

Hétérogénéité des genres textuels produits par la justice : étude textométrique

Bénédicte Diot-Parvaz Ahmad¹, Lucie Gianola²

¹ERTIM, Inalco – benedicte.parvazahmad@inalco.fr

²AGORA, Université de Cergy-Pontoise – lucie.gianola@u-cergy.fr

Abstract

The characterization of text genres in the legal field is of interest to LegalTech as well as to natural language processing and translation professionals. However, legal textual genres relating to justice have so far been neglected by works proposing a typology of the legal field in these disciplines. Based on the legal typology proposed by Bocquet and Monjean-Decaudin, our textometric study of a corpus of documents relating to justice shows that there is a continuum on the morphosyntactic level going from the normative genre (legislative texts) to the oral narrative genre (hearing reports). This continuity in the textual genres of justice reflects the articulation of the judicial process between the theoretical statement of the norm and its actual application. This characterization will make it possible to refine the treatment of the corpora related to justice.

Keywords : genre, textometry, legal, forensic, justice

Résumé

La caractérisation des genres textuels juridiques intéresse aussi bien la LegalTech que le traitement automatique des langues ou les professionnels de la traduction. Or les genres textuels relatifs à la justice ont été jusqu'à présent négligés par les travaux proposant une typologie du champ juridique dans ces disciplines. En partant de la typologie juridique proposée par Bocquet et Monjean-Decaudin, notre étude textométrique d'un corpus de documents relatifs à la justice montre qu'il existe sur le plan morphosyntaxique un continuum allant du genre normatif (textes législatifs) au genre narratif oralisé (comptes-rendus d'audition). Cette continuité dans les genres textuels de la justice reflète l'articulation du processus judiciaire entre l'énoncé théorique de la norme et son application réelle. Cette caractérisation permettra d'affiner le traitement des corpus relatifs à la justice.

Mots clés : genre textuel, textométrie, juridique, judiciaire, justice

1. Introduction

L'émergence de la *LegalTech*, c'est-à-dire le développement des technologies au service du droit, et l'intérêt nouveau du traitement automatique des langues pour le domaine juridique, que reflètent les appels à contribution de revues spécialisées (*TAL*, Nazarenko et Wyner, 2017), nous amènent à questionner la nature du texte produit dans le cadre de la justice et du droit. En effet, Jacques et Aussenac-Gilles (2006) ont souligné l'importance de la prise en compte du genre textuel dans la perspective de l'application de traitements automatiques quelle qu'en soit la finalité. La caractérisation des genres textuels est utile également en dehors du *TAL*, par exemple dans le cadre de la pratique professionnelle d'interprétation et de traduction où elle permet de constituer des corpus au plus près des besoins de traduction.

C'est ainsi qu'en comparant les différentes occurrences d'un motif donné, le traducteur peut vérifier des hypothèses d'équivalence de traduction en contexte et améliorer la qualité de sa production. Or dans le domaine juridique et judiciaire, cette classification en genres n'est pas encore clairement établie. La distinction des genres textuels comporte donc des enjeux linguistiques transdisciplinaires. Après avoir rappelé le contexte de la recherche sur la caractérisation des genres d'un discours, nous présentons le corpus des textes qui a servi à notre analyse. Nous exposons ensuite la méthode d'analyse, en partie fondée sur l'approche textométrique, qui nous permet d'identifier les caractères des genres et de nous prononcer sur l'hétérogénéité du corpus judiciaire.

2. Problématique et contextualisation

Une partie importante de la pratique de la justice en France s'appuie sur l'écrit¹, sous forme de documents papier ou de plus en plus sous forme numérique, à tous les stades qui impliquent de « fixer » des éléments : l'écrit fait foi. Ceci entraîne une production écrite volumineuse et extrêmement variée, avec toutes les problématiques documentaires associées d'exploitation, de conservation et récemment, de numérisation. Sur la variété de la production, le point de vue du traducteur-interprète judiciaire est instructif, comme l'explique Monjean-Decaudin (2012, p.114) qui souligne la très grande hétérogénéité des documents traités par les traducteurs² dans le contexte de la justice, affirmant qu'il est impossible d'en établir une liste exhaustive.

Face à cette situation, une catégorisation en genres de ces documents textuels en apporterait une meilleure compréhension et pourrait améliorer leurs diverses exploitations. Dans la lignée des travaux de Biber (1988), Branca-Rosoff (1999) affirme que la caractérisation des genres textuels mobilise des critères externes (les pratiques sociales significatives) et des critères internes (les fonctionnements linguistiques).

En s'inspirant des travaux de C. Bocquet (2008), S. Monjean-Decaudin (2012) a proposé une classification basée sur la finalité des documents de la justice, différenciant les documents entraînant un effet de droit (documents juridiques), les documents impliqués en procédure judiciaire ayant un effet juridique, par exemple une ordonnance ou un jugement (documents juridico-judiciaires), et les documents impliqués dans une procédure judiciaire qui n'entraînent pas d'effet de droit, par exemple des auditions ou des rapports d'expertise (documents judiciaires).

Du côté de la caractérisation linguistique, nous avons relevé les travaux de Malrieu et Rastier (2001) et Gledhill et al. (2017). Malrieu et Rastier (2001) admettent un « discours juridique »

¹ À l'exception des procédures civiles où la présence d'un conseil n'est pas requise (art. 817 du CPC) et, au pénal, des audiences aux assises.

² « Rapports de toutes sortes (balistique, toxicologique, d'autopsie, d'expertises techniques, médicales, etc.) des comptes rendus d'avarie, d'enquête policière, des transcriptions d'écoutes téléphoniques, des contrats commerciaux, des actes authentiques, des décisions de justice, des procès-verbaux de perquisition, mais plus simplement des extraits de compte bancaire, des cartes grises, etc. » (Monjean-Decaudin 2012, p. 114)

au sein duquel trois catégories sont distinguées : les rapports, codes et lois, sans plus de précisions et en confrontation avec de nombreux genres relevant des discours scientifique, littéraire et essayiste. Leurs recherches concluent notamment à la stabilité des parties du discours en tant que marqueurs du genre textuel. De leur côté Gledhill et al. (2017) ont exploré l'utilisation des schémas lexico-grammaticaux sur deux corpus juridiques comparables : un corpus de 660 directives européennes et leur transposition dans 129 textes de loi et ordonnances en droit français, avec l'hypothèse que ces schémas lexico-grammaticaux permettraient d'identifier des genres et types discursifs.

Cependant, aucun de ces travaux n'a cherché à caractériser les textes de la procédure judiciaire sur le plan discursif, les considérant comme appartenant au domaine plus vaste du juridique. À travers les expérimentations décrites ci-dessous, nous cherchons à caractériser des textes de loi et des textes impliqués dans des processus judiciaires en fonction de la distribution des catégories morphosyntaxiques, afin d'observer leurs proximités et leurs démarcations.

3. Corpus

Nous avons rassemblé un corpus composé de huit sous-corpus de textes relevant au sens large du domaine de la justice, pouvant tous s'inscrire dans le cadre de l'appel du numéro thématique de la revue TAL cité précédemment :

— trois codes : le code pénal, le code civil et le code du patrimoine. Un code est un recueil de lois s'appliquant à un certain domaine ou à certaines situations.

— sept numéros du Journal Officiel de la République Française datant de l'automne 2018. Il s'agit du quotidien officiel en France, entièrement dématérialisé depuis 2016, qui publie et répertorie les événements législatifs, réglementaires, les déclarations officielles et les publications légales

— le « bloc de constitutionnalité » : il s'agit d'un ensemble composé du texte intégral de la Constitution de 1958, du préambule de 1946, de la Charte de l'environnement de 2005, et de la Déclaration des droits de l'homme et du citoyen de 1789. Cet ensemble de textes regroupe les normes constitutionnelles françaises appliquées par le Conseil Constitutionnel.

— un sous-corpus de 190 jugements de cour d'appel. Ces jugements sont rendus après contestation des décisions prises par les juridictions de premier degré et nouvel examen des faits ayant mené à l'ouverture d'une action en justice.

— un sous-corpus de 190 jugements de la Cour de cassation. La Cour de cassation est une cour de dernier ressort, qui examine après le procès initial et l'appel la conformité de l'application des règles de droit. La Cour de cassation ne revient donc pas sur l'examen des faits.

— 370 procès-verbaux d'auditions de témoins réalisées au cours d'une enquête criminelle. L'audition de témoin est une technique d'enquête consistant à recueillir les propos de personnes pouvant apporter des informations relativement à des faits répréhensibles.

— un sous-corpus de 10 ordonnances de renvoi devant le tribunal correctionnel (ORTC). L'ORTC est le document délivré à l'issue d'une instruction par le juge lorsque celui-ci estime que les éléments sont suffisants pour maintenir les charges et présenter l'affaire devant une formation correctionnelle de jugement ;

— un sous-corpus de 3 ordonnances de mise en accusation (OMA). Ce document est l'équivalent du précédent dans la procédure criminelle. Le ou les accusés sont alors renvoyés devant une cour d'assises.

Les textes ont été collectés sur le site LegiFrance³ pour les codes et les jugements, le site du Journal Officiel⁴, et le site du Conseil Constitutionnel⁵. Les auditions ont été mises à notre disposition dans le cadre d'un projet de recherche avec le Pôle Judiciaire de la Gendarmerie Nationale (PJGN). Enfin, les sous-corpus d'ORTC et d'OMA proviennent de différents cabinets d'instruction.

Les sous-corpus qui contenaient des éléments susceptibles d'identifier des situations ou des personnes ont été partiellement ou entièrement anonymisés.

Sous-corpus	Nb. de textes	Nb. de mots	Moyenne
Codes	3	538 000	179 000
Journal Officiel	7	579 000	82 700
Bloc constitutionnel	4	15 000	5 000
Jurisprudence d'appel	190	524 000	2 760
Jurisprudence de Cassation	190	595 000	3 130
Auditions de témoins	370	577 000	1560
OMA	3	25 700	8 600
ORTC	10	133 000	13 300
Total	777	2 986 700	3840

Tableau 1. Statistiques du corpus

4. Méthode et analyse

Le corpus est intégré dans le logiciel de textométrie TXM (Heiden et al., 2010), afin de mettre en évidence les critères internes de distinction des textes judiciaires et des textes juridiques. Nous aurons recours notamment à une méthode d'analyse multivariée, l'AFC.

³ <https://www.legifrance.gouv.fr/> [consulté le 16 janvier 2020]

⁴ <https://www.journal-officiel.gouv.fr/> [consulté le 16 janvier 2020]

⁵ <https://www.conseil-constitutionnel.fr/> [consulté le 16 janvier 2020]

procédure judiciaire, émanant d'organes juridictionnels. Toutefois, on note également l'isolement des ORTC, qui ne semblent caractérisées par aucune catégorie morpho-syntaxique en particulier, alors que l'on aurait attendu une proximité avec les OMA, en raison de leurs fonctions pénales similaires, l'une relevant de la procédure correctionnelle et l'autre de la procédure criminelle. Pour approfondir cette observation, nous avons généré à l'aide de la fonction classification de TXM le cluster présenté à la figure 2. Celui-ci présente un rapprochement entre les OMA et les ORTC.

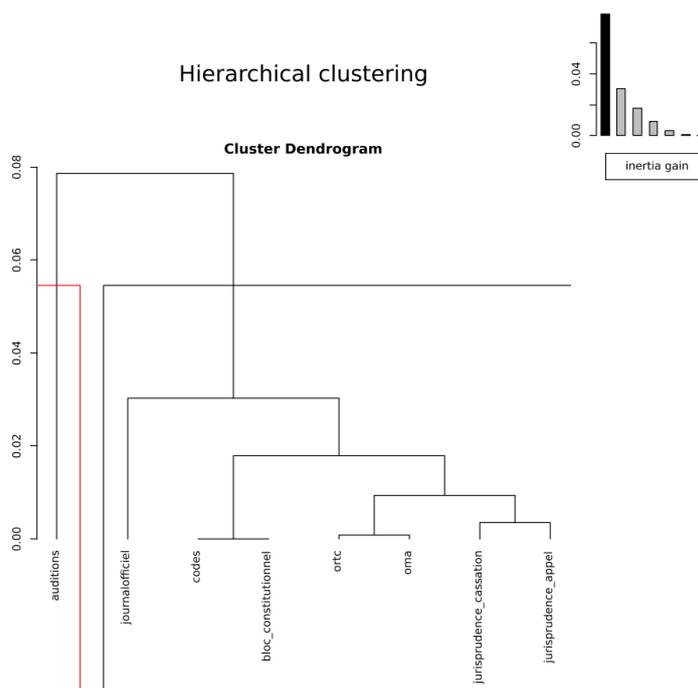


Figure 2. Classification en deux classes

Dans la suite de l'article, nous examinons les parties du discours caractéristiques de chaque sous-corpus.

Journal Officiel. Le sous-corpus Journal Officiel est caractérisé par les ponctuations, en particulier de citation, les valeurs numériques et les symboles. Ceci peut être interprété comme caractéristique des références à d'autres textes (décrets, lois), ainsi que de la mention de dates et de valeurs numériques diverses (pourcentages, sommes d'argent). L'exemple ci-dessous en est représentatif :

Par arrêté du Premier ministre et du ministre de l'action et des comptes publics en date du **24** septembre **2018**, le taux mentionné au dernier alinéa de l'article **11** du décret n° **99-945** du **16** novembre **1999** modifié portant statut particulier du corps des administrateurs civils et au **IV** de l'article **1er** du décret n° **2005-1090** du **1er** septembre **2005** relatif à l'avancement de grade dans le corps des administrations de l'État est fixé à **24 %** pour les promotions prononcées au titre des années **2019**, **2020** et **2021**.

On note aussi la proximité des verbes au futur dans des phrases déclaratives à valeur injonctive. L'injonction est ici bien présente dans la relation d'autorité entre le locuteur (l'État) et les destinataires (les citoyens). Le futur marque la fonction d'annonce publique pour la diffusion de lois et décrets, mais également d'offres d'emplois publics, de modalités de concours de la fonction publique, de dates de réunions parlementaires, etc.

Art. 3. – Toute battue collective au cours de laquelle des chevrotines **seront** employées **devra** être inscrite sur un registre retiré auprès de la fédération départementale des chasseurs.

Le titulaire de l'emploi **assurera** la fonction d'adjoint au directeur de la direction de la législation fiscale.

Jurisprudence. Les deux sous-corpus de jurisprudence, rassemblés en un seul pôle pour l'analyse, se caractérisent par les participes présents et passés, ce que l'on peut interpréter comme la trace de la construction du raisonnement et le rapport des faits investigués, de la procédure déjà réalisée et du jugement initial.

ARRÊT : Contradictoire, prononcé publiquement par mise à disposition de l'arrêt au greffe de la cour, les parties en **ayant été** préalablement avisées conformément à l'article 450 al 2 du CPC.

Il produit en outre un courrier adressé par l'agence régionale de santé, le 26 octobre 2012, **constatant** l'absence de traitement de l'eau de la piscine, ceci **présentant** un risque pour les baigneurs.

En application des dispositions de l'article 945-1 du code de procédure civile, l'affaire a été **débatue** le 29 Octobre 2013 à 14 H 00, en audience publique, les parties ne s'y **étant** pas opposées.

Groupe législatif. Les codes et le bloc constitutionnel sont un « pôle » peu marqué par les parties du discours. Les articles en sont la catégorie la plus caractéristique, mais il est plus intéressant d'examiner les verbes à l'infinitif qui sont un peu plus éloignés. En consultant leurs cooccurrents les plus fréquents, on trouve des verbes porteurs de la modalité déontique : « peut », « peuvent », « pourra », « doit », « doivent » et « ne ». Cette modalité traduit la fonction coercitive de ces deux groupes de textes, précisant ce que chacun peut et doit (ou ne peut et ne doit) faire pour être en accord avec la loi :

Article 4 Toute personne **doit contribuer** à la réparation des dommages qu'elle cause à l'environnement, dans les conditions définies par la loi.

Toutefois, ces autorités **ne peuvent procéder** à la célébration du mariage entre un Français et un étranger que dans les pays qui sont désignés par décret.

On note également la proximité des pronoms indéfinis, dont la fonction référentielle prend une dimension exprimant l'universalité de l'application de la loi :

4. **Tout** homme persécuté en raison de son action en faveur de la liberté a droit d'asile sur les territoires de la République.

5. **Chacun** a le devoir de travailler et le droit d'obtenir un emploi.

Article 221-5 Le fait d'attenter à la vie d'**autrui** par l'emploi ou l'administration de substances de nature à entraîner la mort constitue un empoisonnement.

Auditions de témoins. Les auditions de témoins sont caractérisées par l'imparfait, ainsi que par différents pronoms et par les interjections. L'imparfait, temps du récit, est le signe de l'aspect narratif de ce sous-corpus. Les pronoms correspondent à la pratique de l'audition, au cours de laquelle un témoin relate son expérience. Celle-ci est consignée à la première personne. Enfin, les interjections (essentiellement « oui » et « non ») sont une trace de l'aspect oral transcrit⁷.

Les filles **étaient** maltraités, d'après ce que **je** sais il y avait peut-être de la violence et la mère **allait** travailler en laissant les filles enfermées à la maison.

Question : **Vous** est-il arrivé au cours de cette journée de boire de l'alcool ?

Réponse : **Oui, j'**ai bu deux verres de vin au repas de midi.

OMA et ORTC. Dans la rédaction des ordonnances de mise en accusation ainsi que des ordonnances de renvoi devant le tribunal correctionnel, l'emploi des temps est codifié : on utilisera l'imparfait ou le présent de l'indicatif pour rappeler les faits.

Les sapeurs pompiers **intervenaient** à 17h00. Le foyer de l'incendie se **situait** dans la cave. [OMA].

Les recherches effectuées pour localiser X s'**avéraient** vaines. Un mandat d'arrêt **était** délivré à Y à son encontre. [ORTC]

Le présent de l'indicatif est utilisé pour la partie renseignant la personnalité du justiciable ainsi que pour la partie dite « PCM » (introduite par « par ces motifs ») indiquant la décision :

Le casier judiciaire de X **porte** la mention NEANT [ORTC].

L'examen psychiatrique ne **met** en évidence aucun ensemble psychopathologique significatif [OMA].

[Nous] **ORDONNONS** LE RENVOI DE L'AFFAIRE DEVANT LE TRIBUNAL CORRECTIONNEL [ORTC]

⁷ Les exemples sont rapportés tels que figurant dans les corpus, c'est-à-dire en conservant coquilles et fautes d'orthographe.

Ordonnons la mise en accusation de M. X devant la Cour d'Assises de A (...).
[OMA]

Les pronoms relatifs indiquent des phrases complexes avec des subordonnées décrivant les faits.

Dans la cave se trouvait un lit, **dont** ne subsistait que l'armature métallique suite à l'incendie. Le mur, contre **lequel** était apposé le lit, était le plus abîmé de tout le sous-sol. [OMA]

Le <date>, le chantier " X " à YY, faisait l'objet d'un contrôle au cours **duquel** Y (alias Z) et W étaient trouvés en situation irrégulière et en état de travail pour le compte de la société T), **laquelle** selon témoignage serait en fait dirigée par XX, ancien gérant de V et simple salarié de T. [ORTC]

Les démonstratifs rappellent les protagonistes et les éléments acquis au cours de l'enquête. Ce sont des marqueurs qui structurent la progression de la narration.

Les proches de X témoignaient des bons rapports qu'avait toujours entretenus **cette** dernière avec ses parents ainsi qu'avec son frère et ses sœurs. [OMA]

Les premiers éléments de l'enquête confirmaient la teneur de **ces** propos.
[ORTC]

Les ordonnances de mise en accusation se situent à proximité d'un groupe dense de parties du discours qui semblent à mi-chemin entre le groupe législatif et les auditions, groupe dans lequel on repère la présence des verbes au conditionnel. Dans les auditions, l'emploi du conditionnel correspond au futur dans le passé. Cet usage est caractéristique du discours indirect et correspond à des propos tiers rapportés par le témoin entendu.

Il a dit qu'il n'avait pas le temps mais qu'il **viendrait** à sa pause.

Il m'a dit de ne pas inquiéter les parents, que dès qu'elle **renterait**, il me **rappellerait**.

Le conditionnel est également employé pour formuler des hypothèses sur les faits ou les personnes, le plus souvent dans un emploi d'irréel du passé.

Il ne comprends pas, et s'il avait eu le moindre soupçon il en **aurait** parlé.

Si Christine lui avait annoncé son intention de ne plus se marier, Thierry se **serait** effondré et serait venu nous voir.

Dans les ordonnances de mise en accusation, le conditionnel est employé de la même manière avec une valeur de discours indirect dans lequel le futur est transposé ainsi que pour l'émission d'hypothèses.

Il se souvenait que X lui avait dit que si son frère les voyait ensemble, il la **tuerait**.

Dans l'hypothèse d'un bâillon réalisé avec un adhésif type Chatterton retrouvé sous le lit, des traces **devraient** cependant être visibles au niveau de la sphère buccale de la victime, ce qui n'était pas constaté.

Dans le groupe législatif, l'emploi du conditionnel permet de prévoir les conditions d'une situation et donc les modalités d'application du texte. Il s'agit de nouveau de formulation d'hypothèses.

Dans le cas où la déclaration **aurait** été omise ou **serait** erronée, la rectification de l'acte, en ce qui touche l'omission ou l'erreur, pourra être effectuée conformément à l'article 99-1.

Ceux de ces textes qui **interviendraient** après l'entrée en vigueur de la présente Constitution ne pourront être modifiés par décret que si le Conseil constitutionnel a déclaré qu'ils ont un caractère réglementaire en vertu de l'alinéa précédent.

5. Synthèse

L'étude des parties du discours des différents sous-corpus a permis notamment de constater un continuum entre normativité et oralité du discours dans les textes relatifs à la justice. Ceci est particulièrement visible entre les groupes *législatif*, *ordonnances* et *auditions*. Le *groupe législatif* est marqué par l'aspect coercitif, son but étant d'énoncer la norme. Le *groupe des ordonnances* comporte des éléments narratifs, employés dans la partie d'exposé des faits, et des éléments normatifs relatifs à la qualification des faits. Enfin, les *auditions* se distinguent par des éléments énonciatifs du style oral transcrit. La finalité de ce sous-genre est en effet de recueillir des propos.

Ce continuum manifesté dans les genres textuels reflète l'articulation du processus judiciaire, où la norme du droit (*groupe législatif*) est appliquée (*ordonnances*) à une situation réelle (*auditions*).

6. Conclusion

L'étude que nous avons proposée ne balaie qu'une infime partie des types de documents générés par et pour la justice. Le corpus que nous avons constitué, bien qu'arbitraire et compilé selon les opportunités dont nous disposions, a toutefois permis de démontrer la variété et les proximités entre différents types de textes grâce à un critère quantifiable en la forme des parties du discours. Ainsi, notre étude vient compléter la classification proposée par S. Monjean-Decaudin et C. Bocquet qui distingue entre autres textes judiciaires, textes juridiques, et textes juridico-judiciaires. De leur point de vue, cette classification est basée sur la finalité des textes, or nous avons constaté que cette finalité conditionne également leur contenu. À ce sujet, il serait intéressant d'étudier le lexique et la phraséologie de chaque genre dans des travaux futurs afin d'évaluer le lien entre spécialisation de la langue et distribution des catégories morphosyntaxiques.

Références

- Benzécri, J. P. (1973). *L'analyse des données* (Vol. 2, p. 1). Paris: Dunod.
- Biber, D. (1991). *Variation Across Speech and Writing*. Cambridge University Press.
- Bocquet, C. (2008). *La traduction juridique : Fondement et méthode*. Traducto. Bruxelles : De Boeck.
- Branca-Rosoff, S. (1999). Des innovations et des fonctionnements de langue rapportés à des genres. *Langage & société*, 87(1), 115-129. doi :10.3406/lso.1999.2856
- Gledhill, C., Patin, S. et Zimina, M. (2017). Lexico-grammaire et textométrie : identification et visualisation de schémas lexico-grammaticaux caractéristiques dans deux corpus juridiques comparables en français. *Corpus*, (17).
- Heiden, S., Magué, J.-P. et Pincemin, B. (2010). TXM : Une plateforme logicielle open-source pour la textométrie – conception et développement. *JADT 2010 : 10th International Conference on the Statistical Analysis of Textual Data* (p. 12 p.). Rome, Italie : ENS-Lyon.
- Jacques, M.-P. et Aussenac-Gilles, N. (2006). Variabilité des performances des outils de TAL et genre textuel. Cas des patrons lexico-syntaxiques. *Traitement Automatique des Langues*, 47(1), 22.
- Malrieu, D. et Rastier, F. (2001). Genres et variations morphosyntaxiques. *Traitement Automatique des Langues*, 42(2), 548-577.
- Monjean-Decaudin, S. (2012). *La traduction du droit dans la procédure judiciaire*. Dalloz.
- Nadeau, D. et Sekine, S. (2007). A survey of named entity recognition and classification. *Linguisticae Investigationes*, 30(1), 3-26.
- Nazarenko, A. et Wyner, A. (2017). Legal NLP Introduction. *TAL*, 58(2), 8-18.