

# « La connaissance de la connaissance » : une réflexion sur la triangulation des analyses textuelles à partir d'un corpus spécialisé en gouvernance d'entreprise

Stéphane Trébuq \*

\* Centre de Recherche en Contrôle et Comptabilité Internationale - CRECCI  
IAE de Bordeaux, Rue du Cdt Arould, 33 Bordeaux  
trebuq@u-bordeaux4.fr

**Résumé.** Suite à la survenue récente de scandales financiers, la synthèse des idées mobilisables en gouvernance d'entreprise semble désormais essentielle si l'on veut sécuriser les investisseurs. Dans cette perspective, le présent projet de recherche consiste à mettre en œuvre un panel d'outils d'analyse de données textuelles (Alceste, Syntex, Tropes-Zoom/Decision Explorer, Wordmapper, Weblex) afin d'évaluer les moyens dont peut disposer un analyste désireux d'extraire des connaissances contenues dans un ensemble d'articles académiques. La qualité de représentation du corpus dans sa globalité est tout d'abord testée. L'étude est ensuite centrée sur le concept même de connaissance, mobilisé dans la théorie de la gouvernance des entreprises. La convergence et la complémentarité des approches méthodologiques sont alors explicitées. Il en est de même pour ce qui concerne la capacité d'extraction d'une connaissance pertinente à partir des textes étudiés.

## 1 Introduction

Suite à la survenue de récents scandales financiers, les représentants des entreprises, de même que leurs différentes parties prenantes, ont été conduits à s'interroger quant à l'efficacité des dispositifs de gouvernance<sup>1</sup> actuellement en vigueur. Sur un plan académique, la synthèse des idées mobilisées dans ce domaine est désormais indispensable si l'on veut disposer d'un cadre conceptuel aussi explicite et exhaustif que faire se peut. L'objet de la présente recherche est de mettre en œuvre une triangulation méthodologique, en utilisant un panel d'outils d'analyse de données textuelles. Le corpus retenu pour l'extraction de connaissances est constitué de près de 300 000 mots correspondant à un ensemble de 32 articles publiés dans diverses revues scientifiques et rédigés entre 1985 à 2003 par Gérard Charreaux<sup>2</sup>, dont les travaux font référence dans le domaine de la gouvernance des entreprises.

La démarche expérimentale suivie a été conduite à deux niveaux d'analyse. Elle a tout d'abord offert l'opportunité de tester les capacités de représentation du corpus dans sa globalité. Pour ce faire, nous avons utilisé quatre logiciels spécialisés : Alceste, Syntex,

---

<sup>1</sup> "Le gouvernement des entreprises recouvre l'ensemble des mécanismes organisationnels qui ont pour effet de délimiter les pouvoirs et d'influencer les décisions des dirigeants, autrement dit, qui « gouvernent » leur conduite et définissent leur espace discrétionnaire" (Charreaux 1997).

<sup>2</sup> <http://perso.wanadoo.fr/gerard.charreaux/perso/gcaccueil.html#Sommaire>

Zoom-Tropes, Wordmapper<sup>3</sup>. L'étude a ensuite été centrée sur le concept même de connaissance, et la place qui lui revient dans l'approche théorique relevant de la gouvernance des entreprises<sup>4</sup>. A ces deux niveaux d'analyse, l'étude permet non seulement d'apprécier le degré de convergence des résultats obtenus mais également la complémentarité des approches méthodologiques. Les limites inhérentes aux outils d'analyse textuelle, notamment en ce qui concerne leur capacité à aider un analyste à extraire une expertise pertinente, ont été observées. Il s'en suit une série de propositions concrètes en vue d'améliorer la performance des technologies actuellement disponibles et de parfaire leurs croisements possibles.

## 2 Synthèse globale du corpus

La mise en œuvre des logiciels d'analyse de données textuelles, utilisés dans le cadre de cette recherche, peut être considérée comme une première étape dans la quête de l'analyste souhaitant comprendre la structuration des connaissances en gouvernance d'entreprise. L'objet n'est pas de se livrer à une simple classification des textes, mais d'en comprendre le contenu et d'en extraire des connaissances (voir Dousset 2003). Avant même d'atteindre une compréhension plus fine des articles scientifiques retenus, un premier enjeu se situe au niveau de la capacité à obtenir une représentation globale de l'expertise développée par l'auteur des textes étudiés. Deux approches complémentaires ont permis d'opérer la transition entre la base des textes sélectionnés et l'identification des connaissances mobilisées par l'auteur. Tout d'abord, un premier groupe d'outils, formé par Alceste et Syntex, aboutit à des classifications terminologiques. Le corpus peut ainsi être utilement segmenté, afin d'en révéler sa structuration. Ensuite, un second groupe d'outils, composé de Zoom-Tropes/Decision Explorer et Wordmapper, permet l'obtention d'une cartographie de synthèse du corpus, tout en autorisant des retours au texte d'origine, afin d'explicitier les relations ayant pu être identifiées entre les concepts. Enfin dans la perspective de renforcer la validité des résultats obtenus, les analyses de ces outils peuvent donner lieu à des comparaisons, ainsi qu'à diverses combinaisons.

### 2.1 Structure et catégorisation du corpus

Les logiciels Alceste (Reinert 2003) et Syntex (Bourigault et al. 2000) opèrent à des niveaux différents. Le premier outil est en mesure de repérer les cooccurrences de termes, et ce, à partir d'un découpage du texte en unités de contexte élémentaires (u.c.e). Il s'en suit une classification hiérarchique descendante de ces u.c.e., aboutissant à l'identification de différents « mondes lexicaux » reposant sur des associations entre termes, totalement différentes. Le second outil permet d'extraire la terminologie présente dans les textes étudiés, en isolant notamment l'ensemble des syntagmes nominaux. Il est alors possible de mieux saisir l'organisation de l'ontologie spécifique au domaine étudié. Les résultats de ces deux traitements sont présentés ci-après.

---

<sup>3</sup> Les adresses internet des différents logiciels sont les suivantes. Alceste : <http://www.image.cict.fr/> ; Syntex : <http://www.univ-tlse2.fr/erss/membres/bourigault/> ; Zoom-Tropes : <http://www.acetic.fr> ; Decision Explorer : <http://www.banxia.com/demain.html> ; Wordmapper : <http://www.grimmersoft.com/>.

<sup>4</sup> A ce second niveau d'analyse, l'outil Weblex a également été utilisé. Voir <http://lexico.ens-lsh.fr/local/weblex.html> et <http://weblex.ens-lsh.fr/biblio/slh/>

### 2.1.1 « Mondes lexicaux » identifiés avec Alceste.

En utilisant les paramètres par défaut, procédant notamment à une lemmatisation du texte, le logiciel Alceste identifie 6 classes différentes d'u.c.e., auxquelles s'ajoute une septième composée de fragments de textes dont les cooccurrences ne sont pas clairement caractérisables. Le taux de classement pour l'ensemble des passages du texte analysé atteint, dans le cas présent, 76% (voir Fig. 1 et Tab. 1).

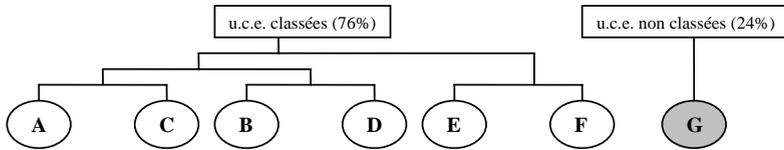


FIG. 1 – Arborescence de la classification obtenue sous le logiciel Alceste

Classe	Principales formes réduites représentatives de la classe (khi2 associé à la forme réduite) <sup>5</sup>	Formes absentes (khi2)
A	valeur+(427,7) ; rente+(409,2) ; creation+(310) ; stakehold+(269) ; enracin+(218,5) ; salarié(214,4) ; latitude(206,8) ; partenaire+(188,4)	societe+(-81,6) ; conseil+(-59,0)
B	theor+(721,2) ; finance+(300,3) ; recherche+(248,5) ; courant+(214,9) ; paradigme+(213,0) ; approche+(212,9)	dirigeant+(-119,6) ; conseil+(-98,7)
C	coût+(376,0) ; acteur+(288,8) ; agent+(197,7) ; risque+(173,9) ; flux(168,3) ; conflit+(153,3) ; gain+(146,7) ; residuel+(143,9)	conseil+(-125,0) ; entreprise+(-99,3)
D	organisationnelle+(557,4) ; forme+(420,1) ; positive+(334,0) ; agence+(300,4) ; theor+(273,5) ; organis+(234,1) ; transaction+(210,6)	dirigeant+(-69,4) ; actionnaire+(-58,6)
E	variable+(805,1) ; privatisatis+(721,3) ; indicateur+(561,0) ; rentabilite+(537,0) ; test+(501,6) ; significative+(497,7) ; resultat+(408,4)	dirigeant+(-78,5) ; mecanisme(-41,7)
F	conseil+(856,3) ; administrat(516,7) ; administrateur+(496,0) ; societe+(308,1) ; contrôle+(252,4) ; etat+(240,5) ; role+(216,2)	theor+(-260,9) ; cout+(-207,7)

TAB 1 – Tableau des « mondes lexicaux » identifiés avec le logiciel Alceste.

Dans le cadre de cette première analyse, deux groupes peuvent être nettement distingués. Les classes A à D représentent des passages théoriques. La présentation des approches contractuelles classiques, dont, entre autres, la théorie de l'agence, est opérée au sein de la classe D. Elle a pour complément le monde lexical B définissant l'appartenance des différentes analyses aux différents champs paradigmatiques présentés par Gérard Charreaux. La prise de connaissance des u.c.e. permet à l'analyste d'identifier les classes A et C comme mondes lexicaux servant à modéliser le comportement des acteurs, notamment à partir des notions centrales de valeur et de coût. Quant aux classes E et F, elles correspondent aux terrains des recherches empiriques de l'auteur, la classe E étant relative à l'étude du cas des privatisations françaises, et la classe F se rapportant plus précisément aux recherches consacrées au fonctionnement des conseils d'administration.

<sup>5</sup> Le khi2 calculé est établi à partir d'une matrice d'occurrences de dimension 2x2, la première ligne considérant les u.c.e (unités de contexte élémentaires) de la classe considérée et la seconde ligne toutes les autres u.c.e., la première colonne la présence de la forme et la seconde son absence.

Cette classification permet un premier repérage des principales thématiques du corpus servant de référence au projet d'extraction des connaissances relatives au champ de la gouvernance. Elle reste sommaire, eu égard à l'absence d'identification des liens pouvant rattacher les termes au sein d'une même classe. De plus, aucune analyse n'est proposée à propos de la catégorie G, correspondant néanmoins à 24% des u.c.e. Nous illustrerons ultérieurement, au point 2.3.2.1, une solution possible de remédiation à ces défauts.

**2.1.2 Classification manuelle des syntagmes obtenus à partir de Syntex.**

Le même corpus, analysé à l'aide du logiciel Syntex, développé par Didier Bourigault, conduit à l'extraction de 4 814 syntagmes nominaux. Les syntagmes les plus cités (voir Tab. 2) permettent bien de confirmer la présence du thème de la gouvernance d'entreprise dans le corpus étudié.

Principaux syntagmes (Fréquence) : conseil d'administration (614), théorie de l'agence/ théorie positive de l'agence (372/246), système de gouvernance/système de gouvernement (242/196), coûts d'agence (240), création de valeur (224), gouvernement des entreprises (192), coûts de transaction (179)
--

TAB 2 – *Tableau des syntagmes les plus fréquents dans le corpus identifiés avec Syntex.*

La liste des syntagmes peut ensuite être traitée par l'analyste à deux niveaux, et ce, manuellement. Le premier tri manuel a permis de repérer sept grandes catégories. Au sein de chaque classe, on peut alors identifier pour chaque syntagme un terme pivot. A titre d'exemples, pour le syntagme « nouveaux actionnaires » on choisira « actionnaire », de même pour « administrateurs indépendants » on retiendra « administrateur ». Ce choix de termes pivots réalisé par l'analyste, bien que discutable, permet d'obtenir une présentation plus synthétique de l'ontologie relative au champ de la gouvernance (voir Tab. 3).

Classe (Nbre de syntagmes)		Principaux termes pivots des syntagmes (nombre de syntagmes nominaux associés)
S1	Acteurs (335)	dirigeant(46) ; actionnaire (43) ; administrateur (27) ; stakeholder ou partie prenante (26) ; conseil d'administration (22)
S2	Jeux entre acteurs (870)	relation (69) ; contrôle (67) ; mécanisme (67) ; rôle (53) ; incitation (36) ; intérêt (30) ; objectif (29) ; conflit (28)
S3	Décisions du dirigeant (387)	décision (39) ; investissement (29) ; financement (21) ; gestion (21) ; acquisition (18) ; conséquences (15) ; stratégie (15)
S4	Cadre qualitatif (1445)	théorie (155) ; analyse (51) ; système (46) ; efficacité (45) ; modèle (34) ; notion (34) ; approche (32) ; étude (32)
S5	Cadre quantitatif (515)	coût (80) ; valeur (58) ; capital (44) ; performance (37) ; mesure (34) ; gain (16) ; critère (15) ; perte (13) ; taux (13)
S6	Cadre juridique (217)	contrat (42) ; marché (39) ; droit (24) ; propriété (19) ; environnement (9) ; protection (9) ; cadre (6) ; capitalisme (6)
S7	Cadre contextuel (803)	forme (42) ; type (41) ; structure (31) ; caractère (30) ; système (30) ; société (28) ; nature (27) ; séparation (27) ; niveau (23)

TAB 3 – *Tableau de classement des syntagmes obtenus sous Syntex.*

Cette classification, lorsqu'elle est exposée dans le détail, peut constituer un puissant outil d'aide à la compréhension de la complexité des analyses menées en gouvernance d'entreprise. Elle permet de lister avec une précision importante l'ensemble des concepts

clés. L'analyste en gouvernance est ainsi conduit à réfléchir aux différents acteurs qu'il souhaite prendre en compte. La connaissance des classes S2 et S3 permet de clarifier les concepts clés sous-jacents à la modélisation du comportement des acteurs. Les classes S4 à S7 conduisent enfin à lister les différentes propriétés pouvant être rattachées aux situations analysées (choix et identification des cadres théoriques, des mesures financières, de l'environnement réglementaire, et du contexte organisationnel).

Toutefois, force est de constater que l'identification de l'ensemble des acteurs de la gouvernance, présents dans les textes, en s'en tenant à la seule liste des syntagmes nominaux identifiés par Syntex, est manifestement insuffisante. A titre d'exemple, aucun syntagme ne comporte le terme d'« avocat » ou de « législateur », alors qu'il s'agit bien d'acteurs potentiels, cités, du reste, dans le corpus. Il s'avère donc nécessaire de recourir à d'autres approches si l'on souhaite obtenir une ontologie complète.

## 2.2 Relations entre concepts et représentations synthétiques du corpus

L'analyse du corpus peut être opérée selon des modalités complémentaires aux deux classifications présentées précédemment. Au lieu de tenter une classification des contextes ou des catégories de concepts mobilisés dans les articles rédigés par l'expert, une représentation de la totalité du corpus sous forme de graphique peut être recherchée. Une telle approche n'abandonne pas totalement l'objectif de classification, mais elle présente l'avantage non négligeable de faire apparaître plus clairement les liens existant entre les concepts. Le logiciel Zoom-Tropes, fondé sur l'analyse propositionnelle des discours (APD) (Ghiglione et al. 1998), peut servir de base à l'identification du lien entre les champs sémantiques présents dans le texte. Ces liens peuvent ensuite être représentés à l'aide d'un outil graphique tel que Decision Explorer. A des fins similaires, il est également possible de recourir à un logiciel tel que Wordmapper. Celui-ci scanne l'ensemble du texte avec une fenêtre glissante de 15 à 20 mots, et mobilise des algorithmes de classification et de MDS (multi-dimensional scaling) pour finalement aboutir à des graphes de cooccurrences entre termes.

### 2.2.1 Cartographie des mises en relations obtenues avec Zoom-Tropes

L'approche de Zoom-Tropes, lorsqu'elle est réalisée au niveau d'agrégation sémantique le plus faible, dénommé dans le logiciel 'références utilisées', permet de lister la quasi-intégralité des occurrences du corpus. La liste des références utilisées, ou référents-noyaux, permet ainsi de compléter l'ontologie obtenue à partir de Syntex, telle qu'elle a été présentée au point 2.1.2. Toutefois, certains regroupements sémantiques opérés automatiquement par le logiciel Tropes peuvent apparaître discutables pour un analyste. Ainsi, le référent-noyau 'patron' regroupe en réalité les termes de « dirigeant », « administrateur », « PDG », « directeur général » et « président du conseil d'administration ». Dans le cas présent, un tel regroupement de termes ne pourrait être retenu en tant que tel par un spécialiste du domaine, du fait même des rôles éventuellement très différents qu'un dirigeant et un administrateur peuvent jouer au sein d'une même entreprise. Il s'en suit la nécessité d'analyser avec minutie la composition des références utilisées, et de proposer à Tropes une liste de termes beaucoup plus détaillées. Une fois la liste des termes à analyser redéfinie, conformément à la méthodologie de l'APD, le texte analysé est découpé en propositions grammaticales, débouchant sur un référencement des cooccurrences entre termes. Le sens des mises en

relations entre termes est conservé, de telle sorte que dans une phrase du type « les actionnaires s'assurent du contrôle des dirigeants » on obtiendra les mises en relations suivantes : 'actionnaire > contrôle', 'contrôle > dirigeant' et 'actionnaire > dirigeant'. De telles relations peuvent, alors, être représentées graphiquement au moyen d'un outil tel que Decision Explorer. L'utilisation d'une telle interface graphique présente deux avantages mais également deux inconvénients majeurs. Le premier avantage concerne la maîtrise parfaite par l'analyste qui construit les liens, au fur et à mesure, à l'aide de Decision Explorer, tout en prenant connaissance avec Tropes des propositions représentatives de ces liens (démarche « ascendante »). Le second avantage réside dans la possibilité de pouvoir modifier le positionnement des termes, afin d'optimiser la lisibilité de la carte. En revanche, au titre des inconvénients, on peut relever tout d'abord la nécessité d'une saisie manuelle des concepts et des mises en relation et de surcroît, l'absence de prise en considération dans le graphe de la fréquence de ces mêmes mises en relations.

### **2.2.2 Graphe des mots significatifs obtenu à partir de Wordmapper.**

Un logiciel tel que Wordmapper présente, comparativement à Zoom-Tropes, plusieurs avantages. En effet, il permet, tout d'abord, sous réserve que l'on sélectionne un nombre suffisant de 'mots significatifs', d'obtenir une série de différentes classes de termes présentes dans le corpus. Comme on peut le constater (voir tableau en Annexe 1), l'algorithme de classification basé sur les 300 premiers mots significatifs fournit près du double du nombre de classes obtenu à partir des logiciels Alceste ou Syntex. On peut cependant noter une certaine convergence entre la classe wordmapper intitulée 'création de valeur' et la classe A obtenue sous Alceste. De même, les classes sous wordmapper intitulées 'théories positives', 'actionnaire dirigeant', 'mesure performance', et 'conseil d'administration' sont respectivement proches des classes B & D, C, E, et F identifiées par Alceste.

Les représentations graphiques obtenues automatiquement fournissent directement un graphe des liens entre les termes. La distance respective entre les termes est cette fois proportionnelle à leur fréquence de cooccurrence. A ce degré d'analyse, le graphe obtenu est assez similaire à celui obtenu à l'aide de Zoom-Tropes et Decision Explorer. Quatre pôles se dégagent (théories mobilisées, contrôle assuré par le conseil d'administration, rôle du dirigeant et création de valeur).

Toutefois, si l'on augmente le nombre de mots pris en considération dans l'analyse, les graphiques deviennent très difficilement lisibles, et s'avèrent d'une aide limitée pour atteindre une meilleure compréhension des relations présentes dans le corpus. En l'occurrence, cela constitue l'une des limites de l'approche proposée par Wordmapper, d'autant plus que le graphe est construit en première phase, et nécessite, dans une deuxième phase, un retour au texte (démarche « descendante »). Il est alors loisible à l'analyste de procéder à d'autres stratégies de représentation du corpus, en recourant par exemple à des approches par triangulation.

## **2.3 Extractions de connaissances supplémentaires par triangulation**

Utilisée afin d'accroître la validité des résultats empiriques obtenus, la triangulation consiste à recourir à plusieurs méthodes (*methodological triangulation*, voir Bryman 2003), et dans notre cas, plusieurs logiciels d'analyse textuelle. Cette triangulation peut être envisagée à deux niveaux. Le premier niveau se situe dans la comparaison des résultats obtenus. Le

second niveau de triangulation consiste à exploiter les extractions d'un logiciel donné à partir d'un autre logiciel. A ce titre, plusieurs croisements originaux peuvent être effectués.

### **2.3.1 Comparaison des résultats obtenus : triangulation de premier degré**

Par triangulation de premier degré, on peut entendre tout d'abord la comparaison entre les sorties des différents logiciels. Comme on a pu le voir, les classifications obtenues à partir d'Alceste et de Wordmapper apparaissent avec un degré de convergence significatif. Wordmapper dispose, à ce titre, d'une supériorité car ses représentations graphiques indiquent explicitement les liens de cooccurrences entre les termes. En ce qui concerne les observations obtenues avec Syntex et Zoom-Tropes, le « dirigeant » et l'« actionnaire » apparaissent bien comme deux acteurs majeurs autour desquels s'organise la modélisation des problèmes de gouvernance. Cette première comparaison permet de conclure à une convergence satisfaisante des analyses. Toutefois, pour l'analyste, des triangulations de second degré peuvent constituer une réelle plus-value en termes d'acquisition de connaissances.

### **2.3.2 Croisement des analyses : triangulation de second degré**

L'intérêt de cette étape est de croiser les analyses des différents logiciels. Une première démarche consiste à croiser les classifications tirées d'Alceste et Syntex. On peut également rechercher à obtenir des représentations graphiques des classes d'Alceste à partir de Wordmapper. Des croisements sont également envisageables entre la classification issue de Syntex et Zoom-Tropes.

#### **2.3.2.1 Croisements entre Alceste et Syntex**

Le croisement entre les classifications Alceste et Syntex pourrait être aisément automatisé. Toutefois, dans le cas présent, il a été réalisé manuellement. On aboutit, de la sorte, à une matrice 6x7 à partir des 6 classes tirées d'Alceste et des 7 classes tirées de la liste des syntagmes sous Syntex. Le résultat s'avère particulièrement intéressant, puisqu'il permet d'isoler, par exemple, certaines catégories d'acteurs (classement Syntex) dans les différents « mondes lexicaux » (classement Alceste). On peut aussi lister les jeux d'acteurs, voire même les concepts qualitatifs, spécifiques à chaque « monde lexical ».

#### **2.3.2.2 Croisements entre Alceste et Wordmapper**

Les croisements d'Alceste avec Wordmapper permettent à l'analyste d'accéder beaucoup plus rapidement à la connaissance des « mondes lexicaux ». En effet, les différentes u.c.e. classées par Alceste peuvent être exportées au format txt, pour être ensuite analysées dans Wordmapper. On obtient alors un réseau de termes associés, non plus sur la totalité du corpus, mais sur une sous partie cohérente, à savoir un « monde lexical ». Une telle démarche peut être appliquée pour chaque classe identifiée grâce à Alceste, mais également pour les u.c.e. non classées. Dans ce dernier cas, cela permet de prendre facilement connaissance des autres thèmes présents dans le corpus, bien que non isolés sous Alceste.

#### **2.3.2.3 Croisements entre Syntex et Zoom-Tropes**

Les croisements possibles entre Syntax et Zoom-Tropes peuvent être illustrés à partir de deux exemples. Un premier type de croisement peut consister à créer un répertoire commun des termes qualifiant les acteurs de la gouvernance à partir de Syntax et Zoom-Tropes. Cette liste devient alors un scénario utilisé dans Zoom-Tropes. Le traitement aboutit à une liste des mises en relations entre acteurs dans le corpus. Ces liens, une fois cartographiés par exemple dans Decision Explorer, permettent de visualiser les interactions potentielles entre les acteurs, mais également d’identifier les acteurs isolés ou marginaux. A titre d’exemple, la figure des comptables, qu’il s’agisse d’experts-comptables ou de commissaires aux comptes, n’est rattachée significativement à aucun autre acteur. L’effort de modélisation du comportement des comptables n’apparaît donc pas central en gouvernance, d’autant que l’accent peut être placé sur des concepts complémentaires tels que la réduction de l’asymétrie d’information et la communication financière des dirigeants. On peut cependant s’interroger sur la place marginale réservée à de tels acteurs, au regard du rôle majeur qui a été le leur dans des affaires telles qu’Enron aux Etats-Unis, ou Parmalat en Italie.

Un second type de croisement, plus fin, part du constat que les deux acteurs fondamentaux de la gouvernance sont le « dirigeant » et l’ « actionnaire ». En travaillant dans Zoom-Tropes à partir de ces termes, il est alors possible d’obtenir la liste des références cooccurrentes que ces deux acteurs partagent. Ces références communes peuvent alors être reventilées en fonction de la classification tirée de Syntax (voir Tab. 4). De tels croisements pourraient naturellement être généralisés afin d’obtenir des bibliothèques complètes.

Classes obtenues avec Syntax		Référents noyaux liés à la fois à « dirigeant » et à « actionnaire », à plus de 10 occurrences (liste obtenue à partir de Zoom-Tropes)
S1	Acteurs	administrateur, apporteur, banque, créancier, groupe, principal, propriétaire, agent, stakeholder
S2	Jeux entre acteurs	relation, risque, conflit, comportement, contrôle, position, remplacement, surveillance, transaction, enracinement, changement, discipline, influence, latitude, pouvoir, rôle
S3	Décisions du dirigeant	choix, création, partage, possibilité, maximisation, investissement, capacité
S4	Cadre qualitatif	analyse, approche, avantage, bénéfice, connaissance, conséquence, hypothèse, mode, perspective, richesse, théorie, vision, séparation, fonction, mécanisme
S5	Cadre quantitatif	coût, efficacité, rente, valeur, mesure, capital, part, taux, performance
S6	Cadre juridique	contrat, propriété, protection, droit, statut, marché
S7	Cadre contextuel	situation, environnement, catégorie, type, contrainte, forme, bourse

TAB 4 – *Concepts médiateurs entre les acteurs « dirigeant » et « actionnaire ».*

Ces différents exemples de triangulation permettent d’illustrer l’intérêt de croiser les différentes grilles d’analyse en vue de mieux saisir les concepts mobilisés par un expert.

### 3 Etude du concept de « connaissance » en gouvernance

Portée à un degré d’analyse encore plus fin, chaque terme peut être l’objet d’approfondissements. Dans le cas présent, le choix s’est porté sur le terme de « connaissance » et l’utilisation dont il fait l’objet dans le cadre du corpus de gouvernance étudié.

#### 3.1 Classification du concept de « connaissance »

Dans une optique de classification, la première interrogation concerne le positionnement du terme « connaissance » dans les classifications obtenues à partir d'Alceste et Syntex.

### **3.1.1 Classification du terme « connaissance » sous Alceste**

S'agissant d'Alceste, le terme de « connaissance » est situé en classe C, laquelle comporte également les notions de coût et d'information. On conviendra qu'un tel classement demeure quelque peu évasif pour l'analyste. C'est pourquoi il s'avère absolument indispensable de revenir aux fragments de texte mobilisant ce terme.

### **3.1.2 Classification des syntagmes comprenant le terme « connaissance »**

Les résultats obtenus dans Syntex sont, comparativement à Alceste, beaucoup plus riches, avec 35 syntagmes comprenant le terme « connaissance », eux-mêmes organisables en deux groupes. Une première série de syntagmes sert à qualifier la connaissance (rôle de la connaissance ; connaissance et décision ; connaissance et droits, connaissance et information ; connaissance spécifique ; pertinence de la connaissance). Une seconde série de syntagmes est liée au processus d'acquisition et d'utilisation de la connaissance (production ; acquisition ; transfert ; localisation ; stockage ; répertoire ; accumulation ; utilisation de la connaissance). Les syntagmes présentant les plus fortes occurrences sont « connaissance spécifique », « répertoire de connaissance » et « base de connaissance », « utilisation de la connaissance » ainsi que « transfert de connaissance » avec respectivement 42, 13, 12, 12, et 10 citations. En termes de classification, les syntagmes associés au terme pivot « connaissance » sont rangés dans la catégorie « cadre qualitatif ».

## **3.2 Représentation du concept de « connaissance »**

Les outils Tropes, Wordmapper et Weblex fournissent un éclairage différent et complémentaire relativement au concept de « connaissance ».

### **3.2.1 Analyse sémantique et représentation longitudinale avec Zoom-Tropes**

L'avantage de Zoom-Tropes est tout d'abord de permettre l'obtention de vues longitudinales du corpus pour un référent-noyau ou des mises en relation entre référents noyaux. Dans le cadre de l'analyse sémantique opérée par défaut, la notion de connaissance est approchée en fonction des termes « compétence » et « connaissance ». Cet éclairage est donc pertinent, si l'on souhaite s'interroger sur les frontières du concept même de connaissance. Cependant, en restreignant l'analyse textuelle au seul terme de connaissance, la vue longitudinale du corpus permet d'observer, durant la période analysée, s'étendant de 1985 à 2003, une augmentation significative des occurrences de ce terme à partir de l'année 1999, et ce, jusqu'à 2003.

### **3.2.2 Représentations en étoile avec Zoom-Tropes et Wordmapper**

Les représentations en étoile obtenues dans Zoom-Tropes et Wordmapper devraient permettre à l'analyste de mieux appréhender les contextes d'utilisation du terme. Ces



### 3.3.2 Croisements entre Syntex, Zoom-Tropes et Wordmapper

Les croisements envisageables entre Syntex et Zoom-Tropes peuvent conduire d'une part à lister dans Syntex les syntagmes associés au terme de « compétence », et d'autre part à étudier via Zoom-Tropes certains syntagmes identifiés par Syntex. On peut, par ailleurs, confirmer la proximité entre les termes de « connaissance » et « compétence » soit au travers des mises en relations dans Zoom-Tropes, soit encore en obtenant dans Wordmapper une cartographie sélectionnant les deux termes. Toutefois, même si ces termes semblent proches, leur utilisation apparaît trop spécialisée pour être confondue. La notion de « connaissance » permet à certains auteurs, cités par Gérard Charreaux, de proposer non seulement une définition de l'entreprise, conçue comme un répertoire de connaissances à gérer de manière optimale, mais aussi d'exposer la relation qui lie les salariés à leur entreprise. Dans ce cas, l'employé prend le risque d'acquérir un savoir-faire difficilement transférable, à partir du moment où celui-ci est lié aux spécificités du cadre de travail offert par l'entreprise.

## 4 Conclusion

La « connaissance de la connaissance », chère à Edgar Morin (1986), est utile en général à tout champ de recherche, et plus particulièrement en gouvernance d'entreprise, domaine qui nous a servi de terrain d'application dans le cadre de cet article. Les analyses de données textuelles, lorsqu'elles sont triangulées, peuvent aboutir à des résultats plus approfondis. Elles permettent ainsi d'obtenir une meilleure catégorisation du corpus, et des analyses plus fines. De telles approches appliquées à des textes scientifiques pourraient également, à terme, permettre aux chercheurs et aux lecteurs d'acquérir une meilleure connaissance des auteurs cités. Ceci impliquerait la constitution d'un web sémantique spécifique ou le croisement de bases de données aussi complètes que possible dans les divers domaines étudiés.

## Références

- Bourigault D., Assadi H. (2000), Analyse syntaxique et analyse statistique pour la construction d'ontologie à partir de textes, in Charlet J. et al. (éds.), *Ingénierie des connaissances. Tendances actuelles et nouveaux défis*. Editions Eyrolles/France Telecom, pp 243-255, 2000.
- Bryman A. (2003), Triangulation, in Bryman A. et al. (ed.), *Encyclopedia of Social Research Methods*, Sage Edition, 2003.
- Charreaux G. éd. (1997), *Le gouvernement des entreprises – Corporate Governance – Théories et faits*, Economica, 1997.
- Dousset B. (2003), *Intégration de méthodes interactives de découverte de connaissances pour la veille stratégique*, Habilitation à diriger des recherches de l'Université de Toulouse 3 (Paul Sabatier), 7 novembre.
- Ghiglione R., Landré A., Bromberg M., Molette P. (1998), *L'analyse automatique des contenus*, Dunod, 1998.
- Heiden S. (2004), Interface hypertextuelle à un espace de cooccurrences : implémentation dans Weblex, in Purnelle G. et al., *Le poids des mots, Actes des 7èmes journées internationales d'analyse statistique des données textuelles*, pp 577-588, 2004.

Morin E. (1986), La connaissance de la connaissance, Seuil, Points Essais, Paris, 1986.  
 Reinert M. (2003), Le rôle de la répétition dans la représentation du sens et son approche statistique par la méthode ALCESTE, Sémiotica, 147 (1/4), pp 389-420, 2003.

## Annexe 1

Titre de la classe attribué par défaut (occurr. de la classe)	Termes associés à la classe
parties prenantes (815)	parties prenantes, nécessairement, différentes
architecture organisationnelle (1649)	architecture organisationnelle, répartition, explication, forme, degré, principe efficience
Demsetz firme (4576)	Demsetz firme, mode - coordination, activités, facteurs production, ressources - opportunités - nouvelles, capacité, contrats, conception
capital humain (5190)	capital humain - salariés, connaissance spécifique, information, allocation - droits propriété - actifs, structure actionnariat - financement - flux investissement - évaluation, accroître - politique
sociétés cotées (4201)	groupe - sociétés cotées - managériales - actions, séparation propriété/décision, rentabilité - croissance, fonds propres
création valeur (4163)	partage - créée - création valeur - partenariale - actionnariale - représentation, objectif maximisation - richesse - critère, comptabilité gestion, origine - compétences rente
actionnaire dirigeant (5946)	actionnaire dirigeant - enracinement stratégie - investisseurs - créanciers financiers - statut - dominant - latitude discrétionnaire - réduire - banques crédit - conflits intérêts
théories positive (7873)	théories positive - relation agence - agents principal - coûts transaction - coopération gains - associés - liés - Williamson, organisations - contractuelles, Jensen Meckling - Fama - recherche
analyse cadre (1296)	analyse cadre - environnement évolution - constitue a priori
discipline marché (6209)	discipline marché - existence principaux , influence réseaux
conseil d'administration (7193)	conseil administration - composition - rôle joue - disciplinaire - présence - administrateurs externes - prises contrôle - surveillance - fonction décisions - risque
système gouvernement (6563)	entreprises publiques, mécanisme confiance, gouvernance - développement économie, différents stakeholders - partenaires, comprendre fonctionnement
mesure performance (3534)	cours incidence - françaises - efficacité - effet privatisation - études résultats - mesure performance - rémunération
modèle (2306)	expliquer comportement - équilibre organisationnel - notion liée - démarche - acteurs - hypothèse rationalité - utilité individus - modèle - complexe
contraintes compte (855)	contraintes compte - dimensions cognitives - approches stratégiques

*Synthèse des classes obtenues avec Wordmapper.*

## Summary

Following the recent financial scandals in the United States and Europe, the synthesis of ideas implemented in governing businesses now appears essential if we wish to have a conceptual framework which can be used by stakeholders. This research project consists of using a set of analytical tools from textual data in order to evaluate an analyst's ability to extract knowledge from a group of academic articles. First we test the capacity of representation of the corpus in its totality. Then the study centers on the very concept of "knowledge" used in the theory of governing businesses. We show the convergence and the complementarity of methodological approaches, as well as the ability to extract pertinent knowledge from the texts studied.