Élaboration du contexte à travers des réticularités

Hadria KHELIFI

Université de Lorraine. France

Résumé : Le présent article vise à montrer comment générer un contexte à partir d'un item favorisé comme mot-pôle parmi tous les mots d'un corpus composé de douze mémoires d'études universitaires de disciplines hétérogènes. Plus précisément, une recherche sur un contexte renforcée par une analyse automatique du corpus, fournissant des résultats qui présentent, selon notre interprétation, l'une des raisons probables du lien entre le contexte et le genre écrit.

Mots clés : Analyse automatique, contexte, genre, hétérogènes, interprétation, réseau, thème.

Abstract: *This article aims to show how to generate a context from a word which is considered as a headword among all the words of a corpus consisted of twelve heterogeneous thesis in various university courses. More specifically, it is a research on context strengthened by an automatic analysis of corpus, providing results which present a potential reason of the link between the context and the written genre.*

Keywords

Automatic analysis, context, genre, heterogeneous, interpretation, network, theme.

Introduction

Afin d'élaborer son discours, le scripteur est tenu à créer une suite de mots, de phrases et de paragraphes qui sont lexicalement, morphosyntaxiquement et orthographiquement appropriés au genre écrit. Mais il doit surtout penser à l'entourage proche et lointain des éléments nécessaires pour l'attribuer un contexte, soit « des tactiques de production et d'interprétation » (Rastier, 2005) que certains chercheurs éprouvent en cas de difficultés à définir le contexte du texte. Notre objectif, à travers cet article, est d'exposer l'une des possibilités d'interprétation du contexte en relation avec le genre du corpus. Comment peut-on concevoir ce contexte à partir d'un mot-axe et son environnement thématique dans un corpus composé de textes de diverses disciplines universitaires, qui n'entretiennent aucune liaison les unes avec les autres, mais qui se sont réunies sous un genre unique ?

À cet effet, nous définissons le genre, le contexte, son interprétation et un exemple d'étude du contexte à travers le logiciel *Hyperbase,* avant de conclure.

1. Le genre

Le concept *genre* a été toujours mis et remis en question malgré les études de catégorisation des textes réalisées autour de son existence. Ayant un usage courant, la notion du *genre* doit être définie de façon très générale, suivant des critères hétérogènes divers (rhétoriques, linguistique, esthétiques, sociologiques…), et changeant d'un écrit à l'autre.

F. Rastier, par exemple, affirme que « Par son genre, chaque texte se relie à un discours » (Rastier, 1987). Cette idée est confirmée par D. Maingueneau qui considère qu'il existe une relation de réciprocité entre les types de genres de discours. Selon ce dernier : « tout type est un réseau de genres : tout genre est rapporté à un type. » (Maingueneau, 2014). Le classement du corpus de cette étude est clair, il se compose de discours appartenant à un seul genre qui est l'écrit universitaire dépendant à son tour du genre scientifique.

1.1. Travaux antérieurs sur le genre assistés par logiciel

Les recherches effectuées via l'analyse automatique sont nombreuses. Nous citons celles d'E. Brunet, le fondateur du logiciel *Hyperbase* qui a réalisé dans le cadre d'une recherche sur le genre une comparaison des textes de grands auteurs : Balzac, Proust, Molière, Flaubert, Corneille, etc. mais a aussi comparé des écrits découlant de deux genres différents écrits par

un seul auteur comme les textes d'Hugo en vers et en prose (Brunet, 2011). Aussi, D. Ablali (2010) a établi une recherche sur la linguistique des genres, et a analysé 700 articles appartenant à sept discours en sciences humaines (linguistique, littérature, philosophie, sociologie, anthropologie, histoire et géographie) à l'aide du logiciel *Hyperbase* et du lemmatiseur *Cordial* (Ablali, 2010). D. Mayffre (2012) a analysé le discours politique français contemporain de l'ex-président de la république française (Nicolas Sarkozy) en recourant au logiciel *Hyperbase* pour une étude de thématique entre autres recherches produites par ce dernier (Mayaffre, 2012).

2. Le contexte

Selon le niveau d'analyse déterminé, le contexte se crée en même temps que le texte. Il est l'entourage à travers lequel l'auteur s'exprime au destinataire. Un environnement qui permet de comprendre le rôle des phonèmes, morphèmes, phrases, périodes, etc. ainsi que des signes typographiques dans un texte. Avant de définir le terme *contexte*, nous précisons qu'il faut distinguer entre le contexte et le co-texte, qui concerne qu'une partie du texte dans le texte lui-même et qui n'est pas le cas du présent exemple qui nécessite le texte intégral pour clarifier le statut du contexte.

Le dictionnaire Larousse définit le contexte comme étant un « ensemble du texte à l'intérieur duquel se situe un élément d'un énoncé et dont il tire sa signification » (Larousse en ligne). C'est à partir de cette définition que nous sommes partie pour le choix d'un mot-vedette et l'étude de son contexte indiqué dans l'exemple plus bas.

Un contexte peut avoir plusieurs catégorisations : linguistiques, non linguistiques, globales, locales, actives et passives, le contexte syntagmatique et le contexte paradigmatique (Rastier ; 2005) sans oublier le contexte culturel. Le même auteur souligne que le contexte ne dispose pas de théorie générale et rarement défini en linguistique, ou il est mal défini, et pourtant il est très souvent sollicité et explicité. Le contexte dépend des conditions spécifiques, et choisi par son auteur selon sa stratégie construite autour du sujet en question avant de l'établir et avant d'interpréter le texte. Pour cela, le scripteur doit prendre en compte des contraintes que C. Garcia-Debanc a appelé « externes » (Garcia-Debanc, 1990) comme le type de mots choisis, orthographe, etc., et des contraintes « internes » (Garcia-Debanc,1990) comme la cohérence textuelle, macro syntaxe, etc. La même auteure rajoute que d'autres scripteurs procèdent à l'écriture après avoir organisé leurs idées selon un plan de travail spécifique. Ceci est possible en collectant par exemple, les notions, les opinions traitant le

thème en question. Ils n'ont qu'à repérer des mots de liaison adéquats au contexte réservé à cet écrit pour établir une cohérence de leur texte. Inversement, le contexte de l'exemple exposé plus bas a été généré et interprété selon une autre démarche. Mais avant de l'exposer, quelles sont les différentes possibilités d'interpréter un contexte ?

3. Interprétation du texte

« L'interprétation est une des façons d'étudier le contenu du texte. Elle renvoie à des notions différentes selon les domaines et les questions traitées, du mot au texte. Selon F. Rastier, « l'interprétation […] se réduit pour nous à une assignation réglée du sens, sans qu'il soit besoin d'évoquer un herméneute, ou un lecteur modèle (cf.Eco, 1979) pour convoquer les divers types de normes dont la rencontre transforme la phrase ou le texte en énoncé pourvu de signification. » (Rastier, 1987).

L'une des démarches proposée par F. Rastier pour l'interprétation d'un texte est d'élargir le corpus soit à d'autres parties du texte, soit à l'ensemble de l'œuvre du même auteur vu qu'elle est considérée comme « la plus grande unité linguistique possible » (Rastier, 1987). L'extension du corpus de la présente recherche jusqu'à douze textes était possible et a constitué une unité linguistique possible, si l'on reprend l'expression de F. Rastier, et pourtant ces textes étaient rédigés par douze auteurs-étudiants différents.

« La théorie des stratégies interprétatives ne relève pas pour autant de la sémantique, et bien plutôt d'une théorie des idéologies, considérées notamment comme systèmes de production et de transformation du sens textuel. » (Rastier, 1987). La sémantique peut par contre montrer comment une stratégie peut engendrer une interprétation. En se basant sur la logique, il faut distinguer les deux conceptions syntaxique et sémantique de l'interprétation. La première quand un mot est interprété selon sa situation syntaxique, et la deuxième basée sur la relation entre les mots, à partir de laquelle nous nous sommes basées pour bâtir le sens de l'exemple du présent article. Une troisième conception qui se rajoute mais elle concerne la rhétorique ou herméneutique, fondée non pas sur la logique mais sur la psychologie, la sociologie, l'anthropologie et les sciences sociales.

Autres orientations sont possibles pour étudier le texte et son contexte selon l'objectif de la recherche. Par exemple, S. Plane propose deux étapes principales pour l'étude d'un texte : une étude générale et une approche détaillée. La première étude porte sur la signification globale du texte qui représente une sorte d'introduction du texte. La deuxième est réservée à

l'étude lexicale, syntaxique et thématique du texte permettant d'établir une cohérence textuelle, un élément essentiel pour l'élaboration du contexte.

Avant d'interpréter un texte, F. Rastier cite quelques principes sémantiques qui concernent le sens appuyant l'idée que ce dernier n'est pas défini selon les représentations mentales ou sur la référence, mais analysé en « *traits sémantiques* ou sèmes » (Rastier, 1989), connu comme des moments stabilisés dans des parcours d'interprétation. Le sens découle plutôt du texte et non des mots isolés. F. Rastier rajoute que le sens d'un signe est défini par son contexte qui, à son tour, identifié par tout le texte contrairement de ce qui concerne la présente recherche où nous nous référons à l'environnement du terme *travail* étudié plus bas. L'entourage de ce lexique se compose des items apparaissant dans des endroits différents du corpus mais qui ont pu se relier, avec ce mot choisi comme axe, générant par conséquence son contexte. Autrement dit, la triade : texte, contexte et sens constituent un cercle vicieux où chaque élément dépend de l'autre pour éliminer toute ambiguïté. F. Rastier précise que les mots ne représentent pas les unités textuelles élémentaires mais des « *passages* » (Rastier, 1989). Par passage, F. Rastier désigne « un extrait dans une expression, et un fragment pour un contenu.» (Rastier, 1989). Effectivement, c'est à partir de cette idée qu'il nous semblait pertinent de ne pas chercher à repérer la phrase complète au sens propre de cette dernière mais de la construire à partir d'un mot choisi selon certains critères, discutés plus bas, et son voisinage qui correspond à son contexte.

Au niveau sémantique, F. Rastier confirme que le sens de chaque unité lexicale est décidé par son contexte. Il rajoute :

> « Le texte isolé n'a pas plus d'existence que le mot ou la phrase isolée : pour être produit et compris, il doit être rapporté à un genre et un discours, puisque tout texte relève d'un genre qui le rattache à un discours et par là un type de pratique sociale. » (Rastier, 1987).

Ainsi J.-M. Adam considère que toutes les connaissances pragmatiques, connaissances du « monde représenté » (Adam, 2005) interviennent dans les opérations de compréhension et de production du contexte du texte.

4. Exemple d'étude du contexte

4.1. L'analyse automatique du texte

Selon M. Tournier, la lexicométrie est aussi appelée :

> « **statistique linguistique** (Guiraud, 1959, 1960), **statistique lexicale** ou **linguistique quantitative** (Muller, 1964, 1967, 1973, 1979), **statistique textuelle** (Salem 1987, 1994), voire **analyse des données en linguistique** (Benzecri 1981), **la lexicométrie** (Tournier 1975, Lafon 1984) n'est pas une théorie mais une méthodologie d'étude du discours, qui se veut exhaustive, systématique est automatisée. » (Tournier, 2002).

D. Mayaffre souligne que : « le numérique complète-augmente la linéarité par des réticularités » (Mayaffre, 2015). À travers l'analyse numérique, la vision du texte et la façon de le lire changent ; nous ne lisons plus ce dernier d'une manière purement linéaire mais plutôt en produisant des liaisons transversales entre les mots fonctionnant en réseau. Appelée aussi logométrie, la lexicométrie est une méthodologie réservée à l'analyse automatique du discours.

4.2. Démarches de recherche

Afin de réaliser une recherche sur le contexte d'un écrit scientifique, nous sommes partie d'un corpus composé de douze mémoires de magister rédigées en français dans des disciplines diverses. Vu la grande taille du corpus, nous avons recouru à une étude assistée par le logiciel *Hyperbase*[1], réservé à l'analyse automatique du discours. Le corpus de référence est général et très varié, il se compose des documents non liés à un domaine spécifique. La comparaison est réalisée par rapport à la banque textuelle ***Frantext*** parce qu'elle regroupe 350 romans et recueils de nouvelles français publiés de 1830 à 1970, le genre le mieux représenté. Grâce à cette fonction, nous avons choisi le mot *travail* qui figure parmi les mots les plus fréquents dans le corpus, pour étudier son contexte. Nous préférons comme mot-pôle les lemmes parce que ces derniers sont les plus performants (Brunet, 2011). Vu que le mot *travail* isolé ne suffit pas lui seul pour son interprétation, le besoin de repérer son contexte et construire un texte autour de ce mot était nécessaire pour l'interpréter. Rappelons que l'objectif de cet article est de montrer que la recherche sur le contexte d'un mot quelconque peut ainsi se réaliser sans appel à la linéarité du texte notamment quand ce dernier est très long. Il faudrait aussi élargir le cercle des liens du mot-pôle pour éviter une interprétation étroite.

[1] Figure parmi les logiciels disponibles pour la plateforme Windows et sous Mac OS. crée par E. Brunet,il est un outil d'analyse statistique et graphique à partir duquel il est possible de définir la quantité de l'écrit, calculer la haute fréquence des mots et des signes de ponctuation, la distance lexicale entre les textes du corpus , la richesse du vocabulaire et son évolution, la typologie du discours, les thèmes transversaux du corpus, la comparaison d'un texte à d'autres textes ou une comparaison interne, etc.

4.2. Contexte par réseau

Après avoir choisi *travail* comme thème, il sera défini grâce à son lien avec tous les autres mots qui se trouvent dans son entourage proche ou lointain comme dans le graphe ci-dessous :

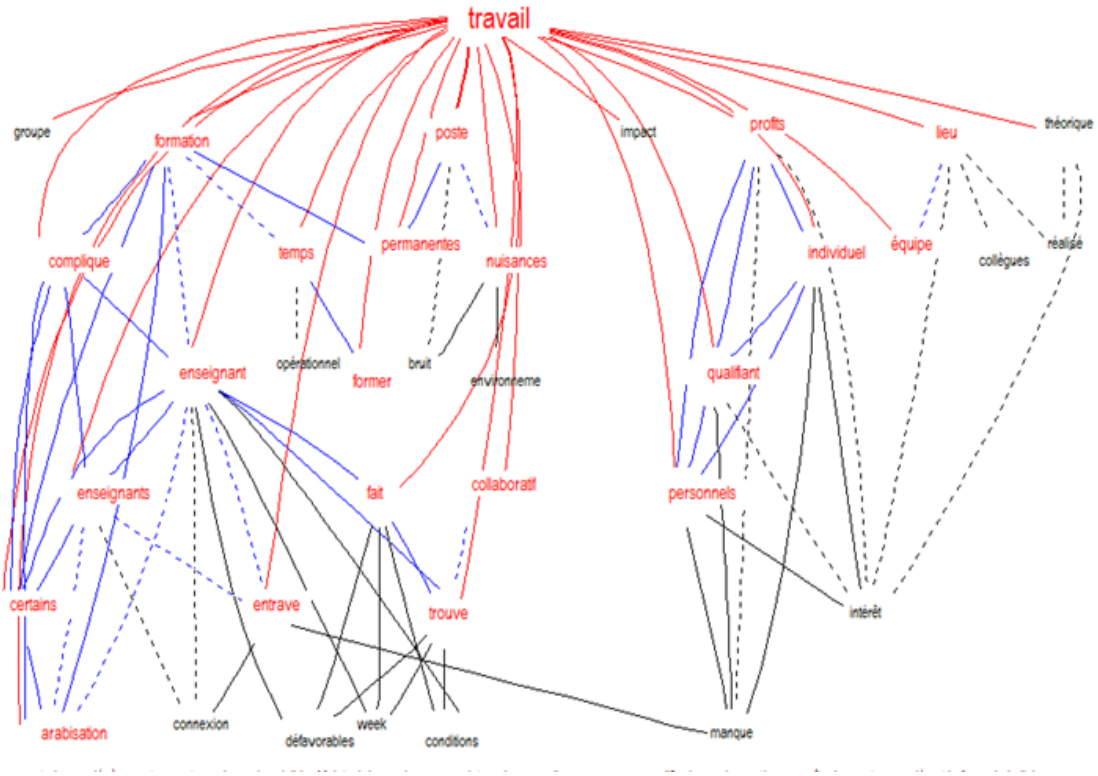

Figure 1 : graphe simplifié du mot travail

Le graphe ci-dessus, est constitué par le programme *contexte* en comptabilisant que les mots en couple ayant une relation directe ou indirecte avec le mot-vedette. Il est un *graphe simplifié*[2] choisi pour éliminer tous les mots outils étant privés de performance sémantique qui permet de définir les thèmes au tour du mot en question comme celle des lemmes qui rendent le graphe plus complexe. Selon ce graphe, on distingue :

[2]Les mots représentés en noir sont les nœuds les plus fréquentés. Les nœuds colorés en rouge sont moins fréquentés et exposent moins de cinq liaisons.
La taille des caractères est relative à la participation du mot au réseau. Les traits tracés en rouge correspondent aux cooccurrences directes avec le mot-axe constituant un cercle des amis.
Les traits tracés en bleu distinguent les relations que les amis du premier niveau ont entre eux. Les traits tracés en noir intéressent le deuxième rang ou les amis des amis.
Les traits pointillé, maigre ou gras en couleur correspond à la force de la liaison.

- Un réseau à droite qui contemple les mots *personnels, qualifiant, profits, lieu* colorés en rouge, dont le lien avec le mot-pôle est fort, et le mot *théorique* en noir attestant d'un lien moins solide avec *travail.*
- Un réseau à gauche faisant appel directement et fortement aux mots *formation, enseignants, enseignant, temps, entrave, poste, nuisance* qui sont en rouge, et au mot *groupe* coloré en noir.
- Le mot pôle a traversé certains items qui sont en relation directe avec lui comme *formation, poste, profits* et *lieu* pour se relier à d'autres termes qui sont en rouge ou en noir se trouvant au deuxième niveau. Des liens qui sont moins puissants en avançant vers le bas du graphe par rapport à ses contacts directs. Cette structure *pôle-amis-proches d'amis* rend la lecture du graphe complexe, mais c'est grâce à cette représentation lexicale qui expose l'environnement complet du mot-axe que le thème choisi est validé. Nous remarquons que le lien direct et fort des mots du réseau indique de diverses manières d'aborder les sujets qui constituent le contexte de ce terme. Par exemple, *la formation des enseignants au travail, comment avoir un poste de travail d'enseignant,* etc. :
- Aussi, nous pouvons extraire à partir des proches des amis *du* mot-pôle *les sujets profits du travail, le lieu du travail, la formation au travail,* etc.
- Notant que le réseau de la droite du graphe est lexicalement plus riche que celui de la gauche formant lui-même un sous-thème autour du domaine de l'enseignement. Quant au réseau de la gauche, dont le lexique est moins nombreux que celui de la droite, nous pouvons l'attribuer le sous-thème *environnement du travail.*
- Les nœuds les plus fréquentés sont, par exemple, *formation, enseignants, profits,* des termes qui vont ensemble pour parler *des formations au profit des enseignants au travail.*

En effet, vu que le mot seul ne peut pas introduire du sens sans son contexte, nous l'avons associé à d'autres mots qui seront un contexte éclairant pour lui, au sein d'une phrase ou d'un paragraphe qui se situe dans un contexte plus grand.

Si nous partons de l'idée de D. Mayaffre**,** un contexte peut être validé à partir de deux termes corrélés ou colloqués. Ce que nous avons constaté selon les données de la présente recherche c'est qu'à partir d'un contexte de deux mots co-occurrents ou deux « formes minimales du contexte » (Mayaffre, 2008) comme *travail* et *formation,* on a pu investir un deuxième contexte plus large (*quelle est la démarche suivie dans la formation au travail dans le cas des enseignants…*) que nous appelons *contexte-phrase* et ainsi de suite jusqu'au *contexte-texte,* construit à partir de tous les items se manifestant autour du mot-vedette. En d'autres termes, à

partir de deux mots, nous pouvons construire une phrase contextualisée, et à partir de cette dernière, nous pouvons bâtir un paragraphe, un texte entier dont le thème général, l'« isotopie générique dominante » (Rastier, 1995) ou simplement le thème est à repérer facilement. Par exemple, en revenant au graphe plus haut, nous remarquons que les dualités que le mot *travail* a construites à partir de ses différents liens ont confirmé que le mot *travail* est un thème générique du corpus qui a émergé à partir de la constitution des liaisons qui représentent ses réseaux thématiques. Enfin, ces mots sont attendus dans le discours des auteurs de ce corpus. Dans un mémoire de fin d'étude universitaire, n'importe quel scripteur fait appel à des termes renvoyant aux difficultés rencontrées lors de sa recherche et sa rédaction de mémoire.

Le terme *travail* dont le chiffre dans le corpus est important (826 occurrences) atteste du besoin à l'utiliser dans les différents textes du corpus. C'est tout à fait normal que *travail* apparait dans un discours des étudiants de parcours universitaire, qui considèrent que le mémoire de fin d'étude est déjà un travail en soi, qui pensent sûrement à leur avenir proche, comment pratiquer leurs compétences acquises à l'université, au terrain professionnel, lieu de leur prochaine étape de vie, etc.

Le nombre du mot *travail* apparait d'une manière disproportionnée dans les textes composant ce corpus et pourtant ses co-occurrents ne montrent qu'un emploi homogène comme s'il s'agit d'un seul texte et d'un seul contexte malgré les différents domaines et les auteurs distincts dont ils dépendent les discours composant ce corpus. Qu'est-ce qu'il renforce l'attirance entre ces résonances collocationnelles construites à partir du contexte ? Probablement c'est leur appartenance à un seul genre d'écrit, celui du discours universitaire, de l'écrit scientifique.

L'environnement thématique du terme *travail* a permis d'examiner ses corrélats qui selon leur échange de lien confirment leur dépendance du même univers lexical et sémantique. En effet, c'est l'interaction entre les différents discours composant le corpus qui a généré un lien connu par l'intertextualité ou le dialogisme (Stolz, 2009). Cette intertextualité est née du fait que ces sous-corpus, descendant de disciplines diverses et partageant une caractéristique dominante qui a pu malgré de nombreuses différences (terminologie disciplinaire, catégorisation textuelle, ponctuation, etc.) influencer le corpus entier créant par là un contexte commun entre ses textes. Cette force éminente est celle du genre.

Conclusion

Le mot peut être considéré comme un contexte minimal à partir duquel une situation plus large se développe. Ce contexte étendu se constitue suite à une sélection des acceptions introduites lors de l'interprétation du sens lexical qui établit la relation entre un mot-vedette et tous les mots qui se trouvent dans son environnement lexical. Le contexte dépend donc du choix du mot-pôle, des stratégies d'interprétation de l'auteur du texte et surtout du genre.

Références bibliographiques

1. ABLALI. D. 2010. Linguistique des genres. Exploration sur corpus, (26/02/2017) <http://www.academia.edu/12231258/Linguistique_des_genres_sur_corpus>
2. ADAM. J-M. 2005, *Les textes : type et prototype*, Paris, Colin, 220 pages.
3. BRUNET. E. 2011. « Hyperbase, manuel de référence », (04/01/2016) <http://ancilla.unice.fr/bases/manuel.pdf>
4. GARÇIA-DEBANC C. 1990. *L'élève et la production d'écrits*, Metz, Centre d'analyse syntaxique de l'Université de Metz.
5. MAINGUENEAU D. 2014. *Discours et analyse du discours*, Paris, A. Colin.
6. MAYAFFRE Damon, 2008, « De l'occurrence à l'isotopie », (31/12/2015) <https://hal.archives-ouvertes.fr/hal-00551114/document>
7. MAYAFFRE, D. 2012. Nicolas Sarkozy, *Mesure et démesure du discours* (2007-2012). Paris, Presses de Sciences Po, 368 pages.
8. RASTIER F. 1987. *Sémantique interprétative*, Paris, Presses Universitaires de France, 276 pages.
9. RASTIER F. 1989. *Sens et textualité*, Paris, Hachette, 286 pages.
10. RASTIER F. 1995. « La sémantique des thèmes ou le voyage sentimental », (18/08/2016) <http://www.revue-texto.net/Inedits/Rastier/Rastier_Themes.html>
11. RASTIER. F. 2005. « La microsémantique », (05/05/2017) <http://www.revue-texto.net/Inedits/Rastier/Rastier_Microsemantique.html>
12. STOLZ. C. 2009. « Dialogisme », (04/03/2017) <http://www.fabula.org/atelier.php?Dialogisme>
13. TOURNIER. M. 2002. Lexicométrie, in CHARAUDEAU Patrick et MAINGUENEAU Dominique, *Dictionnaire d'Analyse des discours*, Paris, Seuil.
14. WILLIAMS. G & MILLON Ch. 2013. « Les verbes et la science : construction d'un dictionnaire organique », p. 45 à 66 in : TUTIN Agnès (dir), GROSSMANN Francis (dir), *L'écrit scientifique : du lexique au discours*. Rennes, Presses universitaire de Rennes, 246 pages.