

# Aides à la compréhension de textes scientifiques par des élèves de onze ans

Brigitte Marin, Patrick Avel, Jacques Crinon, Denis Legros, IUFM de Créteil

Le but de cette recherche est d'étudier des moyens informatiques d'aide à la compréhension des textes scientifiques et de construire une base de données textuelles d'aide à la compréhension de ces textes. Dans cette expérience précise, nous testons deux fonctionnalités hypertextuelles d'aide à la compréhension. Le premier type d'aide fournit des ajouts d'informations par liens hypertextes qui renvoient au contenu du texte (base de texte). Le deuxième type d'aide consiste en ajouts d'informations qui ne figurent pas dans le texte mais sont activées par le texte et renvoient au « modèle mental ».

La première recherche sous forme papier crayon dont nous présentons ici les données principales a pour but de tester et de valider expérimentalement l'effet de ces aides.

## 1. Cadre théorique

Afin d'établir la signification globale du texte, le lecteur doit mettre en cohérence (1) des informations présentes dans le texte, mais qui doivent être explicitées et comprises et donc qui nécessitent des développements explicatifs supplémentaires; (2) des informations absentes du texte, mais qui doivent être inférées pour créer cette cohérence. Or, dans un texte scientifique, l'élève n'a pas toujours les connaissances nécessaires sur le domaine pour faire ces inférences. C'est pourquoi les ajouts d'informations sont nécessaires pour faciliter ce type d'activité inférentielle indispensable à la compréhension scientifique du domaine sur lequel porte le texte.

Les travaux récents sur la compréhension du texte scientifique ont en effet montré l'extrême difficulté pour les non-experts du domaine à comprendre un tel texte (Otero, León & Graesser, 2002). Et de nombreux travaux se sont intéressés à ce type d'activité inférentielle (Van den Broek, Virtue, Everson, Tzeng & Sung, 2002).

La plupart des systèmes hypertextes d'aide à la compréhension utilisent surtout le premier type de fonctionnalité, c'est-à-dire ajoutent des informations qui ne sont pas

forcément pertinentes pour faciliter la compréhension. Le but de notre recherche est donc de valider les types de fonctionnalités les plus utiles à l'élève pour l'aider à comprendre ce qu'il lit, c'est-à-dire construire la cohérence de la représentation sous-jacente au texte.

Cette recherche sur l'aide informatique à la mise en place des processus de construction de la cohérence de la signification du texte et du domaine évoqué par le texte se donne pour objectif à moyen terme de contribuer au développement d'une didactique du texte et de l'apprentissage des connaissances évoquées par le texte qui s'appuie sur le fonctionnement cognitif de l'apprenant. Quand on a compris les difficultés de compréhension d'un texte et les difficultés de compréhension d'un domaine scientifique évoqué par un texte, on est mieux à même de mettre en place des stratégies d'apprentissage qui répondent aux besoins réels de l'élève.

## **2. Méthode**

### **2.1. Participants**

L'épreuve a été passée en avril-mai 2003 dans huit classes de la dernière année de l'école primaire situées dans plusieurs écoles d'une zone d'éducation prioritaire de la banlieue parisienne.

Les élèves concernés présentent pour la plupart des difficultés de compréhension des textes, identifiées lors de tests d'évaluation nationale, liées à un bagage lexical faible et à des connaissances préconstruites limitées ne permettant pas toujours d'établir les inférences nécessaires. Leurs difficultés s'avèrent particulièrement saillantes en présence de textes scientifiques, moins facilement compréhensibles que les textes narratifs où sont mises en œuvre des chaînes d'actions plus en rapport avec l'expérience de la vie quotidienne de l'enfant (Otero, León & Graesser, 2003). Par ailleurs, la spécificité du lexique scientifique, étrangère à l'univers langagier des élèves, crée un obstacle supplémentaire à la compréhension dans la mesure où un terme technique ne peut pas toujours être élucidé par le seul recours au contexte.

Trois groupes de 24 élèves issus de six classes constituent nos trois groupes expérimentaux. Chaque groupe est divisé en deux sous-groupes caractérisés par leur niveau de compétence en compréhension (bons et faibles lecteurs). Ces niveaux sont déterminés par les

résultats à deux tâches de diagnostic cognitif (épreuve de remise en ordre et épreuve de hiérarchisation de l'information en fonction de son importance relative, issues de la batterie Diagnos-Lecture mise au point par Guy Denhière et son équipe).

## **2.2. Tâches et conditions expérimentales**

Tous les participants sont soumis à trois tâches successives, au cours d'une même séance de 45 minutes : lecture d'un texte d'une vingtaine de lignes sur le dérèglement du climat, puis, après une courte tâche distractive, rappel écrit libre du texte, et réponse à cinq questions ouvertes. Seuls les rappels seront analysés dans le cadre de cette communication.

Les tâches sont proposées dans trois conditions expérimentales correspondant aux groupes d'élèves suivants :

- G1 : groupe lisant le texte « authentique » nu ;
- G2 : groupe lisant le même texte accompagné de six notes explicitant le vocabulaire difficile, c'est-à-dire enrichissant la « base de texte »;
- G3 : groupe lisant le même texte accompagné de six notes apportant des informations non présentes dans le texte mais appartenant au « modèle mental du texte ».

## **2.3. Le matériel**

### ***2.3.1. Le texte***

Le dérèglement du climat entraîne une recrudescence de catastrophes naturelles qui se manifestent sous la forme de coups de vent, tempêtes, cyclones, tornades, pluies diluviennes, inondations, éboulements, coulées de boue et sécheresses qui n'épargnent pas les centres urbains. Elles toucheront en particulier les régions équatoriales très peuplées, pauvres, et donc moins aptes à réagir efficacement. L'Europe et la France ne sont pas à l'abri des effets du dérèglement mondial du climat. Les tempêtes de décembre 1999 et les inondations de l'hiver 2000-2001 sont là pour le rappeler. En France, le risque d'inondations touche aujourd'hui plus d'une commune sur quatre et provoque en moyenne plus d'une quinzaine de morts par an.

L'augmentation des sécheresses, selon la WWF, entraîne la perte chaque année de 10

millions d'hectares, en particulier dans les régions tropicales qui deviennent plus chaudes et plus sèches, notamment la zone aride qui s'étend de l'Afrique de l'Ouest à l'Indonésie. La multiplication des sécheresses gagne aujourd'hui l'Europe. Les zones arides du pourtour méditerranéen devraient s'étendre sous l'effet de la combinaison température et évaporation l'été, précipitations massives et érosion l'hiver. En Grèce, le débit du plus grand fleuve du pays, l'Acheloos, a baissé de 40 % en quatre ans. Mais c'est en Afrique, continent le plus vulnérable, où la désertification ne fait que progresser, que les conséquences pour les populations se feront le plus sentir. Selon l'ONU, 250 millions de personnes ont souffert de famine et de malnutrition en 2000. Et ce chiffre pourrait grimper à 900 millions dans quelques décennies.

### *2.3.2. Caractéristiques du texte*

Le texte est un écrit documentaire traitant d'un sujet scientifique, sous la forme d'un article de journal. Il évoque le « dérèglement du climat » et ses effets : catastrophes naturelles diverses, qui ont elles-mêmes des effets sur les populations. Plusieurs exemples de catastrophes sont développés, dans différents pays : tous les pays sont touchés, même si certains (zones arides du pourtour méditerranéen, régions tropicales pauvres) le sont plus que d'autres.

Le texte décrit un ensemble de causes et de conséquences des catastrophes naturelles, ce qui se caractérise par un ensemble d'informations enchaînées causalement. Le texte, pour être compris, nécessite que le lecteur soit capable de comprendre les enchaînements causaux et donc de faire les inférences causales.

Construire cette représentation du contenu du texte nécessite non seulement la prise en compte des relations logiques explicites (les mots « entraînent », « provoquent », « effets », « conséquences » marquent bien l'enchaînement « dérèglement du climat » → « catastrophes naturelles » -> « conséquences pour les populations », « donc » ou « sous l'effet de » aident à repérer des éléments de complexification de ce système causal, le verbe « toucher » introduit explicitement la multiplication des lieux concernés), mais aussi l'activation de connaissances non explicitées dans le texte et nécessaires à la construction de sa cohérence causale. Ainsi, si les catastrophes naturelles « n'épargnent pas les centres urbains », c'est qu'elles atteignent toutes les zones du territoire, et à plus forte raison les campagnes. La pauvreté, notée dans le

texte comme facteur aggravant, ne peut être comprise comme telle que si le lecteur a à l'esprit les préventions possibles, sinon des catastrophes, du moins de leurs conséquences humaines. Le rapport entre la sécheresse, la perte des terres cultivables et les famines ne peut être fait que si l'on sait que les plantes sur lesquelles repose notre alimentation ont besoin d'eau pour pousser. La mise en rapport entre les pluies diluviennes, la perte des terres cultivables et les famines repose sur la connaissance des effets de l'érosion sur les sols (le « trou sémantique » est ici la disparition de la terre emportée par l'eau), alors que les représentations spontanées des lecteurs sont plutôt celles d'inondations rendant les champs inaccessibles...

### **2.3.3. Les notes**

Six notes sont proposées<sup>1</sup>. Dans la condition 2, elles présentent des élucidations lexicales : définitions et reformulations qui développent, décrivent et exemplifient.

Dans la condition 3, elles concernent la chaîne des relations causales et sont constituées d'informations absentes du texte. Elles permettent aux élèves de faire les inférences nécessaires à la construction de la cohérence de la signification du texte. Elles reformulent à chaque fois une relation cause conséquence et/ou introduisent les informations à inférer.

### **2.4. Variables dépendantes et hypothèses**

Les variables dépendantes sont les unités d'information qui constituent le rappel. Elles ont été analysées selon les catégories suivantes :

- informations vraies/fausses ;
- informations catégorisées par leur place dans la chaîne causale :
  - informations ponctuelles renvoyant à la microstructure et qui ne s'insèrent pas dans une chaîne causale ;
  - informations renvoyant à la macrostructure et qui s'insèrent dans une chaîne causale ;
  - informations présentées sous forme générique.

Les différentes catégories d'informations font l'objet d'un traitement différent et nous

faisons les hypothèses suivantes sur les différents traitements qu'elles mettent en jeu.

On s'attend à un effet des informations ajoutées variable selon le type de notes : un rappel plus complet, plus riche en relations causales :

- chez les participants du groupe G2 que chez ceux du groupe G1;
- chez les participants du groupe G3 que chez ceux des deux autres groupes.

On s'attend également à des différences de résultats entre les groupes plus marquées pour les participants de niveau faible que pour ceux de niveau fort.

### **3. Résultats**

Nous avons réalisé trois analyses :

- la première présente les informations rappelées en fonction de leur valeur de vérité dans les différents groupes ;
- la deuxième présente les informations en fonction de leur rapport à la causalité ;
- la troisième présente la valeur de vérité des informations causales (catégorie 3).

#### **3.1. Première analyse**

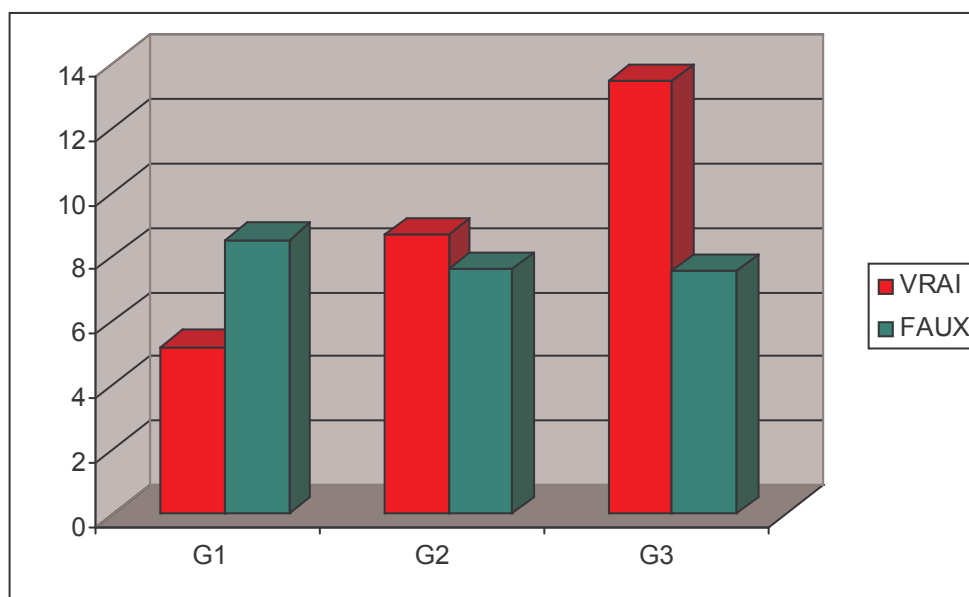
Une première analyse concerne le nombre de propositions rappelées, en fonction de la valeur de vérité (vrai/faux) des informations produites par les élèves.

Les participants du groupe G3 produisent significativement plus d'informations justes (13,5) que ceux du groupe G2 (8,71) et que ceux du groupe G1 (5,21) (voir figure 1). Les deux types de notes ont un effet sur le rappel et, par hypothèse, sur la compréhension. Mais cet effet apparaît plus important lorsque les participants ont à leur disposition des notes qui permettent de faire des inférences.

Cependant, on n'observe pas d'effet significatif du facteur Niveau, tel qu'il a été évalué par nos tâches de diagnostic.

L'interaction entre les facteurs Groupe et Valeur de vérité des informations est significative. Alors que les participants du groupe G1 rappellent plus d'informations fausses,

les participants des deux autres groupes rappellent davantage d'informations vraies.



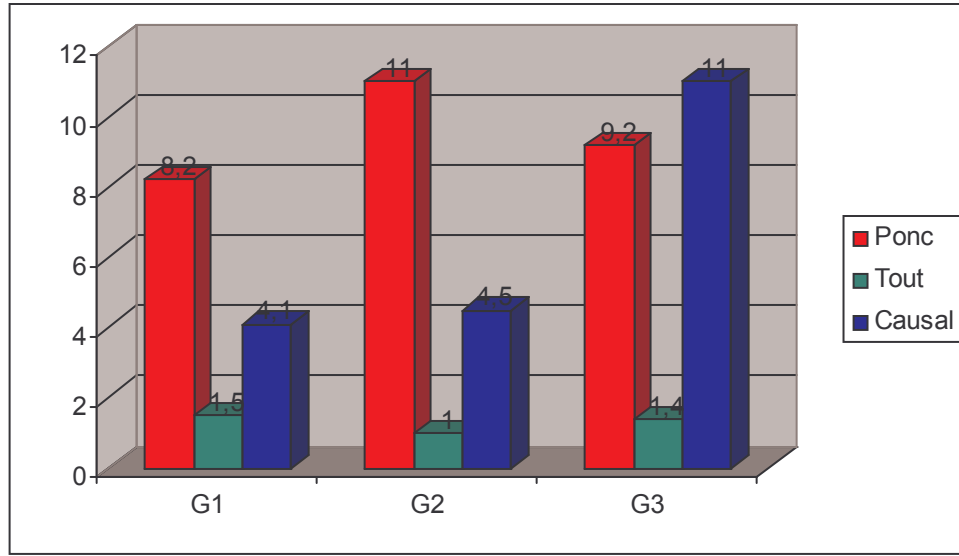
**Figure 1 : Nombre moyen de propositions vraies et fausses rappelées en fonction du groupe**

### 3.2. Deuxième analyse

Une deuxième analyse concerne le nombre de propositions rappelées en fonction de leur type de relations selon les groupes.

Les participants du groupe G3 rappellent significativement plus d'informations renvoyant à la macrostructure et qui s'insèrent dans la chaîne causale (11) que ceux du groupe G2 (4,5) et que ceux du groupe G1 (4,1) (voir figure 2).

On n'observe pas non plus d'interaction avec le facteur Niveau des élèves.

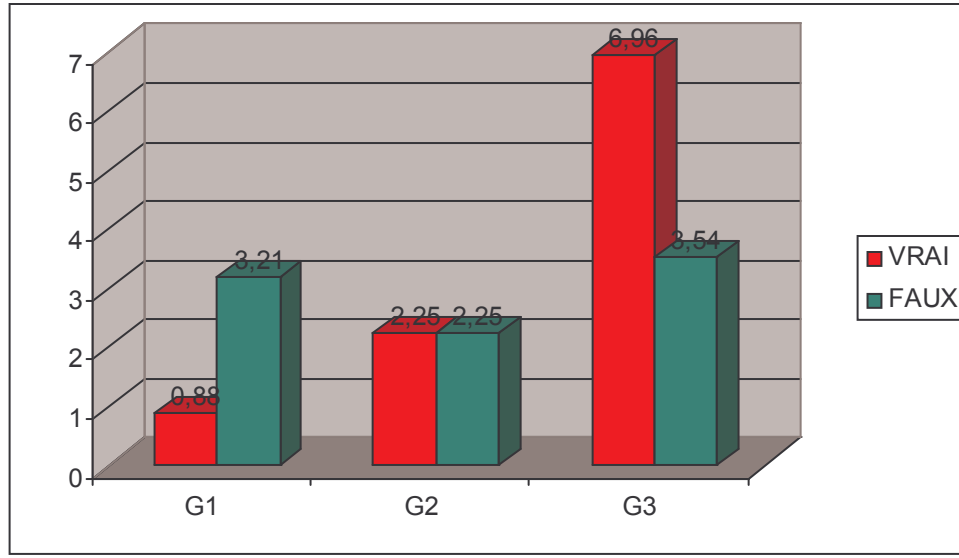


**Figure 2 : Nombre moyen de propositions de différents types rappelées en fonction du groupe**

### 3.3. Troisième analyse

La troisième analyse concerne l'étude de la valeur de vérité des propositions de type causal. Seules les informations de type causal font l'objet de l'étude.

Contrairement aux groupes G1 et G2, le groupe G3 produit plus d'informations mettant en jeu une relation causale vraie que de relations causales fausses (voir figure 3).



**Figure 3 : Nombre moyen de propositions causales vraies et fausses produites en fonction du groupe**

#### 4. Discussion et conclusions

Les résultats présentés confirment les principales hypothèses et indiquent que les informations ajoutées dans des notes et renvoyant au modèle mental sous-jacent au texte favorisent le rappel, améliorent la cohérence du texte rappelé et renforcent la cohérence des informations fournies par le texte et du modèle mental sous-jacent à celui-ci.

Nous situant à l'interface des représentations cognitives et des représentations langagières, nous avons surtout analysé le contenu sémantique du texte rappelé et par hypothèse compris et nous avons mis l'accent sur ses contenus sémantiques et non sur la forme textuelle du rappel. Ce travail de la mise en mots sera analysé ailleurs.

L'aide à l'activation des connaissances causales favorise non seulement la compréhension du texte scientifique, mais elle favorise aussi la qualité des rappels. De plus, dans la mesure où elle favorise l'activation des connaissances renvoyant à la base de texte, mais également au modèle mental et donc aux connaissances évoquées par le texte, elle favorise aussi la construction des connaissances évoquées par le texte.

Cette expérience princeps permet ainsi, dans le cadre de la conception d'une base de données (hyper)textuelles d'aide à la compréhension, à la production des textes scientifiques

et à la construction des connaissances, de vérifier l'efficacité des notes testées. D'autres expériences sont maintenant nécessaires pour valider le logiciel lui-même.

Par ailleurs, les résultats de cette recherche invitent à réfléchir plus avant sur la relation, dans les situations d'apprentissage-enseignement, entre le travail de lecture et les disciplines dont ressortissent les textes lus. Si les rappels des élèves aidés par des notes renvoyant au modèle mental ont gagné en cohérence, celle-ci s'avère essentiellement locale. Même dans le groupe 3, seule une minorité d'élèves réussit à construire une représentation globale cohérente du texte. Sans doute faudrait-il penser cette tâche de lecture comme un élément d'un scénario d'enseignement visant à la construction – ou au moins à l'appropriation – d'un savoir explicatif rigoureux. Ce scénario devrait anticiper sur les représentations communes des élèves pour tenir compte de l'écart entre celles-ci et les savoirs visés ou disponibles dans les textes utilisés. Dès lors qu'une partie importante du travail du lecteur consiste à construire la cohérence causale du texte, à activer et à reconstruire l'univers cohérent de connaissances qui y correspond, l'apprentissage – continu – de la lecture va de pair avec le travail des disciplines, ici les disciplines scientifiques. L'apprentissage de la lecture est transversal aux disciplines d'enseignement, mais il ne peut pas se faire ailleurs que dans le cadre même de chacune des disciplines (voir Baudry, Bessonat, Laparra & Tourigny, 1997).

Enfin, ces résultats ont été obtenus avec des élèves présentant des difficultés de lecture et plus généralement des difficultés scolaires importantes. Cette caractéristique du public doit être prise en compte. L'efficacité de la stratégie didactique qui consiste, dans notre recherche, à aider les élèves à combler les « trous sémantiques » du texte en leur fournissant des notes pourrait être interprétée aussi en termes de signification donnée à la tâche. On fera ici l'hypothèse que des élèves en grande difficulté bénéficient de cette aide notamment parce qu'elle leur permet de comprendre ce qu'ils doivent faire pour comprendre le texte. En effet, les élèves en échec ne le sont pas seulement parce qu'il leur manque certaines connaissances, mais parce qu'ils ne les mobilisent pas de façon opportune : ils se méprennent sur ce qu'on leur demande de faire et sur ce qu'il est nécessaire de mobiliser pour réussir telle ou telle tâche scolaire (Bautier & Rochex, 1998).

## Références

- Baudry, M., Bessonat, D., Laparra, M. et Tourigny, F. (1997). *La maîtrise de la langue au collège*. Paris : CNDP / Savoir Livre.
- Bautier, É. & Rochex, J.-Y. (1998). *L'expérience scolaire des nouveaux lycéens*. Paris : Armand Colin.
- Otero, J., León J. A. & Graesser, A. C., Eds. (2002). *The Psychology of Science Text Comprehension*. Mahwah, NJ : Lawrence Erlbaum Associates.
- Van den Broek, P., Virtue, S., Everson, M. G., Tzeng, Y. & Sung Y. (2002). Comprehension and Memory of science texts : inferential processes and construction of mental representation. In J. Otero, J. A. León & A. C. Graesser (Eds.), *The Psychology of Science Text Comprehension* (p. 131-154). Mahwah, NJ : Lawrence Erlbaum Associates.

**Note :** Les auteurs remercient l'IUFM de Créteil pour son soutien à cette recherche. Ils adressent également leurs remerciements aux enseignants des classes dans lesquelles a été mené le travail, Éric Pihou, Marie-Céleste Da Silva, Laurent Carette, Denis Gaboyer, Frédéric Scheiff, Claire Lacalmontie, Bernard Liberman et Laurence Darnault, les directeurs des écoles, Patrick Dupont, Sébastien Prieur, Laurent Dyrek, Gilbert Tobelem, Myriam Michau, Michel Ruiz et Philippe Quesnel, le conseiller pédagogique de la circonscription, Franck Gillet, et l'inspectrice de l'Éducation nationale, Marie-Christine Hébrard. Ils témoignent en outre leur reconnaissance à Marie-Hélène Lafforgue et à Nadine Chaudun, de l'école Savigny 1 d'Aulnay-sous-Bois, pour leur contribution à la phase préparatoire de la recherche ainsi qu'à tous les élèves ayant participé au travail.

---

<sup>1</sup> Dans les deux conditions, elles comportent le même nombre de propositions sémantiques : la quantité d'informations offerte aux deux groupes est donc la même.