

JAN RADIMSKÝ

PROJET ET CONSTRUCTION D'UN CORPUS DES TEXTES EUROPÉENS (CORTE)

1. Pourquoi un corpus des textes européens ?

Les apports méthodologiques de la linguistique de corpus aux sciences du langage sont, de nos jours, incontestables. En France, il s'agit probablement encore « d'une discipline récente qui remet en cause les longues traditions formalistes ou prescriptives » (Williams G., 2005 : 16), mais en République tchèque, les activités de l'Institut du corpus national tchèque auprès de la Faculté des lettres de l'Université Charles ont déjà abouti à la création des outils de recherche de haut niveau et facilement accessibles au large public scientifique. Si nous parlons d'apports méthodologiques, c'est que la linguistique de corpus n'est pas seulement une *théorie* ou *discipline* linguistique, mais également une *méthode* qui permet dans le domaine des sciences du langage de faire la recherche fondamentale, courante dans les sciences naturelles (cf. Williams G., 2005 : 13, 17 ; Čermák F., 2000 : 16), avec une relative indépendance de la théorie sous-jacente.¹ C'est pourquoi nous considérons que la création des corpus linguistiques est une condition essentielle pour que le caractère scientifique de la linguistique soit une réalité.

C'est dans cette perspective que l'élaboration du Corpus des textes européens (CORTE) a été entreprise au Département de langues et littératures romanes de la Faculté de Pédagogie de l'Université de Bohême du sud, en février 2005. Au départ, notre but était de fournir aux étudiants et aux chercheurs dans le domaine des langues étrangères appliquées (LEA) de notre département un instrument de travail efficace et fiable, mais il se montre que les applications et les domaines

¹ Ainsi, on distingue habituellement deux manières différentes d'exploiter les corpus linguistiques : « l'une déductive, *corpus-based* en anglais, qui utilise le corpus pour confirmer ou infirmer une hypothèse, et une linguistique de corpus inductive, *corpus-driven*, qui cherche à explorer les données sans *a priori*. » (Williams G., 2005 : 13). Pour plus une analyse plus détaillée de la première approche, notamment à l'égard de la validité et de la fidélité de vérification des hypothèses en linguistique, voir Radimský J. (2005). Quant à la seconde approche, cf. l'ouvrage de référence en la matière de TOGNINI BONELLI, E. (2001).

d'utilisation du corpus pourront bientôt dépasser ce cadre étroit. Ainsi, nous nous proposons dans la présente contribution de décrire les étapes et l'état actuel de la construction du corpus d'une part, et d'autre part de s'interroger sur les possibilités de son exploitation et utilisation.

2. Les paramètres de base du CORTE

CORTE est un corpus de textes au sens linguistique du terme, c'est à dire « une collection de textes authentiques, récurrents dans l'usage, en format électronique, représentant un état ou une variété de la langue » (Rossini-Favretti R., 2000 : 41).² Dégageons à présent ses caractéristiques générales à partir des critères courants de classification des corpus.³

Il s'agit d'un corpus de textes écrits, hautement spécialisé, contenant les textes juridiques du droit européen. En effet, l'indication « européen » dans la dénomination du corpus (CORpus des Textes Européens) fait référence aux structures de l'Europe unie, c'est à dire aux Communautés et à l'Union européennes qui sont fondées sur le droit.⁴ Il s'agit par conséquent d'un « corpus linguistique du registre juridique ». Même si sous l'aspect lexical les textes juridiques en question traitent souvent des domaines scientifiques particuliers, tels que la biologie, l'économie, les technologies agricoles, la médecine, etc, il n'en reste pas moins que dans son aspect syntaxique et grammatical, il s'agit à présent purement d'un corpus de textes « juridiques », quoique la possibilité de l'enrichir dans le futur par d'autres modules, relevant des domaines proches à celui du droit (p. ex. le domaine des textes administratifs, etc.), n'est pas exclue.

Quant au nombre des langues impliquées, CORTE est un corpus parallèle. Tous les textes du corpus sont actuellement en version tchèque et française, la préparation des textes en version italienne est en cours et celle des textes en espagnol est prévue. Ainsi, le corpus pourra servir aux chercheurs travaillant sur de différentes langues romanes, éventuellement en confrontation avec le tchèque.

CORTE est un corpus synchronique, en ce qu'il inclue les textes juridiques européens rédigés à partir des années cinquante du siècle dernier (c'est à dire à partir du moment où les premières structures communautaires ont été créées) jusqu'à nos jours. Il s'agit d'une conception légèrement plus large de la synchronie par rapport à celle du Corpus national tchèque par exemple,⁵ mais l'archi-

² «Una raccolta di testi, autentici e ricorrenti nell'uso, in formato elettronico, rappresentativi di uno stato o di una varietà di una lingua.»

³ Cf. p. ex. ROSSINI-FAVRETTI, R. (2000), SINCLAIR J. (1996), et sim.

⁴ Cf. le motif 23 de l'arrêt C-294/83 qui énonce : « Il y a lieu de souligner d'abord, à cet égard, que la communauté économique européenne est une communauté de droit en ce que ni ses états membres ni ses institutions n'échappent au contrôle de la conformité de leurs actes à la charte constitutionnelle de base qu'est le traité. ».

⁵ Dans le corpus tchèque SYN2000 sont inclus les documents à partir de l'année 1990, mais aussi certains ouvrages littéraires importants antérieurs à cette date. Dans ce dernier cas, il

teature variable du CORTE permet, lorsque le besoin s'en fera sentir, de diviser les textes par tranches de temps en fonction de l'année de leur adoption, afin de permettre une interrogation plus spécifique suivant l'axe diachronique. Mais en parlant du caractère «synchronique» du CORTE, il faut également tenir compte de la spécificité du langage juridique qui conserve de nombreux traits anciens, disparus désormais de la langue générale. Ainsi, le participe présent en italien, pour donner un exemple du phénomène, a presque complètement perdu ses caractéristiques verbales et il ne peut être utilisé, dans la langue courante, qu'en tant qu'adjectif. Dans la langue juridique par contre, les phrases à participe présent qui conserve sa valence (sujet, complément d'objet ou phrase complétive) se rencontrent fréquemment.⁶

Du point de vue technique, les textes ne sont pas annotés et ils sont répertoriés sous format de «texte simple». Ainsi, ils sont consultables à l'aide de n'importe quel programme standard d'exploitation des corpus monolingues ou parallèles. Pour les besoins actuels, nous utilisons le programme d'exploitation PARACONC.⁷

3. La représentativité du CORTE

Il est particulièrement important de s'interroger sur la représentativité de notre corpus, car c'est seulement à partir d'un corpus représentatif que le linguiste peut procéder à la généralisation des résultats (cf. Jacques M.-P., 2005 : 26–27). La question de savoir comment assurer la représentativité d'un corpus reste encore ouverte pour les théoriciens du domaine (cf. Rossini-Favretti R., 2000 : 45–48). Cependant, elle sera sans doute plus facile à traiter dans le cas d'un corpus spécialisé comme le nôtre que dans le cas d'un corpus qui se veut représentatif de la langue «générale», car en effet, la représentativité d'un corpus spécialisé est étroitement liée à la manière dont nous arrivons à définir et délimiter d'une part les bornes de la variété linguistique étudiée et d'autre part la structure hiérarchique des textes qui font partie du domaine (cf. Biber D., 2000 : 107–108). Or, il est possible de déterminer les textes qui font partie du droit européen⁸ et d'établir

faut que l'auteur de l'ouvrage ait été né après 1880. Cf. <<http://ucnk.ff.cuni.cz/>> [janvier 2007].

⁶ Cf. par exemple: «(3) considerando che, nella sua risoluzione del 2 giugno 1994 (6), ri-guardante il quadro d'azione comunitario nel settore della pubblica sanità, il Consiglio ha affermato che deve essere attualmente attribuita priorità alle malattie trasmissibili (...)», Decisione n. 2119/98/CE del Parlamento europeo e del Consiglio del 24 settembre 1998.

⁷ Cf. <<http://www.athel.com/para.html>> [janvier 2007]

⁸ Il pourrait sembler que la raison principale pour se limiter uniquement au droit européen lors de la construction d'un corpus parallèle juridique est d'ordre pratique – les documents juridiques sont systématiquement traduits dans toutes les langues du corpus. Mais en réalité, il y a une raison théorique primordiale qui demande de séparer nettement l'étude du registre juridique au niveau national de celle du registre juridique européen : c'est qu'au niveau

leur structuration interne en s'appuyant sur la doctrine juridique. M. Tomášek (1987: 65–75) décrit les limites et la structuration de la langue juridique suivant trois critères qui correspondent aux éléments-clé de la communication, à savoir (1) le locuteur, (2) le récepteur et (3) l'objet de la communication. Selon (1) le locuteur, l'auteur distingue (1-a) le style de la norme juridique, (1-b) le style d'application, (1-c) le style de vulgarisation scientifique, et (1-d) le style rhétorique. Suivant (2) le récepteur la distinction est faite entre (2-a) le registre scientifique, (2-b) le registre pédagogique et (2-c) le registre transgressant la norme linguistique standard. D'après (3) l'objet de la communication, la division se fait selon le critère sémantique, correspondant *grosso modo* aux différentes branches du droit (p. ex. le droit civil, le droit pénal, le droit de commerce, etc.). Ce panorama n'est pas simple, car les trois critères peuvent s'entrecroiser, mais il nous permet de dégager le noyau et la périphérie de la langue juridique de manière stratifiée. Ainsi, nous pouvons proposer un corpus constitué de plusieurs sous-corpus, l'interrogation pouvant s'effectuer par critères croisés. Dans le rassemblement des textes, nous nous proposons de commencer par le noyau, c'est à dire par les normes juridiques (critère 1-a) relevant du registre scientifique (critère 2-a). Dans la typologie des textes (critère no. 3), nous nous appuyons sur la classification thématique des textes législatifs existants, disponible sur le site Eur-Lex.⁹

Une fois que la typologie et la structuration des textes est établie, le vrai questionnement linguistique sur la représentativité du corpus ne fait que commencer. Selon Rema Rossini-Favretti (2000: 45–48), le problème comporte deux volets principaux.

Le premier volet consiste à décider si la représentativité peut être assurée par la taille du corpus ou par la variabilité des textes qu'il contient. Dans la catégorie des «normes juridiques» qui forme le noyau de notre corpus (se situant au croisement des critères 1-a et 2-a), la variabilité des textes est par définition assez limitée et correspond à celle des principaux types de textes juridiques en droit européen. Parmi les normes juridiques européennes (1a), nous aurons donc¹⁰:

I. Le droit primaire (*traités*)

II. Le droit dérivé correspondant aux actes unilatéraux des institutions européennes

II.1. Les actes mentionnés par les traités (*règlements, directives, décisions, avis, recommandations, règlements internes*)

européen, le droit représente une seule réalité, un seul référent ontologique, exprimé par plusieurs langues. Au niveau national par contre, chaque état a sa propre réalité juridique exprimée dans sa propre langue. Par conséquent, l'étude terminologique dans le domaine du droit européen est principalement une question de langue, tandis qu'au niveau national, elle relève non seulement du domaine linguistique, mais également de celui du droit comparé. Cf. RADIMSKÝ, J. (2004) et pour plus de détails cf. GÉMAR, J.-C. (1995).

⁹ <http://eur-lex.europa.eu/RECH_repertoire.do> [janvier 2007].

¹⁰ Cf. FAVRET, J. M., TOULEMON, R. (2001), et MANIN, P. (1999: 297–305).

II.2. Les actes non mentionnés par les traités (par exemple : *déclarations, communications* – dont « *livres blancs* », *conclusions, résolutions, programmes d'action, codes de conduite*, etc.)

III. Les actes bilatéraux : (*accords*)

IV. La jurisprudence (*arrêts*)

L'homogénéité qui en résulte n'est pas un inconvénient pour la recherche, mais une caractéristique du sous-code linguistique étudié. Etendre la variabilité des textes davantage signifierait quitter le sous-code du registre défini, ce qui ne serait pas souhaitable. Mais il est possible d'introduire de nouveaux modules (sous-corpus) correspondants à la stratification du registre étudié, comme nous venons de l'indiquer plus haut. Ainsi, les recherches effectuées séparément sur chaque module individuel permettraient de comparer les sous-registres linguistiques en question, tandis que les recherches effectuées sur l'ensemble des modules seraient en mesure de prendre en compte la variabilité et la richesse de l'ensemble du registre juridique. Dans ce dernier cas, il faudrait cependant déterminer les critères de pondération des différents modules. Quant à l'axe sémantique (critère no. 3), la variabilité peut être très importante, notamment si c'est une recherche terminologique qui est envisagée. Dans ce cas, il faut se poser la question de savoir si le domaine étudié fait effectivement partie du corpus et dans quelle mesure il est représenté ; la variabilité doit, dans ce cas, aller de pair avec la taille, cette dernière étant primordiale. Pour l'étude des autres plans linguistiques, la variabilité des textes sur l'axe sémantique et la taille du corpus pourraient être moins importantes, si nous arrivons à vérifier l'hypothèse que le sous-code linguistique en question est extrêmement uniforme et spécifique. Mais pour ce faire, il faudrait d'abord disposer d'un corpus assez variable et étendu, ce qui fait que la plus grande variabilité possible sur l'axe sémantique est souhaitable en tout cas.

Le second problème théorique lié à la représentativité d'un corpus consiste en la question de savoir si le corpus doit contenir les textes intégraux ou les extraits de taille comparable. Vu que l'application primordiale du CORTE doit être une recherche terminologique parallèle en plusieurs langues, nous considérons que les textes insérés devraient être le plus complets possibles, sans égard à leur taille. Cependant, nous étions parfois obligés d'éliminer les annexes de certains documents juridiques, lorsque ces derniers n'étaient pas disponibles dans l'une des versions linguistiques du corpus. D'autre part, il s'avère inutile d'inclure certains documents extrêmement courts qui contiennent, mise à part l'entête et la finale standardisés, seulement une ou deux phrases générales portant information sur la modification d'une disposition juridique. De telles énoncés se trouvent fréquemment dans tous les textes juridiques traités et n'apportent rien de nouveau au corpus.

Cette dernière remarque nous amène à la question de savoir si le corpus devrait rendre compte des phénomènes linguistiques de manière « proportionnelle » par rapport à leur fréquence réelle dans les textes, ou s'il faut se pencher plutôt du côté de la variabilité, en favorisant l'inclusion des éléments marginaux. La

représentativité proportionnelle est recherchée notamment dans les sciences sociales : une enquête sociologique cherche à relever en premier lieu les tendances générales dans les attitudes, opinions, etc. Nous convenons avec D. Biber (2000 : 115–116) que la représentativité du corpus devrait être d'un ordre différent, car la linguistique ne s'intéresse pas seulement aux phénomènes centraux du système.¹¹ Si nous poursuivons cette réflexion un peu plus loin, nous pourrions même dire que les éléments marginaux (qu'il faut distinguer des hapax ou des erreurs) sont extrêmement importants en ce qu'ils jouent un rôle central lors de la falsification des hypothèses au sens de Popper, ce que nous considérons en effet comme l'un des rôles essentiels de tout corpus linguistique.

4. Une typologie du contenu

Bien que CORTE soit encore en construction, sa taille atteindra bientôt celle des petits corpus de langue spécialisée utilisés couramment pour la recherche. En effet, il compte actuellement 700.000 mots graphiques en version française, et 530.000 mots graphiques correspondants en version tchèque. Les textes qu'il contient peuvent être répartis en fonction de la stratification du registre juridique discutée dans la section précédente ; nous allons présenter d'abord la stratification formelle (critère no. 1 et 2) et ensuite la stratification sémantique (critère no. 3).

Du point de vue formel, CORTE contient les textes de droit dérivé et les actes bilatéraux dans les proportions indiquées dans le tableau 1. C'est notamment le droit primaire (traités) et la jurisprudence qui manquent encore, afin que la variabilité du contenu soit atteinte. Par contre, les principaux actes du droit secondaire sont représentés de manière équilibrée.¹²

Tableau 1. – Typologie formelle

Document	Nombre des documents	Taux	Pages ¹	Taux
Règlement	64	32%	810	38%
Décision	61	31%	264,4	12%
Directive	56	28%	809,3	38%
Accord	15	8%	206,7	10%
Convention	2	1%	28,2	1%
Procès-verbal	2	1%	13,6	1%
Total	200		2132,2	

¹¹ L'auteur (Biber D., 2000 : 116, note 6) précise que c'est notamment pour une recherche lexicographique que la conception proportionnelle d'un corpus est inefficace, car la plupart des unités lexicales rares n'y apparaîtraient jamais. Dans la pratique nous constatons que la recherche des termes spécialisés se fait plus facilement sur un corpus (ou sous-corpus) restreint et ciblé que sur un corpus plus large, où le domaine traité est, proportionnellement parlant, plus faiblement représenté.

¹² Il s'agit d'une page du format A4 qui correspond à 1.800 caractères, espaces compris. Une page correspond à environ 300 mots graphiques.

La stratification sémantique du CORTE est indiquée de manière sommaire au tableau 2. Pour cette typologie, nous nous appuyons sur la classification thématique des textes législatifs existants, disponible sur le site Eur-Lex.¹³ À présent 5 thèmes sur 20 sont représentés, dont 4 de manière équilibrée ; le domaine No. 15 (« Environnement, consommateurs et protection de la santé ») étant actuellement sur-représenté.

*Tableau 2. – Typologie sémantique
(Les chiffres à gauche correspondent à la classification numérique dans la base de données Eur-lex)*

Domaine	Nombre des pages	Taux
05. Libre circulation des travailleurs et politique sociale	384,3	18%
06. Droit d'établissement et libre prestation des services	412,3	19%
11. Relations extérieures	290,4	14%
15. Environnement, consommateurs et protection de la santé	776,3	36%
19. Espace de liberté, sécurité et justice	269	13%
Total:	2132,3	

Il serait cependant précoce de tenter faire une pondération des textes à ce niveau macroscopique, car chaque domaine comporte encore des sous-classes jusqu'au 4^e niveau de l'échelle hiérarchique. En effet, les textes du CORTE relèvent actuellement de 8 domaines très précis, indiquées en caractères gras au tableau 3. Chaque domaine est contextualisé dans le tableau jusqu'au niveau hiérarchique le plus élevé ; les chiffres à droite des thèmes indiquent le nombre des pages A4 correspondants.

*Tableau 3. – Typologie sémantique détaillée.
(Les chiffres à droite indiquent le nombre des pages pour chaque thème)*

05. Libre circulation des travailleurs et politique sociale – 384,3 pages
- 05.20. Politique sociale
- 05.20.40. Sécurité sociale
- 05.20.40.20. Application aux travailleurs migrants – 384,3 pages
06 Droit d'établissement et libre prestation des services – 412,3 pages
- 06.20. Application sectorielle
- 06.20.20. Activités de services
- 06.20.20.20. Banques – 412,3 pages
11. Relations extérieures – 290,4 pages
- 11.70. Politique de développement
- 11.70.20. Aide aux pays en développement
- 11.70.20.10. Aide alimentaire – 290,4 pages

¹³ <http://eur-lex.europa.eu/RECH_repertoire.do> [janvier 2007].

- 15. Environnement, consommateurs et protection de la santé – 776,3 pages
 - 15.10. Environnement – 462,7 pages
 - 15.10.30. Espace, milieu et ressources naturelles
 - **15.10.30.20. Conservation de la faune et de la flore – 119,7 pages**
 - **15.10.30.30. Gestion des déchets et technologies propres – 343 pages**
 - **15.30. Protection de la santé – 235,6 pages**
 - **15.40. Protection des animaux – 78 pages**

- 19. Espace de liberté, sécurité et justice – 269 pages
 - **19.10. Libre circulation des personnes – 269 pages**

Il ressort de cette typologie sémantique que CORTE permet d'effectuer des recherches terminologiques sur des domaines très spécialisés – c'est effectivement la tâche qui est confiée aux étudiantes de la 3^e année du programme d'études «Anglais-français pour l'administration de l'UE», dans le cadre de leurs mémoires de licence. C'est aussi grâce à elles que le corpus ait pu prendre cette ampleur en un espace de temps relativement bref.¹⁴

Par contre, il est assez difficile d'estimer quel est le pourcentage des textes du corpus par rapport à la totalité de la législation européenne traduite en tchèque, car l'ampleur de cette dernière n'est pas retrouvable et varie probablement sans cesse.

5. À propos des possibilités d'exploitation du CORTE

Nous avons déjà évoqué dans le chapitre précédent quelques applications possibles du corpus CORTE; essayons à présent de les systémiser, tout en proposant quelques hypothèses et pistes de recherche nouvelles.

Etant un corpus de textes spécialisés, CORTE est en premier lieu un instrument de recherche du registre juridique. À cette fin, il peut être utilisé aussi bien comme un corpus monolingue (français, tchèque, italien), bilingue ou trilingue (français-tchèque, italien-tchèque, français-italien, français-tchèque-italien).

Sa première application relève du domaine terminologique – il est utilisé par les étudiants dans le cadre de leurs mémoires de licence pour la mise en évidence des termes spécialisés français et leurs versions tchèques. Au cours du suivi de la préparation de mémoires, nous avons pu vérifier l'affirmation de M. Tomášek (2004: 59), selon qui le langage juridique est caractérisé par de nombreuses expressions, termes et tournures figées, allant du mot simple jusqu'à une phrase entière. De plus, nous avons constaté que parmi les termes nominaux complexes, certaines structures sont beaucoup plus fréquentes que d'autres (cf. Radimský, 2007). C'est une hypothèse qu'il serait intéressant de vérifier et d'accompagner éventuellement de propositions du traitement lexicographique des termes juridiques. D'autre part, le domaine de la valence verbale semble présenter également

¹⁴ Pour en savoir plus, cf. l'article RADIMSKÝ, J. (2007) en cours d'impression.

un niveau de figement assez complexe à explorer (cf. également Tomášek M., 2003 : 51–52). Mise à part l'application lexicographique et traductologique de ces phénomènes, nous pourrions en envisager également une application didactique dans l'enseignement de la langue de spécialité.

Les spécificités du langage juridique peuvent être envisagées et étudiées également sur le plan grammatical, syntaxique et textuel. Au niveau syntaxique des textes français par exemple, il semble que certains types de subordonnées sont extrêmement rares,¹⁵ tandis que les propositions implicites contenant un gérondif ou un participe sont au contraire très courantes – dans ce cas, il serait utile d'analyser les procédés de traduction de ces propositions implicites en tchèque. En effet, les anomalies de ce type sont fort nombreuses ; parmi elles, I. Petru (2007) mentionne par exemple encore les problèmes de l'ordre des mots, l'usage simultané de plusieurs synonymes redondants, l'usage de connecteurs spécifiques (*ledit, susdit, etc.*), etc.

6. Perspectives d'avancement du projet

Tous les phénomènes décrits au chapitre précédent peuvent être étudiés avec efficacité sur le corpus CORTE tel qu'il est dans sa forme actuelle – qu'il s'agisse d'une recherche monolingue ou d'une recherche traductologique.

Il est évident que dans certains cas, une annotation grammaticale et syntaxique du corpus, faisant encore défaut, pourrait faciliter considérablement le travail du linguiste. C'est précisément elle qui, allant de pair avec l'augmentation de la taille et de la diversité du contenu, sera envisagée, du moins pour l'une des langues romanes représentées.

Actuellement, la limitation principale pour l'utilisateur du CORTE consiste dans le fait que le corpus est interrogeable uniquement de manière locale, sur un ordinateur de notre département. Sa mise à disposition en ligne est prévue, mais sera conditionnée par l'obtention d'une subvention qui permettrait de créer et de maintenir un logiciel spécifique d'interrogation et de régler certaines questions juridiques liées à cette forme de « publication » des textes. À cette fin, nous pourrions également envisager une coopération plus étroite avec le projet du corpus parallèle INTERCORP, élaboré à la Faculté des lettres de l'Université Charles de Prague. C'est d'ailleurs au directeur du projet INTERCORP, M. le prof. František Čermák, que nous devons un certain nombre de conseils pratiques précieux, sans lesquels CORTE n'aurait jamais vu le jour.

¹⁵ Par exemple les subordonnées concessives sont presque absentes : le corpus de 700.000 mots contient une seule conjonction « bien que », alors que le corpus du journal *Le Monde* de 1998 (cf. ci-dessus pour l'indication bibliographique) de taille comparable (1.100.000 mots) en compte plusieurs dizaines. (*Le Monde* 1998 <http://www.lexutor.ca/concordancers/concord_f.html> [janvier 2007])

Bibliographie

- BIBER, D. (2000), Repräsentativnost v projektu korpusu. In: *Studie z korpusové lingvistiky*, Praha, Karolinum, pp. 107–136.
- ČERMÁK, F., (2000), Jazykový korpus : Prostředek a zdroj poznání. In: *Studie z korpusové lingvistiky*, Praha, Karolinum, pp. 15–38.
- FAVRET, J. M., TOULEMON R. (2001), *L'essentiel de l'Union Européenne et du droit communautaire*. Paris, Gualino, coll. „Les Carrés”.
- GÉMAR, J.-C. (1995), *Traduire ou l'art d'interpréter, tome 2: application. Traduire le texte juridique*, Sainte-Foy (Québec), Presses de l'Université de Québec.
- JACQUES, Marie-Paule (2005), Pourquoi une linguistique de corpus ? In: WILLIAMS G. (sous la direction de), *La linguistique de corpus*. Rennes, Presses universitaires de Rennes, pp. 21–30.
- Le Monde 1998. [online] Accessible sur <http://www.lexutor.ca/concordancers/concord_f.html> [janvier 2007].
- MANIN, P. (1999), *Les communautés européennes : l'Union européenne, droit institutionnel*. Paris, Pedone.
- PETRUŠ, I. (2007), Několik poznámek k překladu francouzského právního textu. [à paraître]. Prešov, Filozofická fakulta Prešovskej univerzity v Prešove, Actes du colloque «Vzťahy a súvislosti v preklade».
- RADIMSKÝ, J. (2004), Dans quelle mesure est-il possible de traduire un terme juridique ? In: *Etudes romanes de Brno*, Sborník prací FF MU, L 25, Brno, 2004, pp. 37–44.
- RADIMSKÝ, J. (2005), Des méthodes de vérification en linguistique. In: ČERMÁK Petr, Tláškal Jaromír (editores): *Las lenguas románicas: su unidad y diversidad*, Praha, Univerzita Karlova v Praze, Filozofická fakulta, pp. 178–184.
- RADIMSKÝ J. (2007), Pour une étude de la norme dans les traductions du droit communautaire. [à paraître], Cracovie, Université Jagellone, Actes du séminaire doctoral.
- ROSSINI-FAVRETTI, R. (2000), Progettazione e costruzione di un corpus di italiano scritto: CO-RIS/CODIS, In: Rema Rossini Favretti (a cura di): *Linguistica e informatica. Corpora, multimedialità e percorsi di apprendimento*. Roma, Bulzoni, pp. 39–56.
- SINCLAIR, J. (1996), Preliminary recommendations on Corpus Typology. [online] *EAGLES Document EAG-TCWG-CTYP/P* [cité le 15 janvier 2007]. Accessible sur World Wide Web: <www.ilc.cnr.it/EAGLES96/pub/eagles/corpora/corpus_typ.ps.gz>
- TOMÁŠEK, Michal (1987), K některým otázkám stylového rozvrstvení právního jazyka, *Právník*, 1, pp. 65–75.
- TOMÁŠEK, Michal (2004), Význam překladu pro poznání evropského práva. In: Hrala Milan, *Český překlad 1945–2003*, Praha, Ústav translatologie FF UK.
- TOMÁŠEK, Michal (2003), *Překlad v právní praxi*. Praha, Linde.
- TOGNINI BONELLI, E. (2001), *Corpus Linguistics at Work*. Amsterdam, John Benjamin's Publishing.
- WILLIAMS, G. (2005), Introduction. In: Williams G. (sous la direction de), *La linguistique de corpus*. Rennes, Presses universitaires de Rennes, pp. 13–20.