La formation discursive eurorégionale. Articulation et approche sémantique d'un corpus multilingue

Euroregional discursive formation. Articulation and semantic approach to multilingual corpus

Marie-Hélène Hermand

Édition électronique

URL : <http://journals.openedition.org/corpus/2899>

ISSN : 1765-3126

Éditeur

Bases ; corpus et langage - UMR 6039

Édition imprimée

Date de publication : 15 janvier 2017

ISBN : 1638-9808

ISSN : 1638-9808

Référence électronique

Marie-Hélène Hermand, « La formation discursive eurorégionale. Articulation et approche sémantique d'un corpus multilingue », *Corpus* [En ligne], 17 | 2017, mis en ligne le 15 janvier 2018, consulté le 08 septembre 2020. URL : <http://journals.openedition.org/corpus/2899>

Ce document a été généré automatiquement le 8 septembre 2020.

© Tous droits réservés

La formation discursive eurorégionale. Articulation et approche sémantique d'un corpus multilingue

Euroregional discursive formation. Articulation and semantic approach to multilingual corpus

Marie-Hélène Hermand

- 1 Le corpus eurorégional est construit dans le cadre d'une recherche menée en analyse du discours pour éclairer le phénomène, toujours plus visible et formalisé, de la décentralisation européenne. Le concept flou et émergent d'eurorégions fait l'objet d'attentions soutenues de la part des géographes, des historiens des relations internationales et des juristes (notamment depuis l'élargissement à l'Est de l'Union européenne au milieu des années 2000). La définition proposée récemment en science politique par Thomas Perrin permet de prendre en compte la diversité des eurorégions et leur degré plus ou moins avancé de formalisation juridique :

Les eurorégions peuvent ainsi se définir comme des organisations européennes de coopération transfrontalière et transnationale, plus ou moins structurées, regroupant des autorités territoriales allant en général de la commune à la région ou à leurs équivalents, associées pour la réalisation d'actions et d'objectifs communs, en fonction d'intérêts partagés et dans le cadre de « territoires de projets ». (Perrin 2013 : 7)

- 2 Sous de multiples dénominations avec ou sans majuscule (euregio, eurégio, eurorégion, europaregion, regio...), l'euro-région recouvre des réalités variées allant de l'association de collectivités territoriales à des organismes de droit privé ou public. Souvent autoproclamées au départ, les eurorégions puisent leurs origines dans les démarches de coopération transfrontalières entreprises à la fin de la seconde guerre mondiale pour maintenir la paix en Europe. Parmi les plus anciennes, on peut citer les euro-régions Enschede-Gronau (1958, partagée entre l'Allemagne et les Pays-Bas) ou

Meuse-Rhin (1976, Belgique-Allemagne-Pays-Bas). De nombreuses autres sont apparues au cours des décennies suivantes : Benego (1980, Belgique-Pays-Bas), Neisse-Nisa-Nysa (1991, République tchèque-Allemagne-Pologne), Barents (1993, Finlande-Norvège-Russie-Suède), Tirol-Südtirol/Alto Adige-Trentino (1998, Autriche-Italie), EuroBalkans (2002, Bulgarie-Serbie-Macédoine), Pyrénées-Méditerranée (2004, France-Espagne), Alpes-Méditerranée (2007, France-Italie), Galicia-Norte de Portugal (2008, Espagne-Portugal), Aquitaine-Euskadi (2011, France-Espagne)...

- 3 Sur 163 régions frontalières dénombrées dans l'Union européenne¹, les recensements font état de plus d'une centaine d'eurorégions (Oliveras González *et al.* 2010 : 29) qui récoltent les fruits du travail d'influence mené au niveau européen par l'Assemblée des régions frontalières européennes (ARFE) depuis 1971 (Alliès 2011). Souligné en 1993 dans la Déclaration de Vienne², le rôle des eurorégions pour la stabilité démocratique de l'espace européen est affirmé et défendu par la Commission européenne qui encourage, depuis la mise en œuvre du programme Interreg (1990), le financement de projets transfrontaliers. Les années quatre-vingt-dix marquent donc à la fois une étape d'encouragement explicite de la politique de coopération transfrontalière, l'autonomisation de la démarche transfrontalière et la multiplication des eurorégions. Au milieu des années 2000, la prolifération des eurorégions s'inscrit en outre dans les logiques observées de régionalisation en Europe de l'Ouest (Cole et Palmer 2009) et de l'Est (Wassenberg 2010).
- 4 La réalité eurorégionale étant très diversifiée, il est possible de l'identifier à partir de critères fonctionnels et structurels³ :
 - du point de vue fonctionnel, les eurorégions peuvent être « soit des associations d'autorités locales et régionales, parfois avec une assemblée parlementaire, ou bien des associations transfrontalières avec un secrétariat permanent et une équipe technique et administrative disposant de ressources propres » ;
 - du point de vue juridique, les eurorégions peuvent être « soit des structures de droit public basées sur des accords interétatiques avec la participation des collectivités territoriales, soit des structures de droit privé, basées sur des associations à but non-lucratif ou des fondations d'un côté ou de l'autre de la frontière en accord avec les juridictions nationales en vigueur ».
- 5 Le développement des eurorégions s'accompagne de nombreux discours en ligne issus des mondes institutionnel, économique et médiatique. Réunis autour du « foyer thématique » eurorégional, ces discours permettent, comme le recommande Maingueneau (2012), de construire et de confronter des sous-corpus homogènes pour l'analyse. L'objectif final de la recherche consiste, d'une part, à caractériser la communication de la régionalisation européenne à un moment où elle passe d'un stade empirique et informel à un stade professionnel et, d'autre part, à analyser les imaginaires politiques mobilisés à l'occasion de la diffusion du modèle de gouvernance transfrontalière.
- 6 Multilingue et multigenre, le corpus eurorégional nécessite tout d'abord une réflexion théorique qui tienne compte des discours et des genres afin de proposer une structuration qui permette d'appréhender leur diversité. Il pose ensuite une question méthodologique relative à la formulation de requêtes et au traitement de résultats multilingues. Enfin, il oblige à concevoir un cadre d'analyse inscrit dans la dynamique d'une sémantique de l'interprétation. Après ces considérations descriptives et critiques, nous appliquerons notre démarche à la *mobilité*, mot-pivot inscrit dans la mémoire du discours eurorégional.

1. Conception critique du corpus

1.1 Mécanisme de sélection des contenus

- 7 Le point de vue à l'origine de la constitution du corpus doit nous permettre d'accéder à des éléments caractéristiques du discours eurorégional indépendamment des lieux (pays) ou des langues d'expression. Exprimé depuis des positions discursives variées en Europe et caractérisé par une importante dissémination sur le web, le corpus a fait l'objet d'une récolte sur une durée d'environ six mois, en vue de constituer un échantillon significatif d'usages discursifs et sémiotiques.
- 8 Regroupant désormais plus de 600 textes différents recueillis en six langues (français, italien, espagnol, anglais, allemand, néerlandais), le corpus compte un peu plus de 600 000 mots⁴ et considère 42 eurorégions sur la centaine recensée. La méthode de collecte du corpus, basée sur le *crawling* thématique (Yapomo 2013), a consisté à identifier des terminologies pertinentes par rapport au thème des eurorégions. Trois types de recherches ont été nécessaires pour répertorier les textes : à partir des noms officiels d'eurorégions, à partir de leurs désignations courantes ou siglées et à partir d'expressions-pivots susceptibles d'évoquer l'univers eurorégional (ex : coopération transfrontalière, régionalisation européenne, cohésion transfrontalière, *cluster* transfrontalier...). Si les sites institutionnels publiés par les eurorégions et les prises de positions des acteurs économiques implantés en régions frontalières ont émergé très rapidement, la découverte de sites dédiés à des activités ou des projets pilotés au niveau eurorégional, d'articles de presse consacrés aux eurorégions et de discours critiques à leur égard a nécessité davantage de persévérance.
- 9 Notre intuition initiale nous portait à attendre une plus forte représentativité de discours disponibles en français et en allemand. La concentration historique des eurorégions le long des frontières de pays ou régions francophones (France, Belgique romane, Luxembourg, Suisse romande, Vallée d'Aoste en Italie) et germanophones (Allemagne, Autriche, est de la Belgique, Luxembourg, Suisse alémanique, Tyrol du Sud en Italie) a en effet très vite offert la possibilité de s'intéresser à de nombreuses eurorégions. Au fil de la collecte, des liens renvoyant vers des textes disponibles en d'autres langues ont permis de voyager dans l'espace européen pour élargir la sélection à des zones où se sont développées de nombreuses eurorégions depuis les années 2000 (notamment en Europe du Sud et en Europe centrale).

1.2 Création de métadonnées

- 10 En raison de la volatilité des sites web et afin de disposer à tout moment de l'identité plurisémiotique du corpus eurorégional, nous avons procédé tout au long de la récolte au figement systématique du corpus de référence dans un bloc-notes virtuel⁵. L'ajout de métadonnées sur les textes et leur description dans une base de données permettent ensuite de nous forger une connaissance approfondie du corpus, condition indispensable à la manipulation ultérieure des données à l'aide d'une quelconque solution logicielle.
- 11 Une première série de métadonnées concerne notamment, pour chacun des textes, l'attribution d'un identifiant normé, l'indication du format informatique disponible à la

consultation, la langue de travail, le site web éditeur et la rubrique d'appartenance, une brève description du contenu textuel, la date de publication lorsqu'elle est disponible, le nom de l'eurorégion émettrice ou concernée et ses pays d'appartenance. Une seconde série de métadonnées concerne les aspects sémiotiques du corpus : la présence d'éléments autres que textuels (images, graphiques, cartes...) et leur format, leur description brève, la mention de *copyrights*, les types d'hyperliens et de destinations vers lesquelles ils pointent. Enfin, une table de liaison, créée en parallèle, permet de se reporter en permanence à la liste de toutes les eurorégions considérées dans le corpus et à leur statut, à leur date de création, à leurs langues officielles et langues de communication ainsi qu'à leur zone géographique d'appartenance. Disposer de ces données permet d'anticiper l'exportation de résultats descriptifs. Nous veillons ainsi à la mise en garde de François Rastier, qui souligne le risque d'une précipitation à utiliser des logiciels de traitement de corpus avant d'avoir cerné préalablement les besoins de l'analyse :

Les sorties logicielles sont souvent ininterprétables : par exemple le résultat graphique d'une analyse factorielle n'obéit à aucune métrique simple, et il faut bien connaître le corpus pour pouvoir l'interpréter. Ce n'est donc pas l'instrumentation qui permet l'interprétation mais l'inverse. (Rastier 2011 : 51)

1.3 Représentativité : couverture géographique, institutionnelle et linguistique

1.3.1 Couverture géographique

- 12 À la fois parce que de nombreuses eurorégions sont concentrées géographiquement et parce que nous avons dû nous limiter aux langues dont la lecture nous est accessible (français, italien, espagnol, anglais, allemand, néerlandais), les zones géographiques dans lesquelles se situent les eurorégions considérées dans le corpus se répartissent de la manière suivante : 39 % des textes du corpus concernent des eurorégions situées en Europe du Nord-Ouest, 21 % dans la zone Alpes-Danube et 17 % en Europe du Sud-Ouest. Viennent ensuite 8 % de textes qui concernent l'Europe centrale et orientale et 7 % qui concernent l'Europe du Nord. Enfin, 8 % des textes concernent les eurorégions en général, mais aucune en particulier (Graphique 1) :

Graphique 1. répartition des textes du corpus eurorégional en fonction des zones géographiques représentées

1.3.2 Couverture institutionnelle

- 13 Pour ce qui concerne la couverture institutionnelle (Graphique 2), la majorité des textes du corpus (52 %) concernent des euro-régions non dotées de statut juridique⁶ ; 29 % des textes concernent des eurorégions dotées d'un statut juridique⁷, 11 % des textes concernent des eurorégions en négociation pour obtenir ce statut et 1 % des textes concerne des eurorégions en cours d'obtention du statut⁸.

Graphique 2. répartition des textes de la formation discursive eurorégionale en fonction du statut juridique des eurorégions considérées dans les textes

* : 7 % des textes concernent les eurorégions en général et aucune eurorégion en particulier.

1.3.3 Couverture linguistique

14 Pour ce qui concerne la mobilisation des langues sur les sites web officiels des eurorégions, la comparaison des langues officielles des eurorégions et de leurs langues de communication permet de dégager différents profils et pratiques d'exposition sur le web (Graphique 3) :

- 69 % des eurorégions considérées dans le corpus utilisent toutes leurs langues officielles sur leur site institutionnel, rendant ainsi leur site accessible en autant de versions que de langues officielles ;
- au sein de cette majorité, 21 % d'eurorégions⁹ utilisent toutes leurs langues officielles (même lorsque l'une d'entre elles est minoritaire, comme le ladin dans le cas de l'*Europaregion Tirol-Südtirol-Trentino*) et y ajoutent d'autres langues (l'anglais systématiquement, et l'occitan en plus pour l'*Eurorégion Aquitaine-Euskadi*¹⁰) ;
- 10 % des eurorégions n'utilisent aucune de leurs langues officielles et privilégient l'anglais pour communiquer : c'est le cas de l'*Euregio Barents* (dont les langues officielles sont le finnois, le norvégien, le russe et le suédois), de l'*Euroregion Baltic* (danois, lituanien, polonais, russe, suédois), de l'*Euroregion EuroBalkans* (bulgare, macédonien, serbe) et de l'eurorégion *Helsinki-Tallinn Euregio* (estonien, finnois) ;
- 7 % des eurorégions utilisent partiellement leurs langues officielles ;
- 2 % des eurorégions utilisent partiellement leurs langues officielles et recourent à d'autres langues¹¹ ;
- 12 % des eurorégions considérées dans le corpus ne disposent pas de site institutionnel.

Graphique 3. répartition des profils d'eurorégions en fonction des langues utilisées sur leur site officiel

2. Trois problèmes principaux

2.1 Faire de l'hétérogénéité un ressort pour l'analyse : proposition de structuration du corpus

- 15 La question qui guide notre structuration du corpus est de savoir si et comment l'on peut regrouper les textes autour du thème eurorégional afin d'observer des régularités au-delà des variations géographico-linguistiques. L'objectif consiste à dégager, s'ils existent, des traits définitoires d'une « culture discursive » (Claudel *et al.* 2013 : 35) eurorégionale, au sens des manifestations discursives circulant sur l'objet « eurorégion » et sur les discours tenus sur cet objet. Pour y parvenir, nous précisons notre perception des faces textuelle et sociale, inséparables et solidaires, de la « pratique discursive » (Maingueneau 1987 : 39) eurorégionale.
- 16 La face textuelle correspond à la définition d'une formation discursive (désormais FD) dans l'acception large d'un « ensemble d'énoncés socio-historiquement circonscrit que l'on peut rapporter à une identité énonciative » (Charaudeau et Maingueneau 2002 : 271) et en adéquation avec le postulat selon lequel le dicible eurorégional forme un système et délimite une identité. Envisagé au sens de Foucault (2008 [1969] : 57) qui voit dans la FD un système de règles, le discours eurorégional présente cependant la particularité de pouvoir surgir en dehors d'un statut contraint (par exemple lorsque l'eurorégion ne possède aucun statut juridique). La définition de Pêcheux offre quant à elle, selon Mayaffre (2004), une occasion particulière de pointer, surtout dans les discours politiques, l'ancrage idéologique du locuteur. Ce trait doit également être

nuancé par les résultats de nos premières analyses (Hermand 2014) : une prise de distance à l'égard de la référence nationale (notamment par la très faible fréquence des noms propres de pays et des noms communs « pays » et « nation ») et une nette conformité du discours eurorégional au programme politique des institutions européennes (par la neutralisation de la polémique) plaident par exemple plutôt en faveur du pragmatisme que d'une véritable prise de position idéologique. Enfin, des traces de discours communautaires (recyclages lexicaux, reprises de mots d'ordre) observables dans le corpus eurorégional renvoient surtout à une conception interdiscursive de la FD et incitent à accorder une importance particulière aux relations de dépendances entre discours :

Une FD n'est pas un espace structural clos, puisqu'elle est constitutivement « envahie » par des éléments provenant d'ailleurs (*i. e.* d'autres FD) qui se répètent en elle, en lui fournissant ses évidences discursives fondamentales (par exemple sous forme de « préconstruits » et de « discours transverses »). (Pêcheux *in* Maldidier 1990 : 297)

- 17 La FD eurorégionale constitue un champ d'observation spécifique car elle produit des discours dans des conditions comparables (sites institutionnels, sites d'entreprises commerciales, sites d'organisations professionnelles), les fait circuler sous les mêmes genres discursifs (message d'accueil, lettre d'information, guides pédagogiques, audits, études prospectives...) et partage des rituels (cérémonies d'inauguration, prise en charge de traditions locales...). En suivant Maingueneau (1987 : 39), nous appréhendons ce concept de formation discursive comme inséparable des communautés discursives qui la produisent et la diffusent. La notion de communauté discursive, « solidaire » (*ibid.*) de la FD et entendue au sens d'« institution qui reçoit sa cohérence de ses pratiques discursives » (Beacco 2004), correspond à la face sociale de la pratique discursive eurorégionale. Elle s'avère particulièrement utile, à condition d'être affinée, pour prendre en compte le bi-/plurilinguisme des locuteurs. Ainsi la notion de « communauté discursive translangagière » (Beacco 1992) permet-elle de prendre en compte « une communauté de communication particulière, dans laquelle les échanges s'effectuent ordinairement en plusieurs langues naturelles » (Charaudeau et Maingueneau 2002 : 107).
- 18 Si l'on considère les différents constituants du corpus euro-régional comme des communautés discursives « nucléaires », il devient possible d'analyser leur tentative d'exister à travers la thématique eurorégionale. À la fois « articulées entre elles mais distinctes », ces communautés « organisent le partage d'un objet du discours commun » (Beacco 1995) : en l'occurrence l'euro-région. Les échanges s'y effectuent en plusieurs langues de travail et avec des locuteurs souvent plurilingues. La notion de communauté discursive translangagière permet surtout de proposer une « typologie situationnelle » fondée sur des « positionnements discursifs » pris dans l'acception d'une « identité énonciative forte » ou d'un « lieu de production discursive bien spécifié » (Charaudeau et Maingueneau 2002 : 453).
- 19 Bien que le graphique 4 soit réducteur de la dispersion et de l'« inconsistance » (Courtine et Marandin 1981 : 24) de notre FD, il permet de distinguer les divers positionnements énonciatifs qui partagent la reconnaissance de l'objet *eurorégion* : les discours de prescription (produits par les institutions européennes), les discours d'appropriation (produits par les institutions eurorégionales et académiques et par les acteurs économiques), les discours de médiation (produits par la presse) et les discours de contestation (produits par les eurosyndicalistes).

Graphique 4. articulation de la formation discursive eurorégionale

- 20 Les discours de prescription regroupent les textes produits par les institutions européennes (surtout la Commission) lorsqu'elles délimitent ce qui peut et doit être dit au sujet des eurorégions et arguent de la proximité de ces entités émergentes avec les citoyens.
- 21 Les discours d'appropriation, qui forment le noyau dur du corpus (53 % des textes à eux seuls), sont composés de discours institutionnels et économiques. D'une part, les discours institutionnels sont entendus comme des « discours produits officiellement par un énonciateur singulier ou collectif qui occupe une position administrativement ou juridiquement inscrite dans l'appareil d'État, qu'il soit fonctionnaire ou représentant politique » (Oger et Ollivier-Yaniv 2003). Il s'agit de textes produits par les structures eurorégionales lorsqu'elles se présentent et se mettent en valeur (par la voix des Présidents d'eurorégions par exemple) et de textes produits par des représentants d'universités qui recrutent des étudiants ou valorisent des partenariats scientifiques dans leur bassin géographique transfrontalier. D'autre part, les discours économiques recouvrent les positionnements :
- de *clusters* économiques transfrontaliers, grappes d'entreprises dédiées à la réalisation de produits et services les identifiant, regroupées le long des frontières par souci d'efficacité économique et dont les discours s'inscrivent à la fois dans une optique de spécialisation et d'innovation pour répondre aux impératifs de reconfiguration économique des zones frontalières ;
 - d'organisations professionnelles en eurorégions, telles que les Chambres de commerce et d'industrie ou les Chambres de métiers et d'artisanat, dont les discours fédèrent une communauté d'acteurs économiques dans une logique de clubs privés et garantissent une offre de services privilégiés en contrepartie d'une adhésion contractuelle ;
 - de sociétés d'audit, dont les discours s'apparentent à ceux d'experts qui posent une évaluation sur la santé (économique, scientifique, technologique) eurorégionale et instaurent par là une voix d'autorité.

- 22 Les discours de contestation, difficiles à réunir (notamment pour des raisons d'accessibilité) et à traiter (pour des raisons de formats), relèvent du positionnement eurosyndical centralisé (textes émis par la CES, Confédération européenne des syndicats) et décentralisé (textes émis par des CSIR, Conseils syndicaux interrégionaux¹²).
- 23 Enfin les discours de médiation, à la fois producteurs et récepteurs d'opinions sur les eurorégions, sont situés à la croisée des autres flux textuels et considérés comme un « lieu de passage » (Moirand et Beacco 1995 : 50) des connaissances sur les eurorégions. Il s'agit de textes produits par des médias régionaux, nationaux et internationaux qui visent à faire savoir ou à faire comprendre l'actualité eurorégionale aux publics médiatiques.
- 24 S'exprimant par le biais de discours multilingues en contact (cérémonies d'inaugurations en plusieurs langues, journaux d'informations transfrontalières alternant les langues...), les communautés discursives eurorégionales nécessitent aussi de préciser les manières de faire, les genres discursifs et les conditions de production. L'attribution de genre, en tant que « dispositif de parole socio-historiquement institué » (Charaudeau et Maingueneau 2002 : 593), résulte de l'observation d'indices variés : l'éditeur, la rubrique, l'auteur, le contenu. Les genres, qui ne sont pas au centre de l'analyse d'une FD, ne sont en outre pas garants d'homogénéité (Cislaru et Sitri 2012 : 62) et leur recensement suggère de porter attention au voisinage seulement apparent de certains d'entre eux. À titre d'exemple, une étude prospective émise par des institutions eurorégionales, par une société d'audit, par un *cluster* économique ou par une organisation professionnelle n'envisage pas forcément la question de la stratégie d'avenir en eurorégion sous le même angle, chaque énonciateur poursuivant ses propres intérêts. Comme d'autres genres apparemment redondants dans le corpus (déclarations d'intentions, rapports d'activités...), elle ne peut donc être considérée *a priori* comme un genre transdiscursif.

2.2 Interroger le corpus dans les six langues : mise en place d'une procédure de traitement

- 25 Au-delà des types ou des genres de discours, il s'agit d'opérer des allers-retours entre les marques linguistiques – les « observables » (Moirand 2007) – et leur portée pragmatique. L'objectif consiste à exploiter le caractère réflexif du corpus (Mayaffre 2002) pour faire émerger un réseau sémantique eurorégional cohérent dont l'analyse pourra ensuite être articulée aux dimensions sémiotique, rhétorique et culturelle du corpus. Dans une optique qui met l'instrumentation au service de l'analyse, nous considérons que ce ne sont pas les données qui doivent s'adapter à la méthode d'analyse mais l'analyse qui doit s'ajuster à la réalité des données. Nous avons dès lors élaboré un modèle « agile » (Vickoff 2003) d'analyse du corpus eurorégional, dans le sens où la souplesse et l'itération du développement découlent d'un besoin affiné tout au long de la démarche de recherche¹³.
- 26 Considérant que notre objectif consiste à fournir une description précise du discours eurorégional disséminé sur le web et à en formuler une interprétation sémantique sur la base d'hypothèses qualitatives, que notre contexte est l'ensemble du millefeuille énonciatif précédemment décrit et que des résultats quantitatifs intermédiaires doivent nourrir l'analyse, nous précisons notre besoin de récolte de données avant tout

choix de logiciel de traitement. Pour baliser l'analyse qualitative, il faut par exemple disposer de résultats sur la productivité lexicologique au sein du corpus, sur les dénominations des eurorégions, sur le fonctionnement énonciatif et sur les segments répétés. La récolte progressive de ces données nous engage dans un processus de va-et-vient entre le corpus (ses positionnements discursifs, son hétérogénéité générique, ses partitions linguistiques) et les résultats extraits à partir d'un traitement informatique adéquat.

- 27 Pour ce qui concerne l'outillage mobilisé pour le traitement du corpus, le logiciel *TreeTagger*¹⁴ fournit l'étiquetage morphosyntaxique multilingue et la lemmatisation de chaque partition linguistique du corpus. Des programmes développés sur mesure¹⁵ optimisent ensuite les requêtes multilingues, créent une base de données des résultats et automatisent leur exportation dans des tableurs. La plate-forme d'analyse ainsi constituée articule un ensemble d'outils éprouvés¹⁶ en vue d'obtenir des résultats statistiques utilisables et d'archiver l'historique des requêtes. Ils permettent surtout un gain de temps substantiel pour tester et grouper des requêtes sur l'ensemble du corpus multilingue, puis visualiser et exploiter les résultats obtenus selon de nombreuses possibilités de tris.

2.3 Contextualiser l'analyse sémantique : élaboration d'un cadre de requêtes

- 28 Selon l'angle d'analyse qualitative choisi (construction du référent eurorégional, traces d'interdiscours, tensions de la FD...), des requêtes automatiques sont effectuées sur des mots-pivots : elles ont notamment pour but de localiser des notions et d'estimer l'ampleur de leur diffusion dans le corpus. Pour limiter les effets liés aux particularités de chaque langue (par exemple la composition des mots en langues allemande et néerlandaise, ou encore les flexions verbales pour marquer la personne et le nombre en italien et en espagnol), nous effectuons des recherches automatiques par regroupements de formes pour obtenir, d'une part, toutes les occurrences non lemmatisées susceptibles d'être en relation avec la notion considérée et, d'autre part, la liste des lemmes associés à ces occurrences. Après un tri manuel des lemmes non pertinents, nous lançons un calcul automatique de pourcentages de répartition des occurrences pertinentes dans les différentes partitions du corpus.
- 29 Les résultats obtenus sont présentés de préférence par sous-corpus émetteur (institutions eurorégionales, européennes et universitaires, acteurs économiques, médias, syndicats) et/ou en fonction des genres discursifs. À partir de la description, de la mise en relation et de l'interprétation de ces différents éléments, l'accès à la culture discursive eurorégionale peut être envisagé par le biais d'une sémantique de l'interprétation. Les résultats sont rarement présentés par langues pour l'analyse (même si c'est techniquement possible) car il ne s'agit pas de mettre en rapport différentes langues, comme le fait traditionnellement la linguistique contrastive, mais les manifestations du thème eurorégional selon des énonciateurs ou des genres discursifs dans des communautés différentes en vue de « décrire et d'interpréter les régularités et les variabilités discursives » (von Münchow 2010). En outre, « lorsque la comparaison est effectuée entre des communautés ethnolinguistiques occidentales, l'impact de la langue est extrêmement réduit » (*ibid.*).

- 30 Le retour à l'observation manuelle du corpus constitue enfin une condition indispensable de traitement : il s'agit de répertorier des constructions, de décrire des usages discursifs et d'orienter leur analyse en fonction de l'inscription du discours dans son contexte social. L'intérêt consiste à faire émerger des observables et à les rapporter aux types et aux genres discursifs présents dans le corpus. À leur tour « générateurs d'hypothèses ou destructeurs d'évidences » (Rastier 2011 : 51), ces observables en contexte élargi ne sont pas seulement envisagés en fonction de leur fréquence : fréquente ou rare, une donnée quantitative peut être intéressante et doit être rapportée au résultat d'une interprétation.

3. Application à la *mobilité*, observable récurrent du corpus

3.1 Cadre de requêtes

- 31 Nous nous interrogeons sur la présence et le parcours d'une « formule » (Krieg-Planque 2009) apparue dans le *Traité de Rome* dès 1957¹⁷ et en circulation dans d'autres discours marqués par l'influence communautaire (Scholz et Angermuller 2013) : la formule de la *mobilité*. Nous faisons l'hypothèse que cette formule se retrouve dans notre corpus et contribue à déployer un argumentaire de séduction en vue de faciliter l'adhésion du citoyen au projet eurorégional. Mais on ne peut saisir le sens de cette formule que si on la considère dans son « contexte situationnel » (Charaudeau 2009), lequel nous dit qui l'emploie, à l'adresse de qui, dans quel dispositif communicationnel, avec quelle finalité et en rapport avec quelle tradition.
- 32 Le cadre de requêtes s'appuie sur l'analyseur morphosyntaxique *TreeTagger* et est construit comme suit :
- un premier programme *Perl* balaie automatiquement tout le corpus pour lister dans les six langues toutes les lemmes comprenant *mobil* (en français, italien, anglais, allemand, néerlandais) et *movil* (en espagnol) ; un nettoyage manuel élimine ensuite les lemmes sans rapport avec le champ de la mobilité ;
 - un second programme *Perl* localise les lemmes en adéquation avec le champ de la mobilité, et obtient :
 - en français : Mobilité | mobilité | mobile
 - en italien : mobile | mobilità
 - en espagnol : movilidad
 - en anglais : Mobile | Mobility | mobile | mobility
 - en allemand : Mobilität
 - en néerlandais : Mobility | arbeidsmobiliteit | mobiliteit
 - Il identifie en outre les cotextes élargis à gauche et à droite de chaque occurrence et permet le retour au texte intégral par un identifiant unique issu des métadonnées ;
 - un troisième programme *Perl* calcule les pourcentages de diffusion de ces occurrences dans l'ensemble des textes du corpus, sous-corpus émetteur par sous-corpus émetteur.

3.2 Résultats obtenus

- 33 En termes de fréquences, les lemmes comprenant le radical *mobil** se placent en bonne position parmi l'ensemble des lemmes repérés dans les différentes partitions linguistiques (Tableau 1) :

Tableau 1. fréquence des lemmes comprenant *mobil** en fonction des partitions linguistiques du corpus eurorégional

Partition linguistique considérée	Rang de fréquence des lemmes comprenant <i>mobil*</i> / nombre total de lemmes
partition francophone	- mobilité : 415 / 9581 - Mobilité : 2621 / 9581 - mobile : 5686 / 9581
partition hispanophone	- movilidad : 270 / 3855
partition italophone	- mobilità : 1210 / 8444 - mobile : 1808 / 8444
partition anglophone	- mobility : 824 / 7236 - mobile : 2503 / 7236 - Mobile : 4565 / 7236 - Mobility : 4566 / 7236 - mobility : 6266 / 7236
partition germanophone	- Mobilität : 1589 / 9883
partition néerlandophone	- mobiliteit : 586 / 6264 - Mobility : 1982 / 6264 - arbeidsmobiliteit : 3843 / 6264

- 34 Il est ensuite intéressant de noter que les occurrences sont inégalement réparties dans le corpus et traduisent une préoccupation très marquée du contre-discours syndical (Tableau 2) :

Tableau 2. mobilisation de la notion de *mobilité* par sous-corpus émetteur dans le corpus eurorégional

Sous-corpus émetteur	Pourcentage de textes comprenant le mot-forme <i>mobilité</i>
institutions eurorégionales	15 %
institutions universitaires	15 %
institutions européennes	9 %
acteurs économiques	9 %
médias	7 %
syndicats	100 %

35 Dans les discours institutionnels, économiques et médiatiques, des constructions-types récurrentes observables dans les différentes langues font de la mobilité :

- un élément d'énumération noyé parmi d'autres préoccupations du discours : *mobilité* est listée avec *économie*, *démographie*, *culture*, *énergie*, *formation*, *services sanitaires et sociaux* ;
- l'objet d'une valorisation par des adjectifs d'évaluation positive ou par la modalisation : c'est *une nouvelle mobilité*, *la mobilité durable*, *la libre mobilité*, *la mobilité importante* qu'il faut *améliorer*, *favoriser*, *stimuler*, *accroître*, *renforcer* ;
- un moyen au service du développement professionnel et de la croissance économique : *mobilité professionnelle*, *mobilité à but économique*, *mobilité du marché du travail transfrontalier*, *mobilité des travailleurs*, *mobilité en première partie de carrière* ;
- un élément de la dénomination de dispositifs : *Schéma stratégique de mobilité transfrontalière*, *Chèque Mobilité*, *Mobilité Chèque EuroCampus*, *Commission Mobilité*, *Schéma de mobilité*, *les aides à la mobilité* ;
- une problématique observable dans les questionnements : *questions de mobilité*, *comment donner le goût de la mobilité*, *comment éduquer à la mobilité* ;
- une composante des professions intellectuelles dans les constructions d'appartenance : *mobilité des enseignants*, *mobilité des chercheurs*, *mobilité des étudiants*, *mobilité étudiante*, *mobilité des connaissances*, *mobilité des professions intellectuelles* ;
- un besoin spécifique de publics délimités : *mobilité des patients*, *mobilité des handicapés*, *mobilité des frontaliers*, *mobilité des apprentis*.

36 Une représentation de ces variables contextuelles permet de visualiser la dimension dominante de la *mobilité* en fonction des sous-corpus émetteurs (Radar 1) :

Radar 1. représentation des variables contextuelles du terme *mobilité* (fréquences relatives) en fonction des sous-corpus émetteurs du corpus eurorégional

sous-corpus institutions eurorégionales

sous-corpus universités

- 37 À l'opposé de la vision positive de ces discours dominants, les occurrences de la *mobilité* relevées dans les discours syndicaux sont entourées des adjectifs *contrainte*, *entravée*, *réduite*, *affectée* et des substantifs *obstacles* et *problèmes*.

3.3 Contextualisation de l'analyse : un imaginaire de mobilité collective

- 38 Présentée comme un droit ouvert à tous à des fins essentiellement professionnelles¹⁸, la *mobilité* vantée dans les discours communautaires dédiés aux eurorégions désigne l'idéal d'une quête à poursuivre depuis l'instauration de la liberté de circulation par le *Traité de Rome*. C'est cet idéal d'une mobilité choisie, à finalité économique ou intellectuelle, qui domine dans les discours d'appropriation et de médiation. Si elle est

déjà commencée et à poursuivre, cette mobilité ne s'adresse pourtant pas à tout le monde. Elle concerne surtout les professions intellectuelles : les *cadres*, les *chercheurs*, les *étudiants*, les *enseignants*, les *artistes*, le *public*, les *œuvres* et la *connaissance* sont les termes récurrents du cotexte. En outre, la mobilité évoquée par les institutions eurorégionales est de préférence *verte*¹⁹, *intelligente* et *forte*²⁰, confortant par là l'image d'eurorégions élitistes.

- 39 Dans un second temps (par le biais du syntagme *mais aussi* dans les énumérations), la mobilité concerne des catégories socio-professionnelles moins aisées ou moins diplômées telles que les *apprentis* et les *personnes en formation professionnelle* en vue de garantir leur insertion professionnelle rapide. Elle peut encore concerner des publics fragiles tels que des *patients* en quête de soins. Cette déclinaison des destinataires de la mobilité, par laquelle le discours tend à rassurer le frontalier « moyen » sur sa capacité à accéder à une mobilité plutôt réservée aux élites, confère aux discours une fonction consolatoire proche de celle qui a été attribuée à la littérature populaire (Eco et Bouzaher 1995).
- 40 Mais cette quête d'une mobilité enrichissante et bienfaitrice ne va pas de soi : elle est parsemée d'embûches et exige des aides très concrètes. Si la mobilité désigne le caractère de ce qui peut simplement être déplacé ou se déplace par rapport à une position d'origine, elle est à la fois encadrée (*Schéma de mobilité transfrontalière*, *Commission Mobilité*) et soutenue (*chèque-mobilité*, *aides à la mobilité*, *favorisant une mobilité*, *il faut améliorer cette mobilité*, *pour accroître la mobilité*). En outre, des questions de mobilité (*comment donner le goût de la mobilité*, *comment éduquer à la mobilité en Europe*) se posent et renvoient à des revendications pressantes (*le droit à une vie personnelle et familiale*, *l'égalité de traitement entre salariés autochtones et salariés mobiles*) exprimées dans les discours de contestation eurosyndicale.
- 41 Le statut argumentatif de la *mobilité* dans les discours de prescription, sa signifiante dans les imaginaires collectifs véhiculés par les discours d'appropriation et de médiation, sa circulation au sein de l'ensemble de la FD eurorégionale et les contre-arguments qui lui sont opposés dans les discours syndicaux confirment l'intérêt qu'il convient de porter à l'examen approfondi de cette formule.

4. Conclusion

- 42 Nous avons présenté le corpus eurorégional à travers trois caractéristiques qui dessinent son originalité. Numérique, multilingue et multigenre, ce corpus a nécessité une réflexion dédiée à l'articulation des discours et à la conception d'une démarche d'analyse adaptée à l'hybridité des données. Inspiré d'une démarche de développement agile (c'est-à-dire évolutive et itérative en fonction des besoins de l'analyse qualitative), le modèle élaboré pour le traitement des données se veut suffisamment léger pour être en permanence (re)connecté au corpus et suffisamment cadré pour revêtir une cohérence. Sans ambition statistique démesurée, il nous permet de nous affranchir de deux difficultés majeures : d'un côté, il optimise la formulation de requêtes chronophages dans les six langues du corpus, et d'un autre côté, il automatise l'exportation contextualisée des résultats.
- 43 Bien qu'elle nous confronte régulièrement à des erreurs de reconnaissance morphosyntaxique et nécessite des opérations manuelles de nettoyage des résultats,

l'association de *TreeTagger* à une base de données permet de tester autant d'hypothèses que nécessaire au fil de la recherche et d'archiver automatiquement les requêtes et les résultats en vue de retracer l'historique de la démarche. Nous espérons aussi limiter les risques induits par la juxtaposition de résultats hétérogènes issus de philosophies différentes de développement logiciel²¹ et par la présentation de statistiques peu appropriées à l'objectif poursuivi.

- 44 Le temps dédié à la conception préalable de ce modèle d'analyse se justifie désormais par la possibilité de manipuler le corpus multilingue et d'envisager l'ajout ponctuel de fonctionnalités utiles à l'identification de parcours de lectures entre mots, passages, textes, énonciateurs et genres. L'aller-retour indispensable entre les textes et les résultats s'en trouve facilité sans être enfermé dans une technique trop contrainte. Appliqué à l'ensemble des textes en vue de tester la formule de la *mobilité*, le modèle a permis, d'une part, de préciser l'ampleur et les usages discursifs de cette formule héritée des textes fondateurs de l'Union européenne et, d'autre part, d'ouvrir la voie à l'interprétation qui fait de la *mobilité* un élément d'idéalisation de la gouvernance transfrontalière et une condition d'appropriation du projet eurorégional par le citoyen.

BIBLIOGRAPHIE

- Alliès P. (2011). « La notion d'Eurorégion et sa mise en œuvre dans l'Union européenne », in C.E. Pachado Amaral (dir.), *Autonomie régionale et relations internationales. Nouvelles dimensions de la gouvernance multilatérale*. Paris : L'Harmattan, pp. 245-255.
- Beacco J.-C. (1992). « Les genres textuels dans l'analyse du discours : écriture légitime et communautés translangagières », *Langages*, vol. 26, n° 105, pp. 8-27.
- Beacco J.-C. (1995). « À propos de la structuration des communautés discursives : beaux-arts et appréciatif », *Les Carnets du Cediscor. Publication du Centre de recherches sur la didacticité des discours ordinaires*, n° 3, pp. 136-153.
- Beacco J.-C. (2004). « Trois perspectives linguistiques sur la notion de genre discursif », *Langages*, vol. 38, n° 153, pp. 109-119.
- Charaudeau P. (2009). « Dis-moi quel est ton corpus, je te dirai quelle est ta problématique », *Corpus*, n° 8, pp. 37-66.
- Cislaru G. et Sitri F. (2012). « De l'émergence à l'impact social des discours : hétérogénéités d'un corpus », *Langages*, vol. 187, n° 3, pp. 59.
- Claudel C., Münchow P. von, Ribeiro M.P., Pugnière-Saavedra F. et Tréguer-Felten G. (2013). *Cultures, discours, langues : Nouveaux abordages*. Limoges : Lambert-Lucas.
- Cole A. et Palmer R. (2009). « Logiques de territorialité et de régionalisation en Europe de l'Ouest », *Revue d'études comparatives Est-Ouest*, vol. 39, n° 3, pp. 19-36.
- Charaudeau P. et Maingueneau D. (dir.) (2002). *Dictionnaire d'analyse du discours*. Paris : Seuil.

- Courtine J.-J. et Marandin J.-M. (1981). « Quel objet pour l'analyse du discours ? », in B. Conein et al. (dir.), *Matérialités discursives*. Lille : Presses universitaires de Lille, pp. 21-33.
- Eco U. et Bouzaher M. (1995). *De superman au surhomme*. Paris : Le Livre de Poche.
- Foucault M. (2008). *L'archéologie du savoir*. Paris : Gallimard.
- Hermant M.-H. (2014). « Le discours eurorégional. Indices convergents de légitimation d'un espace institutionnel », *Mots. Les langages du politique*, n° 106, pp. 75-81.
- Hermant, M.-H., Thouraud, E. (2015). « Traitement de données issues d'un corpus écrit multilingue. Approche agile pour l'analyse du discours eurorégional », *SHS Web of Conferences*, 20, Proceedings of ICODOC 2015 - Colloque « Corpus complexes et enjeux méthodologiques : de la collecte de données à leur analyse », Laboratoire ICAR, Lyon, mai 2015.
- Krieg-Planque A. (2009). *La notion de « formule » en analyse du discours. Cadre théorique et méthodologique*. Besançon : Presses Universitaires de Franche-Comté.
- Maingueneau D. (1987). *Nouvelles tendances en analyse du discours*. Paris : Hachette.
- Maingueneau D. (2012). « Que cherchent les analystes du discours ? », *Argumentation et Analyse du Discours*, n° 9, pp. 2-14.
- Mayaffre D. (2002). « Les corpus réflexifs : entre architextualité et hypertextualité », *Corpus*, n° 1.
- Mayaffre D. (2004). « Formation(s) discursive(s) et discours politique : l'exemplarité des discours communistes versus bourgeois durant l'entre-deux-guerres », *Texte !*
- Moirand S. (2007). *Les discours de la presse quotidienne : Observer, analyser, comprendre*. Paris : Presses Universitaires de France.
- Moirand S. et Beacco J.-C. (1995). « Autour des discours de transmission des connaissances », *Langages*, vol. 29, n° 117, pp. 32-53.
- Münchow P. von (2010). « Langue, discours, culture : quelle articulation ? », *Signes, discours et sociétés [en ligne]*, vol. 4, n° Visions du monde et spécificité des discours.
- Perrin T. (2013). *Culture et eurorégions. La coopération culturelle entre régions européennes*. Bruxelles : Presses Universitaires de Bruxelles.
- Oger C. et Ollivier-Yaniv C. (2003). « Analyse du discours institutionnel et sociologie compréhensive : vers une anthropologie des discours institutionnels », *Mots. Les langages du politique*, n° 71, pp. 125-145.
- Oliveras González X., Durà Guimerà A., Perkmann M. (2010). « Las regiones transfronterizas : balance de la regionalización de la cooperación transfronteriza en Europa (1958-2007) », *Documents d'Anàlisi Geogràfica*, vol. 56, n° 1, p. 21-40.
- Pêcheux M. (1990), « Analyse de discours. Trois époques », in D. Maltidier (dir.), *L'inquiétude du discours. Textes de Michel Pêcheux*. Paris : Éditions des Cendres, pp. 295-302.
- Rastier F. (2011), *La mesure et le grain. Sémantique de corpus*. Paris : Éditions Honoré Champion.
- Scholz R. et Angermüller J. (2013). « Au nom de Bologne ? Une analyse comparative des discours politiques sur les réformes universitaires en Allemagne et en France », *Mots. Les langages du politique*, vol. 102, n° 2, pp. 21-36.
- Vickoff J.-P. (2003). *Systèmes d'Information et Processus Agiles*. Paris : Hermès science publications.
- Wassenberg B. (2010). « Le voisinage de proximité : les eurorégions "géopolitiques" aux frontières externes de l'UE (1993-2009) », *Matériaux pour l'histoire de notre temps*, vol. 97-98, n° 1, pp. 45-49.

Yapomo M. (2013), « Construction de corpus multilingues : état de l'art », Actes de RECITAL'2013 : 15^e rencontre des étudiants chercheurs en informatique pour le traitement automatique des langues, Les Sables d'Olonne, 20 juin 2013.

NOTES

1. http://www.aebr.eu/fr/membres/repartition_regionale.php (consulté le 21 novembre 2015).
2. En ligne sur le site du Conseil de l'Europe : <https://wcd.coe.int/ViewDoc.jsp?id=621763&Site=COE> (consulté le 21 novembre 2015).
3. La distinction qui suit est proposée par l'Association des régions frontalières européennes (ARFE) et reprise dans le Lexique de l'aménagement du territoire européen (Université de Paris 7 - Datar - CNRS) : <http://www.ums-riate.fr/lexique/modeleterme.php?id=21> (consulté le 21 novembre 2015).
4. 621 textes, 603 599 mots. Les textes sont assez courts (près de 80 % des textes entre 100 et 1 000 mots). La moitié d'entre eux est datée et publiée surtout à partir de 2004, date de l'élargissement européen à l'Est qui fait passer l'Union européenne de 15 à 25 États membres.
5. Nous utilisons le logiciel Evernote, développé par la société éponyme (-Redwood City, Californie), avec un compte Premium. Défini comme *cloud-based software service*, Evernote permet notamment l'enregistrement de pages web dans leur intégralité, leur organisation, leur étiquetage, leur édition et la recherche en plein texte.
6. Groupement européen de coopération territoriale (GECT) ou Groupement eurorégional de coopération (GEC).
7. Statut de GECT.
8. La situation décrite est valable en novembre 2015 mais les statuts sont susceptibles d'évoluer.
9. Il s'agit des eurorégions : *EURORegion Elbe-Labe*, *Eems Dollard Regio*, *Euregio Karelia*, *Euregio Rijn-Waal* / *Rhein-Waal*, *Eurométropole Lille-Kortrijk-Tournai*, *Europaregion Donau Moldau*, *Europaregion Tirol-Süd Tirol-Trentino*, *Eurorrexión Galicia-Norte de Portugal*, *Eurorégion Aquitaine-Euskadi*.
10. L'*Eurorégion Aquitaine-Euskadi* utilise trois langues de travail : l'espagnol (castillan), le basque (euskera) et le français. Elle assure en outre sa promotion sur le web en occitan et en anglais.
11. Ainsi l'*Euroregione Adriatico-Ionica* propose-t-elle l'anglais plutôt que le grec aux côtés des versions disponibles en italien et en serbo-croate.
12. Les textes des CSIR (principalement des comptes rendus d'assemblées générales ou de réunions) n'ont pu faire l'objet d'une numérisation exploitable. Ils ne sont donc pas comptabilisés dans le nombre total de mots du corpus et font l'objet d'une analyse qualitative manuelle.
13. Pour une description détaillée et illustrée de la méthode élaborée, nous renvoyons à l'article publié à la suite du colloque ICODOC 2015 « Corpus complexes et enjeux méthodologiques : de la collecte de données à leur analyse », Lyon-Laboratoire ICAR, mai 2015 (Hermand et Thouraud 2015).
14. Mis à disposition par l'université de Stuttgart.
15. Les développements sont réalisés en langage *Perl* par Emmanuel Thouraud, ingénieur informaticien. La base de données est créée en *SQLite*. Les résultats sont exportés automatiquement au format *Excel* et offrent de nombreuses possibilités de tris.
16. *TreeTagger*, *Perl*, *SQLite* et *Excel*.
17. http://www.constitutioneu.eu/cariboost_files/trait_c3_a9_20de_20rome.pdf (consulté le 21 novembre 2015).
18. Les occurrences dans les différentes partitions linguistiques sont nombreuses : *mobilité professionnelle*, *movilidad laboral*, *business mobility*, *labour mobility*, *mobiliteit op de arbeidsmarkt*, *mobiliteit zoals werkgelegenheid...*

19. *green mobility*

20. *forte mobilità*

21. Nous avons ainsi préalablement testé divers outils : citons surtout les apports de *TXM* par rapport à notre démarche (pour son approche textométrique et la mise en forme des résultats), *Tropes* (davantage adapté à l'analyse de contenu qu'à l'analyse du discours), *Iramuteq* (performant pour l'analyse des représentations sociales à partir de questionnaires) et *Lexico3* (en version de test).

RÉSUMÉS

L'objectif de cet article est double : il s'agit, d'une part, de présenter un nouveau corpus permettant d'envisager le phénomène émergent de la communication transfrontalière en Europe et, d'autre part, de formuler trois questionnements utiles au cadrage de son analyse sémantique. À partir du corpus eurorégional – multilingue et multigenre – nous posons les questions de la dispersion des discours en ligne, de l'hétérogénéité des données et de la contextualisation de l'analyse. Notre démarche consiste à construire progressivement un modèle d'analyse adapté à l'appréhension, tantôt automatique et tantôt manuelle, de la diversité des textes. Enfin, nous proposons d'illustrer la démarche en l'appliquant à la *mobilité*, observable récurrent du corpus.

This paper has two objectives: firstly, to present a new corpus to analyze the emerging phenomenon of the cross-border communication in Europe and, secondly, to formulate three questions useful for its semantic analysis. From the Euroregional corpus –multilingual and multi-genre– we ask the questions of the dispersion of the on-line speeches, the heterogeneousness of the data and of the contextualization of the analysis. Our approach consists in gradually building a model of analysis adapted to the appreciation, sometimes automatic and sometimes manual, of the diversity of texts. The analysis of the repeated expression of *mobility* in the corpus will illustrate the subject.

INDEX

Mots-clés : discours eurorégional, corpus multilingue, corpus multigenre, hétérogénéité discursive, traitement agile

Keywords : Euroregional discourse, multilingual corpus, multi-genre corpus, discourse analysis, textometry, discursive heterogeneity, agile process

AUTEUR

MARIE-HÉLÈNE HERMAND

Université libre de Bruxelles, ReSIC