

# **Facteurs textuels dans la compréhension de textes informatifs :**

## **Développement d'un modèle psycholinguistique de la lisibilité textuelle**

### **Résumé du projet**

L'objectif de ce projet est d'étudier les facteurs textuels qui affectent la compréhension afin de disposer d'un cadre théorique qui nous permette de rendre un texte plus compréhensible pour ses lecteurs. En combinant une approche linguistique et une approche psychologique, nous proposons de développer et de valider un modèle de la lisibilité de textes informatifs qui intègre les marques de cohésion référentielle et de cohésion relationnelle.

### 1. Cadre général et importance de ce thème de recherche

Comme le souligne J. Morais, dans *l'Art de Lire* (1994, p.17), on parle actuellement d'une crise de la capacité de lecture alors que jamais avant aujourd'hui un aussi grand pourcentage de la population n'a été " lettrée ". La raison en est que les exigences de notre société en matière d'aptitude à la lecture s'accroissent encore plus rapidement. Si l'illettrisme reste un grave problème, les difficultés que rencontrent nombre de personnes pour comprendre un texte sont bien plus fréquentes et peuvent être gravement handicapantes tant dans la vie professionnelle que dans les activités quotidiennes<sup>1</sup>.

Face à ce constat, que peut-on faire pour améliorer la compréhension de textes? Pour répondre à cette question, commençons par définir ce qu'on entend par « comprendre un texte ».

Comprendre un texte, c'est construire au fur et à mesure de la lecture une représentation cohérente de son contenu (Graesser, Millis et Zwaan, 1997). A cette fin, le lecteur doit déterminer les relations entre les phrases successives (cohérence locale) ainsi que celles qui relient les différentes parties du texte (cohérence globale). Le lecteur établit cette cohérence en s'appuyant sur les relations qui unissent les phrases et les paragraphes. Il s'agit par exemple de relations de *cause-conséquence*, de *contraste*, ou de type *problème-solution*. Ces relations de cohérence sont conceptuelles. Elles peuvent être indiquées explicitement dans le texte par des marqueurs linguistiques comme *parce que*, *mais*, *donc*, *d'une part*, *en résumé*, *par ailleurs*. Lorsqu'elles sont laissées implicites par l'auteur, le lecteur doit les inférer en s'appuyant sur ses connaissances générales.

C'est à ce niveau que se pose le problème. Dans le cas de textes informatifs, le lecteur ne dispose habituellement pas de telles connaissances. Il est donc d'autant plus dépendant des signaux qui indiquent comment la représentation mentale doit être construite (Caron, 1997; Noordman et Vonk, 1997). Améliorer la compréhension passe donc par une description de ces signaux et par la détermination de leur rôle lors de la compréhension. Cette étude est un préalable pour développer des règles de réécriture permettant de rendre un texte plus compréhensible, mais aussi pour proposer un programme d'aide et de formation des lecteurs afin que ceux-ci utilisent efficacement ces signaux, ce qui est loin d'être toujours le cas (Noordman, Vonk et Kempf, 1993).

### 2. Méthodologie

L'objectif de ce projet est de développer et de tester un modèle de la compréhensibilité d'un texte qui intègre ces facteurs textuels. Pour ce faire, nous nous appuyons sur les travaux de Kintsch et de ses collaborateurs qui ont utilisé leur modèle de la compréhension de textes pour mesurer la compréhensibilité de textes informatifs (Kintsch et van Dijk, 1978; Miller et Kintsch, 1980). Ce même modèle a aussi été utilisé pour améliorer la compréhensibilité (Britton, 1994).

Toutefois, comme le souligne Caron (1992, p.217-218), ces travaux abordent uniquement la cohérence locale et seulement d'un point de vue sémantique. Les différents moyens langagiers qui établissent la cohésion référentielle (pronoms, répétition, ...) sont négligés, de même que les marqueurs du discours qui établissent la cohésion relationnelle en spécifiant la nature des connections et en soulignant les relations globales entre des segments de textes.

Nous proposons donc d'étendre ce modèle à l'expression de la cohésion référentielle et de la cohésion relationnelle.

2.1 Pour évaluer la cohésion référentielle d'un texte, nous adapterons au français la procédure de Britton (1994) en l'étendant afin de prendre en compte les dispositifs linguistiques qui signalent ces relations dans le texte, niveau que Britton ne considère pas.

2.2 Les marqueurs du discours, qui implémentent la cohésion relationnelle, font actuellement l'objet d'analyses linguistiques qui soulignent leur importance dans la détermination de la structure d'un texte (Degand, 1997; Grosz, Pollack et Sidner, 1989; Redeker, 1991). Nous commencerons par synthétiser ces travaux.

Nous devons ensuite réconcilier une série de résultats contradictoires à propos de l'impact de ces signaux sur la compréhension. En effet, si les marqueurs du discours semblent jouer un rôle facilitateur pendant le processus de lecture (Bestgen et Vonk, soumis; Sanders, 1992), leur impact sur la représentation textuelle finalement construite par le lecteur semble très variable selon les études (Millis et al., 1993; Sanders, 1992; Sanders et Noordman, soumis). Cette variabilité peut provenir pour une part de la présence d'un certain nombre d'artefacts (textes trop courts, connecteurs inadéquats, variables dépendantes mal choisies), mais aussi d'un effet pervers des marqueurs et plus généralement de toute amélioration d'un texte : accroître la passivité du lecteur qui aurait l'impression de comprendre et donc traiterai moins profondément le texte (Fayol, 1992).

Dans un troisième temps, nous intégrerons ces facteurs dans le modèle de la compréhension.

### 3. Etapes

Nous combinerons une approche linguistique afin de décrire la structure des textes et les marques de celle-ci et une approche psychologique pour étudier l'impact des signaux sur la compréhension et valider le modèle.

3.1 L'approche linguistique confrontera un échantillon de textes informatifs<sup>2</sup> avec une série de modèles linguistiques de la cohérence. De manière générale, ces modèles tentent d'une part de spécifier comment la structure textuelle détermine les relations de cohérence et d'autre part d'identifier les marques de surface de cette cohérence. La classification des relations de cohérence de Sanders, Spooren et Noordman (1992) semble ici particulièrement prometteuse. L'approche de Mann et Thompson (1988) devrait utilement la compléter. Le critère que nous emploierons pour combiner et adapter ces modèles sera le suivant : ils doivent être capables de prédire l'occurrence des signaux effectivement présents dans les textes. Comme on a tout lieu de penser que certaines relations, prédites par les modèles, ne seront pas signalées, l'approche permettra d'identifier les relations qui sont presque toujours marquées et celles qui le sont rarement. Celles qui le sont souvent, mais pas toujours, devraient être les plus utiles pour la seconde phase du projet.

Parallèlement, nous analyserons ces textes au niveau de la cohésion relationnelle comme expliqué en 2.1.

3.2 Lors de la phase II, nous évaluerons l'impact de la présence ou de l'absence de signaux sur la compréhension des textes. Un échantillon des textes analysés lors de la première phase sera utilisé. Nous nous intéresserons tant aux effets globaux mesurés par la qualité d'un rappel libre et d'un résumé, qu'à des effets spécifiques mesurés par des questions portant sur les relations marquées par les signaux manipulés.

Nous examinerons la question de la passivité du lecteur en combinant des informations provenant des temps de lecture, de la qualité de la compréhension, et de jugements métacognitifs portant sur le sentiment d'avoir compris. En effet, si les signaux accroissent la passivité du lecteur, on devrait observer des temps de lecture plus brefs accompagnés d'un sentiment d'avoir compris élevé et une performance plus faible dans des mesures de compréhension.

3.3 Les résultats des deux premières étapes seront combinés afin de développer une équation prédisant la lisibilité d'un texte à partir de la cohésion relationnelle et référentielle<sup>3</sup>. Cette procédure sera validée en comparant nos prédictions avec les performances de lecteurs. Il s'agira tout particulièrement de comparer pour un même texte des versions qui diffèrent quant

à leur lisibilité. Cette étape se conclura concrètement par la proposition de règles de réécriture permettant d'améliorer un texte.

#### 4. Développements ultérieurs

Une limitation des travaux projetés est que l'accent est délibérément mis sur "le texte" dont on tente d'évaluer la compréhensibilité pour un lecteur moyen. Ce point de vue est bien évidemment partiel, mais il est permis de proposer des procédures d'amélioration et de réécriture. C'est d'autant plus important que Traxler et Gernsbacher (1995) et Britton (1994) ont observé que les auteurs surestimaient fréquemment les connaissances des lecteurs et donc n'introduisaient pas les signaux nécessaires. De plus, le modèle que nous comptons développer permettra une étude des différences individuelles entre les lecteurs. En effet, celles-ci devraient se traduire par une modulation du poids des facteurs qui le composent. Par exemple, l'expertise d'un lecteur pour un thème donné devrait le rendre moins dépendant de la présence de signaux textuels explicites.

---

<sup>1</sup> Un enfant sur dix éprouve des difficultés importantes de compréhension alors qu'il ne présente aucune lacune au niveau du décodage des signes graphiques (Stothard, 1994) et, selon une étude américaine, citée par Morais (1994, p. 20) " 60% des adultes entre vingt et un et vingt-quatre ans seraient incapables d'appréhender les idées essentielles d'un article de journal ".

<sup>2</sup> Nous comptons utiliser comme matériel des textes de types informatifs (manuels scolaires, encyclopédies, ...) d'une longueur de 500 à 1500 mots et destinés à des élèves de fin de secondaire ou à des étudiants de candidature.

<sup>3</sup> Il s'agira en quelque sorte d'une extension de la procédure proposée par Britton. Contrairement aux mesures de lisibilité classiques (Henry, 1987) comme celles calculées par le traitement de textes Word, cette approche, implémentée sur ordinateur, est fondée théoriquement (voir à ce sujet Miller & Kintsch, 1980).