



ISSN 1724-0700

ISSN en ligne 2260-8087

## Corpus, *Web* et recherche terminologique : le cas des termes du domaine du bien-être animal

**Cosimo De Giovanni**

Université de Cagliari, Italie

cdegiovanni@unica.it

<https://orcid.org/0000-0003-1976-3166>

Reçu le 10-02-2021 / Évalué le 15-03-2021 / Accepté le 16-04-2021

### Résumé

Dans cet article, nous avons voulu démontrer l'apport des corpus numériques à la recherche terminologique. Nous avons analysé la circulation des deux termes *abattoir mobile* et *unità mobile di macellazione* dans les textes officiels de l'Union européenne (UE), des deux États membres, comme la France et l'Italie, qui réglementent le domaine du bien-être animal (BEA). L'utilisation du corpus numérique a été essentielle à partir du moment où les variantes dénominatives des deux termes pouvaient être extraites et décrites. Ainsi, l'interprétation des contextes d'apparition des candidats termes se révèle fondamentale pour l'analyse du point de vue (PdV) et des choix linguistiques opérés par les locuteurs sur la scène discursive. Les corpus numériques peuvent être utiles pour la construction de nouvelles banques de terminologie et pas pour l'amélioration de la qualité de celles déjà existantes.

**Mots-clés :** *web-corpus*, *corpus-web*, point de vue, corpus heuristique, bien-être animal

### Corpus, *Web* e ricerca terminologica: il caso dei termini nel campo del benessere degli animali

### Riassunto

In questo articolo abbiamo voluto dimostrare l'apporto dei corpora digitali alla ricerca terminologica. Abbiamo analizzato la circolazione dei due termini *abattoir mobile* e *unità mobile di macellazione* nei testi ufficiali dell'Unione europea (UE), dei due Stati membri, quali Francia e Italia, che regolamentano il settore del benessere animale (BEA). Il ricorso al corpus digitale è stato essenziale dal momento in cui si sono potute estrarre e descrivere le varianti denominative dei due termini. Allo stesso modo, l'interpretazione dei contesti in cui compaiono i termini candidati si rivela fondamentale per l'analisi del punto di vista (PdV) e delle scelte linguistiche operate dai relatori sulla scena discorsiva. I corpora digitali possono essere utili per la costruzione di nuove banche terminologiche e non solamente per il miglioramento della qualità di quelle già esistenti.

**Parole chiave:** *web-corpus*, *corpus-web*, punto di vista, corpus euristico, benessere animale

## Corpus, web and terminological research: the case of terms in the field of animal welfare

### Abstract

In this article we wanted to demonstrate the contribution of numerical corpora to terminological research. We have analyzed the circulation of the two terms *abattoir mobile* and *unità mobili di macellazione* in the official texts of the European Union (EU), of the two Member States, i.e. France and Italy, which regulate the animal welfare sector. The use of the numerical corpus was essential from the moment in which the denominative variants of the two terms could be extracted and described. Thus, the interpretation of the contexts in which candidate terms appear is fundamental for the analysis of the point of view (PdV) and of the linguistic choices made by the speakers on the discursive scene. Digital corpora can be useful not only to improve the quality of existing terminology banks but also to build new ones.

**Keywords:** web-corpus, corpus-web, point of view, heuristic corpus, animal welfare

### Introduction

Ces dernières années, on a assisté à une réduction de l'écart entre le corpus et le web dans les études de linguistique. Les raisons qui ont amené à ce rapprochement sont multiples et diversifiées : changement de cap de certains linguistes, ayant condamné l'utilisation du *Web* dans leurs études, prédominance de l'Internet comme média de communication et d'information (Mille, 2014), développement de nouvelles technologies d'analyse des données brutes dans la Toile (*big data*), absence de grands corpus textuels informatisés pour certaines langues et, enfin, liberté accordée par la linguistique de corpus (LdC) à l'utilisation d'outils et de méthodes d'analyse (Teubert, 2009).

En particulier, dans les années 2000, la rencontre du *Web* et de la LdC a entraîné beaucoup de questions, d'incertitudes, de résistances, voire des débats entre divers groupes de linguistes : à la notion de *Web as Corpus* (Kilgariff, 2001) s'opposait le *Web for building corpus* ou le *Web for Corpus* (Hundt, Nesselhauf, Biewer, 2007). D'un côté, à l'instar du corpus, le *Web* est considéré comme « un grand corps de texte en attente d'être exploité, un énorme tissu de données linguistiques souvent entrelacées avec des fils parallèles<sup>1</sup> » (Resnik, Smith, 2003 : 350).

De l'autre, en revanche, le *Web* est vu comme une ressource linguistique, compte tenu du fait qu'il « ne correspond en rien aux caractéristiques habituelles d'un corpus » (Duclaye *et al.*, 2006 : 53).

Cette opposition - qui a entraîné un changement dans la définition traditionnelle du terme de *corpus* et de ses outils (Gatto, 2014 : 149) - a laissé place, au fil

des ans, à deux nouveaux venus, tant sur le plan terminologique que sur le plan conceptuel : le *corpus-web* (Schaeffer-Lacroix, 2020) et le *web-corpus* (Mayaffre, 2010). Tous les deux peuvent être inscrits sous le terme de *corpus numériques*. Les deux approches se prêtent volontiers aux exigences de l'analyse linguistique des données collectées sans opposer les démarches qualitative et quantitative. De plus, elles sont source d'enrichissement et d'intégration de données déjà collectées à partir de corpus traditionnels (parmi les corpus spécialisés, e.g. *Scientext* du *Lidilem*) ou de produits terminographiques déjà existants (e.g. *IATE*) ou en phase d'élaboration.

### 1. Le *corpus-web*

Un *corpus-web* est un corpus constitué à partir des données repérées dans le *Web* à l'aide de logiciels, appelés *crawler*, *web crawler* ou *web spider*<sup>2</sup> (Schäfer, Bildhauer 2013) ayant pour fonction de parcourir périodiquement la toile et d'en collecter le contenu afin de permettre au moteur de recherche de l'analyser et de l'indexer, en identifiant les mots clés et les sujets abordés. Des exemples de *corpus-web* sont le *Web as Corpus Kool Ynitiative (WaCKy)* constitué principalement de trois corpus, respectivement en français, en italien et en allemand, d'environ deux milliards de mots chacun et les corpus *TenTen* repérables dans l'outil en ligne *Sketch Engine*.

L'une des limitations majeures du *corpus-web* est représentée par les droits d'auteur. Pour contourner ce problème, certains corpus contiennent des textes disponibles gratuitement et dépourvus de droits d'auteur ou sous licence *Creative Commons*. Un exemple de cela nous est donné par le corpus italien *Piattaforma per l'Apprendimento dell'Italiano Su corpora Annotati (PAISÀ)* avec environ 250 millions de *token*. D'autres corpus sont composés de textes réduits à des échantillons de phrases, c'est le cas du *Leipzig Corpora Collection*, ou en d'autres unités respectant les limites du droit d'auteur. Il existe aussi des corpus qui contiennent des textes qui ne présentent pas ce type de contrainte ou qui restent dans un cadre légal par rapport aux droits d'auteur : e.g. les corpus publiés sur *Sketch Engine*, le *Hungarian Webcorpus* avec 1,48 milliard de mots ou encore le *Leeds Collection of Internet Corpora*.

### 2. Le *web-corpus*

L'utilisation de moteurs de recherche à des fins linguistiques devient de plus en plus fréquente. Les partisans des deux approches ont toujours reconnu les faiblesses du *Web* dans la recherche linguistique. On peut évoquer, à titre d'exemple, la difficulté de déterminer le degré de représentativité des domaines et des langues,

le caractère incontrôlable des données, la présence de fautes (*dirty corpus*) et, enfin, le caractère anarchique du *Web* (Kilgarriff, Grefenstette, 2003 ; De Giovanni, 2009 ; Maniez, 2011). Ces difficultés sont surmontables à condition d'appliquer des critères linguistiques au cours de la recherche ; dans le cas d'une utilisation à des fins terminologiques, à condition de combiner une analyse par domaines et par usages et, dans les cas extrêmes, d'avoir recours à l'avis d'experts (Lerat, 2003 : 231).

Or, l'introduction du nouveau terme de *web-corpus* sert à démanteler la relation à sens unique existant entre *Web* et *corpus* (*Web-corpus*). De cette manière, il n'est plus question de savoir si le *Web* peut être vu comme un corpus, mais de savoir si le corpus peut exister sans le *Web* (*corpus-Web*) (Mayaffre, 2010).

Au-delà de tous les problèmes de nature épistémologique, on peut affirmer que le *web-corpus* est un corpus constitué des données brutes extraites du *Web* à usages linguistiques. Ces données brutes peuvent être traitées aussi bien *hors-ligne* - à l'aide d'un logiciel externe, gratuit ou payant, (e.g. *AntConc*, *WordSmith*) - qu'*en ligne* - à l'aide d'un moteur de recherche interne au logiciel (e.g. *Webcorp* : Renouf, 2009). Un autre exemple de *web-corpus* est l'outil *Ngram viewer* - un moteur de recherche qui permet de fournir des données fréquentielles sur la distribution chronologique de n-grams (n-mots) - qui repose sur la base des données textuelles de *Google Livres*.

### 3. Corpus et recherche terminologique

L'abandon de l'approche wüsterienne du terme, considéré comme une étiquette du concept, a été décisif lorsqu'une vision textuelle de la terminologie s'est imposée, ouvrant la voie à de nouveaux modèles théoriques (Raus, 2013). L'examen du fonctionnement du terme dans un contexte linguistique est désormais au cœur de l'activité interprétative du linguiste / terminologue. Aujourd'hui, on reconnaît au terme son statut tant d'élément linguistique que d'élément cognitif (Lerat, 1995 ; Zanola, 2018). De plus, dans la perspective d'une terminologie communicationnelle, le terme possède différentes valeurs en fonction du contexte situationnel en vue de répondre à des besoins de communication spécifiques d'un groupe de professionnels. Des changements affectant le monde (sur le plan social, culturel, scientifique, technique, etc.) peuvent entraîner la variation terminologique d'un ou de plusieurs domaines<sup>3</sup>.

En nous inspirant des travaux de Raus (2013) et Condamines *et al.* (2004), nous proposons une analyse, d'ordre qualitatif et quantitatif, des variations dénominatives et conceptuelles dans le domaine du bien-être animal (dorénavant BEA).

Dans le cadre des variations dénominales, nous accorderons une large place à l'analyse des variantes morphologiques des unités repérées. Suivra une analyse sémantique des variantes.

Afin d'observer le phénomène de la variation, une analyse sur corpus s'impose. Un premier corpus servira à l'extraction des termes en langue française et italienne, relevant du domaine du BEA, sur une ligne temporelle restreinte (de 1991 à 2009). Ce corpus parallèle est formé de textes émanant de l'Union européenne (dorénavant UE) repérés sur le site *EUR-Lex*. Un deuxième corpus se composera des textes législatifs promulgués en France et en Italie, entre 1991 et 2009. Enfin, l'utilisation de corpus numériques sera indispensable pour vérifier le comportement des termes en contexte.

Pour nos analyses, nous ferons appel à deux autres notions capitales, celles de *candidat terme* et de *point de vue* (dorénavant PdV). On définit le *candidat terme*<sup>4</sup> comme une occurrence extraite d'un corpus susceptible d'être un terme (Delavigne, 2001 ; L'Homme, 2004). Nous incluons dans cette catégorie non seulement le candidat terme de validation (qui est supposé être terminologique ou non terminologique) mais aussi tout candidat qui entre en concurrence avec les termes officiels ou sont officiellement reconnus par les organismes compétents ou sont traités dans des bases de données terminographiques. La deuxième notion est celle de PdV en langues spécialisées (Boisson 2001) se rapportant à la *perception* d'un objet dans la réalité et dépendant des choix (linguistiques) opérés de la part d'un ou plusieurs locuteurs à des fins spécifiques. Dans le même esprit que Condamines et Rebeyrolle (1997), nous tenterons d'analyser les choix linguistiques adoptés de deux PdV : un premier PdV concernant les choix faits aux niveaux supérieurs, par le terminologue et le législateur ; et, enfin, un second PdV qui concerne les choix linguistiques opérés par un groupe de locuteurs lors de l'actualisation d'une compétence lexicale dans une situation de communication concrète.

#### 4. *Abattoirs mobiles, unità mobili di macellazione* et leurs variantes

##### 4.1. Variabilité dénomminative et conceptuelle dans les textes de l'Union européenne et dans le *Web*

Bien que le BEA reste encore une notion controversée, elle part d'un principe fondateur considérant l'animal comme un être *sensible* (Chessa, De Giovanni, 2019). La présence de sensibilité chez l'animal implique qu'il possède des capacités sensorielles et cognitives et que son bien-être est alors un état mental subjectif, dérivé de son rapport harmonieux avec l'environnement qui l'entoure. Le BEA doit être garanti même au moment de leur mise à mort, surtout dans le cas des animaux

d'élevage destinés à la consommation, en abattoir. Le transport jusqu'aux abattoirs (Le Neindre et al., 2009 : 72) pourrait nuire au BEA du fait de mauvais traitements pendant le chargement et déchargement. Pour cette raison, l'UE a tenté de mieux réglementer les opérations de transport avec une série d'instruments législatifs tout en favorisant, dans le même temps, le déploiement de nouveaux types d'abattoirs afin de réduire la souffrance des animaux pendant la période de pré-abattage.

Le Règlement n°1099 de 2009, sur la protection des animaux au moment de leur mise à mort, prévoit le recours aux *abattoirs mobiles* afin de réduire « la nécessité pour les animaux d'être transportés sur de longues distances [ce qui peut...] contribuer à préserver leur bien-être » et laisse aux États membres le soin d'« établir ou de maintenir des règles nationales pour les abattoirs mobiles ». Le même Règlement distingue les abattoirs mobiles des abattoirs fixes en raison des « techniques de contraintes » et enfin, il estime nécessaire que « les règles techniques [soient] adaptées » (considérant n° 40).

Dans les trois langues de travail du Conseil de l'UE on trouve les termes *abattoir mobile* (fr.), *mobile Schlachthöfe* (all.) et *mobile slaughterhouse* (ang.). Au niveau dénomiatif, le terme italien *unità mobile di macellazione* présente une structure binaire complexe sous forme d'une suite d'expressions enchaînées : {[unità mobile] [di macellazione]}. Au niveau conceptuel, le premier élément *unité* évoque l'objet sous un angle différent : celui d'un groupe d'appareils destiné à effectuer une opération. De même, *macellazione* nous offre un autre PdV : on tient compte de l'ensemble des opérations avec lesquelles les animaux d'élevage, destinés à l'alimentation, sont abattus. Nous pouvons donc voir que le choix du terme italien n'est pas entièrement aléatoire. La présence des deux termes *unità mobili di macellazione* et *unità fissa di macellazione*, nous amène à penser que, sur le plan formel, nous sommes confrontés à une unité plurilexémique semi-ouverte avec des possibilités combinatoires limitées. En effet, une interrogation du moteur de recherche Google (4 février 2021) nous a permis de constater la présence du candidat terme *unità di macellazione* (22 700 résultats)<sup>5</sup>, notamment dans les deux secteurs agroalimentaire et juridique. De plus, nous avons constaté que le terme *unità di macellazione* constitue la base pour trois expressions complexes comme *unità di macellazione autonoma* (gérée directement par une ferme ou une entreprise), *unità di macellazione di animali* et *unità di macellazione annua*. L'interrogation de Google nous a donné aussi une nouvelle dénomination, celle d'*unità di macellazione mobile* (6, 63)<sup>6</sup>. Si dans le terme du Règlement l'élément *mobile* se rattache à la tête de l'expression *unità*, dans le second terme le même élément est placé à la fin : dans ce cas, le déplacement concerne expressément tout le groupe d'appareils destiné à l'abattage. En termes linguistiques, c'est la position des traits sémantiques au sein

de chaque sémème qui motive les deux dénominations. Il est important d'ajouter que l'absence d'un candidat terme comme *unità di macellazione fissa* parmi les données obtenues est due au fait que, à notre avis, l'expression *unità di macellazione* dénomme déjà un abattoir fixe.

Toutefois, les termes *abattoir mobile* et *unità mobile di macellazione* ne sont pas les seuls à circuler dans les textes officiels de l'UE. En fait, par l'interrogation du moteur de recherche du site *EUR-Lex* on découvre qu'il y a au total, pour le français, trois variantes et, pour l'italien, quatre variantes, désignant l'abattoir mobile, distribuées dans les différents types législatifs et réparties sur une période comprise entre 1991 et 2004. En voici quelques exemples (tableau 1) :

Année de publication du texte	fr	It
1991	unités mobiles d'abattage	unità di macellazione mobili
1992	installations mobiles d'abattage	impianti mobili di macellazione
1994	unités mobiles d'abattage	impianti mobili di macellazione
2004	abattoirs mobiles	unità di macellazione mobili / mattatoi mobili / macelli mobili

**Tableau 1 :** Évolution des variations dénominatives dans les textes de l'Union européenne.

Ayant travaillé dans une diachronie restreinte, où la fenêtre temporelle couvre environ dix ans, le résultat que nous avons obtenu est celui d'une variation restreinte.

#### 4.2 Variabilité dénomminative et conceptuelle dans les textes législatifs nationaux : France et Italie

Les actes juridiques nationaux de chaque État membre de l'UE sont un point de départ utile pour la recherche terminologique, surtout lorsqu'il est possible de comparer des méthodes de dénomination et de conceptualisation de deux langues différentes au sein d'un même domaine. Ceci est également possible grâce à l'instrument d'harmonisation législative pour lequel l'UE demande à ses États membres de modifier leur législation nationale pour atteindre un objectif commun. Une harmonisation des pratiques peut entraîner, sous réserve d'exceptions, une harmonisation des systèmes conceptuels et dénominatifs : tout cela dépend surtout de la politique linguistique de chaque pays (Lavault-Olléon, Grossmann, 2008). L'interrogation des portails *Légifrance* et *Normativa*, les sites officiels pour la diffusion et l'accès aux droits français et italien, nous a permis d'obtenir de

premiers résultats, même si l'interrogation du portail italien a été plutôt difficile. En effet, de ce dernier, nous n'avons pu obtenir qu'une seule donnée utile aux fins de notre recherche : le terme *macello mobile* contenu dans le texte d'une loi de 1994<sup>7</sup>. Or, la langue italienne dispose de deux mots, d'usage commun d'après le *Dizionario internazionale* en ligne (De Mauro), pour désigner le lieu destiné à l'abattage des animaux : ceux de *macello* et de *mattatoio*. L'interrogation de la base de données de *Google Books*, à travers le logiciel *N-gram viewer*, nous a permis de vérifier les occurrences des deux mots entre 1500 et 2019 : le mot *macello* est présent à partir de 1502 et continue son évolution jusqu'à nos jours, tandis que l'apparition du mot *mattatoio* remonte aux débuts du XIX<sup>e</sup> siècle<sup>8</sup>. En interrogeant le moteur de recherche *Google*, nous avons pu constater une différence dans la fréquence d'occurrences des termes *mattatoio mobile* (1 680, 1 860) et *macello mobile* (552, 851). Nous n'avons pas d'explication à sur cet écart entre les deux termes. Même *Sketch Engine (itTenTen16)* confirme ces derniers résultats. De plus, à travers l'interrogation de la base de données *Google Books* nous apprenons que le terme *macelli mobili* a fait l'objet d'une mention dans deux documents datant des années 40 (*Bollettino parlamentare* et *Rivista delle colonie*) lors de la gestion de la production et de l'abattage de viande par les Italiens en Érythrée.

Au sein des documents officiels français, on a pu constater que face à la seule attestation de l'expression *installation mobile d'abattage* la dénomination courante est *abattoir mobile* (6 fois). Dans le *Rapport d'enquête sur les conditions d'abattage des animaux de boucherie dans les abattoirs français* de 2016, présenté à l'Assemblée nationale, sont mentionnés deux types d'abattoir mobile : les *caissons d'abattage* (un prolongement des abattoirs traditionnels) destinés uniquement à l'abattage et à la saignée à la ferme ; et les (*camions*) *abattoirs mobiles* totalement autonomes construits sur les modèles de camions (articulés) suédois (de l'entreprise Hälsingestintan). En France, le recours aux abattoirs mobiles a non seulement l'objectif de préserver le BEA, mais aussi celui de réduire les coûts de transport trop élevés pour les fermes dispersées dans les campagnes françaises.

Ce qui est intéressant, en revanche, c'est la présence dans les deux langues des expressions *macello itinerante* (144, 4) et *abattoir itinérant* (2 300, 278), identifiées pour la première fois lors de l'interrogation de *Sketch Engine* (une seule attestation par langue) et employées comme des variantes de *macello mobile* et d'*abattoir mobile*. Dans les deux langues la dénomination est motivée de la même façon : de par son étymologie, le concept d'*itinérant* est lié à celui de voyage, par conséquent les termes concernés renvoient à l'activité de se déplacer d'un lieu à l'autre, tout en accomplissant des fonctions et des opérations liées à l'abattage.



## Conclusion

Les résultats que nous avons obtenus ne sont évidemment pas définitifs, mais ils constituent une première approximation que de nouvelles analyses permettront d'intégrer. Cependant, ces premiers résultats nous amènent à faire quelques considérations finales. Le BEA, étant un domaine interdisciplinaire situé à la frontière entre le droit, la médecine, la psychologie animale, l'éthique et la technique, regorge de terminologies. Malgré la tentative d'harmoniser le système législatif et le système terminologique de tous les États membres de l'UE, le BEA se caractérise par la présence de variantes terminologiques qui circulent dans des contextes différents. Par exemple, dans notre cas, il a été possible de constater que la circulation des termes varie beaucoup, en français et en italien. Ce qui s'explique par l'exploitation de textes différents dans les deux langues au sein du corpus numérique.

Au-delà des questions épistémologiques (Mayaffre, 2010), le corpus numérique constitue une ressource d'une valeur incontestable. Parmi les deux types de corpus numérique, c'est certainement le *web-corpus* qui pose le plus de problèmes, surtout en termes de volatilité des données collectées et de rapidité avec laquelle elles peuvent varier. Cependant, à travers des « méthodes générales d'investigation » (Lerat, 2003 : 231), le *web-corpus* peut très bien se prêter à des analyses en termes quantitatifs et qualitatifs.

En général, le corpus numérique est un support utile non seulement du fait de sa fonction de validation, mais aussi en raison de sa fonction heuristique dans la mesure où il se propose de donner « à un ensemble de faits une structure qui dégage la vue du linguiste dans une situation où celui-ci poursuit une question mais n'a pas de solution » (Scheer, 2004 : 153-154). Cela pourrait contribuer à faire du corpus numérique une ressource à partir de laquelle il est possible de construire directement des banques terminologiques.

Une dernière considération à faire concerne l'interprétation linguistique des données. Elle constitue le résultat auquel le linguiste est parvenu suite à ses analyses. C'est un point de vue (ou un des points de vue) du linguiste / terminologue sur les PdV des autres locuteurs « présents » dans le discours (Kleiber, 2009). Un point de vue discutable, qui ne fait parfois pas l'unanimité, mais qui, dans le cadre des langues de spécialité, pourrait également être valable si une évaluation de l'expert pour comprendre les variations n'est pas requise.

## Bibliographie

- Chessa, F., De Giovanni, C. 2019. « Propositions d'interventions terminologiques dans le domaine du bien-être animal (français-italien) », *Rhesis - International Journal of Linguistics, Philology and Literature*, n° 10.1, p. 34-55.
- Condamines, A., Rebeyrolle, J., Soubeille, A. 2004. « Variation de la terminologie dans le temps : une méthode linguistique pour mesurer l'évolution de la connaissance en corpus », In : *Proceedings of the 11th EURALEX International Congress*, p. 547-557.

- Condamines, A., Rebeyrolle, J. 1997. « Point de vue en langue spécialisée », *Meta : le journal de traducteurs*, n° 42, p.174-184.
- De Giovanni, C. 2009. *Le corpus entre généralités et spécificités*. Bari / Paris : Schena / Alain Baudry et Cie.
- Delavigne, V. 2001. « Repérage de termes dans un corpus de vulgarisation : aspects méthodologiques ». In : *Conférence TIA*, Nancy, p. 33-43.
- De Mauro, T. *Dizionario internazionale*. [En ligne] : <https://dizionario.internazionale.it> [consulté le 04 février 2021].
- Duclaye, F., Collin, O., Petrier, E. 2006. « Fouille du Web pour la collecte de données linguistiques : avantages et inconvénients d'un corpus hors-normes », In : *Actes de l'atelier Fouille du Web des 6èmes Journées Francophones Extraction et Gestion des Connaissances*, p. 53-64.
- Gatto, M. 2014. *Web as Corpus. Theory and Practice*. London : Bloomsbury.
- Hundt, M., Nesselhauf, N., Biewer, C. 2007. *Corpus Linguistics and the Web*. Amsterdam : Rodopi.
- Kleiber, G. 2009. « D'un contexte à l'autre : aspects et dimensions du contexte ». *L'Information Grammaticale*, n° 123, p.17-32.
- Kilgariff, A., Grefenstette, G. 2003. « Introduction to the Special Issue on the Web as Corpus », *Computational Linguistics*, vol. 9, n° 3, p. 334-347.
- Kilgarriff, A. 2001. « Web as corpus ». In : *Proceedings of the Corpus Linguistics Conference (CL 2001)*. Lancaster : University of Lancaster, p. 342-344.
- Lavault-Olléon, E., Grossmann, F. 2008. « Langue du droit et harmonisation terminologique multilingue : l'exemple de LexALP ». *Lidil*, n° 38, p.11-32.
- Le Neindre et al. 2009. *Douleurs animales : les identifier, les comprendre, les limiter chez les animaux d'élevage*. Expertise scientifique collective, Synthèse du rapport, INRA.
- Lerat, P. 1995. *Les langues spécialisées*. Paris : PUF.
- Lerat, P. 2003. « Web et terminologie philologique ». *Linguistica Antverpiensia*, n° 3, p. 231-244.
- L'Homme, M.-C. 2004. *La terminologie : principes et techniques*. Montréal : PUM.
- Maniez, F. 2011. « Quelques tendances récentes de la linguistique de corpus ». In : Maniez, F. (éd.). *Unité et diversité de la linguistique*. Lyon : Université de Lyon, p. 67-85.
- Mayaffre, D. 2010. « Corpus et web-corpus. Réflexion sur la corporalité numérique ». *Cahiers de praxématique*, n° 54-55, p. 233-248.
- Mille, A. 2014. « D'Internet au web », In : Vitali-Rosati, M., Sinatra, Michael E. (éds). *Pratiques de l'édition numérique*. Montréal, Les Presses de l'Université de Montréal, p. 31-48.
- Raus, R. 2013. *La terminologie multilingue. La traduction des termes de l'égalité H/F dans le discours international*. Bruxelles : De Boeck.
- Renouf, A. 2009. « Corpus Linguistics beyond Google : the WebCorp Linguist's Search Engine ». *Digital Studies/le Champ Numérique*, n° 1.
- Resnik, P., Smith, N. A. 2003. « The Web as a Parallel Corpus ». *Computational Linguistics*, n° 3(29), p. 349-380. [En ligne] : <https://www.aclweb.org/anthology/J03-3002.pdf> [consulté le 04 février 2021].
- Schaeffer-Lacroix, E. 2020. « Les corpus web à travers le prisme de l'ALMT ». *Corpus*, n° 20. [En ligne] : <http://journals.openedition.org/corpus/4579> [consulté le 04 février 2021].
- Schäfer R., Bildhauer F. 2013. « Web Corpus Construction ». *Synthesis Lectures on Human Language Technologies*, n° 6(4), p. 1-145.
- Scheer, T. 2004. « Le corpus heuristique : un outil qui montre mais ne démontre pas ». *Corpus*, n° 3. [En ligne] : <http://journals.openedition.org/corpus/210> [consulté le 04 février 2021].
- Simondon, G. 1958. *Du mode d'existence des objets techniques*. Paris : Aubier.

Teubert, W. 2009. « La linguistique de corpus : une alternative ». *Semen. Revue de sémio-linguistique des textes et discours*, n°27. [En ligne] : <https://journals.openedition.org/semen/8923> [consulté le 04 février 2021].

Zanola, M.T. 2018. *Che cos'è la terminologia?*. Rome: Carocci.

### Sitographie<sup>9</sup>

Creative Commons

<https://creativecommons.org/licenses/?lang=fr>

Eur-lex

<https://eur-lex.europa.eu/homepage.html?locale=fr>

Hungarian Webcorpus

<http://mokk.bme.hu/en/resources/webcorpus>

Leeds Collection of Internet Corpora

<http://corpus.leeds.ac.uk/internet.html>

Leipzig Corpora Collection

[https://corpora.uni-leipzig.de/en?corpusId=eng-za\\_web\\_2014](https://corpora.uni-leipzig.de/en?corpusId=eng-za_web_2014)

Ngram viewer

<https://books.google.com/ngrams>

Piattaforma per l'Apprendimento dell'Italiano Su corpora Annotati - PAISÀ

<https://www.corpusitaliano.it>

Sketch Engine

<https://www.sketchengine.eu>

WaCKy - Web as Corpus Kool Ynitiative

<https://wacky.sslmit.unibo.it/doku.php?id=corpora>

### Notes

1. « *As a great big body of text waiting to be mined, a huge fabric of linguistic data often interwoven with parallel threads* » (notre traduction).
2. Le J.O. du 01/01/2013 recommande l'utilisation du terme *collecteur*.
3. La reconstruction du « mode d'existence » des objets techniques et scientifiques et l'identification de la « carrière objectale » aideraient aux processus de conceptualisation et de formation des termes (Simondon, 1958).
4. Le terme « candidat » doit être ici pris dans son sens originel de 'vêtu de blanc' [du lat. *candidatus*]. En effet, à l'époque de la Rome antique, les citoyens, aspirant à une fonction publique, portaient une toge blanche pour indiquer qu'ils étaient au-dessus de tout soupçon.
5. Le recours à l'opérateur *wildcard*, nous a permis de repérer une seule expression correspondant à la séquence *unità.... di macellazione* : *unità produttiva di macellazione*.
6. Entre parenthèses le premier chiffre se réfère au singulier et le deuxième au pluriel.
7. <https://www.gazzettaufficiale.it/eli/gu/1994/12/16/293/so/162/sg/pdf>
8. Selon le De Mauro, le mot *macello* a été introduit en Italie au XIV<sup>e</sup> siècle du lat. *macĕllu(m)* dans le sens de 'marché de viande, poisson et légumes', alors que celui de *mattatoio* est passé en italien, au XIX<sup>e</sup> siècle, par le verbe *mattare* dans le sens de 'tuer, mettre à mort'. On pourrait penser que la différence entre les deux termes était d'abord une affaire étymologique et plus tard une affaire de représentation mentale.
9. La date de dernière consultation des sites est le 4 février 2021.