

Description fonctionnelle et étude expérimentale des marqueurs du discours dans le but d'améliorer l'intelligibilité de textes. Comparaison de la production et de la compréhension, de l'oral et de l'écrit, et du français et du néerlandais.

Functional description and experimental study of discourse markers in order to enhance text readability. Comparison of production and comprehension, spoken and written discourse, and French and Dutch.

Comprendre un texte ou un discours oral, c'est construire une représentation cohérente de son contenu (Graesser et al. 1997). A cette fin, le lecteur/auditeur doit déterminer les relations entre les phrases successives ainsi que celles qui relient les différentes parties du texte. L'auteur peut faciliter cette intégration en signalant ces relations au moyen de marques de cohérence référentielle (anaphores) et de cohérence relationnelle (connecteurs). Ces expressions sont considérées comme de véritables instructions à destination du "comprendre" afin qu'il construise la représentation mentale la plus adéquate.

Même si de plus en plus de chercheurs abordent ces marqueurs, nos connaissances à leur propos restent très partielles. Nous proposons donc d'étudier l'emploi des marques de cohésion à l'oral et à l'écrit en nous attachant à déterminer les conditions qui font que ces marques sont utilisées adéquatement ou non. Nous étudierons aussi l'impact de leur présence, absence ou mauvais emploi sur la compréhension dans ces deux modalités. Afin de déterminer la généralité de nos observations, nous aborderons, dans certaines de nos recherches, deux langues (le français et le néerlandais). Pour mener à bien ce vaste projet, nous comptons sur la composition interdisciplinaire de l'équipe (linguiste et psychologue) et sur l'appui de partenaires déjà impliqués dans ce genre d'études.

Une des principales difficultés que rencontre l'étude des marques de cohérence consiste à identifier et à décrire les expressions linguistiques qui fonctionnent comme marques. Les grammaires de textes ayant fait long feu (Brown & Yule, 1983), les chercheurs qui traquent ces dispositifs se trouvent largement désarmés. Selon nous, seules, les analyses linguistiques de corpus permettent de progresser. En nous appuyant sur nos travaux antérieurs (Degand, 1998, sous presse; Degand & Pander Maat, 1999, sous presse; Pander Maat & Degand, sous presse; Degand & Sanders, 1999) notre premier objectif est de décrire le fonctionnement des marqueurs du discours de manière contrastive en néerlandais et en français dans des textes naturels oraux et écrits. Une telle description fonctionnelle vise à déterminer les contraintes syntaxiques, sémantiques et pragmatiques qui régissent l'emploi de tels marqueurs. Nous concentrerons nos recherches sur quatre types de marqueurs cohésifs (voir projet 1).

Dans ce premier projet, nous étudierons si un marqueur donné est utilisé de manière similaire à l'oral et à l'écrit. Il s'agit de l'approche classique en linguistique, appelée semasiologique ou inductive (Hansen, 1997). Toutefois, cette approche ne nous permettra pas de dire si une fonction discursive est marquée avec la même fréquence et les mêmes dispositifs linguistiques quelle que soit la modalité. Or, rien n'est moins sûr.

— En effet, l'auteur est supposé utiliser ces marques en fonction du principe de coopération (Grice, 1975) ou de l'*optimal design* : "*The speaker designs his utterance in such a way that he has good reason to believe that the addressee can readily and uniquely compute what he meant on the basis of the utterance along with the rest of their common ground*". (Clark et al., 1983, p.246; p. 258 Caron, 1997). Or, ce principe a été formulé spécifiquement pour la communication orale interactive. A l'écrit, la situation est très différente. Il est souvent difficile pour l'auteur d'estimer le terrain commun et il ne dispose pas de feedback (Traxler & Gernsbacher, 1995). On devrait donc s'attendre à un marquage moins efficace à l'écrit.

— Toutefois, on dispose habituellement à l'écrit de plus de temps pour planifier, formuler et éditer sa production. De plus, plusieurs recherches récentes suggèrent que le locuteur n'appliquerait le principe de

l'optimal design que lors d'un processus contrôlé qui ne s'enclencherait que dans un second temps et seulement lorsque des ressources attentionnelles suffisantes sont disponibles (Bard et al., 2000; Horton & Keysar, 1996; Roßnagel, 1999). Cet argument favorise donc l'écrit.

— Enfin, nous avons montré qu'une charge attentionnelle importante pouvait avoir un effet bénéfique en faisant apparaître des marques de cohésion qui sont la trace des difficultés de production (Bestgen, 1998). Ces traces, même si elles ne sont pas intrinsèquement destinées à l'auditeur, indiquent partiellement au moins les relations que celui-ci doit appréhender. Ce dernier argument est, à nouveau, favorable à l'oral.

Pour déterminer dans quelle mesure ces différents facteurs affectent l'emploi des marqueurs, nous proposons deux recherches. Dans la première, nous comparerons l'emploi des marques de cohérence à l'oral interactif, au pseudo-oral (exposé) et à l'écrit au travers d'analyses de corpus. Pour ce faire, nous devons inverser la manière d'aborder les marques. Nous ne pourrons plus procéder en cherchant des marques spécifiques afin d'en déterminer la fonction. Il nous faudra chercher des fonctions spécifiques et voir si elles sont marquées. Une telle approche s'est jusqu'à présent heurtée à des difficultés pratiques et techniques importantes. Dans le deuxième projet, nous proposons une approche qui devrait rendre possible ce genre d'études.

Dans notre troisième projet, nous étudierons expérimentalement les facteurs qui déterminent l'emploi des marques de cohésion. Dans ce but, nous développerons une situation de production orale dans laquelle il sera possible de manipuler le terrain commun, la mise en place du mécanisme secondaire de contrôle de l'énoncé et la production de traces.

L'ensemble de ces recherches serviront de base pour poursuivre notre quatrième objectif : étudier l'impact des marques de cohésion sur la compréhension dans la ligne de nos travaux récents (Bestgen & Vonk, 1995, 2000; Degand et al., 1999; Degand & Sanders, soumis). En effet, notre thèse est qu'un emploi inadéquat de ces marques linguistiques explique une part importante des difficultés de compréhension, un facteur que les travaux classiques à propos de la lisibilité ont largement négligé (Chall & Dale, 1995).

Description des quatre projets

Projet 1 : Analyse de corpus : approche dirigée par les marqueurs.

Les marques de cohésion étant un domaine très vaste, nous nous focaliserons sur quatre dispositifs : les connecteurs causaux et les anaphores, les marqueurs de la segmentation et de la digression. Pour déterminer les contraintes régissant le comportement discursif de ces marqueurs, nous nous baserons sur des analyses de corpus. Celles-ci permettent d'étudier la variabilité d'emploi des marqueurs dans un contexte naturel (oral ou écrit), et de confronter l'usage langagier aux descriptions théoriques, mettant ainsi au jour des mécanismes d'emploi non décelables par l'introspection. C'est ce qui est apparu dans nos travaux antérieurs sur la classification et l'emploi de connecteurs causaux décrits en termes d'implication du locuteur (*Speaker Involvement*, cf. Degand & Pander Maat, 1999, sous presse; Pander Maat & Degand, sous presse). Selon notre hypothèse, les connecteurs peuvent être ordonnés sur une échelle allant d'une implication minimale du locuteur (relation objective: *La voiture a dérapé parce qu'il y avait du verglas.*) à une implication maximale (relation subjective: *Ils ont dû rouler sur une plaque de verglas, car ils ont dérapé sans raison apparente*). Ce type d'analyse est également applicable à d'autres types de réalisations linguistiques, tels les déictiques, proches des anaphores (Hanks, 1992). Nous comptons poursuivre l'étude des marqueurs causaux en termes de *Speaker Involvement* et étendre nos investigations théoriques aux autres types de marqueurs. Cela nous permettra de vérifier la généralisabilité du modèle théorique tant pour le discours écrit qu'oral. Des études cross-linguistiques nous permettront de vérifier notre hypothèse selon laquelle l'échelle est la même pour les différentes langues, mais les marqueurs divergent par les zones qu'ils occupent sur l'échelle.

Pour mener ce genre de travaux, de vastes corpus écrits sont déjà disponibles, et pour ce qui est des corpus oraux, des bases existent pour le néerlandais (Oostdijk, 2000) et le français (Brosens, 1999), mais nous

devrons également construire une partie du matériel. L'usage de corpus suppose que l'on connaît les marqueurs que l'on veut décrire, qu'ils soient assez fréquents, et détectables de manière automatisée. Ceci ne pose pas de problèmes pour les trois premiers types de marqueurs à propos desquels nous avons acquis une expertise certaine (Bestgen, 1998; Bestgen et Costermans, 1994, 1996; Degand, 1998, 2000, sous presse; Degand et al., 1999; Degand & Pander Maat, 1999, sous presse; Pander Maat & Degand, sous presse). En ce qui concerne les marques de digression, un problème particulier se pose. La digression est une rupture particulière qui consiste en un changement de thème avec un retour au thème principal (Lenk, 1998). Si l'on connaît la fonction discursive, on ignore comment elle est marquée lexicalement, et a fortiori si de tels signaux sont ou non bénéfiques à l'auditeur/lecteur. Pour aborder cette question, nous devons faire appel à une approche onomasiologique et déductive (Hansen, 1997).

Projet 2 : Analyse de corpus : approche dirigée par la structure.

Le problème majeur que rencontre l'approche des marqueurs dirigée par la structure réside dans la lourdeur des procédures permettant l'identification de cette structure. Ni les classifications théoriques des relations de cohérence (Mann & Thompson, 1988; Martin, 1992; Sanders, Spooren, Noordman, 1992; Schiffrin, 1987; Sweetser, 1990), ni le recours à l'avis de juges (Bestgen, 1992) ne peuvent être employées à grande échelle. Or, une telle recherche implique l'analyse de vastes corpus. Aussi proposons-nous de tester une technique automatisée permettant de déterminer la structure d'un texte afin d'implémenter une approche onomasiologique et déductive des marqueurs. Nous emploierons l'analyse sémantique latente, une technique permettant de représenter des textes dans un espace sémantique (Landauer et Dumais, 1997; Kintsch, 1998) et de créer un indice de cohérence pour les différentes parties d'un texte. L'adaptation au français de cette technique est l'objet d'un projet de recherches en cours en collaboration avec Guy Lories. Nous validerons la technique sur des sous-ensembles des corpus en comparant les structures obtenues par la procédure automatisée à celles qui peuvent être dérivées d'approches théoriques et de jugements subjectifs. Ensuite, nous déterminerons l'efficacité de cette technique pour identifier les marqueurs. Nous étudierons dans un premier temps les changements de thèmes. On devrait observer en ces lieux une plus grande fréquence d'expressions adverbiales de temps et de lieu (Bestgen et Vonk, 2000; Virtanen, 1992) et d'expressions anaphoriques très explicites (Vonk, Hustinx & Simons, 1992). Nous aborderons ensuite les digressions afin d'identifier les expressions qui les signalent.

Projet 3 : Etude expérimentale des facteurs qui déterminent l'emploi des marqueurs.

Si l'on admet encore généralement que l'auteur d'un message fait de son mieux pour le rendre le plus intelligible possible, des arguments récents suggèrent que fréquemment la communication n'obéit pas au principe de l'*optimal design* (Bard et al., 2000; Horton & Keysar, 1996; Nickerson, 1999; Roßnagel, 1999). Nous proposons d'étudier les implications de ces observations sur l'emploi des marques de cohésion. D'un côté, un locuteur, placé dans une situation cognitivement exigeante, devrait considérer que le destinataire à accès aux mêmes informations que lui et devrait donc négliger de signaler efficacement certaines relations par des anaphores explicites ou des connecteurs adéquats. D'un autre côté, un accroissement de la charge cognitive devrait faire apparaître des traces des difficultés de production et donc des marqueurs sans doute utiles pour l'auditeur.

Concrètement, nous proposons d'utiliser la situation expérimentale suivante inspirée de Horton & Keysar (1996). Le locuteur devra produire une narration sur la base des déplacements de formes géométriques sur un écran d'ordinateur. Afin de manipuler le terrain commun, il sera informé de ce que l'auditeur, un comparse, pourra voir, comme par exemple le décor ou certaines formes. Les scénarios nous permettront de manipuler la structure des narrations à produire. Pour chaque scénario, nous varierons la charge en mémoire de travail et la rapidité avec laquelle le locuteur devra produire la narration. Nous ferons aussi produire ces narrations à

plusieurs reprises. Ces manipulations nous permettront de moduler la possibilité de prendre en compte le terrain commun et la difficulté de la tâche. Les productions seront enregistrées, retranscrites et analysées tant au niveau du débit que de l'emploi des marqueurs et expressions anaphoriques.

Les bénéfices escomptés de ces expériences sont une meilleure compréhension des facteurs qui déterminent un emploi adéquat des marqueurs, mais aussi une catégorisation fonctionnelle de ceux-ci selon qu'ils sont sensibles à l'*optimal design* ou aux difficultés de production.

Projet 4 : Etudes expérimentales en compréhension.

Lorsque les projets "production" auront permis l'identification de nouveaux marqueurs, nous pourrons étudier leur impact sur la compréhension. D'ici là, nous étudierons en compréhension deux ensembles de marqueurs que nous maîtrisons déjà : les marqueurs temporels de la segmentation (Bestgen & Vonk, 1995, 2000) et les marqueurs causaux (Degand, in press; Degand et al., 1999; Degand & Sanders, soumis). Ceci permettra de mieux comprendre l'impact de ces marqueurs, mais aussi de continuer à développer une méthodologie expérimentale adéquate pour l'étude des autres marqueurs (pour l'oral voir par exemple Fox Tree & Schrock, 1999). Faute de place, nous ne détaillerons que les recherches à propos des expressions temporelles.

Notre objectif est de confronter deux conceptions à première vue antagonistes de l'impact des expressions adverbiales de temps sur la compréhension. Selon Bestgen et Vonk (1995, 2000) les expressions temporelles comme *Vers quatre heures* sont des marqueurs de la segmentation. Elles accélèrent la compréhension d'un changement de thème en conduisant le lecteur à court-circuiter l'étape d'intégration qui vise à relier chaque nouvelle phrase à celles qui précèdent. Zwaan (1996; Zwaan et al., 2000) a étudié des expressions temporelles très similaires comme *Une heure plus tard*. Selon son *Event indexing model*, les expressions temporelles qui signalent un changement du cadre temporel de la narration imposent au lecteur la mise à jour de son modèle mental et ralentissent donc la lecture. Ces deux thèses ne sont peut-être pas aussi inconciliables qu'il y paraît. En effet, Zwaan a étudié le rôle des expressions temporelles dans des textes continus, n'incluant pas de changement de thème, alors que notre thèse porte spécifiquement sur la compréhension d'un changement de thème (Bestgen, in press). Afin de déterminer si ces deux thèses ne sont que les deux versants d'une même fonction, nous comparerons l'impact de ces marqueurs dans les mêmes textes qui incluent ou non des changements de thèmes (pour la méthodologie voir Bestgen & Vonk, 2000).

Références citées dans le projet.

- Bard, E.G., Anderson, A.H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*, 42, 1-22.
- Bestgen, Y. (1992). Le textomètre : un outil pour l'étude de la structure du discours et de son marquage. *Psychologica Belgica*, 32, 141-167.
- Bestgen, Y. (1998). Segmentation markers as trace and signal of discourse structure. *Journal of Pragmatics*, 29, 753-763.
- Bestgen, Y. (in press). Au-delà de la phrase : les unités du discours et leur réalité psychologique. *Cahiers de l'Institut de Linguistique de Louvain*.
- Bestgen, Y., & Costermans, J. (1994). Time, space, and action : Exploring the narrative structure and its linguistic marking. *Discourse Processes*, 17, 421-446.
- Bestgen, Y., & Costermans, J. (1996) Temporal markers of narrative structure. Studies in production. In J. Costermans & M. Fayol (Eds.) *Processing interclausal relationships in the production and comprehension of text*. Hillsdale : Erlbaum.
- Bestgen, Y., & Vonk, W. (1995). The role of temporal segmentation markers in discourse processing. *Discourse Processes*, 19, 385-406.
- Bestgen, Y., & Vonk, W. (2000). Temporal adverbials as segmentation markers in discourse comprehension. *Journal of Memory and Language*, 42, 74-87
- Britton, B.C. (1994). Understanding Expository Text. In M.A. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp.641-674). San Diego : Academic Press.

- Brosens, V. (1999), "ELICOP, Etude Linguistique de la Communication Parlée : Bilan et perspectives", in Actes du Colloque TALN'99, Cargèse, Corse, 12-17 juillet 1999, Corpus et TAL : Pour une réflexion méthodologique (Atelier Thématique), pp. 15-25.
- Brown, G., & Yule, G. (1983). *Discourse analysis*. Cambridge: Cambridge University Press.
- Caron, J. (1997). Psychologie cognitive et interactions conversationnelles. In J. Bernicot, J. Caron-Pargue, & A. Trognon (Eds.), *Conversation, interaction et fonctionnement cognitif* (pp.221-237). Nancy: Presses universitaires.
- Clark, H.H., Schreuder, R.; Buttrick, S. (1983). Common ground and the understanding of demonstrative reference. *Journal of Verbal Learning and Verbal Behavior*, 22, 245-258.
- Chall, J.S., & Dale, E. (1995). *Readability revisited*. Cambridge: Brookline Books.
- Degand, L. (1998), Het ideationele gebruik van *want* en *omdat*: een geval van vrije variatie? *Nederlandse Taalkunde*, 3, 309-326.
- Degand, Liesbeth, (2000) Prepositional causatives or causal connectives? Discursive constraints, *Journal of Pragmatics*, 32 (6), 687-707.
- Degand, L. (in press_a) *Form and Function of Causation. A theoretical and empirical investigation of causal constructions in Dutch*, Peeters, Leuven [Studies op het gebied van de Nederlandse Taalkunde].
- Degand, Liesbeth (in press_b) Contextual constraints on causal sequencing in informational texts, *Functions of Language*.
- Degand L., Lefevre, N., & Bestgen, Y. (1999) The impact of connectives and anaphoric expressions on expository discourse comprehension. *Document Design*, 1, 39-51.
- Degand, Liesbeth & Pander Maat, Henk (1999) Scaling causal relations in terms of Speaker Involvement, *Levels of Representation in Discourse, Working Notes of the International Workshop on Text Representation*, Edinburgh University, 45-54.
- Degand, Liesbeth & Pander Maat, Henk (in press) A contrastive study of Dutch and French causal connectives on the Speaker Involvement Scale, A. Verhagen & J. van de Weijer (eds.) *Levels in Language and Cognition: Lexicon, Grammar, Discourse*.
- Degand, Liesbeth & Sanders, Ted (1999) Causal connectives in language use. Theoretical and methodological aspects of the classification of coherence relations and connectives, *Levels of Representation in Discourse, Working Notes of the International Workshop on Text Representation*, Edinburgh University, 3-12.
- Degand, Liesbeth & Sanders, Ted (soumis) The impact of relational markers on expository text comprehension in L1 and L2, submitted to *Reading and Writing*.
- Fox Tree, J.E., & Schrock, J.C. (1999). Discourse Markers in Spontaneous Speech: Oh What a Difference an Oh Makes, *Journal of Memory and Language* 40, 280–295
- Graesser, A.C., Millis, K.K., Zwaan, R.A. (1997). Discourse comprehension. *Annual Review of Psychology*, 48, 163-189.
- Hanks, William F. (1992). The indexical ground of deictic reference. In: Alessandro Duranti and Charles Goodwin (eds.) *Rethinking Context. Language as an Interactive Phenomenon*, Cambridge, Cambridge University Press, 43-76.
- Hansen, M.M (1997). *Alors et donc* in spoken French: A reanalysis. *Journal of Pragmatics*, 28, 153-187.
- Horton, W.S., & Keysar, B. When do speakers take into account common ground?, *Cognition*, 59, 91-117.
- Kintsch, W. (1998). *Comprehension: A paradigm for Cognition*. New York: Cambridge University Press.
- Landauer, T.K., & Dumais, S.T. (1997). A solution to Plato's problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review*, 104, 211-240.
- Lenk, U. (1998). Discourse markers and global coherence in conversation. *Journal of Pragmatics*, 30: (2) 245-257
- Mann, W.C. & Thompson, S.A. (1988). Rhetorical Structure Theory: toward a functional theory of text organization. *Text* 8, 243-281.
- Martin, J.R. (1992). *English Text: systems and structure*. John Benjamins, Amsterdam.
- Nickerson, R.S. (1999). How we know — and sometimes misjudge — what others know: imputing one's own knowledge to others. *Psychological Bulletin*, 125, 737-759.
- Stark, H.A. (1988). What do paragraph markings do? *Discourse Processes*, 11, 275-303.
- Oostdijk, N. (2000) Building a Corpus of Spoken Dutch. In P. Monachesi (Ed.), *Computational Linguistics in the Netherlands 1999. Selected papers from the tenth CLIN meeting*. pp. 147-158.
- Pander Maat, Henk & Degand, Liesbeth (sous presse) Scaling causal relations and connectives in terms of Speaker Involvement, *Cognitive Linguistics*.

- Roßnagel, C. (1999). Cognitive load and perspective-taking: applying the automatic-controlled distinction. *European Journal of Social Psychology*, 29, 318-334.
- Sanders, T., Spooren, W.P.M., & Noordman, L.G.M. (1992). Towards a taxonomy of coherence relations. *Discourse Processes*, 15, 1-35.
- Schiffrin, D (1987). *Discourse Markers*, Cambridge: Cambridge University Press.
- Sweetser, E. (1990). *From Etymology to Pragmatics: Metaphorical and cultural aspects of semantic structure*, Cambridge: Cambridge University Press, [Studies in Linguistics 54].
- Traxler, M.J., & Gernsbacher, M.A. (1995). Improving coherence in written communication. In M.A. Gernsbacher & T. Givon (Eds.), *Coherence in spontaneous text* (pp.215-238). Amsterdam: Benjamins.
- Virtanen, T. (1992). *Discourse functions of adverbial placement in English*. Åbo: Åbo Akademi University Press.
- Vonk, W., Hustinx, L.G.M.M., & Simons, W.H.G. (1992). The use of referential expressions in structuring discourse. *Language and Cognitive Processes*, 7, 301-333.
- Zwaan, R.A. (1996). Processing narrative time shifts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1196-1207.
- Zwaan, R.A., Madden, C.J., & Whitten, S.N. (2000). The presence of an event in the narrated situation affects its availability to the comprehender. *Memory and Cognition*, 28, 1022-1028.