

# COMENEGO (Corpus Multilingüe de Economía y Negocios): Analyse comparée basée sur corpus des genres textuels

*Daniel GALLEGO-HERNÁNDEZ*

Université d'Alicante (Espagne)

## 1 RÉSUMÉ

Le but de ce travail est d'analyser deux catégories de textes contenus dans le corpus pilote COMENEGO (Corpus Multilingüe de Economía y Negocios), à savoir les catégories organisationnelle et légale. Nous commençons par présenter brièvement le corpus en question ainsi que les motivations qui nous mènent à analyser ses contenus. Ensuite, nous sélectionnons une série de types de textes ou genres textuels de ces deux catégories afin de procéder à une analyse plus approfondie de chaque catégorie. Puis nous présentons les résultats obtenus qui montrent une certaine hétérogénéité notamment dans la catégorie organisationnelle du corpus. L'approche suivie ainsi que les résultats obtenus peuvent aider non seulement à reclasser les textes du corpus mais aussi à concevoir la plate-forme qui donnera accès aux textes sur internet.

## INTRODUCTION

COMENEGO (Corpus Multilingüe de Economía y Negocios) est pour l'instant un corpus pilote comparable bilingue (français et espagnol) de textes spécialisés dans le domaine de l'économie et des affaires. Ce corpus a été constitué pour remplir le vide existant en matière de corpus dans ces langues et ces domaines. Il fait partie de l'une des étapes du projet COMENEGO qui vise à stabiliser un corpus sur une plate-forme virtuelle afin de le distribuer, entre autres, aux traducteurs professionnels et en formation afin qu'ils s'en servent comme ressource de documentation.

Ce corpus pilote (Gallego-Hernández & Krishnamurthy, 2013) contient un peu moins de 20 millions de mots distribués équitablement en espagnol et en français et classés en sept catégories : commerciale, didactique, légale, organisationnelle, presse, scientifique et technique. Le tableau suivant montre quelques exemples de types de textes contenus dans chaque catégorie ainsi que le type de site web d'où ils ont été extraits :

**Tableau 1 : classement des textes de COMENEGO**

TYPES DE TEXTES	CATÉGORIES
vente de produits bancaires, d'assurances et de produits financiers, pages corporatives, etc. (sites commerciaux)	commerciale
cours en ligne, guides pour consommateurs, investisseurs et clients, etc. (sites commerciaux et informatifs: pages personnelles d'enseignants, d'universités, d'institutions)	didactique
lois, codes, décrets, etc. (sites informatifs: ministères et organismes)	légale
statuts de sociétés, règlements, procès-verbaux, etc. (sites ministériels et commerciaux)	organisationnelle
communiqués de presse, nouvelles, bulletins, etc. (sites commerciaux et journaux)	presse
articles et travaux, etc. (sites informatifs: revues spécialisées)	scientifique

comptes et rapports annuels, résultats, rapports de gestion, analyses techniques, plans de marketing, etc. (sites commerciaux et informatifs)	technique
---	-----------

Le tableau 2, lui, présente la répartition de ces catégories dans les deux langues selon le nombre de fichiers (fichiers), les mots (tokens), le nombre moyen de mots par fichier (ave), les types de mots (types) et le ratio token/type (ratio) :

**Tableau 2 : répartition des catégories de COMENEGO**

CAT	ESPAGNOL					FRANÇAIS				
	FILES	TOKENS	AVE	TYPES	RATIO	FILES	TOKENS	AVE	TYPES	RATIO
COM	5255	1329915	253	35321	37,65	3909	1325544	339	29316	45,21
DID	1491	1276089	856	40641	31,39	1121	1304585	1164	35937	36,30
LEG	211	1342698	6363	23077	58,18	21	1293704	61605	14772	87,57
ORG	429	1337822	3118	29417	45,47	634	1365468	2154	21885	62,39
PRS	2214	1329029	600	37314	35,61	2859	1308418	458	37928	34,49
SCI	99	1311731	13250	34483	38,03	203	1301102	6409	32710	39,77
TEC	351	1188068	3385	40777	29,13	133	1187806	8931	24646	48,19
TOTAL	10050	9115352	907	113100	80,59	8880	9086627	1023	89133	101,94

Les contenus du corpus pilote COMENEGO ont été recueillis intuitivement, à partir de notre expérience personnelle, par le biais de critères externes (texte-genres) mais aussi grâce au travail de Cassany (2004). Par conséquent, les sept catégories dans lesquelles les ressources textuelles sont actuellement classées peuvent ne pas être objectives et risquent d'être inefficaces pour la plate-forme virtuelle (une fois qu'elles y seront transférées) ou pour la compilation de corpus comparables dans d'autres langues.

Nous croyons que l'analyse du discours basée sur corpus peut nous aider à identifier les caractéristiques linguistiques internes de chaque catégorie, ce qui peut nous amener à confirmer, modifier ou rejeter leur validité taxonomique. Pour ce faire, nous avons choisi de nous baser pour l'instant sur le concept de « métadiscours ».

Le concept de « métadiscours » a été initialement introduit par Crismore *et al.* (1993) et Vandepol (1985), et défini par Hyland (2005) comme « the cover term for the self-reflective expressions used to negotiate interactional meanings in a text, assisting the writer (or speaker) to express a viewpoint and engage with readers as members of a particular community ». Sur le plan fonctionnel, le métadiscours permet à l'auteur d'un texte d'établir des liens avec ses lecteurs (les persuader, les dissuader, etc.) sans ajouter vraiment de nouvelles informations. Sur le plan social, le métadiscours peut varier selon les besoins ou les objectifs que poursuivent les différentes communautés lors de la communication. En fait, Hyland (2005 : 87) souligne que l'un des principaux aspects de ce concept est sa dépendance à l'égard du contexte, de sorte qu'il est étroitement lié aux normes et aux attentes de certains genres textuels.

Hyland (2005 : 48-54) distingue deux grandes catégories de métadiscours : des ressources interactives, qui sont utilisées « to organize propositional information in

ways that a projected target audience is likely to find coherent and convincing », et les ressources interactionnelles, qui « involve readers and open opportunities for them to contribute to the discourse by alerting them to the author's perspective towards both propositional information and readers themselves ». Chaque catégorie est classée en plusieurs marqueurs ayant des fonctions différentes : *transition markers*, *frame markers*, *endophoric markers*, *evidentials*, *code glosses* [respectivement, marqueurs de transition, de cadre, intratextuels, extratextuels et de gloses explicatives (traduction de Tassin & Spanghero-Gaillard, 2012)], pour ce qui est des ressources interactives, et *hedges*, *boosters*, *attitude markers*, *self mention*, *engagement markers* [respectivement : modalisateurs d'incertitude, de certitude, d'opinion, des mentions de soi et des implications du lecteur (traduction de Tassin & Spanghero-Gaillard, 2012)], pour ce qui est des ressources interactionnelles.

Dans d'autres travaux (Gallego-Hernández, 2013a, 2013b et Suau-Jiménez, 2013) nous avons constaté qu'il existe effectivement des différences entre plusieurs catégories du corpus: pour ce qui est des ressources interactives, par exemple, les catégories didactique et scientifique sont caractérisées, entre autres choses, par une présence importante des gloses explicatives qui ne sont pas si présentes dans les autres catégories. De même, pour ce qui est des ressources interactionnelles, la catégorie commerciale est caractérisée par une présence massive de marqueurs d'implication du lectorat, qui ne sont pas si abondants dans les autres catégories. Or, nous avons également constaté que quelques catégories pourraient présenter des similitudes du point de vue tant de marqueurs interactionnels que de marqueurs interactifs, comme c'est le cas des catégories organisationnelle et légale, composées par des genres textuels juridiques et administratifs: des statuts de sociétés, des règlements ou des procès-verbaux, parmi autres, pour ce qui est de la catégorie organisationnelle, et des lois, des codes, des décrets, etc., quant à la catégorie légale.

Dans ce sens, le but de ce travail est d'analyser plus en détail ces deux dernières catégories du corpus pilote COMENEGO afin de donner une réponse aux résultats globaux obtenus dans cette première analyse et de proposer de nouveaux arguments pouvant aider à décider s'il faut préparer une autre sorte de classement du corpus.

## **SÉLECTION DU CORPUS ET MÉTHODOLOGIE**

Pour ce faire, nous n'analysons pas la totalité de chaque catégorie dans chaque langue, car lors des travaux précédents nous avons déjà rencontré des difficultés (surtout de temps) à analyser en détail chaque catégorie. Dans ce sens, nous analysons une partie de chaque catégorie dans chaque langue. Chaque subcorpus contient à peu près 10% du total de nombre des mots de chaque catégorie. De même, nous analysons en particulier cinq types de textes ou genres textuels concernant la catégorie organisationnelle dans chaque langue, à savoir des codes de conduite, des statuts sociaux, des codes éthiques, des règlements intérieurs, des avis de réunion ou des convocations à des assemblées des actionnaires ainsi que des codes éthiques. Pour ce qui est de la catégorie légale, nous analysons quatre types de texte en espagnol, à savoir, une partie de code de commerce, des lois, des décrets et des circulaires. En français, nous n'analysons que quelques parties des différents codes (le code général des impôts et le code de la consommation), ainsi que les annexes du code de la consommation, le code étant *a priori* le seul genre textuel contenu dans cette catégorie en français. Le tableau 3 montre le détail du subcorpus que nous avons choisi d'analyser dans ce travail :

**Tableau 3 : subcorpus de catégories organisationnelle et légale**

TYPE DE TEXTE	CATÉGORIE	ES	FR
codes de conduite	ORG	30064	33272
statuts de sociétés	ORG	34661	33093
avis de réunion	ORG	30283	30917
règlements intérieurs	ORG	16777	13248
codes éthiques	ORG	9851	9606
<i>SUBTOTAL</i>	<i>ORG</i>	<i>121636</i>	<i>120136</i>
codes	LEG	30304	119967
lois	LEG	29067	-
décrets	LEG	31380	-
circulaires	LEG	30887	-
<i>SUBTOTAL</i>	<i>LEG</i>	<i>121638</i>	<i>119967</i>

La méthodologie employée est celle que nous avons déjà présentée dans Gallego-Hernández (2103b): d'une part, nous extrayons des concordances avec le concordancier Antconc à partir des mots clés et des groupes de mots clés (identifiés avec la fonction "clusters" à partir de mots clés précédents), et d'autre part, nous validons manuellement les concordances extraites, ce qui nous permet d'identifier les concordances ayant des mots clés potentiellement métadiscursifs mais n'étant pas réellement métadiscursives, et donc à ne pas prendre en compte dans notre analyse. Par exemple, les trois concordances suivantes contiennent trois mots potentiellement métadiscursifs: *extraordinaire*, *bon* et *voir*. Les deux premiers pouvant être classés dans la catégorie des marqueurs d'opinion, le dernier dans celle des marqueurs intratextuels. Leurs contextes invitent à penser qu'il ne s'agit pas de marqueurs métadiscursifs:

...d'une délibération de l'Assemblée Générale *Extraordinaire* en date du 29 avril 2009, le capital social ...  
 ...tion à ces opérations doit être distinct de tout *bon* de commande de bien ou de service. Article L121-...  
 ...ile du consommateur, en ignorant sa demande de *voir* le professionnel quitter les lieux ou de ne pas y...

Cette méthodologie nous a permis aussi de classer des mots clés métadiscursifs dans des catégories différentes. Par exemple, les deux contextes suivants extraits des statuts contiennent le mot clé *article*, mais dans le premier cas, il s'agit d'un marqueur intratextuel. Par contre, le deuxième cas a à voir avec un marqueur extratextuel:

...veillance, conformément aux dispositions de l'*article* 12 des présents statuts, et que celui-ci la refus...  
 ...t y être mis fin dans les conditions prévues à l'*article* 10 de la loi du 26 juillet 1983 précitée. Aéroपो...

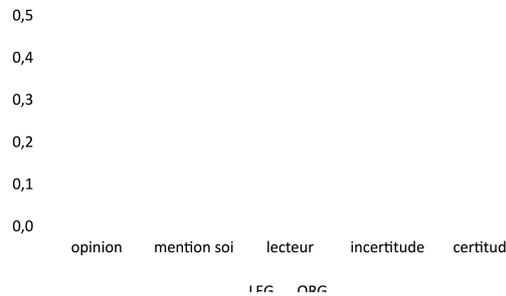
## RÉSULTATS DE L'ANALYSE

Dans cette partie du travail, nous présentons les résultats que nous avons obtenus à partir de notre analyse dans chaque langue. En premier lieu, nous commentons les résultats obtenus en espagnol dans chaque catégorie et analysons en détail les types de textes choisis. En deuxième lieu, nous présentons les résultats obtenus en français et discutons également les détails de chaque genre textuel.

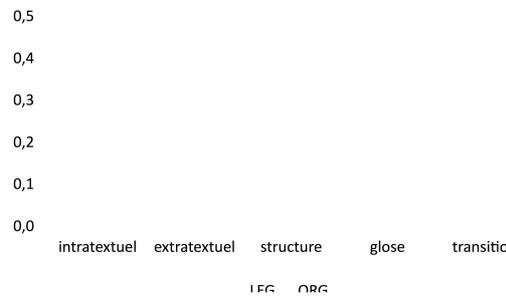
## ANALYSE DU CORPUS ESPAGNOL

Les graphiques 1 et 2 montrent la fréquence (en pourcentage) des ressources interactionnelles et interactives, respectivement, de chaque catégorie du corpus pilote:

**Graphique 1 : marqueurs interactionnels du corpus espagnol**



**Graphique 2 : marqueurs interactifs du corpus espagnol**

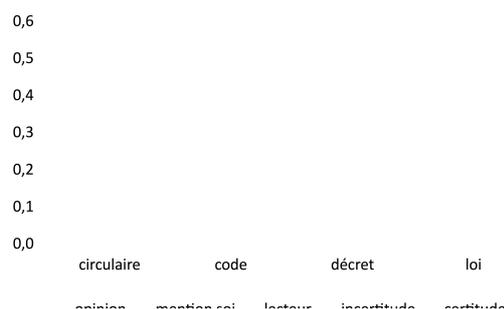


Comme nous l'avions prévu précédemment, il semble exister plus de similitudes que de différences entre ces deux catégories. Les seules différences claires que nous pouvons observer portent sur les mentions de soi, pour ce qui est des ressources interactionnelles, et sur les marqueurs intratextuels, pour ce qui est de la catégorie interactive. Analysons donc chaque type de texte de plus près. Commençons par la catégorie légale.

### Résultats de l'analyse de genres textuels de la catégorie légale en espagnol

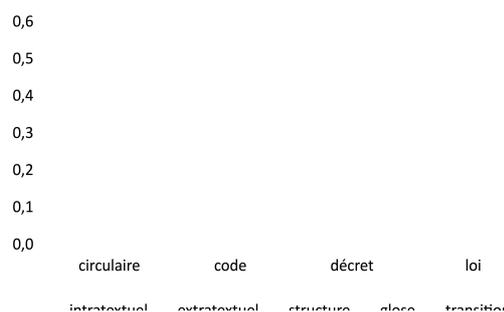
Les graphiques 3 et 4 montrent le schéma métadiscursif des types de textes de la catégorie légale en espagnol. Les résultats de marqueurs tant interactionnels qu'interactifs montrent que les quatre types, à savoir les circulaires, le code de commerce, les décrets et les lois, ne présentent pas de différences métadiscursives excessivement significatives :

**Graphique 3 : marqueurs interactionnels des genres de la catégorie légale en espagnol**



Pour ce qui est des ressources interactionnelles, les marqueurs d'opinion sont les plus fréquents. Il s'agit principalement de mots clés portant sur l'importance et l'obligation des faits tels que *especial++*, *adecuad+*, *mejor++*, *significativ++*, *principalmente* ou *importante+* et *necesari++* ou *obligad++*, respectivement. La catégorie suivante porte sur les marqueurs d'incertitude. Dans ce cas, ce sont les formes du verbe pouvoir *pued\** et *podría\**, ainsi que l'adjectif *posible* qui ont la fréquence la plus élevée.

**Graphique 4 : marqueurs interactifs des genres de la catégorie légale en espagnol**



Par rapport aux marqueurs interactifs, ce sont les marqueurs intratextuels et les marqueurs de transition les plus fréquents. Quant aux gloses explicatives, leur nombre est très réduit dans toutes les catégories, le type de texte *circulaire* étant celui qui a non seulement plus de gloses mais aussi plus de variantes: *en particular*, *es decir*, *denominad++*, *por ejemplo*, *llamad\**, *esto es*, *conocid++ como*, *siguiente modo* et *entendid++ como*. Finalement, pour ce qui est des marqueurs extratextuels, ces textes s'appuient normalement sur d'autres textes de la même catégorie surtout par le biais du mot clé *artículo*, qui sert non seulement de marqueurs intratextuel mais aussi extratextuel comme le montrent ces exemples:

... te del informe anual a que hacen referencia los *artículos* 21, 24 y 25 del Real Decreto 1309/2005, de...  
 ...ara atender a las obligaciones mencionadas en el *artículo* anterior, el Capitán, cuando no tuviere fond...

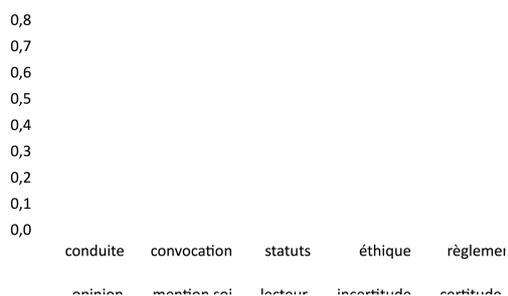
Dans le premier cas, il s'agit d'une concordance extraite d'une circulaire dont le mot clé *artículo* porte sur un autre texte, à savoir, le Real Decreto 1309/2005. Par contre, dans le deuxième cas, extrait du code de commerce espagnol, le mot clé *artículo* fait référence à une partie du même code.

Il est à noter de même le nombre très bas de marqueurs extratextuels du code de commerce, ce qui peut être dû à la date de ce code: 1885.

### Résultats de l'analyse de genres textuels de la catégorie organisationnelle en espagnol

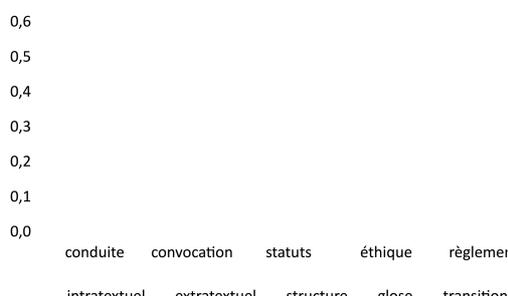
Les graphiques 5 et 6 montrent le schéma métadiscursif des cinq types de textes choisis pour la catégorie organisationnelle en espagnol, à savoir les codes de conduite, les avis de convocation, les statuts, les codes éthiques et les règlements intérieurs des conseils d'administration:

**Graphique 5 : marqueurs interactionnels des genres de la catégorie organisationnelle en espagnol**



Contrairement à ce qui se passait avec les marqueurs interactionnels de la catégorie légale, les résultats montrent quelques différences entre les types de textes, mais aussi quelques similitudes : dans tous les cas, les marqueurs d'opinion sont les plus fréquents, portant également sur l'importance et l'obligation. Or nous observons que le deuxième type de marqueurs le plus fréquent varie selon le type de texte: dans le cas des codes de conduite et des avis de convocation, les mentions de soi (*nuestr++* et des verbes en première personne du pluriel) surpassent clairement les autres marqueurs, dans le cas des codes éthiques, ce sont les modalisateurs d'incertitude, et finalement dans le cas des statuts et des règlements des conseils, les marqueurs interactionnels à l'exception des marqueurs d'opinion sont presque inexistants.

**Graphique 6 : marqueurs interactifs des genres de la catégorie organisationnelle en espagnol**



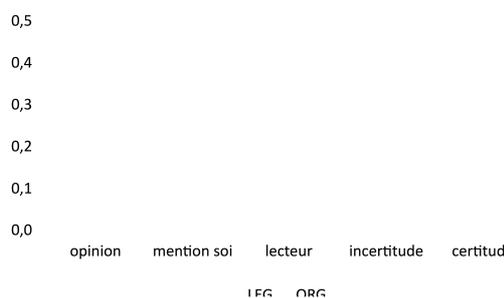
Pour ce qui est de la catégorie interactive, le bilan ne semble pas être toujours le même: les codes de conduite et les statuts ont une proportion similaire des ressources interactives. Par contre, les avis de convocation contiennent plus de mentions de soi, ce qui est normal étant donné la situation communicative de ce type de textes. Quant aux

codes éthiques, ils présentent une structure différente: inexistence des marqueurs intratextuels et une fréquence élevée de marqueurs de structure, de transition et de gloses explicatives. Finalement, les règlements sont caractérisés par une présence de marqueurs intratextuels, de structure et de transition qui surpasse celle des codes de conduite, des avis de convocation et des statuts.

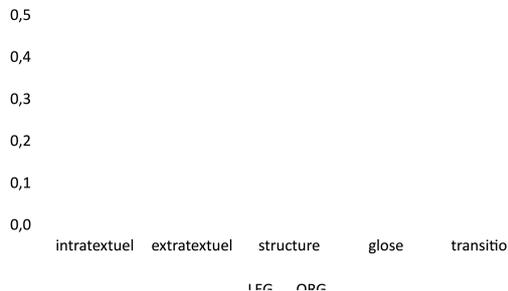
## ANALYSE DU CORPUS FRANÇAIS

Les résultats de l'analyse globale des deux catégories du corpus espagnol montraient notamment deux différences significatives concernant les mentions de soi et les marqueurs intratextuels. Observons dans les graphiques 7 et 8 ce qui se passe avec le corpus français :

**Graphique 7 : marqueurs interactionnels du corpus français**



**Graphique 8 : marqueurs interactifs du corpus français**



Les résultats de l'analyse de ces deux catégories en français montrent non seulement des différences concernant les mentions de soi et les marqueurs intratextuels, mais aussi des différences sur les marqueurs d'opinion. Il est aussi surprenant le nombre élevé de marqueurs d'implication du lectorat concernant la catégorie légale, dont le genre textuel qui la représente, les codes, sont censés être impersonnels. Analysons donc de plus près les types de textes de la catégorie organisationnelle ainsi que la provenance des marqueurs d'implication du lectorat dans la catégorie légale.

### Analyse en détail des codes en français

Les graphiques 7 et 8 montraient non seulement le schéma tant interactif qu'interactionnel de la catégorie légale en français mais aussi celui du seul type de texte contenu dans cette catégorie : les codes. Pour ce qui est des ressources interactives, nous observons une présence élevée des marqueurs intratextuels due surtout au mot clé *article+*, qui contribue aussi à élever la fréquence des marqueurs extratextuels. Quant

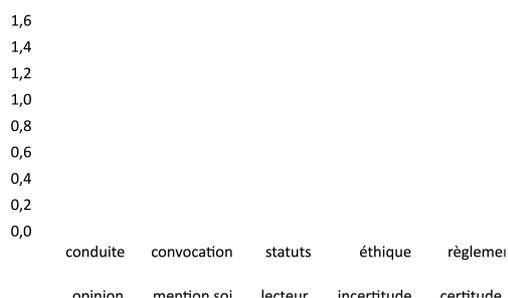
aux ressources interactionnelles, les modalisateurs d'incertitude sont les plus fréquents, représentés par des mots clés tels que *peut*, *aurai\**, *pourrai\**, *serai\** ou *possible+*.

Nous avons également observé avec surprise un nombre élevé de marqueurs d'implication du lectorat dans ce type de texte, représentés par les mots clés *vous* et *votre*. Si nous analysons plus en détail les contextes dans lesquelles ces marqueurs apparaissent, nous apprenons qu'il ne s'agit pas des codes eux-mêmes, mais des annexes aux codes, en particulier, des annexes au code de la consommation. Ce type de texte contient en effet un modèle type relatif à une offre préalable de crédit accessoire à une vente et n'est donc pas du tout caractérisée par l'impersonnalité. Tout au contraire, ce type de texte identifie tant son auteur que son lecteur auquel il s'adresse constamment afin de clarifier les conditions du contrat.

### Résultats de l'analyse de genres textuels de la catégorie organisationnelle en français

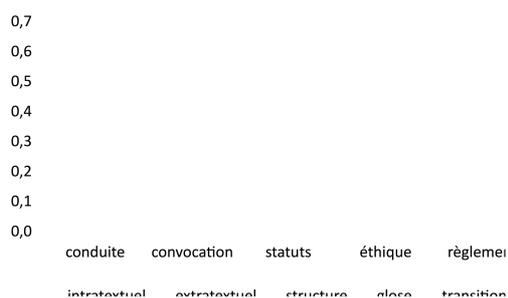
Les graphiques 9 et 10 montrent le schéma métadiscursif des cinq types de textes choisis pour la catégorie organisationnelle en français, à savoir les codes de conduite, les avis de convocation, les statuts, les codes éthiques et les règlements intérieurs des conseils d'administration. Tant les ressources interactionnelles que les interactives montrent non seulement des similitudes mais aussi des différences :

**Graphique 9 : marqueurs interactionnels des genres de la catégorie organisationnelle en français**



Quant aux ressources interactionnelles, les règlements intérieurs et les statuts semblent être les plus proches, surtout pour ce qui est des modalisateurs d'incertitude, représentés essentiellement par le mot clé *peut*, ainsi que des marqueurs d'opinion, représentés également par des mots clés portant sur l'importance ou l'obligation. Les avis de convocation eux ils ont une fréquence très basse de ce type de ressources. Par contre, les codes de conduites et les codes éthiques se caractérisent par une présence de mentions de soi ainsi que de marqueurs d'implication du lectorat.

**Graphique 10 : marqueurs interactifs des genres de la catégorie organisationnelle en français**



Quant aux marqueurs interactifs, nous observons que les codes de conduite et les codes éthiques ont un schéma métadiscursif similaire. Par contre, les avis de convocation sont caractérisés par une présence élevée de marqueurs extratextuels, ce qui est dû surtout au mot clé *article+*, qui est utilisé pour s'appuyer sur des lois ou dans le contexte de modifications des articles des statuts. Les statuts eux, bien qu'ils aient une présence importante de marqueurs extratextuels, ils sont caractérisés surtout par une présence de marqueurs intratextuels. Ceci se passe également avec les règlements intérieurs, bien qu'ils n'utilisent presque pas des références extratextuelles.

## RÉCAPITULATION

Les résultats semblent montrer que la catégorie légale est la plus homogène surtout pour ce qui est de l'espagnol. Par contre, les différents types de textes classés dans la catégorie organisationnelle que nous avons analysés présente quelques différences entre eux tant en espagnol qu'en français. En espagnol, les textes ont presque le même schéma interactionnel caractérisé surtout par de marqueurs d'importance et d'obligation, ainsi que de modalisateurs d'incertitude (mots clé *puede e t posible*). Seul les codes de conduite et les avis de convocation présentent une différence par rapport aux autres textes: une présence importante de mentions de soi. De même, les écarts entre ces textes sont plus importants pour ce qui est des ressources interactives. En français, les différences interactionnelles entre les textes sont causées surtout par l'emploi de mentions de soi et les marqueurs d'implication du lectorat, qui affecte principalement les codes de conduite et les codes éthiques. Quant aux ressources interactives, les différences sont causées par l'emploi de marqueurs extratextuels, affectant surtout les avis de convocation, et par les marqueurs intratextuels, affectant notamment les statuts et les règlements intérieurs des conseils d'administration.

## 2 EN GUISE DE CONCLUSION

Dans ce travail nous avons mené à bien une analyse textuelle du corpus pilote COMENEGO. Nous avons analysé en particulier des genres de texte différents classés dans deux catégories du corpus (la catégorie organisationnelle et la catégorie légale) dans le but de trouver des arguments pouvant aider à décider s'il faut procéder à un nouveau classement de ces textes ou si nous pouvons maintenir ces catégories.

L'objectif global de cette sorte d'analyse est de regrouper dans quelques macro-catégories les différents genres de textes relatifs aux domaines de l'économie et les affaires afin de concevoir une plate-forme virtuelle permettant principalement aux traducteurs travaillant dans ces domaines de filtrer leurs recherches selon leurs besoins.

En ce sens, une réponse claire et bien délimitée à cette question s'avère pour l'instant difficile et semble nécessiter de plus d'analyses basées non seulement sur le métadiscours mais sur d'autres objets. Toutefois, l'hétérogénéité des caractéristiques des genres textuels contenus dans une même catégorie que nous avons pu observer nous invite à considérer la possibilité de permettre aux utilisateurs de la plate-forme de filtrer leurs recherches non seulement par les catégories dans lesquelles sont classés les genres de textes mais aussi par les genres eux-mêmes, de façon à ce que l'utilisateur puisse faire des recherches plus adaptés à ses besoins et aux genres de textes avec lesquels il travail.

Quoi qu'il en soit, nous croyons que le type d'analyse présenté dans ce travail, qui est centré sur les genres eux-mêmes et sur un nombre de mots inférieur à celui de chaque catégorie originelle du corpus pilote peut être non seulement plus abordable mais aussi plus utile dans le sens qu'elle nous permet de connaître plus en détail chaque genre ou type de texte et de découvrir ces textes dont les caractéristiques ne correspondent pas au schéma initial de la catégorie dans laquelle ils sont classés.

## **BIBLIOGRAPHIE**

- Cassany, D. (2004). « Explorando los discursos de las organizaciones ». Dans van Hooff Comajuncosas, A. (ed.). *Textos y discursos de especialidad. El español de los negocios*. Amsterdam/New York: Rodopi, p. 49-70
- Crismore, A., Markkanen, R. & Steffensen, M. (1993). « Metadiscourse in persuasive writing: A study of texts written by American and Finnish university students ». *Written Communication*, 10/1, p. 39-71.
- Gallego-Hernández, D. & R. Krishnamurthy (2013). « COMENEGO (Corpus Multilingüe de Economía y Negocios): design, creation and applications ». *ELR Journal*, 7.
- Gallego-Hernández, D. (2013a). « A comparative corpus-based analysis of metadiscourse in COMENEGO ». *7th edition of the International Contrastive Linguistics Conference (ICLC 7) / 3rd edition of Using Corpora in Contrastive and Translation Studies (UCCTS 3)*, Université de Gand.
- Gallego-Hernández, D. (2013b). « COMENEGO (Corpus Multilingüe de Economía y Negocios): A Metadiscursive Analysis Approach ». *Procedia - Social and Behavioral Sciences*, 95C, p. 146-153.
- Hyland, K. (2005). *Metadiscourse*. Continuum: London/New York.
- Suau-Jiménez, F. (2013). « Propuesta de análisis metadiscursivo interpersonal del corpus COMENEGO con implicaciones para la traducción y la caracterización genérica », *II Seminario sobre traducción e interpretación económica e institucional: docencia, investigación y profesión*, Université d'Alicante.
- Tassin, A. & N. Spanghero-Gaillard (2012). « Le développement des habiletés rédactionnelles à l'université : des marqueurs linguistiques de l'expertise discursive. Former à l'écrit universitaire : un terrain pour la linguistique? » *Colloque co-organisé par le laboratoire MODYCO de l'Université Paris Ouest Nanterre LaDéfense et l'Université de Chicago*, Paris.
- Vande-Kopple, W. J. (1985). « Some exploratory discourse on metadiscourse ». *College Composition and Communication*, 36, p. 82-93.