CONDITIONS POUR CONSTITUER LES CORPUS DISCURSIFS DE SPECIALITE

Ioana-Crina Coroi, Assist. Prof., PhD, "Ștefan cel Mare" University of Suceava

Abstract: This communication aims to present an approach to discursive, cognitive and linguistic conditions involved in the process of establishing a specialized corpus. The aim of our work is to emphasize that, to constitute a corpus of discursive analysis, is required to take into account the empirical data as well as the theoretical constraints that gives corpus homogeneity and functionality.

Keywords: corpus, discourse, scientific, homogeneity, functionality.

I. Pour une approche sommaire de la linguistique de corpus

Constituer tout corpus signifie faire confiance à une discipline exploratoire, à une discipline dont le caractère pratique ne peut pas être nié ou ignoré. Il s'agit de la linguistique de corpus. En fait, la linguistique de corpus, en tant que discipline développée assez récemment, issue de l'école contextualiste, vise le dépassement des formalismes pour mettre en évidence les mécanismes de la communication que les locuteurs utilisent dans des contextes réels. Ses caractéristiques définitoires l'inscrivent dans la sphère de la linguistique appliquée, ayant des instruments qui puissent la lier à l'informatique et à la terminologie, à la traduction, à l'étude des langues étrangères, à la lexicographie, restant quand même fidèle aux sciences humaines, en dépit de son adaptabilité évidente dans plusieurs domaines de recherche.

Un corpus est forcément constitué tout en respectant des séries de critères établis par les chercheurs, c'est-à-dire des critères de choix plutôt sociolinguistique. La structure de tout corpus marque le dépassement d'un simple texte à analyser, l'orientant vers une analyse plus complexe et plus ordonnée selon des paramètres bien déterminés pour construire, en réalité, un ensemble de textes. Sans doute, les raisons de communication placées à la base des choix des locuteurs renvoient aux productions à caractère pragmatique où l'artificialisation ne puisse pas trouver sa place, puisque les corpus de référence ou d'exploitation doivent soutenir entièrement les besoins des chercheurs.

En 2005, dans la préface du volume qu'il a coordonné, Geoffrey Williams affirmait que « la linguistique de corpus est à la fois une discipline et une méthodologie, il faut donc pouvoir mettre en lumière les aspects techniques » (Williams, 2005 : 17) et c'est justement ce que nous nous proposons de faire dans la partie suivante de notre recherche qui porte sur l'illustration de la constitution des corpus à caractère scientifique.

II. Constitution et exploitation des corpus scientifiques

Les corpus représentent de véritables instruments pour l'investigation de tout discours et pour en extraire les informations et les structures linguistiques qui offrent la validité scientifique d'une certaine démarche. Le chercheur opère avec des corpus pour illustrer ses propos, il ne fait pas une analyse sans appui, mais une investigation concrète et opérationnelle par le biais de laquelle il formule et vérifie la réalisation des hypothèses scientifiques énoncées au début de sa démarche.

BDD-V640 © 2014 Arhipelag XXI Press
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 22:55:12 UTC)

Toute une série de questions principales doivent être envisagées pour la constitution des corpus : *l'homogénéité, la représentativité, la généralisation, la fonctionnalité* etc. Les données des corpus doivent être soumises à des conditions concrètes pour répondre à la réalisation des objectifs de la recherche, des conditions qui offrent une image cohérente et articulée sur toute la démarche d'investigation. Ainsi, pour illustrer une partie des conditions nécessaires à la constitution des corpus, nous allons focaliser notre attention sur des corpus scientifiques du domaine forestier. Les échantillons de corpus, des fragments de textes de spécialité, constituent quelques exemples à discuter selon les critères de l'objectivité et de la subjectivité.

En général, le discours scientifique est objectif, ses limites sémantiques étant très bien esquissées. Les productions de discours scientifique sont ancrées dans des réalités vérifiables, les perspectives discursives sur les textes offerts ne présentent pas une pluralité de choix à interpréter, ni les moyens stylistiques qui puissent intervenir dans la bonne réception des messages.

Par exemple, il y a des auteurs d'ouvrages à caractère scientifique qui préfèrent exprimer leurs points de vue dans la préface du livre de spécialité, tout en établissant une sorte de liaison subjective avec l'interlocuteur, une sorte d'invitation à une lecture qui réponde à des conditions concrètement découpées et organisées pour donner de la cohérence discursive. Ce type de préambule, réalisé dans le présent de la communication, crée une implication directe de l'énonciateur subjectif dans la lecture de ses interlocuteurs :

[T1] « Toutes les plantes de cet ouvrage ont quelque chose à nous proposer pour nous aider à apprivoiser la vie. Vous serez étonnés de voir à quel point elles prennent soin de nous. Elles nous montrent la voie, et croyez-moi, elles savent de quoi elles parlent. Depuis si longtemps qu'elles sont sur terre, elles en ont vu des gens se fourvoyer. » (Funiel, 2005 : 8-9)

Ce texte [T1] représente un bon exemple de subjectivité locutoriale, une image de l'intention auctoriale de devenir partie intégrante dans la perspective de ses lecteurs. Le choix des verbes et de leur emploi temporel, la présence des pronoms personnels *nous, vous*, de même que la mobilisation de la fonction conative dans le discours y font témoignage.

Pour la plupart des préfaces des textes scientifiques, les verbes utilisés dans la construction du discours sont intégrés dans la catégorie du passé (le passé composé y est dominant, mais il existe aussi des situations discursives lorsque les temps verbaux sont l'imparfait et le plus que parfait) :

[T2] « L'habitat des espèces est détaillé, comme dans toute flore. Mais chacune, avec d'autres espèces, appartient à des *unités de végétation* bien définies ; nous avons précisé cette appartenance à une communauté végétale particulière en faisant appel à plusieurs systèmes d'analyse du tapis végétal par :

- *formation végétale* (unité définie à partir de la physionomie d'ensemble de la végétation et / ou des espèces dominantes), système le plus simple (pelouses, marais, prairies, lisières forestières, fruticées, forets : chênaies, hêtraies-sapinières …, laies ou coupes forestières, etc.) ;
- *série de végétation*, dans le cas où la plante se rencontre dans presque tous les stades d'évolution conduisant à un climax donné ;
- *alliance phytosociologique* (parfois ordre ou classe) ; on trouvera en annexe 2 le détail de la classification phytosociologique adoptée avec les espèces (décrites dans le

BDD-V640 © 2014 Arhipelag XXI Press
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 22:55:12 UTC)

présent tome « montagnes ») caractéristiques ou différentielles, ainsi que les compagnes principales, de chaque unité. (Rameau *et alii*, 1993 : 11)

Ce type de préface, avec des verbes au passé, offre une clé de lecture, un répertoire de conditions de lecture, tout à fait indisponibles pour une certaine catégorie d'utilisateurs qui préfèrent avoir, dès le début, le schéma discursif complet sur la structure de l'ouvrage à parcourir. Il s'agit également d'un bon instrument d'investigation mis à la disposition des chercheurs qui s'attardent sur des concepts, des notions ou des syntagmes de spécialité pour y trouver des réponses à leurs objectifs énoncés. En voici une telle situation discursive :

[T3] « À un *index des noms botaniques* latins et français des espèces décrites, nous avons joint un vocabulaire des termes techniques et scientifiques. Ce dernier pourra paraître un peu développé, mais nous avons pensé qu'il aidera une gamme d'utilisateurs aussi large que possible, forestiers ou non, à tirer le meilleur profit de cet ouvrage ». (Jacamon, 1996 : XII)

Un autre type de situation discursive est présent dans des discours scientifiques préfacés par un autre spécialiste qui offre une perspective globale sur le livre en discussion, fait intéressant à analyser dans la situation où l'on est intéressé à construire un corpus de textes qui illustrent l'intention directe de l'auteur par rapport à son produit discursif. Comme exemple, nous avons découpé un petit échantillon textuel de la présentation réalisée par R. Rol, Directeur de l'Ecole Nationale des Eaux et Forêts, pour le livre de Pierre Cochet :

[T4] « L'auteur a adopté, dans la rédaction de certaines parties de son livre, un style dialogué qui, au premier abord, surprendra peut-être un peu. Il se substitue en quelque sorte au lecteur pour soulever les objections, les discuter et envisager les solutions possibles. Cette forme rend plus personnelle l'étude des techniques à adopter et permettra sans doute au lecteur de mieux saisir l'importance de ses interventions dans les peuplements ». (Cochet, 1997 : 14)

Sans doute, dans [T4] la perspective discursive change et le lecteur ne peut pas savoir si la constitution des produits discursifs - qu'il a sous les yeux - respectent les conditions de constitution réclamées par l'auteur proprement-dit. L'objectivité et la subjectivité dans la réception des informations scientifiques s'appuient alors sur d'autres éléments, tels la présence des définitions, des classifications, des perspectives objectives qui présentent des situations réelles et correctes de point de vue scientifique.

Un corpus de définitions [T5-T7] peut offrir une réponse honnête à la question de la véridicité des informations présentées, du sérieux de l'acte de documentation entreprise par le chercheur :

[T5] « Un ARBRE (sapin, chêne rouvre) est un végétal à tige ligneuse, atteignant, à l'état adulte, une hauteur de 10 mètres, au moins. Si la hauteur atteinte est inférieure à 10 mètres, il s'agit d'un ARBUSTE (noisetier). Lorsque cet arbuste ne possède pas une tige élaguée, c'est-à-dire lorsque sa tige est ramifiée dès sa base, il prend le nom d'ARBRISSEAU (troëne). Les arbrisseaux qui n'atteignent pas une hauteur de 1 mètre sont appelés SOUS-ARBRISSEAUX (myrtille, callune) ». (Cochet, 1997 : 38)

[T6] « […], on groupe souvent sous le nom de MORT-BOIS, les arbustes, arbrisseaux et sous-arbrisseaux qui existent en forêt » (*idem*)

[T7] « […], on appelle ESSENCE FORESTIERE, ou simplement ESSENCE, toute espèce d'arbre susceptible de croître en forêt » (*idem*)

BDD-V640 © 2014 Arhipelag XXI Press
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 22:55:12 UTC)

En même temps, dans une catégorie similaire, on peut intégrer les classifications scientifiques, des informations qui se constituent dans des corpus véridiques et réels qui décrivent des situations discursives concrètes et qui les distribuent sur des niveaux perceptibles :

[T8] « Le martelage d'une coupe d'éclaircie comporte trois aspects :

- le choix des arbres dont la suppression va améliorer la composition ou la croissance du peuplement,
- la désignation explicite de ceux-ci de telle manière que l'exploitation porte bien sur eux et sur eux seuls,
- l'estimation en matière puis en argent des tiges qui vont être vendues et abattues, dans le but d'en réaliser la vente au plus juste prix » (Cochet, 1997 : 99)

Le cas de la subjectivité dans le discours scientifique concerne une problématique très intéressante qu'on peut aborder par toute une série d'instruments d'investigation discursive qui, par exemple, appartiennent à la théorie de l'Imaginaire linguistique (IL). N'oublions pas que cette théorie possède un vaste répertoire normatif pour analyser un corpus du type suivant :

[T9] « L'alkékenge est de la famille des solanacées, celle où figure la belladone, plante extrêmement toxique. C'est bien la seule chose délicate qui est liée à la plante, encore que les alcaloïdes que renferme l'alkékenge ne sont pas vraiment dangereux ». (Fuinel, 2005 : 18)

[T10] « Maintenant, la bourrache croît un peu partout mais elle est peu utilisée. Il faut dire que c'est une plante un peu mystérieuse. Sa tige est creuse, elle se développe facilement. Elle a des fleurs bleues absolument merveilleuses mais elle est très désagréable au toucher ». (Fuinel, 2005 : 24)

[T11] « Ce qui caractérise l'hysope c'est son odeur agréable, envoutante, séduisante. Les Hébreux, d'ailleurs l'appellent *ezob*, ce qui se traduit par herbe de bonne odeur, et pourquoi pas, nous suggère la plante *herbe de bonne humeur*. En effet, l'hysope n'est pas une triste. Mais pour les Hébreux, elle était avant tout une plante sacrée et ça ne les incitait pas à rigoler ». (Fuinel, 2005 : 38)

[T12] « Le mélilot, belle plante jaune, égaie le bord des routes et des sentiers dans les campagnes délaissées. Il est toujours bien mis, bon chic, bon genre, mais sans prétention, il est toujours prêt à la discussion, à la rencontre. Il constate malheureusement que les promeneurs passent à côté de lui sans le remarquer. Dommage, il a plein de choses à nous dire ». (Fuinel, 2005 : 61)

[T13] « La sauge porte dans son nom l'excellence de sa qualité médicinale, *salvia* : celle qui sauve. Et la grande considération qu'ont eu toutes les civilisations vis-à-vis de cette plante témoigne des immenses services qu'elle rend à l'humanité » (Fuinel, 2005 : 72)

Les textes qui constituent le corpus présenté [T9-T13] peuvent facilement constituer un bon exemple de corpus propre à une analyse discursive appuyée sur la catégorie des normes subjectives dans la théorie de l'IL. Il est bien évident le fait que l'auteur de ce type de discours scientifique préfère exprimer son opinion en ce qui concerne le référent scientifique, parfois même faisant appel aux figures de style et aux repères historiques et culturels pour

BDD-V640 © 2014 Arhipelag XXI Press
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 22:55:12 UTC)

soutenir l'importance et la réalité consacrée de ses produits discursifs. Ce qui démontre également que le locuteur qui opère avec des réalités scientifiques n'est pas seulement un représentant de l'objectivité discursive qui se borne à transmettre des informations à caractère général, mais aussi un représentant de la bonne coopération entre les domaines scientifiques et humanistes.

III. En guise de conclusion

Les exemples de corpus offerts dans notre démarche discursive constituent quelques intentions d'offrir de possibles axes d'investigation du domaine scientifique à partir de différentes conditions de production. Il ne s'agit pas d'un schéma standard qui soit opérationnel dans l'analyse de tout corpus de langue, puisque le choix langagier de celui qui construit un corpus peut changer d'un cas à l'autre, d'un contexte à l'autre. La langue est dynamique, elle se transforme continuellement sous l'impact des réalités discursives extérieures et nous soutenons également les perspectives discursives selon lesquelles : « Ce que l'on nomme vérité à caractère général conservera toujours ce caractère. L'idée restera toujours la même. Seule la forme qu'elle épousera sera différente à cause de, ou plutôt grâce à ce processus continuel de renouvellement de la langue » (Moroșan, 2006 : 377)

La démarche de constitution des corpus doit avoir quelques points de repère qui orientent d'une manière concrète la production discursive pour acquérir de l'homogénéité et de la fonctionnalité dans tout contexte d'analyse approfondie. Il faut explorer des contextes réels, opérer avec des instruments conceptuels adéquats aux disciplines qui coordonnent la recherche, réaliser des structures cohérentes qui puissent valider les objectifs de la recherche. Sans doute, les corpus y jouent un rôle définitoire, ils désignent l'adéquation des instruments d'analyse aux contextes envisagés, ils marquent un certain positionnement des locuteurs par rapport aux champs discursifs soumis à l'analyse.

Les modes d'organisations discursifs et textuels offrent également des conditions appropriées pour la mise en pratique de l'acte de constitution des corpus, ils mettent à la disposition des chercheurs les structures, voire les superstructures, qui déclenchent et valident des phénomènes linguistiques extrêmement intéressants dont la contribution pour le domaine des sciences du langage est définitoire. Tout corpus représente une entité en soi, un mélange représentatif d'éléments discursifs qui puissent transformer radicalement les perspectives focalisées sur les discours. Il s'agit également d'une marque d'originalité des locuteurs ou des chercheurs, d'une façon inédite de faire preuve d'un imaginaire élargi, adapté à différentes situations discursives, généreuses pour toute investigation à caractère pragmatique.

***Note**: Cet article a été financé par le projet « SOCERT. Société de la connaissance, dynamisme par la recherche», n° du contrat POSDRU/159/1.5/S/132406, cofinancé par le Fonds Social Européen, par le Programme Opérationnel Sectoriel pour le Développement des Ressources Humaines 2007-2013. Investir dans les Gens!*

Bibliographie

Charaudeau, P., Maingueneau, D. (coord.), *Dictionnaire d'analyse du discours*, Seuil, Paris, 2002.

BDD-V640 © 2014 Arhipelag XXI Press
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 22:55:12 UTC)

Moroșan, N., „Les propriétés du discours scientifique pascalien”, in *Le français et les sciences*, Editions Universitaires „Alexandru Ioan Cuza”, Iași, 2006.
Rovenţa-Frumuşani, D., *Analiza discursului - ipoteze şi ipostaze*, Editura Tritonic, Bucureşti, 2005.
Williams, G. (coord.), *La linguistique de corpus*, Presses Universitaires de Rennes, 2005.

Corpus d’analyse

Cochet, P., *Etude et culture de la forêt. Manuel pratique de gestion forestière*, ENGREF, Nancy, 1977.
Fuinel, G., *La bonne humeur & Les plantes*, Editions Amyris, Bruxelles, 2005.
Jacamon, M., *Guide de dendrologie. Arbres, arbustes, arbrisseaux des forêts françaises*, ENGREF, Nancy, 1996.
Rameau, J.C., Mansion D., Dumé G., Lecointe A., Timbal J., Dupont P., Keller R., *Flore forestière française. Guide écologique illustré*, Institut pour le développement forestier, 1993.

BDD-V640 © 2014 Arhipelag XXI Press
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 22:55:12 UTC)