

**Analyser automatiquement
les auto-explications de compréhension
d'élèves de primaire**

Aurélié Nardy, Philippe Dessus, Maryse Bianco &
Françoise Toffa
Laboratoire des sciences de l'éducation, UPMF

•
Communication à la Journée Langage du Pôle
Grenoble Cognition, 3 avril 2012



Plan

- I. Objet et enjeux
- II. Méthode
- III. Analyse automatique des paraphrases

Journée Langage – PGC – 3/4/12 2

I. Objet

- o La compréhension des textes (en lecture)
 - o Etude de son développement et des conséquences pour l'enseignement

Journée Langage – PGC – 3/4/12 3

I. Enjeux sociétaux

→ Très fort enjeu actuel dans le champ de l'enseignement

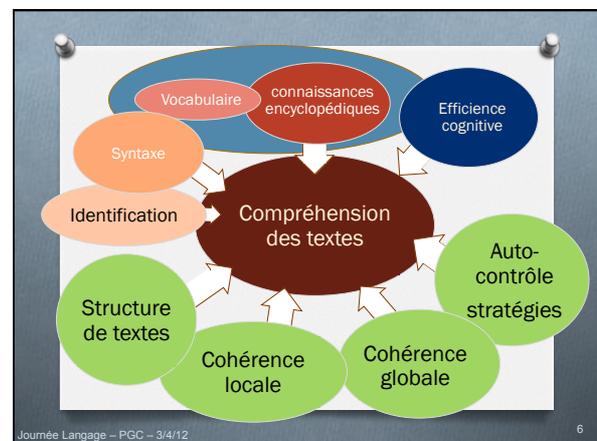
- o Comprendre ce qu'on lit, condition
 - o d'une scolarité réussie
 - o de l'insertion professionnelle
- o PISA 2006 – 2009
 - o Les élèves français de 15 ans ont des résultats très moyens aux épreuves de lecture, tout particulièrement en compréhension
- o Evaluations nationales de CM2
 - o Sur la période 1998-2008, les performances moyennes ont chuté en compréhension de textes. La chute des performances affecte spécifiquement les élèves issus de milieux modestes [Rocher 2008]

Journée Langage – PGC – 3/4/12 4

I. Enjeux théoriques

- o La compréhension des textes :
 - o une habileté spécifique ou une expression de capacités cognitives générales ?
- o Tradition des recherches sur le langage marquée par les conceptions de Fodor,
 - o Le langage recèle cette caractéristique particulière d'avoir « d'un côté, une nature suffisamment organisée pour se laisser prendre aux filets de la science, ...; de l'autre, un magma insaisissable, ... trop désordonné pour faire l'objet d'une science » [Andler 2004]

Journée Langage – PGC – 3/4/12 5



I. Objet

→ ANR « DEVCOMP »

- Mieux comprendre le développement des habiletés de compréhension
- Fournir **une analyse dimensionnelle** de cette activité
- Produire un **outil d'évaluation-diagnostic** pour les acteurs de l'éducation

→ **Présentation d'aujourd'hui**

The diagram illustrates the components of text comprehension. At the center is a dark red oval labeled 'Compréhension des textes'. To its left are two light green ovals: 'Structure des textes' (top) and 'Cohérence locale' (bottom). To its right is a dark green oval labeled 'Auto-contrôle stratégies'. Below the central oval are two more light green ovals: 'Cohérence globale' (left) and 'Cohérence locale' (right). Arrows point from the outer ovals towards the central 'Compréhension des textes' oval.

Journée Langage – PGC – 3/4/12 7

II. Méthode

Journée Langage – PGC – 3/4/12 8

II. Auto-contrôle et stratégies

Le lecteur expert est un lecteur stratégique

- évalue sa compréhension au fur et à mesure de la saisie des informations
- met en œuvre un ensemble de stratégies et procédures pour construire la cohérence
 - Paraphraser
 - Faire des relations entre les informations (*Bridging*)
 - Faire des relations entre les informations et ses propres connaissances (Elaborations)
 - Prédire, faire des hypothèses

(Bianco & al., 2004; Magliano & Graesser, 1991; Magliano & Millis, 2003; Van den Broek et al., 1994; Kendeou & van den Broek, 2005)

Journée Langage – PGC – 3/4/12 9

Méthode: auto-explication

- **Verbalisations (protocole de pensée à voix haute)**

Auto-explication :

- permet d'accéder à la nature de la stratégie
- un indicateur pour étudier/ évaluer les stratégies
- une méthode utilisable pour l'enseignement

- **Échantillon** : 45 élèves de cycle 3 (23 CE2 & 22 CM2)
 - Consigne

Tu vas lire une histoire et chaque fois que tu verras ce personnage, tu arrêteras de lire et tu diras ce que tu as compris, juste à ce moment-là. Moi je cacherai ce que tu viens de lire avec une feuille.

- 1 texte: Matilda (453 mots), 6 paragraphes

Journée Langage – PGC – 3/4/12 10

II. Méthode

- **Analyse causale du texte**

Ce soir-là, la famille de Matilda dînait comme d'habitude devant la télévision, **quand ils entendirent une voix forte venant du salon dire : « salut, salut, salut »**. La mère devint toute blanche. Elle dit à son mari « il y a quelqu'un dans la maison ». Ils arrêtaient tous de manger. Ils étaient tous sur le qui-vive. **La voix reprit « salut, salut, salut »**. Le frère se mit à crier « ça recommence ! ». Matilda se leva et alla éteindre la télévision.

The diagram shows a causal network for the text passage. It consists of 11 numbered nodes (circles) and several arrows indicating causal relationships. Node 3 is highlighted in yellow. Node 8 is also highlighted in yellow. The nodes are arranged in a roughly circular pattern with arrows pointing from 1 to 2, 2 to 3, 3 to 4, 3 to 5, 3 to 6, 3 to 7, 3 to 8, 4 to 5, 5 to 6, 6 to 7, 7 to 8, 8 to 9, 8 to 10, 8 to 11, 9 to 10, and 10 to 11.

Journée Langage – PGC – 3/4/12 11

II. Méthode

- Codage des verbalisations (Lynch & Van den Broek, 2007; McNamara, 2009; Kendeou et al., 2011)
- Codage en stratégies par des juges humains
 - **Contrôle**
 - **Paraphrases (+/-)**
 - **Inférences fondées sur la connaissance (+/-)**
 - **Inférences fondées sur le texte (+/-)**
- Dans cette étude: centration sur les paraphrases

Journée Langage – PGC – 3/4/12 12

III. Analyse automatique des paraphrases

Journée Langage – PGC – 3/4/12 13

III. Spécificités de notre étude

- *iStart* [McNamara 2006], système automatisé d'enseignement des stratégies, focalisé sur le secondaire => Répliquer ces études au primaire
- Utilisation intensive de LSA (*Latent Semantic Analysis*) pour comparer sémantiquement les verbalisations au texte
- Apport d'un espace de connaissances prenant en compte le niveau primaire ? => Corpus de 3,2 millions de mots [Denhière et al. 2007] :
 - contes, productions écrites d'enfants, manuels de lecture, encyclopédies
 - validé par comparaison à des normes d'association

Journée Langage – PGC – 3/4/12 14

Comparaisons de segments de *Matilda* avec LSA (C & V)

◦ T : ils étaient tous sur le qui vive
 ◦ V : et ben il y a quelqu'un qui fait des bruits qui dit salut salut et la famille elle arrête de manger et après y a mathilde elle matilda elle a elle va elle a arrêté de la télévision et puis. -0.046

◦ T : matilda dit alors que c'était un fantôme le salon est hanté je croyais que vous le saviez
 ◦ V : et ben matilda elle dit que le salon il est hanté et qu'elle le savait et puis les parents et son frère ils le savaient pas puis du coup ben les parents y ressortent très pâles très pâles du salon suivis des enfants voilà 0.305

◦ T : les parents très pâles sortirent du salon suivis par les enfants
 ◦ V : ben Matilda elle dit que le salon il est hanté et qu'elle le savait et puis les parents et son frère ils le savaient pas puis du coup ben les parents y ressortent très pâles très pâles du salon suivis des enfants voilà. 0.801

Journée Langage – PGC – 3/4/12 15

Comparaisons LSA segments du texte lu vs. verbalisations

L1

L2

L3

V1

L4

L5

L6

L7

L8

V2

L9

- Comparaison avec le segment précédant la verbalisation (**récence**)
- Comparaison avec les segments de **causalité locale** (fixés *a priori*)
- Comparaison avec les segments de **causalité distale** (fixés *a priori*)

Journée Langage – PGC – 3/4/12 16

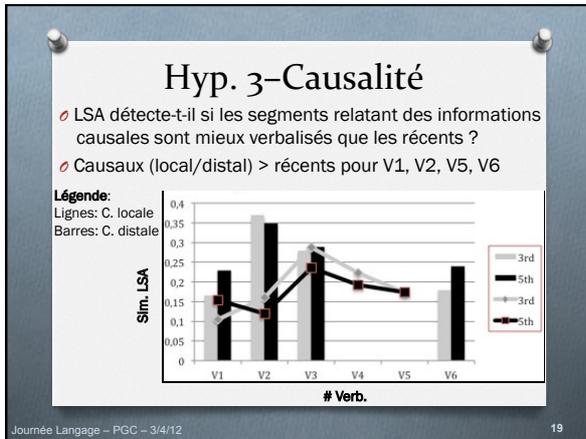
Hyp. 1 – LSA vs. Juges humains

Journée Langage – PGC – 3/4/12 17

Hyp. 2 – Effet de récence

- LSA détecte-t-il si les informations du segment précédant chaque verbalisation sont mieux rappelées ?
- Valeurs de proximités Verb.-phrase récente par niveau
- Pas d'effet de récence ; peu d'effet de niveau (V6)

Journée Langage – PGC – 3/4/12 18



Conclusion

- Résultats à améliorer, en testant d'autres techniques de traitement automatique de la langue
- LSA pourrait être utilisé pour analyser le contenu des auto-explications
- À terme, implémentation d'un tuteur d'aide à la compréhension de textes

Journée Langage – PGC – 3/4/12 20

Merci de votre attention, des questions ?

- Travail financé par l'ANR « DEVCOMP »...
- ... et en partie mené en partenariat avec des collègues de l'université Polytechnique de Bucarest, Roumanie

Journée Langage – PGC – 3/4/12 21

Analyse sémantique latente (Landauer & Dumais 97)

- LSA méthode d'analyse statistique de grands corpus textuels. Part du principe que
 - deux mots ont un sens similaire s'ils apparaissent dans des contextes similaires
 - deux contextes (paragraphes, phrases) ont un sens similaire (contiennent des informations similaires) s'ils contiennent des mots de sens similaire
- Approche «paquet de mots», sans prise en compte de la syntaxe

Journée Langage – PGC – 3/4/12 22