



Ghislaine Chartron, Benoît Epron et Annaïg Mahé (dir.)

Pratiques documentaires numériques à l'université

Presses de l'enssib

Lumière sur le Web universitaire : l'influence de la culture disciplinaire sur les représentations en ligne

Jenny Fry

DOI : 10.4000/books.pressesenssib.1172

Éditeur : Presses de l'enssib

Lieu d'édition : Presses de l'enssib

Année d'édition : 2012

Date de mise en ligne : 20 juillet 2017

Collection : Papiers

ISBN électronique : 9782375460511



<http://books.openedition.org>

Référence électronique

FRY, Jenny. *Lumière sur le Web universitaire : l'influence de la culture disciplinaire sur les représentations en ligne* In : *Pratiques documentaires numériques à l'université* [en ligne]. Villeurbanne : Presses de l'enssib, 2012 (généré le 01 février 2021). Disponible sur Internet : <<http://books.openedition.org/pressesenssib/1172>>. ISBN : 9782375460511. DOI : <https://doi.org/10.4000/books.pressesenssib.1172>.

+++++

LUMIÈRE SUR LE WEB UNIVERSITAIRE : L'INFLUENCE DE LA CULTURE DISCIPLINAIRE SUR LES REPRÉSENTATIONS EN LIGNE ¹

+++++

INTRODUCTION

+++++

Cet article a pour but d'établir un pont entre les caractéristiques culturelles fondamentales de divers domaines d'études et la façon dont les activités de recherche sont représentées sur le Web universitaire, en mettant l'accent sur l'analyse qualitative des liens hypertextes². Pour ce faire, nous nous demanderons si la théorie de Whitley [1984] concernant l'organisation intellectuelle et sociale des sciences, bien connue dans le champ des études des sciences et des techniques, peut s'appliquer au contexte du Web. La théorie de Whitley s'appuie sur deux concepts capitaux : le degré de *mutual dependence* qui existe entre les chercheurs ou les disciplines en matière de contribution au corps des connaissances, et le degré de *task uncertainty*, qui touche à la coordination et à l'interprétation des méthodes de recherche et leurs résultats. L'application de ces deux concepts, intimement liés et relatifs, constitue un véritable défi car on ne peut les mesurer en termes absolus. En examinant le comportement des communautés de chercheurs en mode « connexion » ou « déconnexion », cet article fournit un cadre explicatif à la compréhension des phénomènes intellectuels et sociaux qui sous-tendent les relations hypertextuelles dans le Web universitaire.

Comme le soulignent Harries *et al.* [2004] ainsi que Thelwall [2006], malgré l'intérêt actuel pour les relations hypertextuelles et les nombreuses statistiques indiquant que l'on peut associer les liens Internet à l'importance et à la productivité d'une discipline, aucun lien de cause

1. N.D.T. : cet article a été publié sous le titre "Studying the scholarly Web: how disciplinary culture shapes online representations", dans *Cybermetrics: International Journal of Scientometrics, Informetrics and Bibliometrics*, 10 (1), 2006. [En ligne] < <http://cybermetrics.cindoc.csic.es/articles/v10i1p2.pdf> >.

2. Pour une typologie des analyses de liens, voir Thelwall [2006].

à effet n'a été démontré, et les efforts pour fournir un cadre théorique permettant de donner un sens probant à ces résultats en sont restés au stade de tentatives. Pour Harries et ses collaborateurs [2004], il est urgent de remédier à cette situation car la compréhension lacunaire de la création des liens Internet constitue un obstacle au développement du champ émergent qu'est la webométrie.

Dans ses recherches sur les motivations sociales qui sous-tendent la création des liens Internet, Thelwall [2006] signale les limites d'une analyse de liens à grande échelle et prône la triangulation des méthodes qualitative et quantitative. Cet article tente de pallier les lacunes concernant la théorie sociale de la webométrie en puisant uniquement dans les techniques de l'ethnographie virtuelle [Hine, 2000 ; Marcus, 1995], en vue de contribuer au concept de sphère Internet³ développé par Schneider et Foot [2002]. Schneider et Foot ont développé ce concept dans un certain nombre de contextes non universitaires, en particulier dans le cadre d'une étude sur les campagnes électorales en ligne. Ils définissent une sphère Internet comme étant un « ensemble de ressources numériques définies recouvrant plusieurs sites Internet jugés pertinents ou en rapport avec un thème ou un objet central ». Dans cette étude, nous avons pris pour « thème central » les frontières épistémologiques de deux champs d'études précis.

Ces domaines sont : la théorie de l'argumentation d'une part, champ interdisciplinaire au confluent des études en philosophie, linguistique et communication ; et la linguistique pragmatique d'autre part, champ également interdisciplinaire regroupant des aspects de la linguistique computationnelle, tels que le traitement automatique du langage naturel, et des domaines de la linguistique plus traditionnelle, tels que la sociolinguistique, la linguistique contrastive et la morphologie.

La linguistique pragmatique se caractérise par un très haut degré de travail collaboratif, lequel se réalise en termes de projets à l'échelle nationale et dont les résultats aboutissent, entre autres, à l'élaboration de corpus [Becher, 1989]. L'étude de cas de linguistique pragmatique s'appuie sur un important projet de corpus national aux Pays-Bas : le *Corpus Gesproken Nederlands* (CGN)⁴, le corpus de néerlandais parlé.

La théorie de l'argumentation se caractérise, quant à elle, par des écoles de pensée plutôt que par des projets [Becher, 1989]. Notre étude

3. N.D.T. : nous choisissons de traduire *websphere* par « sphère Internet » dans cet article.

4. CGN : < <http://lands.let.kun.nl/cgn/> >.

de cas reflète cette donnée et se concentre sur ce que les chercheurs de ce domaine nomment « l'École d'Amsterdam », groupe de recherche basé dans le département de communication du langage, de théorie de l'argumentation et de rhétorique de l'université d'Amsterdam.

Si ces deux études de cas tombent sous la coupe de la même discipline mère – la linguistique –, il n'en demeure pas moins que les objectifs, l'identité culturelle et la visibilité sur Internet de leur communauté universitaire respective, diffèrent sur de nombreux points. Les sphères Internet imbriquées ont été construites en se basant sur des entretiens avec les deux chefs de projet du CGN et un membre de l'École d'Amsterdam ainsi que sur des données collectées par le biais de Microsoft Site Analyst et de moteurs de recherche⁵.

Les résultats pointent vers des indicateurs qualitatifs potentiellement étalonnables qui rendent ce que Cronin et ses collaborateurs [1998] décrivent comme les traces liminales de l'estime de ses pairs, d'influence sur la réputation et les ressources, et la légitimation des problèmes, approches et techniques de recherche.

L'ÉTUDE DES PRATIQUES DES CHERCHEURS UNIVERSITAIRES SUR INTERNET

 Les études sur la communication universitaire ont utilisé divers niveaux de granularité comme unités d'analyse. Cela reflète la production du savoir dans une multitude de configurations sociales et institutionnelles se chevauchant parfois, comme les groupes de chercheurs, les projets, les collègues invisibles, les domaines spécialisés, les laboratoires, les départements et les disciplines universitaires, dont les frontières manquent parfois de précision.

Price [1963] et Crane [1972] ont utilisé la notion de communautés scientifiques orientées vers la recherche se regroupant autour de problèmes ou de méthodes de recherche comparables. Le concept de communauté scientifique est souvent décrit comme idyllique et consensuel et en tant que tel a été critiqué [Swales, 1998]. Price a inventé le terme de collègue invisible en 1963 : il s'agit pour lui d'un réseau social d'une capacité d'environ 200 membres regroupés autour d'un noyau d'une vingtaine de chercheurs. Price a expliqué qu'au-delà de cette capacité, le domaine de

5. Voir les annexes : [en ligne] < http://cybermetrics.cindoc.csic.es/articles/v10i1p2_Appendices.html > (consulté le 3 mai 2004).

recherche arrivait à saturation et que les groupes se divisaient afin de former de nouvelles disciplines spécialisées. Des chercheurs ont développé ce concept dans un certain nombre de disciplines. En 1983, Kochen a, par exemple, eu recours à l'estimation de la taille d'un groupe de recherche de pointe en modélisation mathématique afin de simuler l'évolution des disciplines spécialisées. En sociologie, Crane [1972] s'est quant à lui appuyé sur la notion de collège invisible de Price pour exemplifier la diffusion des connaissances *via* les réseaux des communautés scientifiques. Dans des travaux plus récents, Bruckner, Ebeling et Scharnhorst [1989] ont utilisé la théorie de Price pour simuler des instabilités dans les systèmes scientifiques évolutionnistes et ont poursuivi le modèle de Kochen.

S'orientant vers une vision moins idyllique des communautés scientifiques où la compétition et la répartition du travail sont capitales, Bourdieu a introduit la notion de champ intellectuel en 1988. Selon la définition de Bourdieu, les champs sont des espaces fluides basés sur le principe de la compétition – plutôt que sur le consensus, comme il en est question dans le concept des « communautés de pratique » [Wenger, 1998] –, principe selon lequel des communautés organisées en réseau sont en compétition pour les mêmes problèmes et les mêmes ressources.

La notion de discipline constitue, quant à elle, une façon plus institutionnelle d'envisager les communautés scientifiques. Les disciplines scientifiques sont indissociables des institutions pédagogiques et de formation, et représentent des monopoles de marché de certains types de connaissances [Whitley, 1984]. Salter et Hearn [1996, p. 23] ont réuni les thèmes récurrents qui tentent de définir ce qui caractérise une discipline : constellation de sujets, perspectives et méthodes ; approche dominante, avec des perspectives critiques ; reconnaissance institutionnelle sous la forme de départements, journaux et conférences ; communauté autoproclamée d'universitaires ; et méthodes pour inculquer ou imposer l'adhésion à la culture de ladite discipline. Mais le concept de discipline académique est loin d'être simple. Le rapport de Heilbron [2004] montre que l'histoire des disciplines est différente selon les pays. Lenoir [1997] souligne que la majeure partie de la recherche littéraire ne se limite pas à une seule discipline, mais fait appel aux travaux de plusieurs disciplines. L'existence de départements universitaires pertinents ne définit qu'en partie une discipline, et il ne s'ensuit pas automatiquement que chaque département représente une discipline [Becher, 1989]. Dans l'étude de la visibilité sur Internet, l'analyse hypertextuelle a tendance à associer les départements universitaires aux domaines de recherche apparentés. Par exemple, une

étude de Heimeriks, Hörlesberger et Van den Besselaar [2003] portant sur la communication et la collaboration dans les relations entre université, industrie et gouvernement, a utilisé les pages d'accueil de départements universitaires comme représentations de champs de connaissances de mode 2 [Gibbons *et al.*, 1994] ; pourtant, un seul département de cette étude de cas représentait en réalité les activités de recherche apparentées. De telles associations peuvent mener à des résultats trompeurs et à des erreurs d'interprétation dans les interrelations en jeu dans l'étude des liens hypertextes. Afin de saisir au plus juste la représentation numérique des sociétés de la production du savoir, cet article se concentre sur un niveau d'agrégation moindre que celui des départements universitaires, optant plutôt pour les projets et les groupes de recherche en tant qu'études de cas imbriquées dans un champ intellectuel.

Bien que Whitley [1984] ait basé sa théorie sur la comparaison des disciplines, Fry [2006] a démontré qu'elle pouvait être utilisée pour étudier les pratiques de chercheurs dans un domaine spécialisé, dans la mesure où les concepts de *mutual dependence* et de *task uncertainty* sont valables aussi bien pour l'organisation interne (répartition du travail et compétition) d'une communauté de spécialistes, l'environnement scientifique plus large de la discipline dont ils dépendent, que pour le monde de la science. Par exemple, la position stratégique qu'occupe une spécialité au sein de sa discipline mère peut constituer un facteur majeur dans la détermination de son autonomie, de sa cohérence et de la direction de ses recherches. Suivant le degré de granularité – discipline ou spécialité –, les communautés scientifiques peuvent différer dans leur degré relatif de *mutual dependence* et de *task uncertainty*, tout comme des spécialités peuvent s'avérer plus ou moins coordonnées, intégrées et standardisées que leur discipline mère.

Les études sur la communication universitaire à l'intérieur des réseaux numériques tendent à privilégier l'analyse du social sur le cognitif ou l'inverse. Par exemple, Kling et McKim [2000] se sont concentrés sur l'influence qu'ont des structures sociales telles que la reconnaissance ou la récompense sur les publications universitaires sur le Web, sans prendre en compte les caractéristiques de la structure des connaissances en question. En raison de leur caractère interdépendant, les deux dimensions de Whitley permettent à ces deux éléments du travail universitaires d'être envisagés comme valides du point de vue de l'analyse. Ils se subdivisent en quatre éléments analytiques qui se recoupent parfois et sont liés soit au contrôle de la réputation et à la coordination des stratégies de

recherche et des priorités intellectuelles, à la « dépendance stratégique » et à l'« incertitude stratégique », soit à la coordination des standards de compétence, des techniques de recherche et des résultats, à la « dépendance fonctionnelle » et à l'« incertitude technique ».

Bien que cette étude ait porté sur un phénomène Internet, il était important de convoquer l'environnement plus général des universitaires de chaque étude de cas afin d'interpréter les traces révélées par la collecte des données en ligne et leur analyse. Comme le signalent Wouters et de Vries [2004], les études qui se focalisent sur les caractéristiques numériques de la communication universitaire ont tendance à négliger les aspects intellectuel, social et institutionnel. Les recherches ayant le Net pour objet ont montré que les frontières entre mondes réel et virtuel sont souvent indistinctes [Hine, 2000 ; Haythornthwaite et Wellman, 1998], ce qui implique qu'une distinction entre mondes « connecté » et « déconnecté » soit problématique. Le concept de monde « déconnecté » est utilisé ici avec prudence. Sa fonction principale est d'indiquer que les chercheurs ont passé du temps à examiner les espaces traditionnels alloués à la communication [Marcus, 1995], tels que littérature grise, conférences et articles de revue.

CRÉATION D'UNE SPHÈRE INTERNET

+++++
 En raison de l'amélioration qualitative du concept de sphère Internet, notre méthodologie n'a pas visé à saisir dans sa globalité une sphère Internet pour la linguistique pragmatique et la théorie de l'argumentation au niveau de la discipline. Nous avons privilégié une approche intégrée de l'étude de cas, en nous appuyant sur les répartitions du travail typiques de chaque domaine, ce qui a créé la base d'une sphère Internet égocentrée.

Le choix de ces deux études de cas est fonction de plusieurs critères, dont l'accessibilité et la renommée hors ligne de leurs membres. L'URL qui représentait collectivement chaque étude de cas au niveau institutionnel, comme le serveur de l'université, a été utilisée en tant qu'URL centrale de chaque sphère Internet⁶.

Le concept de sphère Internet se base sur la production du Web considérée du point de vue du producteur. Étant donné qu'il s'agissait d'une étude sur l'intégration et la coordination de la recherche et la reconnaissance de

6. Pour la première URL de chaque sphère Internet, voir les annexes : [en ligne] < http://cybermetrics.cindoc.csic.es/articles/v10i1p2_Appendices.html > (consulté le 10 novembre 2001).

ses pairs, il était nécessaire d'étendre la notion de production au contexte des sphères Internet universitaires afin de prendre en compte l'interconnectivité entre plusieurs sphères. Les liens dirigés vers les URL centrales ont donc également été inclus dans la collecte et l'analyse des données.

Étant donné la différence d'approche dans la production numérique de ces deux études de cas, il fallait aussi une approche différente dans la construction de leur sphère Internet respective. Par exemple, une différence majeure réside, dans le cas du CGN, dans la production « directe » d'une sphère Internet par le biais d'une URL centrale < <http://lands.let.ru.nl/cgn> > hébergée par l'université catholique de Nimègue, dont le contenu, produit par le chef de projet, est la représentation numérique « officielle » du projet. Le moteur de recherche Google.nl a ensuite été utilisé pour identifier les liens entrants accumulés au fil du temps. Ainsi, l'URL « officielle » du projet CGN était le centre de la sphère Internet.

Le contraste avec l'École d'Amsterdam, dont la production en matière de sphère Internet est principalement indirecte, est important. Ses membres ont d'ailleurs, dans un premier temps, décliné l'invitation à participer à cette étude au motif qu'ils ne se considéraient pas comme producteurs ni même utilisateurs de contenu numérique. Les pages d'accueil institutionnelles des membres du groupe hébergées par l'université d'Amsterdam < <http://cf.hum.uva.nl/data/afd/neerlandistiek/tar-english/> >⁷ ne contenaient que très peu d'informations en rapport avec les activités de recherche de l'École d'Amsterdam. Cette URL a initialement été choisie comme nœud central de la sphère Internet, mais elle ne contenait qu'un lien externe et ne possédait de lien entrant sur aucune autre page, et n'était donc pas reliée à un réseau d'URL externes. Il a donc fallu chercher un remplaçant à l'URL centrale de la sphère Internet de l'École d'Amsterdam plus à même de refléter la présence du groupe de chercheurs sur Internet. En 1986, suite à la première conférence internationale sur la théorie de l'argumentation qui s'était tenue à l'université d'Amsterdam, le directeur de l'École d'Amsterdam et d'autres membres ont fondé l'International Society for the Study of Argumentation (ISSA). C'est suite à l'entretien au cours duquel cette information a été révélée que l'URL du site Internet de l'ISSA < <http://cf.hum.uva.nl/issa/> > a été utilisée comme URL centrale de la sphère Internet. En sus de cette conférence

7. N.D.T. : cette URL qui a servi de base pour le corpus ne fonctionne plus. Celle du Department of Speech Communication, Argumentation Theory and Rhetoric de l'université d'Amsterdam est < <http://www.hum.uva.nl/argumentation> > (consulté le 10 novembre 2011).

annuelle, l'École d'Amsterdam a créé une revue internationale, *Argumentation: an International Journal on Reasoning* < <http://www.kluweronline.com/issn/0920-427X> >⁸, et cela joue également un rôle important dans la présence de l'École d'Amsterdam sur Internet. Les sites Web de l'ISSA et de la revue internationale dirigée par le directeur de l'École d'Amsterdam diffèrent de celui du projet « officiel » du CGN dans la mesure où ils n'ont pas été produits par des membres de l'École d'Amsterdam en tant que représentation directe des activités de recherche du groupe, mais plutôt comme représentation d'activités professionnelles et de leurs publications. En construisant une sphère Internet ethnologiquement renseignée, qui n'en demeure pas moins fonctionnelle (dépendant des liens sortants de l'URL centrale), il a été possible de distinguer une certaine représentativité de l'École d'Amsterdam, bien que ses membres aient perçu leur groupe comme non utilisateur et non producteur de contenu numérique⁹.

Pour les besoins de ce projet, j'ai défini une sphère Internet et sa présence sur le Web comme incluant les types d'inscriptions et d'actions numériques suivants :

URL centrale d'une page ou d'un site produit(e) par les membres d'un groupe / projet représentant la recherche collective ou les activités professionnelles du groupe / projet (telles qu'identifiées en entretien)	nœud central	sphère Internet produite directement
URL sortantes du nœud central (identifiées par Microsoft Site Analyst)		
URL entrantes vers le nœud central (identifiées via l'outil de recherche de liens entrants de Google.nl)		sphère Internet étendue produite indirectement
URL non reliées au nœud central, mais jugées, en entretiens et lors d'échanges d'e-mails avec des participants à l'étude, pertinentes pour la recherche et les activités professionnelles du groupe/projet		
URL au concept proche contenant des évocations du contenu de chaque étude de cas, mais non reliées par un lien hypertexte (identifiées en recherchant des extraits exacts via Google.nl)		présence sur le Web – sphère Internet potentielle

8. Le décompte des liens entrants figurant sur les pages en ligne de revues papier est compliqué par le fait que certains éditeurs, tels que Kluwer, fournissent de nombreuses options en matière de liens. N.D.T. : depuis 2004, la revue est publiée chez Springer : < <http://www.springerlink.com/openurl.asp?genre=journal&issn=0920-427X> > (consulté le 10 novembre 2011).

9. Ma requête initiale pour leur participation ayant été déclinée au motif de cette perception.

Ainsi définie, chaque sphère Internet a été enrichie de données ethnologiques, construite longitudinalement et validée en envoyant la liste des URL constituant chaque sphère Internet initiale aux membres sélectionnés pour l'étude de cas. Les URL centrales initiales ont été confirmées lors d'entretiens avec les membres clés de chaque étude de cas. Puis l'on a eu recours à Microsoft Site Analyst pour identifier les pages vers lesquelles renvoyait chaque URL centrale. Les liens entrants¹⁰ ont, quant à eux, été identifiés par des recherches régulières *via* l'outil de Google.nl prévu à cet effet, environ six fois en douze mois, et les résultats ont été révisés afin de supprimer les liens autoréférentiels et les URL lourdes¹¹. Ces opérations ont mis au jour de multiples sites pour le projet du CGN hébergé par diverses universités partenaires aux Pays-Bas et en Flandre, mais tous étaient reliés au site Internet « officiel » hébergé aux Pays-Bas, ainsi qu'un intranet protégé par mot de passe pour le projet hébergé par l'université d'Utrecht. Elles ont également identifié deux liens réciproques entre le projet du CGN et la Fondation néerlandaise pour la science (NWO), cofondatrice du projet, et l'Union de la langue néerlandaise (Nederlandse Taalunie).

La représentativité de chaque cas sur Google.nl a également fait partie de l'analyse. Pour ce faire, des expressions précises ont été recherchées afin de trouver des pages en rapport avec le sujet mais non reliées *via* un lien hypertexte. Dans un contexte de recherches liées à la discipline, cette étude de la présence sur le Web a abouti à la représentation d'une sphère Internet potentielle pour chaque étude de cas, qui était bien plus importante dans les deux cas que la sphère Internet existante. Par exemple, la recherche des termes *Corpus Gesproken Nederlands* (CGN), corpus de néerlandais parlé, aboutit à environ 900 URL, alors que le nombre d'URL dans la sphère Internet du CGN est de 63. De même, la recherche du mot « pragma-dialectique », nom familier et champ de recherche de l'École d'Amsterdam, sur Google.nl a abouti à 860 URL¹² (pour une liste des 30 premiers résultats de chaque recherche), comparé à 28 URL sur la

10. Voir Björneborn et Ingwersen [2004] pour une typologie des liens.

11. Il est possible d'exclure les liens autoréférentiels sur Alta Vista en ayant recours aux opérateurs booléens (voir Rousseau [1997] pour une utilisation de l'outil de recherche avancée d'Alta Vista concernant les liens entrants) et, au cours de cette étude, il semblerait que Google ait reconfiguré son algorithme et exclut dorénavant automatiquement les liens autoréférentiels des résultats de recherche.

12. Voir les annexes V et VI : [en ligne] : < http://cybermetrics.cindoc.csic.es/articles/v10i1p2_Appendices.html >.

sphère Internet. Ces recherches ont également été menées *via* le moteur de recherche Alta Vista en cas de couverture limitée.

Pour comparer : le cas de la linguistique pragmatique concerne une sphère Internet de projet, qui reflète la structure des sites Web construits autour de projets dans le domaine de la linguistique comparée, tandis que le cas de la théorie de l'argumentation s'appuie sur une sphère Internet d'école de pensée, qui reflète aussi sa présence sur le Net au sens large.

MESURE DE LA *MUTUAL DEPENDENCE* ET DU *TASK UNCERTAINTY*

+++++

Les degrés relatifs de *mutual dependence* et de *task uncertainty*, caractéristiques de l'identité culturelle de chaque discipline, ont été estimés grâce à des entretiens approfondis avec deux participants du projet CGN et un participant de l'École d'Amsterdam¹³ ainsi qu'à une analyse de littérature spécialisée par des participants à l'étude de cas. Ces dimensions de *mutual dependence* et *task uncertainty* au moyen desquelles notre étude a jugé les disciplines concernées incluent la taille de la communauté nationale et internationale, le degré du contrôle sur les divers canaux de diffusion, les genres prédominants de communication, la position stratégique au sein de la discipline mère, la présence d'un auditoire profane, des questions d'ordre pédagogique (telles que la discipline fait-elle partie d'un programme d'études ?), et le rôle des gardiens de l'information tels que les éditeurs.

Le tableau 1, inspiré de Whitley [1984], illustre les degrés relatifs de *mutual dependence* de la linguistique pragmatique et de la théorie de l'argumentation en s'appuyant sur ce qui a été découvert à propos de l'organisation intellectuelle et sociale de chaque discipline.

13. Voir les annexes I et II sur le détail des questions : [en ligne] : < http://cybermetrics.cindoc.csic.es/articles/v10i1p2_Appendices.html >.

Tableau 1 : différences de degré de dépendance stratégique et fonctionnelle entre la linguistique pragmatique et la théorie de l'argumentation

		Dépendance fonctionnelle	
		Basse	Haute
Dépendance stratégique	Basse		<i>Linguistique pragmatique</i> : « Groupes de chercheurs spécialisés poursuivant des objectifs distincts au moyen de procédures standardisées. Importante coordination des résultats et haut degré de spécialisation des sujets, mais peu de considération pour la hiérarchie des objectifs. » [Whitley, 1984]
	Haute	<i>Théorie de l'argumentation</i> : « Écoles de recherche entretenant des liens étroits et poursuivant des objectifs distincts au moyen de procédures différentes. Haut degré de coordination au sein des écoles, mais faible entre elles. Importante compétition pour dominer la discipline. » [Whitley, 1984]	

Selon Whitley, un degré de dépendance fonctionnelle croissant débouche sur un haut degré de spécialisation et de différenciation des problèmes et des objectifs, et donc à la création de procédures spécifiques pour les gérer. Un haut degré d'interdépendance au sein d'une discipline spécialisée mène, quant à lui, à une importante coordination des résultats et des domaines spécialisés. À mesure que le degré de dépendance stratégique augmente, la coordination des problèmes, objectifs et procédures s'intensifie elle aussi. La compétition augmente également lorsque des scientifiques issus de différentes sous-disciplines tentent de convaincre leurs collègues de l'aspect crucial de leurs problèmes et de leurs objectifs pour leur domaine.

L'environnement hors ligne de la linguistique comparée indique un degré de dépendance fonctionnelle plus important que dans la théorie de l'argumentation. Ceci est sûrement dû en grande partie au fait que ses objectifs s'orientent vers la production, ce qui requiert l'intégration des techniques. Et cela va à l'encontre du pluralisme intellectuel des disciplines axées sur la théorie telle que la théorie de l'argumentation.

Généralement parlant, le chevauchement disciplinaire entre différents domaines d'effort intellectuel convergeant au sein du champ, tels que la linguistique computationnelle, la sociolinguistique, la phonologie et la lexicographie, est désormais considéré comme admis, non sans controverse préalable [Sampson, 2003 ; Lawler et Dry, 1998 ; Borsley et Ingham, 2002]. En conséquence, tandis qu'en théorie de l'argumentation, la rhétorique tend à avoir recours à des expressions conciliantes comme « se conformer » ou « envers », les textes de linguistique pragmatique tendent à employer des métaphores guerrières – « gagné », « bataille » – lorsqu'il s'agit de refléter la position du domaine au sein de la discipline mère [Sampson, 2003]. Cette donnée suggère que si le degré de dépendance fonctionnelle est plus important en linguistique pragmatique, celui de dépendance stratégique est plus bas en théorie de l'argumentation.

Certaines frontières intellectuelles et sociales de la linguistique pragmatique sont plus contestées que d'autres. L'interface avec la linguistique théorique est la frontière la plus controversée et la plus discutée dans la littérature de linguistique générale¹⁴. Dans le contexte plus général de la discipline mère, le degré de dépendance stratégique n'a pas encore été déterminé avec certitude, mais peut être considéré comme beaucoup plus faible que dans un champ bien établi tel que la physique des particules.

Depuis la fin des années 1990, la littérature produite par les membres de l'École d'Amsterdam contient la rhétorique d'intégration des deux principaux paradigmes contradictoires [Van Eemeren et Houtlosser, 2001]. Afin de déterminer le degré de *mutual dependence* dans ce domaine, la volonté de consensus décelée dans la littérature a été interprétée comme une indication que le degré de dépendance stratégique est relativement bas, mais que l'École d'Amsterdam et ses alliés travaillent à l'augmentation du degré de dépendance stratégique, en particulier entre les groupes, organisations et institutions qui représentent les diverses écoles de pensée au sein de la théorie de l'argumentation. Cela est nécessaire afin d'exploiter les ressources telles que la réputation, le public et le personnel, et de faire progresser le programme de recherche encore naissant vers l'institutionnalisation et la survie en tant que discipline.

Les remarques de Whitley sur les conséquences culturelles d'un degré de dépendance stratégique croissant trouvent un écho prolongé dans l'étude de cas de la théorie de l'argumentation. Au début des années 1990,

14. Voir Borsley et Ingham [2002] pour une critique typique dont la linguistique pragmatique fait l'objet.

les membres de l'École d'Amsterdam ont produit un certain nombre de monographies [Van Eemeren et Grootendorst, 1994] proposant un programme de recherche interdisciplinaire au sein de la théorie de l'argumentation, en se basant sur l'approche pragma-dialectique. La création de l'International Society for the Study of Argumentation (ISSA) en 1986 peut également être considérée comme un moyen de placer l'École d'Amsterdam, et donc l'étude de la théorie de l'argumentation telle qu'elle se pratique aux Pays-Bas, dans une position essentielle au sein de la communauté internationale de la théorie de l'argumentation. L'ISSA a également mis sur pied un programme de bourses, qui renforce le positionnement stratégique de l'École d'Amsterdam au sein de la discipline mère en termes de contrôle de réputation. Le financement par le Netherlands Institute for Advanced Study in the Humanities and Social Sciences d'un petit groupe de recherche axé sur « Des sophismes comme violations des règles du discours argumentatif » confirme d'autant plus l'impression d'une École d'Amsterdam entourée d'un réseau international aux mailles serrées. Parmi les experts de l'argumentation qui faisaient partie de ce programme figurent le président du département de la communication orale, de la théorie de l'argumentation et de la rhétorique ainsi que d'autres partisans importants de l'approche pragma-dialectique qui, eux-mêmes, étaient, pour beaucoup, membres de ce département à l'époque.

Tableau 2 : différences de degré d'« incertitude technique » et d'« incertitude stratégique » entre les deux études de cas

<i>Technical task uncertainty</i>			
		Bas	Haut
<i>Strategic task uncertainty</i>	Bas		
	Haut	<i>Linguistique pragmatique :</i> « Prévisibilité, stabilité et visibilité des résultats importantes. Consensus sur la façon d'interpréter et de coordonner ces résultats. Problèmes et objectifs assez divers, instables et non clairement organisés. »	<i>Théorie de l'argumentation :</i> « Contrôle technique des phénomènes limité, résultats instables et difficiles à interpréter. Opinions divergentes sur les implications des résultats et peu de coordination et de comparaison entre eux. Problèmes et objectifs divers, instables et incompatibles. »

Selon Whitley, un degré croissant d'« incertitude technique » débouche sur une organisation de travail des communautés intellectuelles caractérisée par une forte dépendance au contrôle direct et personnel du travail

effectué, ce qui mène à de trop grandes fluctuations dans les méthodes de travail et à un manque de fédération des objectifs de recherche au niveau international. L'« incertitude stratégique » inclut l'incertitude quant aux priorités intellectuelles, l'importance des sujets de recherche et les méthodes privilégiées de traitement de ces sujets, les récompenses potentielles de différentes stratégies de recherche en matière de renommée, et la pertinence des résultats pour les objectifs intellectuels collectifs [Whitley, 1984]. Un degré d'« incertitude stratégique » croissant débouche donc sur une plus grande dépendance à un petit groupe de collègues pour l'accès au matériel et à la renommée, d'où la nécessité d'aboutir à un consensus sur les priorités intellectuelles et une hiérarchie des sujets de recherche.

Le degré de « dépendance stratégique » de la linguistique pragmatique est mitigé par un degré d'« incertitude technique » relativement élevé. Les tâches qu'implique typiquement la construction d'un corpus incluent le plan, l'enregistrement, la transcription, l'indexation, l'annotation et la mise au point d'un logiciel d'exploitation. Il y a beaucoup de décisions à prendre sur la façon d'aborder ces tâches, et la littérature foisonne de débats sur la meilleure approche à adopter. Il existe, par exemple, plusieurs niveaux d'annotation de données empiriques : syntaxique, prosodique, ou encore phonétique, et chacun d'eux offre différentes possibilités analytiques aux futurs utilisateurs inconnus. Beaucoup de tâches nécessitent la création de critères techniques. En 2000, Atwell et ses collaborateurs ont mené une étude sur l'usage d'outils normés d'analyse grammaticale en linguistique et ont découvert une grande fluctuance dans l'emploi des schémas d'analyse.

La tendance vers un degré plus élevé de « dépendance stratégique » au sein de la communauté de la théorie de l'argumentation signifie que dans un mouvement simultané le degré d'« incertitude stratégique » diminuera.

PRINCIPAUX CONTRASTES DANS LES CARACTÉRISTIQUES DU CONTENU DES DEUX SPHÈRES INTERNET

+++++

Une différence capitale entre les sphères Internet du projet CGN et de l'École d'Amsterdam réside dans la taille en termes de : nombre d'URL, identification du contenu des URL, destination des liens sortants de

chaque URL centrale en termes de caractéristiques nationale (identifiés par le code pays et le nom du domaine de premier niveau – *country code Top Level Domain* ou ccTLD¹⁵), organisationnelle (reconnue, entre autres, grâce au nom du domaine de premier niveau – TLD¹⁶) et de contenu ; et l'origine des URL entrantes vers l'URL centrale en termes de caractéristiques nationale, organisationnelle et de contenu (la liste des URL qui constituent chacune des deux sphères Internet¹⁷).

Caractéristiques de la sphère Internet du CGN produite directement et indirectement

+++++
 Au sein du site Web « officiel » du CGN, représenté par l'URL < <http://lands.let.ru.nl/cgn> >¹⁸, figuraient 55 pages contenant une description détaillée du projet, la justification de ses objectifs, des échantillons de données, des références bibliographiques, des rapports techniques et d'autres sur l'état d'avancement des recherches.

Des documents étaient également disponibles au format PDF, PPT et postscript. Le tableau 3 résume les principales dimensions de la sphère Internet du CGN.

Tableau 3 : résumé de la sphère Internet du CGN

Nombre de pages de l'URL centrale	55	
Nombre de liens sortant de l'URL centrale	8	
URL sortantes identifiées par les noms ccTLD et TLD	.nl	5
	.fr	1
	.de	1
	.it	1
	.at	1
Nombre de liens entrant vers l'URL centrale	39	

15. La liste des noms de domaine de premier niveau code pays est disponible sur < <http://www.iso.org/iso/en/prods-services/iso3166ma/02iso-3166-code-lists/index.html> >.

16. De plus amples informations sur les noms de domaine de premier niveau sont disponibles sur < <http://www.icann.org/en/tlds/> > (consulté le 10 novembre 2011).

17. Voir les annexes III et IV pour la liste des URL : [en ligne] < http://cybermetrics.cindoc.csic.es/articles/v10ilp2_Appendices.html >.

18. Le 3 mai 2004.

URL entrantes identifiées par les noms ccTLD et TLD	.nl	20
	.be	9
	.org	5
	.de	2
	.dk	1
	.ie	1
	.at	1
Nombre d'URL ajoutées par les participants		14
Nombre d'URL de la plateforme		62

Il y avait 8 liens sortants du site et 39 liens entrants (à l'exclusion des liens internes). Sur les 8 liens sortants, 4 appartenaient au ccTLD **.nl** :

- la Fondation néerlandaise pour la science (NWO), qui a financé le projet conjointement avec le gouvernement flamand ;
- l'Union de la langue néerlandaise (Nederlandse Taalunie), à qui revient la propriété du corpus ;
- l'intranet du CGN hébergé par l'université d'Utrecht ;
- PRAAT, outil phonétique utilisé pour le projet et mis au point par le département de phonétique de l'université d'Amsterdam.

Les 4 liens sortants restants étaient des noms ccTLD européens :

- ELDA (Evaluations and Language Resources Distribution Agency), basée en France et en charge de la diffusion du corpus du CGN ;
- EAGLES, norme internationale du génie de la langue basée en Italie ;
- Negr@corpus, corpus allemand de textes de presse ;
- le dernier, enregistré **.at (Autriche)**, n'est plus en service et redirige l'internaute vers Nederlandse Taalunie ; le lexique de l'URL suggère également que la destination du lien était à l'origine néerlandaise.

Sur une période d'un an, d'août 2003 à juillet 2004, 39 URL entrantes vers l'URL du projet CGN ont été identifiées grâce à l'outil de recherche

de Google.nl. Vingt de ces URL appartenait au ccTLD.**nl**¹⁹. Ces URL incluent des sous-pages de la page « officielle » du CGN produites par des institutions partenaires du projet, des pages créées par des utilisateurs du corpus, des assistants-chercheurs d'anciens projets, le département du langage de l'université catholique de Nimègue où travaille le chef de projet de la section néerlandaise du corpus, des listes de liens qui dirigent les gens vers des ressources sur la technologie du langage, des ateliers de recherche, l'Union de la langue néerlandaise (Nederlandse Taalunie) et la Fondation néerlandaise pour la science (NWO).

Sur les 19 URL entrantes en dehors du nom ccTLD.**nl**, 9 appartenait au ccTLD.**be**. Les URL du domaine **.be** représentent en partie la participation des partenaires du projet basés en Belgique en charge de la section flamande du corpus. Toutefois, les URL des principaux acteurs belges impliqués dans la mise au point et l'utilisation du corpus ne sont pas représentées dans la sphère Internet produite de façon directe. Ces URL ont été ajoutées à la sphère Internet étendue produite de façon indirecte par le chef de la section flamande du projet au cours de l'étape de validation de la construction de la sphère Internet.

Cinq des URL entrantes appartenait au nom de domaine **.org** et étaient hébergées par deux organisations néerlandaises : DBNL (Digitale Bibliotheek voor de Nederlandse) (1 lien) et l'Union de la langue néerlandaise (Nederlandse Taalunie) (3 liens). Le dernier lien provenait de l'HLT (Human League Technologies), basé en Europe.

Les 5 URL entrantes restantes étaient aussi hébergées en Europe. Deux sous le nom de domaine **.de** : une page sur la philologie aux Pays-Bas hébergée par l'université libre de Berlin et comportant du contenu coécrit par le chef de projet en charge de la section néerlandaise ; une autre hébergée par un centre d'étude de la langue néerlandaise basé à Leipzig. Le lien entrant appartenant au domaine **.dk** représentait le site Web Danish Dependency Treebank, le *treebank* étant un arbre syntaxique, une technique d'analyse utilisée par le CGN.

19. Voir l'annexe III pour une liste complète des URL entrantes et des intitulés de pages : [en ligne] < http://cybermetrics.cindoc.csic.es/articles/v10i1p2_Appendices.html >.

Caractéristiques de la sphère Internet de l'École d'Amsterdam produite directement et indirectement

+++++

- Au cours de la période d'étude, l'URL centrale de l'École d'Amsterdam ne contenait que 3 pages (une page d'accueil décrivant les objectifs de l'ISSA, une page contenant des détails sur la conférence annuelle, et une page recensant les récipiendaires de la bourse de l'ISSA) et 2 liens sortants (un pointait vers un associé de l'École d'Amsterdam basé dans le département de communication de l'université de Louisville [.edu], et l'autre vers la revue *Argumentation: an International Journal on Reasoning*, dirigée par le président de l'École d'Amsterdam et publiée par Kluwer, maison basée aux Pays-Bas [.nl]²⁰). Malgré le manque de contenu en rapport avec la recherche, d'informations concernant les cours, d'articles et de dossiers téléchargeables, l'URL de l'ISSA profitait de 17 liens entrants. En général, les pages Web universitaires bénéficient de peu de liens entrants [Thelwall et Harries, 2004] ; c'est pourquoi il est surprenant que l'URL de l'ISSA compte plus de liens entrants que de pages de contenu. Le tableau 4 résume les principales dimensions de la sphère Internet de l'École d'Amsterdam.

Tableau 4 : résumé de la sphère Internet de l'École d'Amsterdam

Nombre de pages de l'URL centrale	3	
Nombre de liens sortant de l'URL centrale	2 (en rapport avec la recherche) 36 (informations touristiques)	
URL sortantes identifiées par les noms ccTLD et TLD	.edu	1
	.nl	1
Nombre de liens entrant vers l'URL centrale	17	

20. Trente-six liens sortants figurant sur les pages de l'ISSA étaient sans rapport avec la recherche, il s'agissait de renseignements touristiques destinés aux participants à la conférence annuelle de l'ISSA.

URL entrantes identifiées par les noms ccTLD et TLD	.edu	5
	.com	4
	.ca	2
	.ru	1
	.org	1
	.nl	1
	.uk	1
	.ch	1
	.de	1
Nombre d'URL ajoutées par les participants	6	
Nombre d'URL de la plateforme	28	

Par opposition à l'orientation nationale de la sphère Internet du CGN, 16 des liens entrants de la sphère Internet de l'École d'Amsterdam émanaient de l'extérieur du domaine **.nl** et étaient de nature internationale. L'unique lien entrant émanant du domaine **.nl** appartenait au site Web d'un membre actif de la liste de discussion ARGTHRY (principale liste de discussion de la communauté de la théorie de l'argumentation) ayant créé un site Internet personnel conséquent consacré à l'étude de la théorie de l'argumentation. Le contenu de la page d'accueil du lien entrant a trait au 9^e Atelier international sur le raisonnement non monotone et le lien référant la conférence de l'ISSA de 2002 est intégré dans une liste de liens en rapport avec le sujet.

D'autres pages extérieures au domaine **.nl** qui pointaient vers le site de l'ISSA incluent : "Philosophy resources on the Internet: EpistemeLinks" (**.com**), site commercial de vente de produits en rapport avec la philosophie ; "Rhetoric resources on the Web" (**.edu**), qui fait partie des pages d'accueil d'un membre du département d'anglais de l'université de Wisconsin-Madison ; "Tim Gelder's critical thinking on the Web: Institutes, Centers and Societies" (**.org**) ; "The Reasoning Page" (**.edu**), liste annotée de ressources en ligne dans l'étude de l'argumentation.

Le site Web de Vale Press (**.com**) pointait 3 liens sortants vers l'URL de l'ISSA. Vale Press est un éditeur spécialisé dans le champ de la pensée critique. Ils publient notamment la collection « SIC SAT » en théorie de l'argumentation, support de publication des monographies écrites par les membres de l'École d'Amsterdam. Les liens sortants qui pointaient vers le site de l'ISSA apparaissaient à deux reprises sur la page d'accueil de Vale

Press : dans le calendrier, avec la référence à la conférence annuelle de l'ISSA, et dans la liste des liens les plus populaires. Ces liens étaient tous les deux intégrés à proximité de 2 liens dirigeant vers l'Association pour la logique informelle et la pensée critique (AILACT), qui possède sur son site un lien pointant vers la revue de l'ISSA. Le troisième lien sortant figurant sur le site Web de Vale Press pointant vers l'URL de l'ISSA était intégré à une liste de titres de monographies.

Il y avait 5 liens entrants émanant du domaine **.edu** (dont deux ont déjà été décrits plus haut). L'un d'eux émanait du site Web de la Stanford Encyclopedia of Philosophy sous l'entrée *Informal Logic* (logique informelle). Les 2 URL entrantes restantes du domaine **.edu** sont des pages personnelles d'universitaires des départements de rhétorique ou de communication et apparaissent en tant que ressources utiles, d'une part, et associations et conférences en rapport avec la communication, d'autre part.

Il y avait 2 URL entrantes appartenant au domaine **.ca** (Canada). L'une dirigeait vers le site Web du *Journal of Informal Logic: Reasoning and Argumentation in Theory and Practice*, dans une liste de liens pointant vers d'autres revues et ressources, qui contient également un lien sortant vers l'URL de la revue de l'ISSA. Le second lien entrant du domaine **.ca** pointait vers la Wetaskiwin Telephone Company, qui contient un lien vers un article publié dans les actes de la cinquième conférence de l'ISSA, à Amsterdam, du 25 au 28 juin 2002.

Un membre de l'École d'Amsterdam publie en ligne une revue intitulée *Argumentation, Interpretation and Rhetoric*, en russe et en anglais. L'Institut néerlandais de Saint-Petersbourg est le lien qui relie l'École d'Amsterdam à la Russie et l'URL entrante du domaine **.ru** (Russie) appartenait à une page d'archives référençant la revue *Argumentation, Interpretation and Rhetoric*.

Un ancien étudiant de troisième cycle du département de la communication orale, de la théorie de l'argumentation et de la rhétorique avait créé un lien entrant dans ses pages personnelles, appartenant au domaine **.de** (Allemagne). Le domaine **.uk** (Royaume-Uni) comptait un lien entrant, figurant dans les pages Web d'un atelier de pré-conférence de l'ECIA 2002 (Conférence européenne sur l'intelligence artificielle) sur les modèles computationnels de l'argumentation naturelle. Le dernier lien entrant, du domaine **.ch** (non listé dans la liste des codes pays ISO 3166), n'a en réalité qu'un lien trivial avec les activités de l'École d'Amsterdam, étant donné qu'il figure sur les pages d'un centre international de la santé et de la coopération, listé sous diverses associations industrielles et universitaires.

Les URL ajoutées par le participant à l'étude de cas sont toutes en rapport avec les ateliers et conférences majeurs sur la théorie de l'argumentation auxquels les membres de l'École d'Amsterdam ont participé. Étant donné que les pages Web de Kluwer, ancien éditeur de la revue de l'ISSA, étaient intimement liées à celles du site de l'ISSA dans la mesure où il existait un lien réciproque entre elles et où les éditeurs jouent un rôle primordial dans la sphère Internet de l'École d'Amsterdam, l'on a également procédé à une recherche de liens entrants à partir de l'URL de la revue < <http://www.kluweronline.com/issn/0920-427X> >²¹ en utilisant Google.nl²². Les 23 liens entrants pointant vers la revue de l'ISSA, tous extérieurs au domaine .nl, ne seront pas évoqués en détail mais fournissent un contexte plus large et illustrent les différences qui existent lorsqu'une sphère Internet est construite avec pour unité d'analyse différentes représentations d'une communauté universitaire. La figure 1 résume les principaux aspects des sphères Internet de chaque étude de cas :

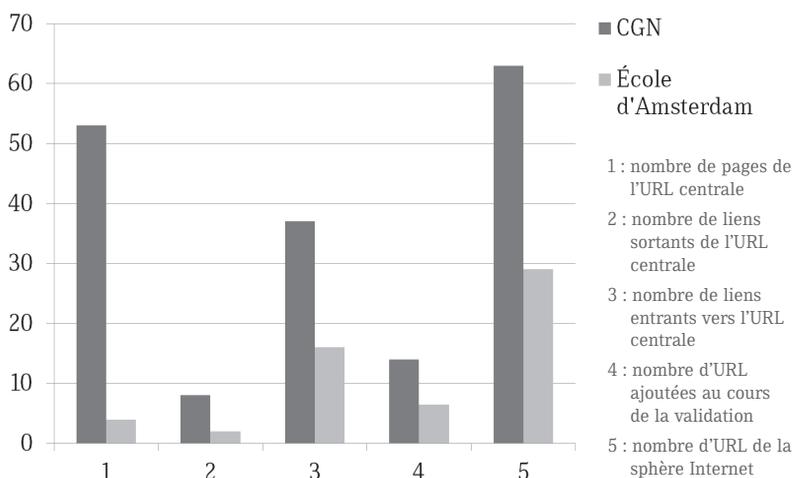


Figure 1 : comparaison des principaux aspects des sphères Internet du CGN et de l'École d'Amsterdam

21. N.D.T. : voir note 8.

22. Voir l'annexe IV pour une liste des URL entrantes : [en ligne] < http://cybermetrics.cindoc.csic.es/articles/v10i1p2_Appendices.html >.

Présence sur le Web : une sphère Internet potentielle

+++++

La présence sur le Web de chaque étude de cas telle que représentée sur Google.nl a fourni un contexte lié à la discipline. En tant que telle, elle donne une image de ce que pourrait être la sphère Internet de chaque étude de cas, en termes de nombre d'URL et de genres prédominants, bien plus importante dans les deux cas que la sphère Internet existante. Par exemple, la recherche de l'expression exacte « corpus gesproken nederlands » aboutit à environ 900 URL, alors que le nombre d'URL dans la sphère Internet du CGN n'atteint que 63. De même, la recherche du mot « pragma-dialectique », nom familier et champ de recherche de l'École d'Amsterdam, sur Google.nl a abouti à 860 occurrences, comparé à 28 URL sur la sphère Internet. Cet important contraste soulève la question de la pertinence du concept de sphère Internet comme outil d'étude de la production du Web universitaire. Les recherches les plus récentes ont été menées le 1^{er} mai 2005 et les 30 premières occurrences sont recensées dans les annexes V et VI²³.

Il existe des problèmes de comparabilité entre les représentations numériques des deux disciplines obtenues par le biais de moteurs de recherche, pour des raisons linguistiques [Vaughan et Thelwall, 2004 ; Bar-Ilan, 2005], et les problèmes associés aux requêtes multilingues [Moukdad, 2002]. Pour le projet du CGN, les résultats sont différents lorsque l'on entre la traduction anglaise de « corpus de néerlandais parlé » dans la barre de recherche, mais étant donné la faible représentation de l'anglais dans la sphère Internet du CGN, il a été estimé plus pertinent d'avoir recours au néerlandais. Cependant, malgré l'aspect rudimentaire de la méthode, nous avons pu observer que le nombre d'URL faisant référence dans le contenu de leurs pages au projet du CGN ou aux membres de l'École d'Amsterdam était plus important que celles qui contenaient véritablement un lien sortant pointant vers l'URL centrale de chaque étude de cas. Cette donnée souligne les limites de la prise en compte exclusive des liens hypertextes comme critère de l'interconnectivité entre les acteurs en jeu.

Un échantillonnage manuel des 30 premiers et 10 derniers résultats de chaque série de recherches a été réalisé afin de vérifier la pertinence des occurrences. On a jugé de l'identification de chaque URL en s'appuyant sur le résumé informationnel contenu dans la structure de l'URL, comme

23. Voir les annexes V et VI : [en ligne] < http://cybermetrics.cindoc.csic.es/articles/v10i1p2_Appendices.html >.

les caractéristiques de l'institution, le titre de la page, et le contenu des pages lui-même lorsque les métadonnées ne fournissaient pas suffisamment de renseignements pour pouvoir estimer l'identification.

Dans le cas du *Corpus Gesproken Nederlands* (CGN), ce sont des liens pointant vers des ressources telles que des corpus, des outils, des chercheurs, des organisations, des mémoires, des rapports et du matériel pédagogique qui arrivaient en tête des résultats de recherche. Les librairies et bibliothèques en ligne représentant les publications de la linguistique pragmatique n'apparaissaient qu'après les 60 premiers résultats. Ce qui contraste avec les recherches de la « pragma-dialectique », où les équivalents numériques des monographies papier arrivaient en tête de liste.

L'identification des URL obtenues pour les requêtes de la théorie de l'argumentation s'orientait vers les versions numériques de monographies trouvées sur les pages Web d'éditeurs, de librairies électroniques et de bibliothèques en ligne. Elles comprenaient celles écrites ou dirigées par des membres de l'École d'Amsterdam.

TRACES DE *MUTUAL DEPENDENCE* ET DE *TASK UNCERTAINTY* DANS LA REPRÉSENTATION NUMÉRIQUE DES DISCIPLINES INTELLECTUELLES

+++++

Il est donc clair que la production numérique au sein de la linguistique pragmatique et de la théorie de l'argumentation possède des caractéristiques bien distinctes. Dans l'étude de cas du CGN, nous observons une sphère Internet locale et décentralisée (locale dans le sens où pages et sites sont organisés au niveau du projet et non de la discipline ; décentralisée dans la mesure où l'impératif organisationnel de la sphère Internet obéit davantage à la notion de dépendance fonctionnelle qu'à celle de dépendance stratégique), qui s'adresse en conséquence à un public national de chercheurs, du fait que la discipline ne soit pas dominée par une approche unique et ne soit pas régie par une hiérarchie de problèmes. Cela signifie qu'il est moins nécessaire de démontrer au domaine scientifique l'importance des problèmes, techniques et aboutissements en matière de recherche. Dans le cas de la théorie de l'argumentation, en revanche, la production numérique est centralisée au niveau de la discipline, avec des pages et des sites organisés autour de l'activité professionnelle des écoles de recherche et des communautés de chercheurs représentant un paradigme dominant. L'impératif organisationnel est centralisé et s'appuie ici sur l'idée de « dépendance stratégique », où ce sont la hiérarchisation des

questions à traiter et l'édification d'une réputation qui priment. La sphère Internet du CGN a, par exemple, une orientation nationale très marquée, avec un nombre négligeable d'URL extérieures au domaine **.nl**. La sphère Internet de l'École d'Amsterdam a, au contraire, une orientation internationale, avec une majorité d'URL extérieures au domaine **.nl**.

En termes de public scientifique hors ligne pour le CGN, les résultats de recherche ont été disséminés au fil de conférences internationales et de revues spécialisées, et des articles publiés par des membres du projet ont été cités dans des revues répertoriées par l'ISI Web of Knowledge. Toutefois, la linguistique pragmatique est dominée par la recherche en langue anglaise, et l'analyse hypertextuelle de la sphère Internet du CGN semble refléter cette donnée. Des corpus de langue très influents, tels que le corpus national britannique ou le corpus international d'anglais, n'ont pas sur leur site Internet de lien pointant vers l'URL du projet CGN. Et bien que le contenu du projet CGN évoque le corpus national britannique sur la question des critères internationaux à atteindre dans la création des corpus, aucun lien ne pointe vers ce dernier, ni vers aucun autre corpus national.

Cela peut constituer un autre signe de haut degré de *strategic uncertainty* pour le projet CGN, dans le sens où il n'est pas nécessaire que les autres acteurs de la linguistique pragmatique s'alignent sur ce projet pour gagner en renommée ou en légitimité dans la discipline. L'inverse se vérifie dans le cas de la sphère Internet de l'École d'Amsterdam, où la grande proportion de liens entrants comparée à la quantité relativement limitée de pages et de contenu (en termes de diversité des types de dossiers, outil utilisé par la mesure de la représentativité sur Internet) indique la position politique dominante du groupe : les autres chercheurs doivent reconnaître le groupe afin d'être crédibles. Thelwall [2006] a découvert que plus une page Internet est la cible de liens, plus il sera probable qu'elle soit ciblée par des liens nouvellement créés. Donnée qui fait écho au schéma de la naissance de l'estime que Merton [1968] a décrit sous la dénomination « effet saint Matthieu ».

Bénéficier d'une forte représentativité numérique peut s'avérer important pour les groupes appartenant à une discipline émergente, où ils ont besoin d'accroître leur degré de *strategic dependence*, comme c'est le cas pour l'École d'Amsterdam. Si, en revanche, leurs objectifs intellectuels sont mieux servis par une communication classique, comme c'est souvent le cas dans les domaines théoriques, qui nécessitent de longues explications de concepts et d'arguments, alors la communication de style

monographique [Swales, 1998] tend à être le mode de communication dominant. Cela signifie que la représentation numérique d'un courant ou d'une discipline peut être dominée par les gardiens de la communication publiée de façon classique, tels que les bibliothèques et les éditeurs, plutôt que par la communication informelle orientée vers la recherche – comme on le voit dans la prédominance des éditeurs et bibliothèques en ligne dans la sphère Internet de l'École d'Amsterdam. Il est intéressant de constater que bien que l'École d'Amsterdam n'ait pas produit directement une sphère Internet hypertextuelle, sa position stratégique dans le monde hors ligne est reflétée par une sphère Internet étendue produite indirectement comme l'indique le nombre important de liens entrants vers les pages Web de l'ISSA. Le fait que ce site ne contienne que trois pages de contenu en rapport avec la recherche (le reste du site étant consacré au tourisme local et aux informations pratiques concernant le logement en vue de la conférence de l'ISSA), sans lien sortant vers d'autres groupes de recherche ou organisations, et dont les liens entrants dépassent le nombre total de pages et de liens sortants, démontre la position politique dominante du groupe : les autres chercheurs doivent s'aligner sur le groupe afin d'être crédibles, et non créer seulement un lien informationnel [Rousseau, 1999].

Tout comme la sphère Internet de l'École d'Amsterdam, le projet CGN est présent dans des sources secondaires d'information telles que portails thématiques et bulletins d'information. Les deux disciplines ont un public profane, mais en raison de la spécificité de la langue dans le projet CGN et le jargon technique abscons inhérent au domaine lui-même, il est probable que le public profane soit restreint. Lorsqu'on lui a demandé à qui se destinaient les résultats des recherches effectuées par son groupe, un membre de l'École d'Amsterdam a répondu qu'en plus des universitaires des domaines de la logique, de la rationalité, de la rhétorique, de la persuasion et de la dialectique, les étudiants de deuxième et troisième cycles étaient également importants, tout comme les non-universitaires intéressés par les rapports de science vulgarisés.

La communication informelle est un mode de communication dominant dans le milieu de la linguistique pragmatique, où, dans les domaines techniques de la discipline tels que la technologie du discours, le modèle de communication s'apparente davantage à la science computationnelle. Beaucoup d'informations techniques sont communiquées sous la forme de rapports sur l'avancée des recherches, de manuels et de comptes rendus de conférences, et les traces de ces documents non publiés hors ligne

sont très visibles sur Internet. Ceci est également dû au fait que le degré de « dépendance fonctionnelle » est plus élevé qu'en théorie de l'argumentation, ce qui signifie qu'il y a un plus grand besoin de coordination des techniques et des résultats à l'échelle des groupes, de la discipline et du pays. La prédominance du nom de domaine **.nl** est surprenante au regard de cette interdépendance, surtout dans la mesure où des membres du groupe avaient publié des articles dans des revues internationales et donné des conférences internationales sur la création du corpus. Inversement, bien que le corpus national britannique soit référencé dans des publications et articles produits dans le contexte du projet, il n'existait aucun lien pointant vers ses pages Web. Une attention particulière a en revanche été accordée à la visibilité de références internationales telles qu'EAGLES et PRAAT *via* des liens sortants, afin de s'attribuer une certaine autorité sur le sujet. Un corpus peut s'envisager à la fois comme outil et comme sujet de recherche. Pour les développeurs de corpus tels que le CGN, le projet constitue un sujet de recherche, qui soulèvera des questions de recherche, mais pour ceux qui utiliseront ce corpus, il deviendra un outil pour répondre à des questions. Peut-être qu'à mesure que le CGN deviendra un outil, les réseaux qui se sont formés autour de lui le placeront plus au centre de la scène internationale de la linguistique pragmatique. Il est également possible que sa focalisation sur la langue néerlandaise l'enferme dans un réseau national, malgré sa nouveauté et son intérêt technique potentiel en tant que rare exemple de corpus oral.

Dans la sphère Internet du CGN, la recherche menée par les membres du projet à titre individuel est noyée par la recherche collective, à laquelle on donne l'avantage. Par exemple, le chef du projet et développeur du site officiel du CGN n'a intégré aucun lien sortant vers les pages personnelles des membres du projet, ni même vers la sienne. Bien qu'il existe plusieurs liens pointant vers l'URL centrale du projet sur les sites Web d'anciens étudiants et d'assistants-chercheurs, aucun lien n'émane des pages Web des universitaires reconnus qui ont participé au projet.

Les contributions personnelles des membres de l'École d'Amsterdam sont, elles aussi, submergées par le flot d'informations concernant la recherche collective. Les activités des membres de l'École d'Amsterdam en tant qu'enseignants sont plus visibles que leurs activités de recherche. Il existe diverses raisons capables d'expliquer pourquoi si peu de pages Web sont consacrées aux ressources de la recherche dans le domaine de la théorie de l'argumentation. Il se peut que le degré élevé de « dépendance stratégique », entraînant une compétitivité accrue, dissuade les

chercheurs de partager leurs informations et leurs données. À l'inverse, le degré plus bas de « dépendance fonctionnelle » peut signifier qu'il est moins impératif de partager les ressources.

Des degrés relatifs de *mutual dependance* et de *task uncertainty* peuvent avoir une influence importante sur la taille des sphères Internet (en termes de nombre d'URL), car un degré décroissant de *mutual dependence* (tant fonctionnelle que stratégique) indique que le besoin de coordination et de partage des ressources est moindre. Cela peut se combiner à un degré croissant de *task uncertainty*, ce qui rend la coordination des problèmes, techniques et résultats de recherche plus difficile. Ces différences culturelles rendent problématique la création d'indicateurs numériques standardisés visant à mesurer la productivité des disciplines.

Ces résultats indiquent aussi que lorsque l'on a recours à différentes unités d'analyse, différents niveaux de proéminence et de connectivité peuvent être observés. Dans le cas de l'École d'Amsterdam, s'appuyer sur la communauté professionnelle pour créer une sphère Internet révèle des indications d'évaluation par ses pairs, d'estime et de pouvoir, alors que ce ne serait pas le cas pour une sphère Internet construite autour des pages de leur département.

CONCLUSION

+++++

Comme l'a démontré cet article, les différences entre ce que les disciplines rendent visible sur Internet et ce qui demeure invisible sont très révélatrices. La théorie de Whitley [1984] s'est avérée très utile pour la compréhension de ces différences. Par exemple, dans le cas de l'École d'Amsterdam, le taux important de liens pointant vers le site de l'ISSA peut être interprété comme un degré croissant de « dépendance stratégique » (avec ses exigences inhérentes en matière d'alliances et de construction d'une communauté), alors que le CGN ne partage pas cette position stratégique centrale au sein de sa discipline, ce qui pourrait expliquer le taux relativement bas de liens pointant vers le site Web du projet. En termes de contenu des URL de chaque sphère Internet, le plus haut degré de *strategic uncertainty* au sein du projet CGN, qui suit une tendance générale de la discipline, est reflété par l'emploi d'un langage justificatif pour le plan du corpus et par la façon ostentatoire dont les liens sortants pointant vers des références techniques internationales sont intégrés dans cette justification.

Voici quelques aspects particuliers de l'identité culturelle d'une discipline intellectuelle qui pourraient apporter une contribution importante à la création d'un cadre théorique pour l'interprétation des relations hypertextuelles ainsi qu'un moyen d'identifier les unités d'analyse adéquates :

- objectifs de la communauté de chercheurs, en termes de théorie, production, recueil et partage des données, etc. ;
- méthodes, techniques et outils ;
- public : étudiants de deuxième cycle, public profane, universitaires issus d'autres disciplines, etc. ;
- maturité de la discipline ;
- orientation nationale ou internationale de la discipline ;
- envergure de la discipline.

Capables de conférer un pouvoir interprétatif aux phénomènes sociaux, intellectuels et institutionnels (voir également Beaulieu et Simakova [2006] sur le sujet), les études hypertextuelles ethnographiques sont d'autant plus précieuses qu'elles permettent d'identifier une unité d'analyse plus subtile pour l'étude des différences qui opposent les disciplines en matière de production et de pratiques numériques. Il est toutefois nécessaire d'augmenter proportionnellement ces études afin que le cœur pédagogique ou les activités de recherche des disciplines intellectuelles soient immédiatement identifiables et que des schémas comportementaux plus généraux se fassent jour. Un faible taux de liens universitaires dans la littérature, en particulier en sciences humaines, peut être le résultat d'un problème de méthodologie plutôt que le signe d'une déficience au niveau de la superstructure des universités [Harries *et al.*, 2004] ou de la recherche institutionnelle. Afin d'éviter tout contresens dans l'interprétation des résultats, il est indispensable, au niveau méthodologique, d'être clair sur l'unité d'analyse employée – bien faire la distinction notamment entre département universitaire et discipline afin d'éviter tout amalgame. Si, par exemple, une étude exclusivement quantitative des relations hypertextuelles aux Pays-Bas prenait les départements universitaires pour nœud central, alors les pages Web de l'École d'Amsterdam < <http://cf.hum.uva.nl/data/afd/neerlandistiek/tar-english> >²⁴ feraient apparaître l'École comme un groupe isolé (en raison de l'absence d'URL entrantes ou sortantes). Dans la théorie des réseaux, l'étendue de l'interconnectivité entre les nœuds (par exemple une URL comptant beaucoup de liens entrants) est utilisée pour interpréter des facteurs sociaux tels que la dose de critique

24. N.D.T. : voir note 7.

que s'autorise un chercheur ou un groupe de chercheurs, faisant autorité en tant que producteurs de connaissances, à l'égard d'une communauté scientifique. Utilisé comme indicateur de performance quantitative, le modèle hypertextuel qui se dégagerait de l'École d'Amsterdam en tant que département serait très loin de refléter l'envergure internationale du groupe et le rôle capital qu'il joue au sein de la théorie de l'argumentation.

Les résultats exposés dans cet article démontrent que des approches méthodologiques de l'étude des liens hypertextes et des modèles hypertextuels différentes aboutissent à des représentations des domaines de recherche et des communautés universitaires qui les peuplent différentes. Dans notre cas, c'est le recours aux techniques ethnographiques (entretiens et analyse du discours) allié à l'utilisation d'outils tels que le moteur de recherche de Google et Microsoft Site Analyst (qui s'appuient tous les deux sur l'exploration automatique du Web), qui ont fourni un apport particulier au concept de sphère Internet : celui de rendre les traces de l'identité intellectuelle et sociale de disciplines universitaires spécialisées plus visibles, sur l'Internet s'entend. Dans une étude à plus grande échelle que celle rapportée ici, il aurait peut-être été impossible de trianguler autant de techniques différentes, d'où l'importance de rester vigilant et de bien comprendre les limites que comporte toute approche lorsqu'il s'agit d'interpréter les résultats.

REMERCIEMENTS

++++
 Cet article s'appuie sur des recherches que j'ai menées au sein du groupe Networked Research and Digital Information (NERDI) à l'Institut néerlandais des sciences de l'information et a été financé par l'Académie royale des arts et des sciences néerlandaise. J'aimerais remercier mes collègues du NERDI (devenu le Virtual Knowledge Studio des sciences humaines et sociales) pour leur soutien inestimable au cours de ces recherches et leurs commentaires avisés sur les versions préliminaires de cet article. Je voudrais aussi remercier mes deux correcteurs, le professeur Peter van den Besselaar et le D^r Viv Cothey, dont les commentaires détaillés m'ont permis d'aboutir à une version définitive améliorée.

RÉFÉRENCES BIBLIOGRAPHIQUES

++++
 Atwell Eric, Demetriou George, Hughes John, Schiffrin Amanda, Souther Clive et Wilcock Sean, "A comparative evaluation of modern English

corpus grammatical annotation schemes”, *ICAME Journal*, n° 24, 2000, pp. 7-23.

Bar-Ilan Judit, “How do search engines respond to some non-English queries?”, *Journal of Information Science*, 31 (1), 2005, pp. 13-28.

Beaulieu Anne et Simakova Elena, “Textured connectivity: an ethnographic approach to understanding the timescape of hyperlinks”, *Cybermetrics*, 10 (1), 2006.

[En ligne] < <http://cybermetrics.cindoc.csic.es/articles/v10i1p5.html> > (consulté le 3 juin 2004).

Becher Tony, *Academic Tribes and Territories: Intellectual Enquiry and the Cultures of Disciplines*, Buckinghamshire, Open University Press, 1989.

Björneborn Lennart et Ingwersen Peter, “Toward a basic framework for webometrics”, *Journal of the American Society for Information Science and Technology*, 55 (14), 2004, pp. 1216-1227.

Borsley Robert D. et Ingham Richard, “Grow your own linguistics? On some applied linguists’ views of the subject”, *Lingua*, 112 (1), 2002, pp. 1-6.

Bourdieu Pierre, *Homo Academicus*, Traduit par Peter Collier, Stanford (Ca), Stanford University Press, 1988.

Bruckner Eberhard, Ebeling Werner et Scharnhorst Andrea, “Stochastic dynamics of instabilities in evolutionary systems”, *Systems Dynamics Review*, 2, 1989, pp. 176-191.

Crane Diana, *Invisible Colleges: Diffusion of Knowledge in Scientific Communities*, London, Chicago, The University of Chicago Press, 1972.

Cronin Blaise, Snyder Herbert W., Rosenbaum Howard, Martinson Anna et Callahan Ewa, “Invoked on the Web”, *Journal of the American Society for Information Science*, 49 (14), 1998, pp. 1319-1328.

Eemeren Frans H. van et Grootendorst Rob, *Argumentation, Communication, and Fallacies: A Pragma-Dialectical Perspective*, Mahwah, New Jersey, Lawrence Erlbaum Associates, 1992.

Eemeren Frans H. van et Grootendorst Rob (dir.), *Studies in Pragma-Dialectics*, Amsterdam, Sic Sat, International Centre for the Study of Argumentation, 1994.

Eemeren Frans H. van et Houtlosser Peter, "Rhetoric in pragma-dialectics", *Argumentation, Interpretation and Rhetoric* [Electronic journal University of St. Petersburg & University of Amsterdam], 1 (1), 2001.

Fry Jenny, "Scholarly research and information practices: a domain analytic approach", *Information Processing and Management*, 42 (1), 2006, pp. 299-316.

Gibbons Michael, Limoges Camille, Nowotny Helga, Schwartzman Simon, Scott Peter et Trow Martin, *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies*, London, Sage, 1994.

Harries Gareth, Wilkinson David, Price Liz, Fairclough Ruth et Thelwall Mike, "Hyperlinks as a data source for science mapping", *Journal of Information Science*, 30 (5), 2004, pp. 436-447.

Haythornthwaite Caroline et Wellman Barry, "Work, friendship and media use for information exchange in a networked organization", *Journal of the American Society for Information Science*, 49 (12), 1998, pp. 1101-1114.

Heilbron Johan, "A regime of disciplines: toward a historical sociology of disciplinary knowledge", in Camic Charles et Joas Hans (dir.), *The Dialogical Turn: New Roles for Sociology in the Post-Disciplinary Age*, Lanham, Rowman & Littlefield Pub Inc, 2004, pp. 23-42.

Heimeriks Gaston, Hörlesberger Marianne et Besselaar Peter van den, "Mapping communication and collaboration in heterogeneous research networks", *Scientometrics*, 58 (2), 2003, pp. 391-413.

Hine Christine, *Virtual Ethnography*, London, Sage, 2000.

Kling Rob et McKim Geoffrey, "Not just a matter of time: field differences and the shaping of electronic media in supporting scientific communication", *Journal of the American Society for Information Science*, 51 (14), 2000, pp. 1306-1320.

Kochen Manfred, "Mathematical model for the growth of two specialties", *Science of Science*, 3 (11), 1983, pp. 199-217.

Lawler John M. et Dry Helen Aristar (dir.), *Using Computers in Linguistics: a Practical Guide*, London, Routledge, 1998.

Lenoir Timothy, *Instituting Science: The Cultural Production of Scientific Disciplines*, Stanford (Ca), Stanford University Press, 1997.

Marcus George E., "Ethnography in/of the world system: the emergence of multi-sited ethnography", *Annual Review of Anthropology*, 24, 1995, pp. 95-117.

Merton Robert K., "The Matthew effect in science", *Science*, 159 (3810), 1968, pp. 56-63.

Moukdad Haidar, "Language-based retrieval of Web documents: an analysis of Arabic-recognition capabilities of two major search engines", *Proceedings of the American Society for Information Science and Technology*, 39 (1), 2002, p. 551.

Price Derek J., *Little Science, Big Science*. New York, Columbia University Press, 1963.

Rousseau Ronald, "Sitations: an exploratory study", *Cybermetrics*, 1 (1), 1997.

[En ligne] < <http://cybermetrics.cindoc.csic.es/articles/v1i1p1.html> > (consulté le 21 juin 2004).

Rousseau Ronald, "Daily time series of common single word searches in Alta Vista and NorthernLight", *Cybermetrics*, 2/3 (1), 1999.

[En ligne] < <http://www.cindoc.csic.es/cybermetrics/vol2iss1.html> > (consulté le 10 mai 2005).

Salter Liora et Hearn Alison, *Outside the Lines: Issues in Interdisciplinary Research*, Montréal, McGill-Queen's University Press, 1996.

Sampson Geoffrey, *Are we nearly there yet, Mum?*. Corpus Linguistics 2003 conference, Lancaster, mars 2003.

[En ligne] < <http://www.grsampson.net/index.html> > (consulté le 3 janvier 2004).

Schneider Steven M. et Foot Kirsten A., "Online structure for political action: exploring presidential Web sites from the 2000 American election", *Javnost (The Public)*, 9 (2), 2002, pp. 43-60.

Swales John M., *Other Floors, other Voices: A Textography of a Small University Building*, London, Lawrence Erlbaum Associates, 1998.

Thelwall Mike, "Interpreting social science link analysis research: a theoretical framework", *Journal of the American Society for Information Science and Technology*, 57 (1), 2006, pp. 60-68.

Thelwall Mike et Harries Gareth, "Can personal Web pages that link to universities yield information about the wider dissemination of research?", *Journal of Information Science*, 30 (3), 2004, pp. 243-256.

Vaughan Liwen et Thelwall Mike, "Search engine coverage bias: evidence and possible causes", *Information Processing and Management*, 40 (4), 2004, pp. 693-707.

Wenger Étienne, *Communities of Practice: Learning, Meaning and Identity*, Cambridge, Cambridge University Press, 1998.

Whitley Richard, *The Intellectual and Social Organization of the Sciences*, Oxford, Clarendon Press, 1984.

Wouters Paul et Vries Repke de, "Formally citing the Web", *Journal of the American Society for Information Science and Technology*, 55 (14), 2004, pp. 1250-1260.