt que narrative (cf. *supra*, p.14).

En ce qui concerne l'impact des différents formats sur le niveau de difficulté de l'item, nous allons retenir deux critères principaux : la qualité et la quantité des éléments qui fonctionnent comme étayage de la question.

En effet, le type de format n'est pas « en soi » un facteur qui permet d'évaluer la difficulté de l'item – car il y a toujours interaction avec tous les autres aspects qui caractérisent l'item. Ainsi, un format qui fournit moins d'étayage aux élèves demande un travail cognitif plus important et, par conséquent, suggère un plus grand niveau de difficulté. En ce sens, un QCM qui propose des choix parmi lesquels les élèves doivent trouver la bonne réponse devrait être moins exigeante qu'une question à réponse brève. Toutefois, ce critère quantitatif ne suffit pas car, si les choix proposées sont trop proches l'un de l'autre et ne constituent pas vraiment des éléments facilitateurs, mais

¹⁴ En fait, la présence d'un contexte donne du sens à la tâche dans son ensemble et, de ce point de vue, elle concerne également le contenu. Mais c'est en lien avec la consigne qu'il nous a paru le plus simple de la prendre en considération.

plutôt des facteurs de complexification, l'étayage, apparemment important, ne facilite pas vraiment la tâche.¹⁵

c) Matériel complémentaire

Nous faisons référence ici à la possibilité de faire appel, dans le cadre d'un item, à des ressources telles que :

- dictionnaires,
- correcteur orthographique,
- tableaux de conjugaison,
- *Mémento* grammatical,
- cahier de constats,
- cahier personnel, etc.

Ces documents sont censés faciliter la résolution de l'item, car ils fournissent un étayage supplémentaire. Ils représentent une « ressource externe » (Candelier, 2007) que l'élève peut mobiliser pour réussir l'activité. La possibilité de consulter ce type de matériel constitue donc, *a priori*, un élément facilitateur.

3.3 Vers un système d'estimation de la difficulté des items

Pris en compte tous ensemble, les facteurs retenus nous permettent de penser un système de classement – simplifié – de la difficulté des items. Nous proposons d'attribuer à chaque facteur une valeur qui correspond à son niveau de difficulté estimé; nous n'avons retenu finalement que trois valeurs :

-1 : facile

0 : intermédiaire

1 : difficile

Puis de faire, d'une part, la moyenne des valeurs pour le contenu, d'autre part la moyenne des valeurs pour l'enveloppe. Chaque valeur moyenne ainsi obtenue doit ensuite être mise en relation avec la table ci-dessous qui donne un indice de difficulté *qualitatif*.

¹⁵ En fait, le rôle des distracteurs est crucial pour définir l'impact du format de questionnement sur le niveau de complexité d'un item. Une tâche, quelle qu'elle soit, peut *a priori* être considérée comme plus simple quand il n'y a aucune information « de distraction »; elle devient d'autant plus compliquée que le nombre de distracteurs augmente, surtout si les distracteurs partagent plusieurs caractéristiques avec la bonne réponse et si – dans le texte – il y a peu de distance entre la bonne réponse et les autres informations fonctionnant comme distracteurs (Kirsch, 2001, b, p. 17).

Tableau 2 : table pour l'estimation de la difficulté

Valeur	Indice de difficulté
-0.61 à -1	Facile (F)
-0.21 à -0.6	plutôt facile (f)
-0.2 à 0.2	Intermédiaire (I)
0.21 à 0.6	plutôt difficile (d)
0.61 à 1	Difficile (D)

En reprenant l'exemple d'item traité précédemment (p. 16), nous pouvons maintenant présenter la manière dont nous estimons la difficulté du contenu et de l'enveloppe d'un item.

Tableau 3 : exemple d'estimation de la difficulté du contenu et de l'enveloppe de l'item

Item								
Contenu				Enveloppe			Texte	
Opération		Objet langagier		Formulation de la consigne	Format de questionnement	Matériel à disposition	cf. chapitre 4 Un indice concernant la difficulté du texte devra également être intégré à l'estimation de la difficulté de l'item.	
Type	Quantité	Ampleur	Statut					
Repérage	Une seule opération requise	Fragment de texte	Information littérale (syntagme situé au début du texte)	Syntaxe simple, mais termes techniques	QCM, mais avec options relativement proches	Absence de matériel		
Valeurs	-1	-1	0	-1	0	0	0	cf. chapitre 4
Valeurs moyennes	-3/4 = -0.75			0/3 = 0			cf. chapitre 4	
Indices de difficulté	Facile			Intermédiaire			cf. chapitre 4	

En ce qui concerne le contenu, l'opération mise en jeu est donc considérée comme simple – il s'agit du repérage (-1) d'informations littérales (-1). Une seule opération est requise (-1) qui nécessite de prendre en compte un fragment du texte pour s'assurer de la bonne réponse (0). Pour l'enveloppe, tous les facteurs ont été considérés comme de difficulté moyenne (0). Pour calculer l'indice de difficulté du contenu et de l'enveloppe de l'item, il s'agit alors, pour chacun d'entre eux, de faire la somme des diverses valeurs obtenues et ensuite de calculer la moyenne (en divisant cette somme par le nombre de critères pris en considération): -0.5 pour le contenu et 0 pour l'enveloppe. Il faut

ensuite se référer à la table de conversion présentée précédemment (tableau 2) pour obtenir les indices de difficulté : plutôt facile pour le contenu, intermédiaire pour l'enveloppe.¹⁶

Ainsi, sur la base de l'estimation de la difficulté des différents éléments qui, selon la littérature et selon nos observations, influent sur la difficulté de l'item, on peut conclure qu'il s'agit ici d'une question plutôt facile. Il faut préciser, cependant, que les valeurs définies ici peuvent soulever diverses questions et que notre procédure d'analyse demande encore une mise à l'épreuve sur un corpus de plusieurs items évalués par différents « juges » afin de tester son efficacité et sa validité. De plus, et surtout, nous devons encore prendre en considération le texte lui-même (*cf.* chapitre 4), avec lequel tous les facteurs analysés ici sont en interaction. La complexité du texte à comprendre devient donc un facteur clé dans l'estimation de la difficulté d'un item et d'une tâche en général, en particulier bien sûr lorsque cet item concerne la compréhension.

¹⁶ La question de savoir s'il faudrait pondérer ces critères et donner plus de poids à certains d'entre eux reste pour le moment ouverte. Elle pourra être tranchée lorsque nous expérimenterons nos outils.

4. Mesurer la difficulté des textes

De nombreuses études portant sur les processus d'apprentissage et sur l'évaluation de la compréhension écrite ont soulevé l'importance cruciale des caractéristiques des textes utilisés pour mesurer les performances des élèves.

Les implications du choix d'un texte pour l'enseignement ou l'évaluation de la compréhension écrite ne sont pas toujours pris suffisamment en considération. Les enseignants – qui eux doivent quotidiennement effectuer de tels choix sans disposer d'outils sophistiqués et sans avoir le temps pour des analyses approfondies – déterminent généralement la difficulté d'un texte par un examen superficiel (Mesnager 2002). La sélection des textes se base souvent sur leurs intuitions ou sur leur expérience – à moins qu'ils se contentent tout simplement d'utiliser les textes proposés par les moyens d'enseignements, sans les mettre en question. En dépit des contraintes et limites de leur pratique, il serait intéressant d'analyser et mettre en évidence les critères qui les guident dans leurs choix et de tenter d'en tirer des critères opératoires – à savoir à la fois compatibles avec ceux établis par les chercheurs et praticables pour les enseignants.

Mais quels sont les critères qui sont habituellement pris en compte par la recherche ? Une première orientation revient à insister sur la complexité du texte en soi – en liant alors la compréhension du texte aux capacités langagières du lecteur – mais sur l'interdépendance « texte et lecteur », la compréhension dépendant aussi – surtout ? – de ses connaissances préalables par rapport au contenu du texte, son intérêt pour la thématique, etc. En effet, les caractéristiques du lecteur – sa mémoire, son niveau d'attention, son développement cognitif, ses connaissances sur un domaine particulier – modifient son approche du texte. Par ailleurs, certains facteurs contextuels – la communauté d'appartenance, le contexte socioculturel, le « projet » dans lequel la lecture s'inscrit¹⁷, etc. – peuvent également avoir une incidence sur la compréhension du texte et sur la manière d'interpréter les tâches demandées. Une telle orientation revient donc à mettre prioritairement l'accent sur le lecteur envisagé dans sa globalité et sur le contexte pragmatique de la lecture.

Toutefois, il n'en reste pas moins qu'il serait absurde de considérer que tous les textes ont le même niveau de complexité et que les éventuelles difficultés rencontrées s'expliquent sans prendre cet élément en considération et en tenant compte uniquement du contexte pragmatique de la lecture et des caractéristiques « non-linguistiques » du lecteur (Mesnager, 2002). Même s'il n'y a jamais qu'un seul facteur pour expliquer les difficultés que les lecteurs peuvent rencontrer lors d'une tâche de compréhension, le texte – et sa complexité – constitue l'un de ces facteurs, et l'un des plus importants (Mesmer, Cunningham & Hiebert, 2012). Il y a donc, clairement, des caractéristiques intrinsèques aux textes qui rendent la compréhension plus difficile, qui permettent de modéliser – et d'anticiper – les obstacles et qui, par conséquent, permettent d'esquisser des niveaux de difficulté des textes et, *a fortiori*, des items portant sur ces textes. C'est cette seconde orientation qui nous paraît la plus opératoire et que nous privilégions ici – tout en la modulant notamment,

¹⁷ Ainsi, une recette de cuisine sera lue de manière très différente si on est appelé ensuite à la réaliser, d'autant plus encore si cette mise en œuvre est collective !

comme nous le verrons, par une prise en compte de certaines caractéristiques non linguistiques du lecteur.

Il existe de nombreuses études qui visent à identifier ce qui rend un texte plus ou moins complexe ou, autrement dit, ce qui en définit le niveau de « lisibilité ». Pourtant, entre autre en raison des éléments mentionnés ci-avant, il n'est pas facile de définir précisément ce qui est « lisible » pour tel ou tel (type de) lecteur (Mesnager, 2002)¹⁸. La recherche a plutôt exploré les éléments qui rendent un texte « difficile » en soi : la fréquence et le niveau de familiarité et d'abstraction des mots employés (Beck, Perfetti & McKeown, 1982), les marques de cohésion utilisées, l'organisation de l'information, la familiarité de la thématique, la structure du texte (Caccamise, Friend, Kintsch & Kintsch, 2013) et le genre textuel (Best, Ozuru, Floyd & McNamara, 2006), ainsi que l'articulation et les points de contact entre ces différents critères (Entin & Klar, 1985, cité par Boyer, 1992; Hiebert & Pearson, 2010, pour une expérience empirique de comparaison de formules mesurant des aspects différents, mais complémentaires).

4.1 Les formules de lisibilité

La préoccupation pour mesurer la facilité de compréhension d'un texte a donné lieu au développement d'une multiplicité d'indices et de formules de lisibilité, qui essaient d'estimer la complexité syntaxique, lexicale, voire typographique, et ce faisant la complexité globale des textes.

Dans le monde anglophone, depuis les années 40, plusieurs formules ont été développées, sur la base de deux paramètres : la longueur des phrases et la longueur des mots. L'indice de Gunning (1952), par exemple, un des indices le plus connus et répandus, permet d'estimer le niveau général de complexité syntaxique et lexicale, avec une échelle qui va de 6 (facile) à 20 (difficile). Il se calcule avec la formule suivante :

$$(\text{moyenne de mots par phrases}) + (\% \text{ de mots de trois syllabes ou plus}) \times (0,4)$$

Le niveau résultant indique les années de scolarité nécessaires pour comprendre le matériel (d'après le système éducatif des Etats-Unis). Ainsi, un niveau 7 ou 8 est requis pour un texte proposé aux élèves de fin d'école primaire, tandis qu'un texte avec un score supérieur à 12 risque de s'avérer trop difficile pour le même élève. L'indice se base sur les hypothèses suivantes : plus les mots sont longs, plus ils sont complexes; et, à leur tour, plus les phrases sont longues, plus elles sont complexes. Ainsi, en calculant la longueur moyenne des phrases et des mots on peut estimer le niveau de difficulté du texte. Le même principe se trouve à la base de la formule de Flesch (1948)¹⁹, qui va un peu plus loin en ajoutant un deuxième score, « d'intérêt humain », qui met en évidence le degré de personnalisation et de dramatisation du texte²⁰. La formule de Dale et Chall (1948), pour estimer la difficulté du vocabulaire, prend quant à elle en compte le pourcentage de mots du texte qui ne figurent pas dans une liste de vocabulaire fondamental. À l'heure actuelle, quelque

¹⁸ Les termes « intelligibilité » et « compréhensibilité » semblent être en concurrence avec le terme « lisibilité ». Pourtant, il y a des nuances, car les deux premiers traduisent davantage « le degré d'accessibilité au sens du texte » et mettent l'accent sur la relation texte-lecteur (pour une clarification terminologique, voir Baudet, 2001). Nous allons donc conserver le terme « lisibilité », en l'utilisant souvent comme synonyme de « niveau de complexité ».

¹⁹ Formule de Flesch : score de 206,84 (= constante) - 1,815 LP (longueur moyenne des phrases) - 0,85LM (longueur moyenne des mots selon le nombre de syllabes). Autrement dit : $206,84 - (x+y)$, avec x = nombre de mots par phrase X 1,815, et y = nombre de syllabes pour 100 mots X 0,85.

²⁰ Pour calculer le score d'intérêt humain, les « mots personnels » sont d'abord pris en compte (les prénoms, les noms communs désignant une personne, les pronoms personnels et les adjectifs qui se réfèrent à une personne). Dans un deuxième temps, les « phrases personnelles » (discours directs, ordres, prières s'adressant au lecteur) sont répertoriées.

200 formules de lisibilité ont ainsi été relevées (cf. Beaudet, 2001), dont certaines peuvent être calculées à l'aide de logiciels disponibles en ligne.

Dans le monde francophone, il y a eu diverses adaptations de ces formules (voir synthèse chez François, 2009, et chez Boyer, 1992) et la création de formules spécifiques pour la langue française (de Landsheere, 1963; Mesnager & Bres, 2008). Parmi ces dernières, la formule de Henry (1975) a été la première élaborée de manière spécifique pour le français et elle combine plusieurs variables linguistiques. Le travail de Mesnager et Bres (2008), de son côté, a la particularité de viser les enfants. L'indice signale un niveau de difficulté estimé en fonction du niveau de scolarisation et tient compte du pourcentage de mots absents de listes spécialement élaborées pour l'école française.

Ces formules sont certainement pertinentes, mais elles s'avèrent, de manière générale, toujours insuffisantes pour définir concrètement la difficulté d'un texte. Les critiques se centrent sur le fait que la plupart d'entre elles prennent en compte seulement le lexique et la syntaxe, et laissent de côté de nombreux autres aspects (pour une synthèse, François, 2009). De plus, les valeurs utilisées ne sont que des approximations de facteurs en jeu et non des mesures absolues. Finalement, ces formules ne constituent pas des outils prédictifs, mais des évaluations négatives par rapport aux aspects analysés (Mesnager, 2002). Au-delà de ces critiques, une question plus générale se pose : peut-on vraiment proposer des indices similaires pour tous les genres de texte et pour tous les niveaux de scolarité ? Cette possibilité a été souvent mise en question (Mesnager, 2002; Labasse, 1999), ce qui suggère que les formules peuvent certes être pertinentes, mais qu'elles ne pourront jamais saisir dans toute sa complexité ce qui fait qu'un texte s'avère plus ou moins difficile.

Vers une approche élargie

En fait, la possibilité d'estimer le niveau de difficulté d'un texte pour identifier les potentielles difficultés des élèves doit prendre en compte d'autres variables. Ces variables ne concernent pas seulement la complexité syntaxique et le vocabulaire – la fréquence et la longueur des mots employés –, mais également la présence d'organiseurs textuels, le nombre d'inférences à faire ainsi que la typographie et la mise en page des textes (pour une synthèse, voir McNamara *et al.*, 2012). Nous allons à présent examiner quelques-unes.

Le vocabulaire

D'après les recherches, pour bien comprendre un texte de manière autonome, il est nécessaire de comprendre, au minimum, entre 84% et 95% des mots (voir Matsuoka et Hirsh, 2010, pour un état des lieux). L'estimation du poids du vocabulaire dans la compréhension réside donc dans l'identification des mots qui peuvent poser problème au lecteur.

Mais que signifie « poser problème » ? Les essais pour définir la difficulté d'un mot d'après la quantité de syllabes qu'il contient ne semblent pas concluants. Bien sûr, un mot plus long est plus difficile à retenir qu'un mot plus court. Mais il y a d'autres aspects à prendre compte, comme la familiarité du mot, son niveau d'abstraction et sa fréquence.²¹ De nombreuses études ont montré l'importance du facteur *fréquence* (Just, Carpenter & Wolley, 1982). Pourtant, là encore, il n'y a pas un consensus absolu par rapport à ce qu'on peut considérer comme un mot fréquent : sur quelles bases établir la fréquence ? Dans quelle mesure est-elle liée au contexte socio-culturel, au contexte

²¹ En outre, les mots courts sont souvent plus polysémiques que les mots longs, ce qui peut aussi rendre la compréhension plus difficile...

d'usage, au genre de texte ?... En ce sens, la variable « niveau de familiarité du mot » a pris de l'importance. La familiarité est définie en fonction du degré de contact de l'individu avec le mot (Guzmán & Jiménez, 2001). Cette variable devient particulièrement pertinente en contexte didactique et doit prendre en compte également les différents niveaux de scolarité : ce qui peut être considéré comme facile ou connu pour un enfant de 8^h (enfants âgés de 11-12 ans) peut se révéler compliqué pour un élève plus jeune.

Dans cette perspective, une solution consiste à calculer le pourcentage de mots d'un texte qui seraient absents d'une liste de mots « fréquents », « courants », ou « fondamentaux » du français (la proposition de Gérard et Roegiers, 2009, est particulièrement restrictive : seulement 570 mots pour le contexte scolaire !²²). Il existe d'ailleurs des listes en ligne ainsi qu'un logiciel spécialement développé pour l'évaluation de textes, qui permet d'estimer de manière générale le niveau du vocabulaire des textes pour l'enseignement du français (Mesnager et Bres, 2008)²³.

La cohésion

La cohésion textuelle joue également un rôle important dans la facilitation de la compréhension du texte, rôle qui s'avère crucial pour les lecteurs avec des difficultés (McNamara, Graesser & Louwerse, 2012; Best *et al.*, 2006). En effet, la cohésion est – via les marques langagières qui l'assurent – une caractéristique objective du texte qui permet de guider l'interprétation des idées exprimées dans le discours, d'établir les connexions entre elles et avec le thème global. La cohésion est conçue comme un *constructo* textuel (Louwerse & Graesser, 2005) qui fait appel à plusieurs ressources, notamment les mécanismes de cohésion référentielle et les organisateurs textuels. Quand ces mécanismes sont absents, le lecteur doit parfois faire des inférences qui rendent la tâche de compréhension plus complexe. En particulier en ce qui concerne les textes informatifs, les organisateurs de cause facilitent de manière significative la compréhension, tant en L1 qu'en L2 (Degand & Sanders, 2002).

Tout en sachant qu'une présence plus forte d'organisateur ne veut pas toujours dire que les liens entre les phrases sont plus simples, on retiendra donc qu'il y a des marques langagières qui, en renforçant la cohésion du texte, clarifient la structure textuelle et rendent donc plus facile la compréhension.

Mise en page et typographie

On pourrait penser que seuls les éléments « linguistiques » ont une incidence sur la lisibilité du texte. Mais des éléments de type graphique – en rapport avec les processus de perception – ont ou peuvent également avoir une influence sur le niveau de complexité du texte. La typographie – la police et la taille des caractères – ainsi que l'espacement des caractères (macrotypographie)

²² Gérard et Rogiers proposent des critères pour la sélection de manuels et la conception de matériels. C'est dans ce contexte qu'ils font appel à certains critères qui, indirectement, peuvent être conçus comme des pistes pour l'évaluation de la difficulté textuelle. Parmi ces critères, ils estiment important de « pondérer le nombre de mots différents d'un texte par rapport au nombre total de mots du texte », de préférer les mots « dont la fréquence dans la langue parlée ou écrite est importante », d'« utiliser des mots actifs et imagés qui constituent des analogies directement compréhensibles par les élèves », de « limiter le nombre de notions nouvelles – mots nouveaux – sur une page »; ils soulignent également l'importance des exemples pour introduire de nouveaux concepts, etc. (2009, 243 et ss.)

²³ L'indice de Mesnager et Bres (2008) est formulé de la manière suivante : $\text{indice} = (\text{abs} \times (3) + \text{LP}) / 4$. « Abs » : nombre de mots absents d'une liste de base; « LP » : longueur moyenne des phrases. À noter : le vocabulaire pèse trois fois plus que la syntaxe. La proposition de Mesnager et Bres (2008) est particulièrement intéressante parce que les niveaux de difficulté sont estimés d'après le système scolaire français.

s'avèrent importants (voir synthèse chez Mesnager, 2002). De manière similaire, la mise en page d'un texte peut faciliter la compréhension, par exemple dans le cas où le texte s'organise en paragraphes clairement séparés avec des sous-titres et espacements importants. Par contre, la mise en page peut devenir un obstacle pour la perception lorsque le texte est organisé en colonnes très étroites, avec une police trop petite et serrée.²⁴ Toutefois, comme nous le verrons (*cf. infra*, p. 28), certains éléments de la mise en page se retrouvent également dans le domaine de la structure textuelle et nous les rediscuterons à ce moment.

Les genres textuels

Dans le contexte de la création de la base de données, en accord avec le PER, le principal critère retenu pour une caractérisation des textes est l'appartenance à un (regroupement de) genre(s) textuel(s). En effet, le PER distingue plusieurs regroupements de genres de textes, définis par l'action langagière qu'ils réalisent : *textes qui transmettent des savoirs*, *textes qui racontent*, *textes qui relatent*, *textes qui règlent des comportements*, *textes qui jouent avec la langue*, *textes qui argumentent*. Certaines études prennent parfois en considération le rôle du genre dans l'estimation de la difficulté textuelle. Et on trouve certains logiciels qui calculent la difficulté des textes en se concentrant sur un regroupement de genre en particulier (*cf. Mesnager et Bres, 2008*, pour les textes narratifs).

Mais est-ce qu'il y a des genres de textes plus simples que d'autres ? Nous serions plutôt prudents pour faire des affirmations si générales. C'est vrai qu'il y a des genres de textes qui semblent plus familiers pour le lecteur. Dans ce sens, les textes qui racontent ou relatent peuvent sembler plus simples que les textes qui transmettent des connaissances scientifiques éloignées du contexte proche du lecteur. Diverses études ont montré que les textes narratifs (*qui racontent* ou *qui relatent*) sont lus deux fois plus vite que les textes expositifs (*les textes qui transmettent des savoirs*) et qu'ils sont, de plus, mieux reformulés (Haberlandt & Graesser, 1985; Graesser, Hauff-Smith, Cohen, & Pyles, 1980, cités par McNamara et al, 2012) et globalement mieux compris (Best, Ozuru, Floyd & McNamara, 2006). Les textes narratifs tendent en effet à présenter des expériences de vie, dans un langage familier, tandis que les textes expositifs mobilisent des concepts et des mots moins connus et demandent plus d'inférences de la part du lecteur.

Toutefois, à notre avis, ce n'est pas « le genre » *a priori* qui peut définir la difficulté du texte, mais les caractéristiques linguistiques – fréquence des mots, structure et mécanismes de cohésion, style / registres de langue – qui lui sont associées et que nous retrouvons ailleurs dans nos analyses. Nous estimons par conséquent qu'un *texte qui raconte* n'est pas *a priori* plus difficile ou plus simple que, par exemple, un *texte qui argumente*; ce sont les caractéristiques intrinsèques à chaque texte – notamment liées à son appartenance générique – qui doivent être prises en considération pour apprécier les possibles obstacles à sa compréhension. Les caractéristiques des genres de textes joueront par exemple un rôle dans l'appréciation de la structure textuelle : est-elle conforme à ce qui est attendu étant donné le genre textuel dont relève le texte ? On peut en effet considérer que lorsque la structure du texte correspond sans surprise à celle d'un genre connu, il devient plus facile à comprendre.

Par conséquent, la question de l'appartenance générique du texte ne constitue pas un critère *a priori* pour l'estimation de la difficulté. En revanche, ce qu'il importe de retenir à ce propos, c'est que l'évaluation de la compréhension d'un texte doit impérativement tenir compte de son appartenance à un genre et des éléments constitutifs propres à chaque genre : on ne lit pas – et on

²⁴ Voir à ce propos Bautier, 2015, et Bautier et al., 2012.

n'évaluera donc pas – de la même manière un texte encyclopédique qui « transmet des savoirs » qu'un récit d'aventure ou une recette de cuisine !

La structure textuelle

Dans le but de décrire la structure des textes, les distinctions et les classements se multiplient. Dans le domaine de l'évaluation de la compréhension en lecture, une distinction clé est faite entre textes continus et discontinus (Mosenthal & Kirsch, 1998; OCDE, 2013; Kirsch, 2001 a et b).

Les textes continus présentent une organisation des phrases en paragraphes, avec des alinéas et éventuellement des sous-titres. Comme nous venons de le voir, ils peuvent aussi être classés selon leur appartenance générique, liée à la visée du texte. Les textes non continus, de leur côté, sont organisés différemment, selon des formats allant de simples tableaux ou graphiques aux cartes²⁵.

Il s'agit donc de nous interroger sur le possible impact de la structure et de la composition des textes sur le processus de compréhension, notamment chez les enfants.²⁶ Par exemple, Englert et Hiebert (1984) ont observé que, parmi les différents textes expositifs analysés, les structures de type « séquence » et « énumération » étaient plus faciles à comprendre pour les enfants que celles contenant une description ou une comparaison. En ce qui concerne les textes argumentatifs, diverses études montrent que l'information est plus difficile à retenir lorsqu'ils proposent plusieurs arguments nouveaux. Il est surtout bien établi aujourd'hui que les textes sont plus faciles à comprendre, à mémoriser et à reproduire lorsqu'ils possèdent une organisation claire et canonique, autrement dit typique pour les genres concernés.²⁷

Bien qu'il n'existe pas, à l'heure actuelle, d'algorithme pour mesurer automatiquement le poids de la macrostructure sur la compréhension, on trouve diverses alternatives, par exemple la proposition de calculer le nombre de thèmes regroupés dans un même texte (« topic cluster ») et le nombre d'interruptions de thème (« topic cluster break ») (Caccamise *et al.*, 2013).

La familiarité du contenu

De nombreuses études ont montré l'incidence des connaissances « déjà-là » du lecteur sur les processus de compréhension textuelle.²⁸ Elles mettent en évidence le rôle « facilitateur » des connaissances préalables par rapport au sujet du texte, en ce qui concerne la compréhension, la réalisation d'inférences et la résolution de problèmes.

Rawson & Kintsch (2002) ont notamment montré que disposer d'informations préalables à la lecture d'un texte facilite également la construction d'une bonne représentation du texte – qui sert alors de base pour la reformulation ou la restitution du texte. Adams, Bell & Perfetti (1995) relèvent pour leur part que les connaissances sur le sujet peuvent interagir avec les compétences générales en lecture – des habilités pour le décodage à la capacité de faire des inférences – et se compléter mutuellement : les lecteurs tendent à compenser leurs difficultés en lecture par leurs connaissances sur le domaine, et inversement.

²⁵ D'autres critères ont également été évoqués : le support – numérique ou imprimé sur papier –, la mise en page (surtout pour les textes non continus) et les caractéristiques des textes numériques (OECD, 2012).

²⁶ Pour un état des lieux, voir Caccamise, Friend, Kintsch & Kintsch, 2013.

²⁷ Plusieurs recherches portent d'ailleurs sur l'effet de la modification de la structure textuelle pour rendre les textes plus compréhensibles et accessibles aux enfants (Mayer, Bove, Bryman, Mars, & Tapangco, 1996; McNamara, Kintsch, Songer, & Kintsch, 1996; Hynd & Guzzetti, 1998; Guzzetti, Snyder, & Glass, 1992; Glynn, Law, & Doster, 1998).

²⁸ Pour une synthèse, voir Mannes & St. George, 1996.

Il semble ainsi ressortir que moins le contenu du texte est familier pour l'élève, plus il devient difficile pour la compréhension. Cela, bien sûr, ne veut pas dire qu'il faut éviter tout texte dont la thématique ne serait pas familière aux élèves, mais plutôt qu'il faut être attentif au fait qu'on leur demande alors un effort supplémentaire, qui doit être compensé par les habilités générales en lecture.

Et encore...

Au-delà de tous les aspects que nous venons de présenter, on observe encore de nombreux autres facteurs qui semblent clarifier le texte, en faciliter la compréhension, et qui pourraient être pris en considération. Parmi eux, on peut citer l'introduction d'explications pour les concepts peu clairs, la présence d'exemples ou d'analogies... De même, l'interpellation directe au lecteur, l'humour et l'introduction d'expressions de la vie quotidienne rendent le texte plus proche du lecteur, attirent son attention et dynamisent la lecture. Certains éléments, comme les ressources utilisées pour s'adresser directement au lecteur, ont par exemple été pris en considération dans le test d'intérêt humain de Flesch (*cf.* Sánchez Abchi & De Mier, 2014).

En plus de leur intérêt intrinsèque, la multiplicité des facteurs qui jouent un rôle dans le processus de compréhension d'un texte met en évidence la complexité de notre tâche consistant à formuler des propositions pour « mesurer », *a priori*, la difficulté des textes utilisés ou à utiliser dans des tâches d'évaluation. Il est par conséquent temps maintenant de nous interroger sur la manière dont nous avons essayé de résoudre ce redoutable problème...

4.2 Quels critères peut-on conserver pour l'analyse des textes ?

Idéalement, la prise en considération de plusieurs indicateurs de diverse nature de manière simultanée (syntaxe, lexique, cohésion, organisation, familiarité, etc.) constituerait la meilleure manière de cibler le niveau de difficulté d'un texte. Récemment, des logiciels prenant en considération plusieurs critères à la fois ont été développés dans le contexte anglophone (Graesser, McNamara & Kulikowich, 2011). Cependant, ces outils n'ont pas été adaptés pour le français. Et, surtout, même ceux qui semblent les plus performants restent limités lorsqu'ils sont confrontés à certaines caractéristiques spécifiques des textes.

Selon nous, toutefois, bien que l'obtention d'un indicateur global ne semble pas toujours possible, il nous a paru pertinent de tenter d'identifier des critères bien ciblés, qui peuvent orienter à la fois les praticiens et les concepteurs. Il s'agit ainsi de retenir quelques indices « techniques », certes, mais aussi de mettre en évidence des critères non explicités, qui sont en fait à la base des processus de sélection plus intuitifs mis en œuvre par ces usagers.

Dans cette section, nous présentons notre sélection des critères – repris parmi ceux que nous avons rencontrés dans nos lectures. Ces critères visent avant tout à l'élaboration d'une sorte de *questionnaire-guide*, qui devrait orienter les enseignants et/ou les concepteurs dans leurs pratiques d'enseignement et d'évaluation.²⁹ À chacun de ces critères nous attribuerons une valeur :

²⁹ La rédaction de ce *Guide* devrait faire l'objet de travaux à venir de l'IRDP. Rappelons que notre étude concerne les textes pour la 8^e. Nous partons par conséquent du principe que les élèves ont des capacités de décodage suffisantes pour assurer la lecture des textes.

-1 : facile

0 : intermédiaire

1 : difficile.

Pour calculer la difficulté estimée du texte, il s'agit alors de faire la somme des diverses valeurs obtenues, de calculer la moyenne (en divisant cette somme par le nombre de critères pris en considération). On peut ensuite se référer à la table p. 20 pour obtenir l'indice de difficulté du texte.³⁰

En tout, nous avons retenu six critères :

1. Indice de complexité lexico-syntaxique

L'indice de Gunning nous fournit une première orientation sur le niveau de complexité syntaxique et lexicale. Nous avons donc adapté les résultats de l'échelle de Gunning pour les faire correspondre à notre propre échelle.

Valeurs de Gunning 6, 7, 8 : -1

Valeurs de Gunning 9, 10, 11 : 0

Valeurs de Gunning 12, 13, 14 15, 16, 17, 18, 19, 20 : 1

Pour établir cette échelle, Gunning a estimé que le résultat devait correspondre à la quantité d'années de scolarité nécessaires pour comprendre le matériel. Dans cette échelle, un texte simple se situe idéalement entre 6 et 8, et un score de 12 est considéré comme difficile (voir p. 24).

2. Indice de complexité lexicale

L'indice fourni par le logiciel de Mesnager et Bres (2008) nous semble approprié, car il vise les enfants et il est élaboré sur la base de listes de mots fréquents dans le contexte scolaire. Nous estimons donc la difficulté du vocabulaire en fonction du pourcentage de mots absents des listes fournies par le logiciel. Sur la base de cet indice et de nos observations, nous avons fixé les valeurs de référence suivantes :

- un texte avec moins de 5% de mots peu fréquents est censé être facilement accessible pour les enfants concernés : valeur -1;
- un texte dont le pourcentage de mots peu fréquents se situe entre 6% et 9% peut être considéré comme de difficulté intermédiaire : valeur 0;
- et, finalement, un texte dont le pourcentage de mots peu fréquents dépasse le 10% peut être considéré comme difficile : valeur 1.

3. Indice de cohésion

Nous centrons cet indice sur le taux de présence des organisateurs textuels (rate organisateurs / phrase) et leur qualité. Nous considérons ce critère de manière prudente, car, bien que les organisateurs textuels tendent *a priori* à réduire le travail du lecteur, en particulier en diminuant la nécessité d'*inférer* les liens entre les énoncés, l'effet réel des organisateurs dépend en fait de plusieurs facteurs : complexité sémantique et pragmatique de chaque organisateur (*justement est*

³⁰ La question de savoir s'il faudrait pondérer ces critères et donner plus de poids à certains d'entre eux reste pour le moment ouverte. Elle pourra être tranchée lorsque nous expérimenterons nos outils.

bien plus complexe que *donc*, par exemple), pertinence de l'emploi dans le texte concerné, etc. Nous prendrons par conséquent uniquement en considération, de manière qualitative, la présence des organisateurs textuels *facilitant la tâche de compréhension*. Pour simplifier, nous pouvons résumer cela de la manière suivante :

- au moins un organisateur adéquat, « facilitateur », par paragraphe → texte plus simple : valeur -1;
- tendance intermédiaire (ou absence de pertinence de ce critère) : valeur 0;
- absence d'organiseurs ou présence d'organiseurs non pertinents : valeur 1.

4. Indice de structuration

Quel que soit le genre concerné, nous considérons qu'un texte est d'autant plus facile à approcher (et, *a fortiori*, à comprendre) que sa structure correspond à celle du genre dont il relève, qu'elle est « canonique »; à l'inverse, les structures « non conventionnelles » demandent un travail de compréhension plus important. Autrement dit, plus la structure est proche du schéma du genre, plus il est simple.

Nous proposons donc d'examiner si les textes présentent une structure textuelle canonique et familière pour le lecteur³¹.

Des critères plus classiques de structuration, tels les paragraphes, intitulés, etc., sont donc pris en compte de manière indirecte, dans la mesure où ils correspondent à ce qui est attendu par la structure du genre dont relève le texte.

- structure générique « standard » : valeur -1;
- structure peu éloignée de la structure générique standard : valeur 0;
- structure générique particulière, non canonique : valeur 1.

5. Indice de familiarité du contenu traité (contenu nouveau vs contenu familier)

Ici encore, nous prenons en compte ce facteur de manière qualitative :

- contenu très proche pour les enfants : valeur -1;
- contenu moyennement proche ou introduction d'explications pour les concepts ou les idées moins familiers : valeur 0;
- contenu éloigné du contexte des enfants avec absence ou très peu d'explications : valeur 1.

6. Divers

Divers autres éléments du texte, difficiles à prévoir, peuvent fournir des ressources facilitant la lecture et la compréhension (exemples, comparaisons, analogies, reformulations, explications, humour, etc.) ou, au contraire, la rendre plus difficile. Nous considérons que de tels éléments doivent, le cas échéant, être pris en compte, mais de façon évidemment qualitative :

- Présence d'éléments divers facilitant la lecture et la compréhension : -1;

³¹ Cf. section 4.4 pour la distinction entre textes continus et discontinus.

- Absence d'éléments ou présence d'éléments qui n'ont pas d'influence sur la lecture et la compréhension : valeur 0;
- Présence d'éléments divers rendant la lecture et la compréhension plus difficiles : valeur 1.

En relation à ce dernier indice, signalons que, vu la centration de l'analyse sur les éléments linguistiques des textes, les aspects « graphiques » ou « typographiques » ne sont pris en compte qu'indirectement, lorsqu'ils jouent un rôle particulier et entrent alors dans notre indice « Divers ».

En outre, bien que la longueur d'un texte ne puisse être considérée comme un facteur de difficulté en soi, elle peut le devenir dès qu'une contrainte temporelle est présente, comme c'est souvent le cas dans le contexte de l'évaluation (*lire le texte et répondre aux questions en 30 minutes...*). Comme notre propos, ici, est d'analyser le texte indépendamment d'une tâche et, donc, d'un temps de passation défini, nous ne prendrons cependant pas en compte cette variable.

4.3 Exemples d'analyse des textes

Pour exemplifier la procédure d'analyse de leur difficulté, nous présentons ici trois textes, issus des épreuves HarmoS pour la 8^e.

Pour ce faire, nous avons d'abord pris en considération les critères présentés précédemment (cf. section 4.2). Puis, dans un deuxième temps, nous avons demandé à trois didacticien-ne-s de procéder à une évaluation « intuitive », sans l'appui de critères explicites, de la complexité des textes proposés. Finalement, nous avons comparé les résultats de ces évaluations « intuitives » et de notre évaluation critériée.

Nous avons fait le choix de prendre des textes appartenant à des regroupements de genres différents, car cette diversité permet de mieux représenter la réalité des textes utilisés dans le cadre des évaluations HarmoS ou des épreuves cantonales. Le tableau 4 synthétise l'analyse d'après les différents critères retenus.

Tableau 4 : synthèse de l'analyse des textes

	Texte 1 Huile de Millepertuis	Texte 2 Recette de cuisine	Texte 3 L'arbre aux hérissons
Genre textuel	Notice d'utilisation	Récit dialogué (qui « raconte » le contenu d'une recette de cuisine)	Narration / relation d'expérience personnelle
Longueur du texte	282 mots	350 mots	511 mots

Critères retenus pour le calcul de l'indice de difficulté			
	Texte 1 Huile de Millepertuis	Texte 2 Recette de cuisine	Texte 3 L'arbre aux hérissons
1. Indice de complexité lexico-syntaxique	10.1 (0)	6.4 (-1)	9.8 (-1)
2. Indice de complexité lexicale³²	La langue de ce texte est difficilement accessible avant la 4 ^e (13 ans). Pourcentage de mots absents : 12.9% (1)	La langue de ce texte est accessible à la plupart des enfants de CM1 (9 ans) et de CM2 (10 ans). Pourcentage de mots absents : 3.7% (-1)	La langue de ce texte est accessible à la plupart des enfants de CM2 et de 6 ^e . Pourcentage de mots absents : 6.2% (0)
3. Indice de cohésion	Absence d'organiseurs textuels, sauf deux « pour ». (1)	Présence d'organiseurs du registre oral (voilà, voici, au bout de, et). (-1)	Il y a 19 organisateurs textuels pour 20 phrases. La plupart sont des organisateurs temporels typiques de la narration, mais on trouve également des organisateurs de cause et d'addition. (-1)
4. Indice de structuration textuelle	Texte qui correspond bien à la structure du regroupement de genres <i>textes qui règlent des comportements</i> : organisation en paragraphes, à l'aide de questions qui jouent le rôle des sous-titres. (-1)	Structure typique d'une expérience, construite de manière dialogique. Le fait d'utiliser un titre qui renvoie à un autre genre pourrait poser des problèmes aux élèves pour comprendre cette structure « innovante ». (0)	Structure emboîtée : - partie introductive qui contextualise le texte; - texte qui raconte avec structure canonique. (0)
5. Indice de familiarité du contenu traité	Thématique très éloignée du quotidien des enfants. (1)	Thématique proche des enfants. (-1)	Thématique proche des enfants. (-1)

³² L'indice proposé pour cette catégorie résulte de l'utilisation du logiciel de Mesnager & Bres (2008); la formulation des commentaires en est directement issue.

Critères retenus pour le calcul de l'indice de difficulté			
	Texte 1 Huile de Millepertuis	Texte 2 Recette de cuisine	Texte 3 L'arbre aux hérissons
6. Divers	un exemple; deux explications entre parenthèses; questions pour guider la compréhension (0)	présence d'exclamations, d'onomatopées, d'expressions de la vie quotidienne, humour, appel au lecteur tout au long du texte... Mais le registre parodique rend le texte moins simple (0)	absence d'éléments (0)
Total (= moyenne des valeurs)	2/6 = 0.33	-4/6 = -0.66	-3/6 = -0.5
Indice de difficulté du texte	Plutôt difficile	Facile	Plutôt facile

Dans le tableau 5 ci-après, nous synthétisons l'évaluation des didacticien-ne-s qui ont explicité leurs critères de manière subjective et relativement générale.

Tableau 5 : évaluations subjectives de trois didacticien-ne-s

	Texte 1 Huile de Millepertuis	Texte 2 Recette de cuisine	Texte 3 L'arbre aux hérissons
Évaluation 1	Texte difficile (genre textuel, vocabulaire, structures syntaxiques) malgré quelques éléments facilitateurs (structure).	Texte facile (genre bien connu, schéma narratif typique, humour, contexte d'énonciation clair, vocabulaire simple, structure simple des phrases).	Niveau intermédiaire (vocabulaire accessible, malgré quelques exceptions; texte narratif simple, mais avec structure emboîtée). Niveau très simple s'il s'agit de prendre en compte seulement la narration.

	Texte 1 Huile de Millepertuis	Texte 2 Recette de cuisine	Texte 3 L'arbre aux hérissons
Évaluation 2	Texte difficile (contenu, format, taille de la police, organisation en 3 colonnes, vocabulaire – nombre de mots techniques ou inconnus, mots longs). L'organisation du texte en questions-réponses peut produire des confusions avec le genre.	Texte pas facile malgré le vocabulaire connu et la présence d'expressions orales ou soulignant l'intonation. Difficultés : début du texte abrupt, contexte d'énonciation, structure pas commune, nombreuses métaphores, absence d'adéquation entre le titre et le texte.	Visuellement, le texte est plutôt agréable. Deux analyses : - partie introductive : difficile (vocabulaire compliqué, structure textuelle compliquée, informations données dans une logique peu commune, source du texte placée à un endroit peu habituel); - partie récit : structure narrative habituelle [donc facile] soulignée par l'organisation en paragraphes, quelques mots longs, phrases très longues, beaucoup de phrases relatives, d'énumérations. La plupart des paragraphes sont constitués d'une seule phrase.
Évaluation 3	Texte difficile : le thème n'est pas évident, le vocabulaire est difficile bien que parfois précisé; cela demande de faire pas mal d'inférences.	Texte assez facile malgré le titre qui pourrait être mal interprété; le caractère « allégorique » n'est pas évident.	Catégorie intermédiaire entre assez facile et assez difficile; la partie du début n'est pas évidente (vocabulaire).
Estimation moyenne de la difficulté des textes	Difficile	Facile	Intermédiaire

Comme on peut le voir, les résultats montrent finalement une similitude relativement forte entre l'ensemble des évaluations spontanées et l'évaluation notée, tant au niveau des critères pris en compte qu'au niveau de l'indice de difficulté auquel elles conduisent.

Dans le cas des évaluations « intuitives », trois critères récurrents ressortent fortement, qui se retrouvent également dans l'analyse notée : le vocabulaire, le contenu et la structure. Ces trois critères sont systématiquement évoqués par tous les évaluateurs – ce qui illustre le « consensus » existant par rapport à ces facteurs. Par ailleurs, la syntaxe, la situation d'énonciation et la mise en page sont parfois aussi prises en considération par les didacticien-ne-s, mais avec moins de consensus. Le genre textuel, quant à lui, est en tous les cas associé soit à la description de la structure textuelle, soit à la thématique et au vocabulaire (*cf.* La notice d'utilisation), ce qui nous

rassure dans notre décision de ne pas considérer l'appartenance générique comme un indice de difficulté en soi.

Les résultats auxquels aboutissent les analyses – critériée ou « intuitives » – sont également très similaires. En effet, dans tous les cas, le texte *Huile de Millepertuis* a été estimé comme le plus difficile. Pareillement, le texte *L'arbre aux hérissons*, considéré comme *entre plutôt facile et intermédiaire* dans l'échelle notée, a été estimé comme intermédiaire par les évaluateurs. Cette estimation découle essentiellement du fait que le texte combine deux parties : une partie d'introduction – plutôt difficile, surtout à cause du vocabulaire et des références culturelles – et une partie narrative assez simple. Les évaluateurs ont par conséquent considéré qu'il y avait là deux niveaux de difficulté différents, selon qu'on prend en compte tout le texte ou seulement la narration. Les calculs « automatiques », par contre, comme celui de Gunning ou Mesnager, ne prennent pas en compte cette distinction : ces indices établissent une moyenne sur l'ensemble du texte. Cette différence attire notre attention sur le fait que les logiciels ou les formules automatiques ne se préoccupent généralement pas de telles particularités à l'intérieur des facteurs qu'ils sont en train de mesurer.

Le seul désaccord entre les évaluations concerne la « *recette de cuisine* », qui est considérée comme facile dans tous les cas, sauf par un évaluateur qui l'estime « pas facile » – malgré la simplicité du vocabulaire –, principalement à cause de la structure et de la situation d'énonciation, qui sont perçues comme peu communes.

4.4 Des propositions pour la pratique

De cette confrontation des évaluations ressortent ainsi de nombreux points de convergence qui nous confortent dans notre approche. Les dimensions et critères d'évaluation communs – autrement dit qui s'avèrent justifiés à la fois par les études existantes et par les jugements « intuitifs » des didacticiens – peuvent être pris en compte pour une estimation de la difficulté des textes et, *a fortiori*, pour une sélection des textes à utiliser dans l'évaluation. La coïncidence des évaluations intuitives avec notre évaluation critériée montre en effet que les critères utilisés par les didacticien-ne-s peuvent être justifiés à l'aide d'outils théoriques et empiriques. Néanmoins ces critères doivent être systématisés pour gagner en stabilité dans les pratiques d'analyse et de sélection de textes.

Nous sommes conscients, pourtant, de la difficulté d'application d'une grille critériée trop rigide et d'avoir toujours « à portée de main » les outils qui seraient alors nécessaires pour évaluer de manière précise la difficulté des textes. Pour ces raisons, nous proposons ici des pistes plus « légères » pour orienter la pratique et, en même temps, alléger le processus d'évaluation de la difficulté des textes. La démarche proposée se déploie en deux temps à partir de questions qui orientent l'évaluateur pour l'estimation :

1. D'abord, nous proposons de retenir trois critères centraux à mesurer de manière précise et quantitative, mais sans faire appel à des outils trop complexes.
2. Dans un deuxième temps, nous proposons d'identifier des facteurs généraux qui peuvent constituer des « bonus » ou des « malus » dans l'estimation finale de la difficulté du texte, mais qui ne seront pas chiffrés.

Critères à mesurer de manière précise

Pour l'évaluation notée, nous proposons, en conservant les valeurs de notre échelle (-1, 0 et 1), de retenir les trois facteurs suivants :

Vocabulaire³³ : comme indiqué dans la section 4.2, pour pouvoir comprendre un texte, il faut comprendre un pourcentage minimum, relativement élevé (situé à 84%), des mots du texte. Nous proposons donc de calculer, pour l'ensemble des mots du texte, le pourcentage de mots censés être peu fréquents ou difficiles pour les enfants d'un niveau donné de scolarité. Pour l'indice de complexité, nous retenons les mêmes valeurs que celles définies au point 4.3 : un texte dont le pourcentage de mots difficiles dépasse le 10% peut être considéré comme difficile (valeur : 1); un texte dont le pourcentage de mots peu familiers se situe entre 6% et 9% comme de difficulté intermédiaire (0); finalement, un texte avec moins de 5% de mots peu connus comme facilement accessible pour les enfants concernés (-1).

Complexité syntaxique : pour évaluer la complexité syntaxique on peut certes faire appel aux formules disponibles en ligne pour la calculer (par exemple, l'indice de Gunning ou de Flesch). Mais nous avons choisi de nous inspirer de guides proposant des procédures simplifiées³⁴. Il en découle dès lors un mode de calcul certes approximatif mais simplifié de la difficulté, que nous représentons de la manière suivante :

- Peu de phrases subordonnées dans le texte, nette majorité de phrases simples : valeur -1;
- Plusieurs phrases subordonnées dans le texte: valeur 0;
- Majorité de phrases complexes, présence élevée de subordonnées dans le texte : valeur 1.

Structure textuelle (et organisateurs textuels) : nous avons décidé de regrouper ces deux aspects pour les évaluer de manière intégrée. Les critères suivants, déjà établis au point 4.3, sont d'abord retenus pour attribuer une valeur à chaque dimension :

- structure générique « standard » : valeur -1;
- structure peu éloignée de la structure générique standard : valeur 0;
- structure générique particulière, non canonique : valeur 1.

Et, en rapport avec la structure textuelle, mais pour les organisateurs éventuellement présents dans le texte :

- Structure rendue plus claire par les organisateurs textuels : valeur -1;
- Structure peu influencée par les organisateurs textuels : valeur 0;
- Structure rendue plus complexe par les organisateurs textuels : valeur 1.

L'indice retenu correspond à la moyenne des deux indices.

Cette première phase de l'estimation peut être résumée ainsi :

³³ Nous considérons l'indice de Mesnager comme un outil tout à fait applicable au contexte suisse. La formule que nous proposons est une alternative simplifiée, qui peut être calculée de manière indépendante des logiciels.

³⁴ Nous nous sommes en particulier appuyés sur l'indice de complexité syntaxique de Hunt (1970), fortement répandu, qui consiste à calculer la quantité de subordonnées par unité terminale (T-unit). L'unité terminale correspond à l'unité grammaticale « phrase » et inclut donc les phrases subordonnées qu'elle comporte. Nous avons également consulté divers guides destinés aux enseignants pour évaluer la complexité des textes : Hess et Hervey (2011), Schumacker & Lomax (2004).

Tableau 6 : questions qui permettent de donner une valeur chiffrée aux critères

Critères	Questions	Valeurs -1 / 0 / 1
Vocabulaire	Est-ce que vous trouvez plusieurs termes qui pourraient poser des difficultés aux élèves ? Si oui quel est le pourcentage des termes qui risque d'entraver la compréhension ?	
Complexité syntaxique	Est-ce qu'il y a un grand nombre de constructions subordonnées ? Si oui combien ?	
Structure textuelle	Est-ce que la structure est « typique » du genre dont relève le texte ? Est-ce que les organisateurs textuels présents facilitent la lecture ?	

La valeur finale, pour ces critères, sera calculée comme pour les autres composantes – en faisant la moyenne des trois valeurs retenues et en se référant ensuite à la table (cf. p. 20) pour obtenir un premier indice, intermédiaire, de difficulté du texte.

Facteurs généraux à estimer de manière qualitative

Les questions ci-après, qui ne sont de loin pas exhaustives, parcourent les principaux autres facteurs concernés dans l'estimation de la difficulté d'un texte. Elles permettent une estimation qualitative amenant à prêter attention à des éléments divers qui peuvent jouer le rôle de *facilitateurs* ou, au contraire, de *perturbateurs* de l'activité de compréhension. Nous avons en effet observé, dans certains textes, des éléments – difficilement prévisibles – qui entraînent un changement qualitatif dans le processus de compréhension et s'avèrent dès lors « décisifs » dans l'estimation de la difficulté du texte.

Pour cette deuxième analyse, tous les éléments cités dans la section 4.3, voire d'autres, peuvent être pris en considération. Cette procédure est envisagée comme une attribution de « bonus » et de « malus », que nous n'intégrons pas dans l'évaluation « quantitative », mais que nous tenons à rendre visibles, pour que les enseignants et autres utilisateurs soient attentifs aux facteurs signalés.

Tableau 7 : questions qui permettent d'attribuer un « bonus » ou un « malus »

Catégories	Questions	Bonus (B) / malus (M)
Contenu	Est-ce que le contenu est familier (ou non) pour le lecteur ?	
Syntaxe	Est-ce que les phrases sont trop longues ?	
	Est-ce que le texte comporte, en trop grand nombre, des constructions syntaxiques « complexes » (inversions sujet-verbe, etc.) ?	

Catégories	Questions	Bonus (B) / malus (M)
Structure textuelle	Est-ce que l'organisation du texte (paragraphe, sous-titres, etc.) aide à visibiliser l'appartenance générique et orienter la compréhension ?	
Éléments paratextuels	Est-ce qu'il y a des éléments paratextuels (graphiques, tableaux, images, etc.) qui facilitent ou compliquent la compréhension ?	
	Est-ce qu'il y a <i>trop</i> d'éléments paratextuels, rendant ainsi difficile l'articulation avec le texte en prose ?	
Éléments typographiques ³⁵	Est-ce que la mise en page est claire ?	
	Est-ce que la police sélectionnée est lisible ?	
Éléments stylistiques / rhétoriques	Y a-t-il des explications, des exemples, des analogies, des reformulations... qui réduisent la difficulté des mots peu connus et/ou qui aident à la compréhension générale du texte ?	
	Est-ce qu'il y a de l'humour, des expressions de la vie quotidienne, des onomatopées qui rendent le texte plus dynamique et qui en facilitent la compréhension ?	
	Est-ce que le texte comporte des éléments figurés ou des figures rhétoriques difficiles à comprendre ?	
Autre...		

Cette analyse nous donne donc une estimation qualitative. Elle oriente le regard et permet d'identifier de possibles obstacles. Et elle influence ainsi l'estimation de la difficulté du texte.

Enfin, l'estimation de la difficulté du texte donne lieu à 2 indices :

- un indice quantitatif, situé entre -1 et 1, résultant de la moyenne des critères chiffrés;
- un indice qualitatif pouvant prendre les valeurs D (= difficile), N (= intermédiaire) ou F (= facile); cet indice est défini en tenant compte de l'indice quantitatif et en pondérant ensuite ce résultat de manière subjective par la prise en compte des bonus et malus qui auront été retenus. Pour ce deuxième indice, c'est donc l'évaluateur qui – en dernière instance – va décider de l'importance à accorder aux bonus et malus.

Par ailleurs, lorsqu'il n'est pas nécessaire d'aboutir à une estimation explicite de la difficulté du texte, l'identification de ces bonus et malus peut également s'avérer utile en facilitant, par exemple, le choix de textes à travailler dans le cadre de l'enseignement et la mise en place de dispositifs didactiques. Il suffit alors à l'utilisateur de passer en revue cette liste de questions, en leur ajoutant,

³⁵ Pertinent seulement pour la compréhension écrite. Mais on pourrait proposer des questions comparables pour la compréhension orale : la voix est-elle audible ? La prononciation est-elle compréhensible ? L'intonation facilite-t-elle la compréhension ?...

sous une forme qualitative, celles qui sont traitées ici de manière quantitative : le vocabulaire présente-t-il des difficultés particulières ? Le texte comporte-t-il beaucoup / trop de subordonnées ? La structure du texte est-elle transparente ? Et correspond-elle à un genre connu du récepteur ? Les organisateurs présents dans le texte facilitent-ils la compréhension ?...

Retour au texte « Recette de cuisine »

Pour illustrer notre procédure, reprenons le texte « Recette de cuisine » de Bernard Friot, qui figure en annexe.

En ce qui concerne le vocabulaire, pour un total de 350 mots, nous en avons identifié 9, moins fréquents, qui pourraient s'avérer difficiles pour les élèves³⁶, soit 2,5% du total de texte. D'après notre échelle, le vocabulaire du texte est donc facile et obtient la valeur - 1.

Pour l'indice de complexité syntaxique, nous avons trouvé une moyenne de 0.15 subordonnées par phrase, ce qui nous permet d'estimer que le texte est facile du point de vue de la syntaxe. Nous lui attribuons par conséquent la valeur -1.

Pour le dernier indice, à propos de la complexité structurelle, nous attribuons la valeur -1, car la structure du texte peut être considérée comme canonique et les organisateurs textuels, peu nombreux, ne paraissent pas avoir une influence particulière sur l'accessibilité du texte (valeur de 0).

De la moyenne qui résulte de cette analyse (-1), le texte sera donc estimé comme facile. Petit bémol cependant : la nature « parodique » du titre – qui fait référence au genre « recette de cuisine » alors que le texte est en fait un récit d'expérience – pourrait être considérée comme un possible obstacle à prendre en compte par l'enseignant ou l'évaluateur, donc comme un « malus ». Nous considérons toutefois que ce « malus » ne suffit pas à lui seul à rendre le texte difficile et nous attribuons donc à ce texte, pour le second indice, la valeur F.

* * *

Globalement, les deux outils d'évaluation proposés – l'estimation quantitative et la mise en évidence d'éléments « facilitateurs » ou « perturbateurs » à partir de questions – constituent une démarche à la fois systématique, simple et formative, pour l'estimation de la complexité des textes en vue de travailler ou évaluer la compréhension orale et écrite.

4.5 Comment traiter les textes discontinus ?

Ce que nous avons vu pour le moment concerne essentiellement les textes continus au sens de PISA (OCDE, 2013). La question des textes discontinus soulève un certain nombre de questions spécifiques.

Dans le cadre de PISA, on les teste le plus souvent de manière isolée, ce qui pourrait contribuer à leur difficulté. Normalement, les textes dits discontinus, comme les illustrations, les cartes ou les tableaux, font partie d'un texte global qui les inclut au même titre que les parties en prose, qu'ils complètent, précisent, exemplifient... Dans ce sens, la situation de communication dans laquelle le texte discontinu s'insère – susceptible de lui donner un sens plus aisé à appréhender – pourrait être un facteur important à considérer pour évaluer sa complexité³⁷. Nous n'avons pas véritablement

³⁶ Cannelle, épouvantable, émouvante, emparés, golden, indignée, reinette, sauna, sanglots.

³⁷ Nous remercions particulièrement Martine Wirthner pour ses remarques à ce sujet.

approfondi la question des textes *discontinus* mais nous sommes d'avis que certains éléments de notre procédure 4.3 dans son principe, restent valables pour cette catégorie de textes également.

Mosenthal & Kirsch (1998) ont analysé de manière plus précise comment mesurer la difficulté de ces textes. Ils distinguent les *documents matrices* et les *documents graphiques*. Pour la première catégorie, l'information est organisée en lignes et/ou en colonnes, comme par exemple dans les tables, listes, tables des matières, etc. Pour les documents graphiques, ils citent la ligne du temps, les documents de repérage ou les cartes, l'histogramme, le diagramme circulaire, etc. Et, pour tous ces documents discontinus, les auteurs ont proposé une formule de lisibilité qui est basée sur deux facteurs : la structure et la densité de l'information.

Mosenthal & Kirsch (1998) proposent, sur la base de leur analyse, de retenir 2 critères – à savoir, un indice lié à la structure du texte et un lié à la densité d'information contenue dans le document. Deux autres facteurs pourraient être envisagés : la familiarité du contenu et la familiarité avec la forme des textes discontinus.

Finalement, comme cette distinction n'a guère été explicitement exploitée dans les épreuves cantonales, nous allons quant à nous retenir un critère plus large : le cas échéant, nous nous demanderons si les graphiques ou les éléments paratextuels contribuent à la compréhension du texte dans sa globalité ou si, à l'inverse, ils rendent la tâche plus complexe.

Pour le moment, nous ne nous sommes pas occupés des images que comportent parfois certaines épreuves, et qui peuvent être « impliquées » dans certains items. Il existe une très abondante littérature à ce propos³⁸, mais la question ne nous a pas paru prioritaire. Nous la reprendrons par conséquent au moment opportun.

³⁸ L'analyse de l'incidence des images sur la compréhension, ainsi que la manière de prendre en compte les éventuelles difficultés de compréhension qu'elles sont susceptibles de provoquer, dépasse les buts de notre travail. Pour une synthèse, voir Glenberg, 1992 et Brookshire, Scharff & Mosses, 2002.

5. Les items d'évaluation de la production et du fonctionnement de la langue

Bien que la focale de ce travail soit mise sur les items pour l'évaluation des activités de compréhension, les réflexions qui concernent la complexité de l'item peuvent également être appliquées à l'analyse des items qui touchent les autres axes du PER, tels la production orale et écrite et le fonctionnement de la langue.

Les critères développés ici à propos de *l'enveloppe* restent, à l'évidence, valables. Quel que soit l'axe considéré, la formulation de la consigne, le format de questionnement et l'éventuelle présence d'un matériel complémentaire à disposition contribuent à l'estimation de la difficulté des items. *Grosso modo* nous devrions pouvoir les prendre en compte de la même manière que nous l'avons fait ici.

La situation est plus complexe pour ce qui concerne la dimension *contenu*. Pour chacun des autres axes, il va être nécessaire tout d'abord de préciser les opérations et les objets en jeu, puis de définir la manière d'estimer la complexité de leur rencontre dans les contenus. Pour la production écrite, par exemple, on aura des opérations telles que *choisir, appliquer, relire, etc.*, à appliquer à des objets multiples. Pour le fonctionnement de la langue, des opérations telles que *analyser, comparer, classer, etc.*

Bien qu'une description approfondie de ces opérations et du niveau de difficulté pour les domaines de la production écrite et du fonctionnement de la langue dépasse le but de ce travail, nous présentons ici une analyse des enjeux que cela représente.

Dans le contexte suisse romand, étant donné la difficulté de mettre en place des outils pour évaluer la production orale individuelle des élèves, l'évaluation de la production se centre surtout sur la production écrite. Pour définir les opérations visées dans les items de production écrite, nous pouvons nous inspirer à la fois des études sur le processus de la production écrite (voir synthèse chez Alamargot et Chanquoy, 2001), d'HarmoS et du PER, qui nous aident à préciser les opérations visées par l'enseignement. De manière synthétique, les opérations concernant la production écrite peuvent être regroupées d'après trois ou quatre processus principaux : situer / contextualiser, planifier, « mettre en texte » et réviser l'écriture. Le tableau 8 présente des exemples d'opérations tirés du PER, regroupés en quatre catégories.

Tableau 8 : exemples d'opérations concernant la production écrite

Catégories	Exemples d'opérations tirées du PER
Situer	Identification de la situation de communication en fonction du projet d'écriture (lieu social, émetteur, destinataire, but du texte). → identifier
Planifier l'écriture	Sélection du genre textuel → sélectionner / choisir Choix d'un support adéquat (<i>lettre, affiche, fiche signalétique, feuille,...</i>) → sélectionner / choisir Élaboration du contenu à l'aide de documents écrits et/ou audio (<i>recherche d'idées, d'informations, de mots-clés,...</i>) → élaborer Organisation du texte selon un modèle donné (anticipation globale de la mise en page du texte) → organiser
Mettre en texte/ réaliser	Utilisation d'un vocabulaire adéquat → activer + sélectionner / choisir + utiliser /réaliser Production d'énoncés syntaxiquement et orthographiquement corrects → produire / réaliser Utilisation d'organiseurs verbaux (<i>temporels : le jour suivant,...</i>) et non verbaux (<i>titre, sous-titre, paragraphe, ponctuation,...</i>) → activer + sélectionner / choisir + utiliser /réaliser Utilisation de reprises pronominales et nominales → analyser + sélectionner / choisir + utiliser / réaliser (notamment accorder) Utilisation adéquate des temps des verbes → analyser + sélectionner / choisir + utiliser / réaliser (notamment accorder)
Réviser/ évaluer/ réparer	Vérifie et améliore l'utilisation des temps verbaux, la syntaxe, la ponctuation, l'orthographe lexicale et grammaticale → vérifier (+ réparer / corriger) EVALUATION → évaluer

Au-delà des difficultés intrinsèques aux opérations d'écriture, il y peut avoir dans l'évaluation de la production écrite une complexité supplémentaire : les tâches de production écrite sont souvent liées à une tâche de compréhension et présupposent alors la compréhension d'un texte de base. Plus généralement, certains genres de textes demandent de lire pour écrire, de lire pour parler, d'écrire pour lire, etc.

C'est par exemple le cas de consignes qui proposent de continuer ou de développer un évènement ou un extrait d'un « texte qui raconte » utilisé pour évaluer la compréhension. Dans ce cas, il y aura certes une certaine réduction des tâches liées à la planification textuelle – car l'élève peut s'inspirer des idées du texte source –, mais il faudra d'abord avoir compris le texte ou le passage du texte concerné.

Si la tâche d'écriture présuppose donc la compréhension préalable d'un texte, il y a deux observations à prendre en compte :

- a) l'évaluation de la production écrite n'est plus un *constructo* évalué de manière indépendante (une mauvaise compréhension du texte de base pourrait entraîner des problèmes dans la tâche de production);
- b) la complexité de la tâche d'écriture est par conséquent plus ou moins élevée en fonction de la difficulté de la tâche de compréhension (écrite ou orale).³⁹

Ceci soulève par conséquent la question de la séparation, dans l'évaluation, des quatre « compétences » que constituent la compréhension écrite, la compréhension orale, la production écrite et la production orale. En fait, nous sommes d'avis que rien n'empêche de mélanger ces compétences dans une même épreuve, soit qu'on regroupe ensuite les items selon la compétence dont ils relèvent pour évaluer spécifiquement chacune des quatre compétences, soit que l'évaluation porte non plus sur l'une de ces quatre compétences, mais sur la compétence générique à laquelle renvoie l'épreuve : compétence à débattre, compétence à présenter un exposé, etc.

Le troisième élément à prendre en compte pour estimer la difficulté des items – le *texte* – pose encore d'autres problèmes, encore plus délicats.

Les items concernant le fonctionnement de la langue portent souvent sur les mêmes textes que ceux servant à évaluer la compréhension. Mais, généralement, ces textes changent en fait totalement de statut : d'objet visé par le processus de compréhension ils deviennent simple « prétexte » aux activités; et bien souvent il n'est absolument pas nécessaire d'avoir compris le texte pour effectuer les tâches demandées. La notion de difficulté du texte conserve-t-elle alors un sens ? Nous ne le pensons pas : c'est la difficulté du contenu – donc, en gros, une opération appliquée à un matériel langagier – qui semble dans ce cas décisive. Mais il nous faudra encore approfondir cette question pour apporter une réponse plus affirmée. Dans d'autres cas, les items concernant le fonctionnement de la langue portent sur des « corpus », à savoir des ensembles de mots, de phrases, voire de textes. Qu'en est-il alors ? Est-il pertinent, et possible, d'établir un indice de difficulté des corpus ? À première vue, nous répondrions positivement. En effet, par expérience, nous observons que certaines tâches d'analyse grammaticale portent sur des corpus « simples » – c'est-à-dire constitués d'exemplaires typiques du problème traité – alors que d'autres portent sur des corpus « complexes », contenant des cas particuliers, voire même, parfois, insolubles... Or, ceci ne nous paraît pas si différent de la situation que nous avons rencontrée avec les textes utilisés dans l'évaluation de la compréhension, où certains correspondent de manière évidente, immédiate, à la forme « typique » d'un genre alors que d'autres présentent des structures bien moins stéréotypées (cf. section 4.2, La structure textuelle). Cette question de la complexité des corpus devra donc être examinée ultérieurement car elle exerce indéniablement une influence sur la difficulté de la tâche demandée aux élèves, quelle que soit par ailleurs l'opération concernée.

Dans la production écrite, la situation est encore différente puisqu'il n'y a tout simplement pas de texte – si ce n'est celui produit par l'élève... Il ne saurait donc y avoir estimation *a priori* de la difficulté du texte. Cependant, on pourrait se demander si la complexité du texte produit – estimée au moyen des critères que nous avons présentés ici – ne devrait pas être prise en compte dans l'évaluation : en effet, placé devant une tâche d'écriture, un élève peut *prendre le risque* de produire un texte complexe et d'être alors confronté à une multitude d'obstacles qu'il ne

³⁹ Ce problème n'est pas propre à la production écrite : nous avons également vu, en effet, que la compréhension était fréquemment évaluée au moyen de tâches d'écriture : résumer, reformuler, etc.

parviendra pas toujours à surmonter; mais un autre élève pourra « se contenter » de répondre à la consigne en produisant des banalités, des idées auxquelles il n'adhère pas vraiment, pour autant que celles-ci soient formulables avec les moyens langagiers qu'il maîtrise... Pour l'évaluation, *a posteriori*, ne faudrait-il pas alors tenir compte de la prise de risque du premier élève ?... C'est dans ce sens que les critères d'analyse de la difficulté des textes pourraient être un indicateur supplémentaire pour l'évaluation de la production des élèves.

Une réflexion quant au rôle joué, dans les autres axes du PER, par le *matériel langagier* (textes ou corpus) devra donc encore être menée.

6. Vers une mesure de la difficulté des items : synthèse

Nos analyses pour estimer la complexité des items nous ont menés, d'abord, à analyser la multiplicité des facteurs en jeu et, ensuite, à prendre des décisions pour estimer leur incidence sur le niveau de difficulté des items. Il en découle une procédure concrète permettant :

- a) d'estimer et/ou de calculer – sur une échelle comportant trois valeurs : -1 (= facile), 0 (= intermédiaire) et 1 (= difficile) – la difficulté de chacun des facteurs finalement retenus pour le contenu, l'enveloppe et la partie mesurable de ceux concernant le texte;
- b) de faire la moyenne des valeurs de manière séparée pour chacun de ces trois éléments (contenu, enveloppe, partie mesurable du texte);
- c) de définir pour, respectivement, le contenu de l'item, son enveloppe et le texte (ou matériel langagier) qui lui est lié, un indice qualitatif de difficulté obtenu par la conversion de la moyenne des valeurs des différents facteurs pris en compte pour les deux premiers et la pondération par les bonus et malus pour le troisième;
- d) de définir un indice qualitatif de difficulté de l'item : nous pensons finalement qu'il n'est pas possible de fondre ces 3 indices en un seul qui résumerait l'estimation de la difficulté de l'item; nous nous contentons par conséquent de considérer comme indice de la difficulté de l'item la simple juxtaposition des indices de ses 3 composantes (contenu, enveloppe et texte).

Au final, cette procédure peut être schématisée de la manière suivante:

Tableau 9 : calcul de l'indice de difficulté de l'item

Calcul de l'indice de difficulté d'un item									
Contenu				Enveloppe			Texte		
Opération		Objet langagier		Consigne	Format de quest.	Matériel à disposition	Voc.	Complexité syntaxique	Structure textuelle
type	nb.	ampleur	statut						
Va	Vb	Vc	Vd	Ve	Vf	Vg	Vh	Vi	Vj
Valeur moyenne Cont = Va + Vb + Vc + Vd 4				Valeur moyenne Env = Ve + Vf + Vg 3			Valeur moyenne Texte = Vh + Vi + Vj 3		
							Indice intermédiaire qualitatif (cf. table, p. 20) = valeur moyenne convertie en indice qualitatif		
Indice qualitatif I_{Cont} (valeur moyenne 1 convertie en indice qualitatif en se référant à la table, p. 20)				Indice qualitatif I_{Env} (valeur moyenne 2 convertie en indice qualitatif en se référant à la table, p. 20)			Indice qualitatif I_{Texte} (indice qualitatif pondéré par les bonus et malus)		
Indice de l'item = juxtaposition des 3 indices qualitatifs $I_{Cont} - I_{Env} - I_{Texte}$									

Si nous reprenons maintenant l'exemple traité précédemment (« Qu'utilise-t-on pour fabriquer l'huile de millepertuis ? »), nous obtenons le tableau suivant :

Tableau 10 : calcul de l'indice de difficulté pour l'item « Qu'utilise-t-on pour fabriquer l'huile de millepertuis ? »

Item										
Contenu				Enveloppe			Texte			
Opération		Objet langagier		Formulation de la consigne	Format de questionnement	Matériel à disposition	Vocabulaire	Complexité syntaxique	Structure textuelle	
Type	Quantité	Ampleur	Statut							
Repérage	Une seule opération requise	Fragment de texte	Information littérale (syntagme situé au début du texte)	Syntaxe simple, mais termes techniques	QCM, mais avec options relativement proches	Absence de matériel	Vocabulaire difficile (12.9% mots absents)	Structures syntaxiques complexes	Structure correspond au genre; Organisation en paragraphes	
Valeurs	-1	-1	0	-1	0	0	0	1	0	0
Valeurs moyennes	-3/4 = -0.75			0/3 = 0			1/3 = 0.33			
							Plutôt difficile (indice intermédiaire)			
Indices de difficulté	Facile (F)			Intermédiaire (I)			Plutôt difficile (d) (2 malus + 2 bonus) ⁴⁰			
Indice de difficulté de l'item	F – I – d									

Une telle procédure, simplifiée dans la mesure du possible, devrait permettre aux acteurs concernés (concepteurs d'épreuves, enseignants, etc.)

- de mieux comprendre ce qu'ils font, la nature de ce qu'ils enseignent ou évaluent;
- de faire des choix en meilleure connaissance de cause, à la fois en ce qui concerne les contenus langagiers à évaluer, l'enveloppe des items d'évaluation (formulation des consignes, formats de questionnement) et les textes.

Ces pistes de travail pourraient donc guider les différents acteurs concernés : les enseignants, par exemple, pourraient s'en servir de manière générale pour analyser la nature des items d'évaluation qu'ils proposent à leurs élèves. Cette procédure pourra également être utilisée pour l'élaboration d'une base de données d'items, dans laquelle l'indication d'un degré de difficulté des items pourrait orienter les usagers. Cette base de données devrait ainsi aider, dans un deuxième temps, les personnes responsables de la conception et de l'élaboration de futures épreuves cantonales.

⁴⁰ Nous avons attribué 2 malus respectivement pour le contenu (peu familier pour les élèves) et pour la syntaxe (constructions syntaxiques plutôt complexes : beaucoup de phrases passives), et 2 bonus pour la structure textuelle (paragraphes organisés autour de questions clairement délimitées) et pour les éléments stylistiques / rhétoriques (nombreuses explications pour les termes difficiles). Au final, l'impact des malus et bonus est nul.

7. En guise de conclusion

Revenant au titre de notre travail, il s'agissait ici de prendre en compte la difficulté des items et des textes, dans le cadre d'une réflexion plus large sur l'évaluation du français. Dans le but d'évaluer, il est nécessaire de bien comprendre, décrire et définir les outils employés – soit, en l'occurrence, les questions proposées et les textes utilisés. En effet, la manière d'évaluer définit, dans une certaine mesure, l'objet à évaluer / l'objet évalué. En ce qui concerne l'évaluation de la compréhension, « *l'évaluation, par son contenu, ses modalités, ses principes organisateurs, véhicule des modèles implicites de ce qu'est la lecture; à ce titre, elle constitue un enjeu stratégique essentiel.* » (Lafontaine, 2003).

En estimant la difficulté que les outils représentent *en eux-mêmes*, nous arrivons à ce qui, à vrai dire, constitue le cœur de nos réflexions : savoir si on évalue vraiment ce qu'on veut évaluer. De fait, la manière d'évaluer entraîne une sorte de « reconstruction » des contenus à évaluer / évalués, une reconfiguration de ce qui est proposé par les prescriptions officielles et qui a déjà été réélaboré et transformé dans l'enseignement. Est-ce qu'on évalue les mêmes contenus si on propose aux élèves des items de repérage d'information ou si on leur demande d'expliquer des inférences de causalité ? Si on leur demande l'interprétation d'une nouvelle ou si on vise l'identification de l'antécédent d'un pronom ? Certainement pas. Toutefois, tous ces « objets » font partie du domaine de la compréhension orale et écrite, et y sont évalués.

Pareillement, les caractéristiques des textes utilisés peuvent faire appel, comme on l'a vu, à une diversité de ressources et de connaissances dont l'évaluateur doit être conscient.

L'estimation de la difficulté des items et des textes va donc bien au-delà d'une mesure sur le niveau de réussite atteint par un élève, une classe, un canton. L'estimation *a priori*, qui nous concerne ici particulièrement, suppose, d'un côté, de mieux cibler les objets d'évaluation en jeu; d'un autre côté, la prise en compte de la difficulté des textes et des items permet d'identifier les obstacles potentiels d'apprentissage pour les élèves, en lien avec les pratiques d'enseignement.

Les analyses pour estimer la complexité des items nous ont menés, d'abord, à analyser la multiplicité des facteurs en jeu et, ensuite, à prendre des décisions pour estimer comment ces facteurs avaient une incidence sur le niveau de difficulté des items. De cette procédure ressortent des pistes concrètes et, dans la mesure du possible, simplifiées pour la conception et l'application d'épreuves pour évaluer le français.

Au final, nous proposons donc une procédure résumée (*cf.* chapitre 6), qui consiste à attribuer une valeur de difficulté à chacun des aspects du contenu, de l'enveloppe, et à une partie de ceux concernant le texte puis à en déduire une valeur moyenne. Chaque valeur est mise en relation avec une table pour en faire un indice de difficulté qualitatif, sauf dans le cas du texte où on passe par un indice intermédiaire qualitatif (déterminé également grâce à la table) qui est ensuite pondéré par des bonus et malus pour au final obtenir l'indice 3.

Nous avons renoncé à définir un indice unique pour la difficulté de l'item en tant que tel et préféré conserver, en les juxtaposant, les indices des 3 composantes de l'item, car cela nous a paru plus pertinent d'un point de vue didactique et docimologique.

Les pistes de travail que nous proposons pourraient ainsi guider les différents acteurs concernés : les enseignants, par exemple, pourraient s'en servir de manière générale pour analyser la nature des items d'évaluation qu'ils proposent à leurs élèves. Cette procédure pourra également être utilisée pour l'élaboration d'une base de données d'items, dans laquelle des indices de difficulté des items pourront orienter les usagers. Cette base de données devrait ainsi aider, dans un deuxième temps, les personnes responsables de la conception et de l'élaboration de futures épreuves cantonales.

Par ailleurs, il reste bien du travail à faire. Dans cette contribution nous nous sommes centrés sur le domaine de la compréhension écrite et orale et nous n'avons qu'esquissé certaines observations concernant les autres axes du PER – production écrite et fonctionnement de la langue en particulier. Dans des travaux à venir il faudrait prendre en compte ces domaines de manière plus précise, notamment dans le but de définir le contenu des items : quels sont les opérations et les objets langagiers concernés ?

En ce qui concerne l'évaluation des textes, nous sommes partis du principe que « *pouvoir évaluer la difficulté est de la première importance dans la mise en œuvre d'une pédagogie de la compréhension* » (Mesnager, 2002, p.30). Nous avons trouvé intéressant – et rassurant ! – de découvrir que les évaluations dites « spontanées » se rapprochaient de l'évaluation basée sur des critères précis, émergeant de la littérature. Toutefois, ce serait une tromperie de penser que n'importe quelle évaluation spontanée suffirait à clarifier tous les éléments susceptibles de poser des problèmes pour la compréhension d'un texte. L'intuition qui émerge de l'expérience de la pratique est précieuse. Mais il est nécessaire de montrer les bases de cette « intuition » et de pouvoir la transformer en une procédure, disponible pour tout le monde, et qui suit des étapes claires. Il s'avère donc nécessaire de systématiser la méthode de travail, afin de stabiliser les pratiques de conception d'épreuves, d'évaluation et même d'enseignement. C'est donc là le but des propositions pour la pratique esquissées (cf. section 4.4).

Bien que, actuellement, aucune « formule de lisibilité » ne semble complètement définitive et absolue pour estimer la difficulté d'un texte, nous espérons avoir, dans ces quelques pages, rendu attentif le lecteur sur l'importance de la prise en compte des facteurs ayant une incidence sur la complexité des textes utilisés pour l'évaluation. Dans de futurs travaux, il s'agira également d'examiner si les « corpus » – souvent liés aux items qui évaluent le fonctionnement de la langue – peuvent être abordés d'une manière comparable à celle présentée ici pour les textes et, en particulier, étudier si on peut envisager une mesure de la complexité des corpus.

Ces pages proposent un essai de systématisation de nos réflexions, dans le cadre de la problématique de l'évaluation en Suisse romande. Une mise en pratique de nos propositions pourrait mieux nous renseigner sur leurs forces et leurs limites et pourrait rendre visibles les ajustements nécessaires aux besoins des différents acteurs. Il s'agit bien ici d'un premier pas.

8. Références bibliographiques

- Adams, B.C., Bell, L.C. & Perfetti, C.A. (1995). A trading relationship between reading skill and domain knowledge in children's text comprehension. *Discourse processes*, 20(3), 307-323.
- Alamargot, D. & Chanquoy, L. (2001). *Through the models of writing*. Dordrecht: Kluwer Academic Publishers.
- Bautier, E. (2015). Quand la complexité des supports d'apprentissage fait obstacle à la compréhension de tous les élèves. *Spirale*, 55, 11-20.
- Bautier, E., Crinon, J., Delarue-Breton, C. & Marin, B. (2012). Les textes composites : des exigences de travail peu enseignées? *Repères : recherches en didactique du français langue maternelle*, 45, 63-80.
- Beaudet, C. (2001). Clarté, lisibilité, intelligibilité des textes : un état de la question et une proposition pédagogique. *Recherches en rédaction professionnelle*, 1(1), 1-19.
- Beck, I.L., Perfetti, C.A. & McKeown, M.G. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of educational psychology*, 74(4), 506-521.
- Best, R., Ozuru, Y., Floyd, R.G. & McNamara, D.S. (2006). Children's text comprehension: effects of genre, knowledge, and text cohesion. In S.A. Barab, K.E. Hay & D.T. Hickey (Eds), *Proceedings of the 7th international conference on learning sciences* (pp. 37-42). Mahwah: Erlbaum.
- Bourquin, J., Carulla, C., Corti, D., De Pietro, J.-F., Kassam, S., Roth, M. & Sánchez Abchi, V. (2015). *Projet EpRoCom : rapport scientifique intermédiaire*. IRDP : Neuchâtel.
- Boyer, J.-Y. (1992). La lisibilité. *Revue française de pédagogie*, 99, 5-14.
- Brookshire, J., Scharff, L.F. & Moses, L.E. (2002). The influence of illustrations on children's book preferences and comprehension. *Reading psychology*, 23(4), 323-339.
- Caccamise, D., Friend, A., Kintsch, W. & Kintsch, E. (2013). Measuring text complexity: poster presented at the Twenty-third annual meeting of the society for text & discourse. Valencia: Universitat de València.
- Candelier, M. (Éd.). (2007). *À travers les langues et les cultures : cadre de référence pour les approches plurielles des langues et des cultures*. Graz : Conseil de l'Europe, Centre européen pour les langues vivantes.
- Conférence intercantonale de l'instruction publique de la Suisse romande et du Tessin (CIIP). (2010). *Plan d'études romand (PER)*. Neuchâtel : CIIP [Page Web]. Accès : <http://www.plandetudes.ch/> (Page consultée le 17.04.2015).
- Conférence suisse des directeurs cantonaux de l'instruction publique (CDIP). (2011). *Compétences fondamentales pour la langue de scolarisation : standards nationaux de formation adoptés par l'Assemblée plénière de la CDIP le 16 juin 2011*. Berne : CDIP.
- Dale, E. & Chall, J.S. (1948). A formula for predicting readability. *Educational research bulletin*, 27(2), 37-54.
- De Landsheere, G. (1963). Pour une application des tests de lisibilité de Flesch à la langue française. *Le travail humain*, 26(1/2), 141-154.
- De Mier, V. & Sánchez Abchi, V. (2014). *Texts to work reading comprehension: challenges for text selection in two socio-cultural contexts with different teaching methods: work presented at the APREC conference*. Strasbourg: APREC.
- Degand, L. & Sanders, T. (2002). The impact of relational markers on expository text comprehension in L1 and L2. *Reading and writing*, 15(7/8), 739-757.
- Englert, C.S. & Hiebert, E.H. (1984). Children's developing awareness of text structures in expository materials. *Journal of educational psychology*, 76(1), 65-74.
- Enright, M., Grabe, W., Koda, K., Mosenthal, P., Mulcahy-Ernt, P. & Schedl, M. (2000). *TOEFL 2000 reading framework*. Princeton: Educational Testing Service.

- Entin, E.B. & Klar, G.R. (1985). Relationships of measures of interest, prior knowledge, and readability to comprehension of expository passages. In B.A. Hutson (Ed.), *Advances in reading/language research* (pp. 9-38). Greenwich: JAI Press.
- Fisher, D.L. (1981). Functional literacy tests: a model of question-answering and an analysis of errors. *Reading research quarterly*, 16(3), 418-448.
- Flesch, R. (1948). A new readability yardstick. *Journal of applied psychology*, 32(3), 221-233.
- Fletcher, J.M. (2006). Measuring reading comprehension. *Scientific studies of reading*, 10(3), 323-330.
- François, T. (2009). Modèles statistiques pour l'estimation automatique de la difficulté de textes de FLE. In *Actes de la 11e Rencontre des étudiants chercheurs en informatique pour le traitement automatique des langues (RECITAL), Senlis, 24-26 juin 2009* (pp. 1-10). Paris : Université, Laboratoire d'informatique de Paris-Nord.
- Gérard, F.-M. & Roegiers, X. (2009). *Des manuels scolaires pour apprendre : concevoir, évaluer, utiliser*. Bruxelles : De Boeck.
- Glenberg, A.M. & Langston, W.E. (1992). Comprehension of illustrated text: pictures help to build mental models. *Journal of memory and language*, 31(2), 129-151.
- Glynn, S.M., Law, M. & Doster, E.C. (1998). Making text meaningful: the role of analogies. In C. Hynd (Ed.), *Learning from text across conceptual domains* (pp. 193-208). Mahwah: Lawrence Erlbaum.
- Goodman, M., Finnegan, R., Mohadjer, L., Krenzke, T. & Hogan, J. (2013). *Literacy, numeracy, and problem solving in technology-rich environments among U.S. adults: results from the Program for the international assessment of adult competencies 2012: first look*. Washington: National Center for Education Statistics [Page Web]. Accès : <http://nces.ed.gov/pubs2014/2014008.pdf> (Page consultée le 24.11.2015).
- Graesser, A.C., Hauff-Smith, K., Cohen, A.D. & Pyles, L.D. (1980). Advanced outlines, familiarity, and text genre on retention of prose. *The Journal of experimental education*, 48(4), 281-290.
- Graesser, A.C., Singer, M. & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological review*, 101(3), 371-395.
- Graesser, A.C., McNamara, D.S. & Kulikowich, J.M. (2011). Coh-metrix providing multilevel analyses of text characteristics. *Educational researcher*, 40(5), 223-234.
- Guay, A. (2011, non publié). *Analyse de la formulation des consignes des épreuves adaptatives (EA) de la 3e année primaire (5e année HarmoS) en français et en mathématiques*. Neuchâtel : IRDP (Rapport de stage).
- Gunning, R. (1952). *The technique of clear writing*. New York: McGraw-Hill.
- Guzmán, R., Jiménez, J.E., Ortiz, M.R., Hernández-Valle, I., Estévez, A., Rodrigo, M. & Hernández, S. (2004). Evaluación de la velocidad de nombrar en las dificultades de aprendizaje de la lectura. *Psicothema*, 16(3), 442-447.
- Guzzetti, B.J., Snyder, T.E. & Glass, G.V. (1992). Promoting conceptual change in science: can texts be used effectively?. *Journal of reading*, 35(8), 642-649.
- Haberlandt, K.F. & Graesser, A.C. (1985). Component processes in text comprehension and some of their interactions. *Journal of experimental psychology: general*, 114(3), 357-374.
- Henny, A.-F. (1998). *De la difficulté de compréhension de la lecture de consignes et de quelques solutions*. Lausanne : Centre vaudois de recherches pédagogiques (CVRP).
- Hess, K. & Hervey, S. (2011). *Tools for examining text complexity*. Dover: NCIEA [Page Web]. Accès : http://www.nciea.org/publication_PDFs/Updated%20toolkit-text%20complexity_KH12.pdf (Page consultée le 24.11.2015).
- Hiebert, E.H. & Pearson, P.D. (2010). *An examination of current text difficulty indices with early reading texts*. Santa Cruz: TextProject (Reading research report 10-01).
- Hunt, K.W. (1970). Recent measures in syntactic development. In M. Lester (Ed.), *Reading in applied transformational grammar* (pp. 179-192). Nueva York: Holt, Rinehart and Wiston.
- Hynd, C. & Guzzetti, B.J. (1998). When knowledge contradicts intuition: conceptual change. In C. Hynd (Ed.), *Learning from text across conceptual domains* (pp. 139-164). Mahwah: Lawrence Erlbaum.

- Just, M.A., Carpenter, P.A. & Woolley, J.D. (1982). Paradigms and processes in reading comprehension. *Journal of experimental psychology: general*, 111(2), 228-238.
- Kandel, L. & Moles, A. (1958). Application de l'indice de Flesch à la langue française. *Cahiers études de radio-télévision*, 19, 253-274.
- Keenan, J.M. & Betjemann, R.S. (2006). Comprehending the gray oral reading test without reading it: why comprehension tests should not include passage-independent items. *Scientific studies of reading*, 10(4), 363-380.
- Kirsch, I.S. (2001a). The framework used in developing and interpreting the International adult literacy survey (IALS). *European journal of psychology of education*, 16(3), 335-361.
- Kirsch, I.S. (2001b). *The international adult literacy survey (IALS): understanding what was measured*. Princeton: Educational Testing Service.
- Kirsch, I.S. (2003). Measuring literacy in IALS: a construct-centered approach. *International journal of educational research*, 39(3), 181-190.
- Labasse, B. (1999). La lisibilité rédactionnelle : fondements et perspectives. *Communication et langages*, 121(1), 86-103.
- Lafontaine, D. (2003). *Comment faciliter, développer et évaluer la compréhension des textes aux différentes étapes de la scolarité primaire* (document envoyé au PIREF en vue de la Conférence de consensus sur l'enseignement de la lecture à l'école primaire, les 4 et 5 décembre 2003). Liège : Université.
- Lavigne, J. (2008). *Les mécanismes d'inférence en lecture chez les élèves de sixième année du primaire* (Doctoral dissertation). Québec : Université Laval.
- Lecocq, P., Leuwers, C., Casalis, S. & Watteau, N. (1996). *Apprentissage de la lecture et compréhension d'énoncés*. Villeneuve-d'Ascq : Presses universitaires du Septentrion.
- Louwerse, M.M. & Graesser, A.C. (2005). Coherence in discourse. In P. Strazny (Ed.), *Encyclopedia of linguistics* (pp. 216-218). Chicago: Fitzroy Dearborn.
- Mannes, S. & St George, M. (1996). Effects of prior knowledge on text comprehension: a simple modeling approach. In B.K. Britton & A.C. Graesser (Eds), *Models of understanding text* (pp. 115-139). Hillsdale: Lawrence Erlbaum.
- Marc, V. & Wirthner, M. (2013). *Développement d'un modèle d'évaluation adapté au PER : rapport scientifique du projet d'épreuves romandes communes*. Neuchâtel : IRDP.
- Matsuoka, W. & Hirsh, D. (2010). Vocabulary learning through reading: does an ELT course book provide good opportunities?. *Reading in a foreign language*, 22(1), 56-70.
- Mayer, R.E., Bove, W., Bryman, A., Mars, R. & Tapangco, L. (1996). When less is more: meaningful learning from visual and verbal summaries of science textbook lessons. *Journal of educational psychology*, 88(1), 64-73.
- McNamara, D.S., Kintsch, E., Songer, N.B. & Kintsch, W. (1996). Are good texts always better?: interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and instruction*, 14(1), 1-43.
- McNamara, D.S. & Kintsch, W. (1996). Learning from texts: effects of prior knowledge and text coherence. *Discourse processes*, 22(3), 247-288.
- McNamara, D.S., Graesser, A.C. & Louwerse, M.M. (2012). Sources of text difficulty: across genres and grades. In J.P. Sabatini, E. Albro & T. O'Reilly (Eds), *Measuring up: advances in how we assess reading ability* (pp. 89-116). Lanham: Rowman and Littlefield Publishers.
- Mesmer, H.A., Cunningham, J.W. & Hiebert, E.H. (2012). Toward a theoretical model of text complexity for the early grades: learning from the past, anticipating the future. *Reading research quarterly*, 47(3), 235-258.
- Mesnager, J. (2002). Pour une étude de la difficulté des textes : la lisibilité revisitée. *Le français aujourd'hui*, 137, 29-40 [Page Web]. Accès : <http://www.cairn.info/revue-le-francais-aujourd-hui-2002-2-page-29.htm> (Page consultée le 24.11.2015).
- Mesnager, J. & Bres, S. (2008). *Évaluer la difficulté des textes : cycle 2-3, 6e-5e* (CD-Rom). Paris : Nathan.
- Mosenthal, P.B. & Kirsch, I.S. (1998). A new measure for assessing document complexity: the PMOSE/IKIRSCH document readability formula. *Journal of adolescent & adult literacy*, 41(8), 638-657.

- OCDE. (2013). *Cadre d'évaluation et d'analyse du cycle PISA 2012*. Paris : OCDE [Page Web]. Accès : http://www.keepeek.com/Digital-Asset-Management/ocd/education/cadre-d-evaluation-du-pisa-2012_9789264190559-fr#page1 (Page consultée le 24.11.2015).
- OECD. (2012). *Literacy, numeracy and problem solving in technology-rich environments: framework for the OECD survey of adult skills*. Paris: OECD [Page Web]. Accès : http://www.keepeek.com/Digital-Asset-Management/ocd/education/literacy-numeracy-and-problem-solving-in-technology-rich-environments_9789264128859-en#page1 (Page consultée le 24.11.2015).
- Pini, G., Reith, E., Weiss, L. & Bugniet, F. (2006). *Guide méthodologique pour l'évaluation et la mesure en éducation*. Genève : Développement et innovation pédagogique au cycle d'orientation(DIPCO).
- Rawson, K.A. & Kintsch, W. (2002). How does background information improve memory for text content?. *Memory & cognition*, 30(5), 768-778.
- Schumacker, R.E. & Lomax, R. (2004). *A beginner's guide to structural equation modeling*. East Sussex: Psychology Press.
- Vrignaud, P. (2006). La mesure de la littéracie dans PISA : la méthodologie est la réponse, mais quelle était la question?. *Revue française de pédagogie*, 157, 27-41 [Page Web]. Accès : <http://rfp.revues.org/409> (Page consultée le 24.11.2015).

Annexe 1

Textes analysés

(tirés du Projet HarmoS, Langue de scolarisation, 6^e, 303, 411 et 440)

1. Huile de millepertuis – médicament phytothérapeutique – Information destinée aux patients

Huile de millepertuis – médicament phytothérapeutique – Information destinée aux patients

1. Qu'est-ce que l'huile de millepertuis A. Vogel et quand est-elle utilisée?

Pour fabriquer l'huile de millepertuis A. Vogel, les fleurs et les boutons de millepertuis fraîchement récoltés et broyés sont mélangés avec de l'huile de tournesol et longuement exposés au soleil. Ce procédé donne à la préparation cette couleur rouge lumineuse si typique, qui devient phosphorescente sous l'influence de la lumière. Ce sont des substances telles que l'hypéricine qui donnent à l'huile sa couleur rouge. L'huile de millepertuis A. Vogel est indiquée pour les applications externes suivantes: coupures superficielles, lésions cutanées (écorchures, cicatrices) et brûlures de surface réduite (premier degré); prévention des escarres (blessures) dues à un

séjour au lit prolongé des patients (décubitus). Anti-inflammatoire.

2. De quoi faut-il tenir compte en dehors du traitement?

En cas d'aggravation locale ou d'absence d'amélioration, consulter un médecin! En cas de dégradation de l'état général (p.ex. fièvre), consulter immédiatement un médecin.

3. Quand l'huile de millepertuis A. Vogel ne doit-elle pas être utilisée ou seulement avec précaution?

L'huile de millepertuis A. Vogel ne doit pas être utilisée sur des plaies sanglantes étendues. Elle ne doit pas être utilisée en cas d'allergie connue à l'*Hypericum perforatum* (millepertuis perforé). La préparation n'est pas adaptée à un usage interne (voie orale).

4. Comment utiliser l'huile de millepertuis A. Vogel?

Usage externe: frictionner délicatement plusieurs fois par jour les zones touchées avec quelques gouttes d'huile. Pour les petites blessures, tamponner 2 à 3 fois par jour les bords enflammés de la plaie avec de l'huile de millepertuis A. Vogel.

5. Quels effets indésirables l'huile de millepertuis A. Vogel peut-elle avoir?

En cas d'emploi approprié, aucun effet secondaire n'a été observé à ce jour.

Source: notice d'emballage huile de millepertuis, A. Vogel, Bioforce AG, CH-9325 Roggwil, TG

2. Recette de cuisine

J'ai pu enregistrer, dans le bac à légumes de mon réfrigérateur, une conversation émouvante entre une pomme golden et une pomme de terre. Voici ce document étonnant:

5 «Ah, chère madame», dit la pomme golden à la pomme de terre, «il faut que je vous raconte ce qui est arrivé à ma meilleure amie, une pomme de reinette que je connais depuis l'école maternelle. C'est absolument é-pou-van-ta-ble! Figurez-vous qu'on en a fait de la marmelade! Deux individus se sont emparés d'elle, un homme tout en blanc et une jeune femme avec un grand tablier bleu. La femme a pris un couteau spécial et elle a déshabillé complètement ma copine. Imaginez un
10 peu: toute nue sur une table de cuisine! L'homme, lui, l'a découpée en quatre, comme ça, zic zac, en deux coups de couteau. Et il lui a arraché le cœur avec tous les pépins.»

«Arrêtez, arrêtez, c'est horrible!» s'écria la pomme de terre en se bouchant, stupidement, les yeux.

15 «Ce n'est pas fini», poursuivit la pomme golden. «Ils ont jeté la malheureuse dans une casserole, avec plein d'autres copines. Ils ont ajouté un tout petit peu d'eau et, hop! ils ont allumé le gaz. Au bout de deux minutes, avec la vapeur, c'était pire que dans un sauna.»

«Oh, un sauna», dit la pomme de terre, «c'est bon pour la santé.»

20 «Eh bien», répliqua la pomme golden, «je voudrais bien vous y voir!
Au bout de vingt minutes environ, les copines étaient toutes fondues, une
vraie bouillie. Alors l'homme a pris une cuillère en bois, il a rajouté 50
grammes de sucre et un peu de cannelle et il a bien remué le tout.»

 «Hm hm», murmura la pomme de terre, «ça devait sentir bon!»

25 «Oh, vous! vous n'avez pas de cœur!» s'écria, indignée, la pomme
golden.

 Et elle éclata en sanglots.

 «Vous savez», répondit la pomme de terre, «je pourrais vous raconter
des choses plus horribles encore. Figurez-vous que mon fiancé a été
30 transformé en purée! Voilà comment ça s'est passé: un homme est venu le
chercher...»

 Malheureusement, l'enregistrement s'arrête là. Une panne de courant,
probablement.

In Histoires pressées, Bernard Friot
Editions Milan, 1999

3. L'arbre aux hérissons

L'arbre aux hérissons

A:

L'épisode présenté ci-dessous est un souvenir authentique de l'auteur qui le raconte à son fils parce qu'il sait qu'il adore les animaux. Il s'agit de l'extrait d'une lettre adressée en 1932 par Antonio Gramsci, homme politique, journaliste et lettré, antifasciste, à son fils Delio, qui était âgé de huit ans à l'époque et qui n'avait que deux ans lorsque son père fut jeté en prison. Cette lettre a été d'ailleurs écrite en prison: nous sommes à l'époque du régime fasciste, quand le gouvernement emprisonnait ses adversaires politiques. Gramsci avait été condamné à vingt ans de réclusion; les souffrances et privations endurées pendant son incarcération ont entraîné sa mort en 1937. Au cours de cette période, il n'a eu de cesse de rester le plus proche possible des personnes qui lui étaient chères en écrivant des lettres qui ont été réunies et publiées.

(Extrait de l'anthologie *La Lecture 1* (Bologne, Zanichelli, 1969) établie par Italo Calvino – Giambattista Salinari)

B:

Un soir d'automne, il faisait déjà nuit, mais l'obscurité était illuminée par une pleine lune resplendissante; je suis allé avec un de mes amis dans un champ plein d'arbres fruitiers et surtout de pommiers. Nous nous sommes cachés dans un buisson, contre le vent.

5 Tout à coup des hérissons sont arrivés, cinq en tout: deux gros et trois petits. En file indienne, ils ont avancé en direction des pommiers et après avoir un peu roulé dans l'herbe, ils se sont mis au travail: s'aidant du museau et des pattes, ils faisaient rouler les pommes que le vent avait fait tomber des arbres, pour ensuite les rassembler bien rangées au même endroit.

10 Les pommes tombées ne semblaient toutefois pas suffire; le plus grand des hérissons, le nez en l'air, a regardé autour de lui et a jeté son dévolu sur un arbre bien incliné. Suivi de sa femme, il a grimpé jusqu'aux premières branches.

Ils se sont posés sur une branche bien lourde de fruits et ont commencé à se balancer en rythme: leurs mouvements ont entraîné la branche qui a commencé à osciller de plus en plus fort, avec de brusques saccades, ce qui a fait tomber beaucoup de pommes.

15 Après les avoir rassemblés près des autres, tous les hérissons, grands et petits, se sont mis en boule en hérissant leurs piquants pour ensuite s'étendre sur les fruits qui restaient accrochés aux pointes: certains n'en avaient que peu (les bébés hérissons), mais le père et la mère avaient réussi à embrocher chacun sept ou huit pommes.

Pendant qu'ils retournaient à leur tanière, nous sommes sortis de notre cachette, nous avons mis les hérissons dans un petit sac et les avons emportés chez nous.

20 J'ai gardé le père et deux bébés hérissons pendant plusieurs mois. Ils vivaient en liberté dans la cour où ils chassaient toutes sortes de bestioles, blattes, coccinelles, etc. et mangeaient des fruits et des feuilles de salade.

Ils aimaient beaucoup les feuilles fraîches ce qui m'a permis de les apprivoiser un peu; à la fin, ils ne se roulaient plus en boule quand ils voyaient des gens.

Annexe 2

L'indice de difficulté des items et des textes : présentation synthétique de la procédure d'estimation

Le calcul d'un indice décrivant, a priori, la difficulté d'un item ou d'un texte s'inscrit dans le cadre du projet de développement d'une base données mettant à disposition des matériaux d'évaluation didactiquement fondés et validés – à savoir des unités d'évaluation et, dans la plupart des cas, des matériaux langagiers sur lesquels portent ces unités d'évaluation⁴¹ – et d'une description de ces matériaux.

L'item comme unité de la base de données

L'item constitue (avec les textes, cf. *infra*) l'unité centrale d'une telle base de données. Cette notion n'est pas simple à définir. À des fins opératoires, nous considérons qu'un item est constitué par une ou plusieurs questions dépendant d'une même consigne et liées à un même matériel, visant à mesurer des objets globalement « de même nature ». L'item est autonome, c'est-à-dire qu'il peut être utilisé indépendamment de tout autre item. Il correspond ainsi à une tâche qui permet de donner des indices (traces, données) sur ce que les élèves savent et savent faire (compétences, connaissances, attitudes ou références culturelles).⁴²

La description des items

Chaque item doit faire l'objet d'une description afin que les utilisateurs (concepteurs de tests ou de moyens d'enseignement, enseignants...) puissent effectuer des choix en fonction de leurs besoins. La description doit prendre en compte les différents éléments qui caractérisent un item : l'objet langagier qu'il permet d'évaluer, la consigne et le format de questionnement, les matériaux langagiers sur lesquels il porte ainsi que son contexte d'utilisation et l'estimation de sa difficulté.

L'estimation de la difficulté de l'item

L'objectif consiste par conséquent à proposer un outil opératoire qui guide les enseignants et les concepteurs d'épreuves, à quelque niveau que ce soit (tests de référence nationaux, EpRoCom, épreuves cantonales...), dans l'analyse et l'estimation de la difficulté des items et qui les aide à les élaborer ou les choisir.

Nous considérons que trois composantes distinctes de l'item – dans lesquelles les éléments à prendre en compte qui en relèvent seront intégrés – doivent être prises en considération : le contenu de l'item, son enveloppe et, s'il y en a un, le matériel langagier (le plus souvent un texte, parfois un corpus) qui lui est attaché. Nous parlons de contenu pour décrire les opérations

⁴¹ Plusieurs unités d'évaluation peuvent de plus être réunies au sein d'« épreuves ».

⁴² Nous sommes conscients des limites de cette définition qui a avant tout, pour nous, une valeur heuristique et doit nous permettre d'élaborer une base de matériaux opératoires pour une évaluation au niveau romand (cf. chapitre 1).

mobilisées et les *objets* langagiers sur lesquels elles portent et pour nous référer aux objectifs du plan d'études qui sont concernés (cf. annexe 3). L'enveloppe concerne plusieurs facteurs, tels que la consigne et le contexte de la tâche qui est dévolue à l'élève, le format de questionnement et le matériel complémentaire mis à disposition pour résoudre la tâche. Le matériel langagier est traité de manière différente, partiellement autonome, dans la mesure où il possède un statut un peu particulier : en effet, un même matériel langagier (texte ou corpus) est le plus souvent lié à plusieurs items et il possède par conséquent une relative autonomie par rapport à ceux-ci.

Nous avons de plus renoncé, finalement, à fournir un indice unique de difficulté et préféré juxtaposer, pour l'item, les indices retenus pour chacune de ces trois composantes. De plus, pour l'indice concernant le texte, nous distinguons deux modalités de calcul : l'une, plus « lourde », destinée plutôt aux concepteurs de tests et/ou de moyens d'enseignement; l'autre, d'un usage plus simple, destinée plutôt aux enseignants désirant se faire une idée de la complexité des matériaux qu'ils utilisent dans leur enseignement, notamment au moment d'évaluer leurs élèves. On notera d'ailleurs que le calcul de l'indice de difficulté des textes peut être intéressant pour lui-même, lorsqu'il s'agit par exemple de choisir un texte à travailler en classe et qu'on souhaite se faire une idée de sa difficulté.

Tableau 1 : calcul de l'indice de difficulté de l'item

Calcul de l'indice de difficulté d'un item		
Contenu	Enveloppe	Texte
V1	V2	V3
Indice de l'item = juxtaposition des 3 indices : V1-V2-V3		

Dans chacune des trois composantes, différents facteurs sont évalués, soit via un calcul soit via une estimation qualitative (facile, intermédiaire, difficile)⁴³. L'estimation globale, pour chaque composante, se fait alors en trois étapes :

- a) dans un premier temps, les estimations qualitatives sont transformées en valeurs chiffrées afin de pouvoir calculer ensuite, pour cette composante, une valeur moyenne :

facile	→	-1
intermédiaire	→	0
difficile	→	1

- b) dans un 2^{ème} temps, la valeur moyenne est calculée en faisant la somme des valeurs de chacun des facteurs divisée par le nombre de facteurs :

$$\frac{\text{[somme des valeurs des facteurs]}}{\text{Nb de facteurs}} = \text{valeur moyenne (située entre -1 et 1)}$$

⁴³ Pour certains facteurs de la complexité du texte notamment.

- c) enfin, dans un 3^{ème} temps, la valeur moyenne obtenue est retransformée en valeur qualitative – celle qui sera finalement retenue pour qualifier chacune des trois composantes – selon une nouvelle table de conversion, plus fine :

Tableau 2 : table pour l'estimation de la difficulté

Valeur	Indice de difficulté
-0.61 à -1	Facile (F)
-0.21 à -0.6	plutôt facile (f)
-0.2 à 0.2	Intermédiaire (I)
0.21 à 0.6	plutôt difficile (d)
0.61 à 1	Difficile (D)

Pour une présentation détaillée du mode de calcul pour, respectivement, le contenu, l'enveloppe et le matériel langagier – en l'occurrence le texte – voir les chapitres correspondants dans le document.

En résumé

La procédure permettant d'estimer la difficulté d'un item – ou, plutôt, de ses trois composantes retenues ici – peut être synthétisée ainsi :

- estimer et/ou calculer – sur une échelle comportant trois valeurs : -1 (= facile), 0 (= intermédiaire) et 1 (= difficile) – la difficulté de chacun des facteurs finalement retenus pour le contenu, l'enveloppe et la partie mesurable de ceux concernant le texte;
- faire la moyenne des valeurs de manière séparée pour chacun de ces trois éléments (contenu, enveloppe, partie mesurable du texte);
- définir pour, respectivement, le contenu de l'item, son enveloppe et le texte (ou matériel langagier) qui lui est lié, un indice qualitatif de difficulté obtenu par la conversion de la moyenne des valeurs des différents facteurs pris en compte pour les deux premiers et la pondération par les bonus et malus pour le troisième;
- définir un indice qualitatif de difficulté de l'item : comme nous considérons finalement qu'il n'est pas possible de fondre ces 3 indices en un seul qui résumerait l'estimation de la difficulté de l'item, nous définissons l'indice de la difficulté de l'item par la simple juxtaposition des indices de ses 3 composantes (contenu, enveloppe et texte).

Tableau 3 : calcul de l'indice de difficulté de l'item

Calcul de l'indice de difficulté d'un item									
Contenu				Enveloppe			Texte		
Opération		Objet langagier		Consigne	Format de quest.	Matériel à disposition	Voc.	Complexité syntaxique	Structure textuelle
type	nb.	ampleur	statut						
Va	Vb	Vc	Vd	Ve	Vf	Vg	Vh	Vi	Vj
Valeur moyenne Cont = Va + Vb + Vc + Vd 4				Valeur moyenne Env = Ve + Vf + Vg 3			Valeur moyenne Texte = Vh + Vi + Vj 3		
							Indice intermédiaire qualitatif (cf. table, p. 20) = valeur moyenne convertie en indice qualitatif		
Indice qualitatif I_{Cont} (valeur moyenne 1 convertie en indice qualitatif en se référant à la table, p. 20)				Indice qualitatif I_{Env} (valeur moyenne 2 convertie en indice qualitatif en se référant à la table, p. 20)			Indice qualitatif I_{Texte} (indice qualitatif pondéré par les bonus et malus)		
Indice de l'item = juxtaposition des 3 indices qualitatifs $I_{Cont} - I_{Env} - I_{Texte}$									

Annexe 3

Lexique

Opérations : activités cognitives de l'élève appliquées à un objet langagier et hiérarchisées selon leur difficulté intrinsèque (repérer x < inférer x < interpréter x...).

Objets langagiers : objets auxquels s'appliquent les opérations, caractérisés par leur ampleur (mot < segment < texte...) et leur statut (littéral < figuré / implicite).

Formulation de la consigne : estimée selon sa complexité syntaxique et lexicale, ainsi que par la présence ou non d'un contexte facilitant / complexifiant la compréhension.

Format de questionnement : forme du questionnement appliqué à l'objet évalué, de complexité intrinsèque variable (QCM < questions à réponse brève < questions à réponse textuelle...).

Matériel complémentaire à disposition : présence ou non d'un matériel qui facilite / complexifie la compréhension.