

Quel rôle pour le corpus dans la modélisation ontoterminologique multilingue : l'exemple de la balance des paiements

Role of the corpus in ontoterminological modelling: the case of the balance of payments

Stéphane Carsenty

Université Savoie Mont Blanc (Listic, Equipe Condillac)
et Banque nationale suisse (Service linguistique)

stephane.carsenty@univ-smb.fr

ORCID : 0000-0002-9767-3492

Abstract

This paper presents the role a specialised corpus can play in the ontoterminological modelling (Roche 2007b) of the balance of payments (IMF, 2009). It presents the domain of the balance of payments and the characteristics of a corpus of texts produced in English, in French and in German by thirteen central banks and statistics offices, which will be used for terminological work. The analysis leads to a typology of the texts grouped in the corpus. Far from only providing examples for knowledge collected from experts, this corpus will be analysed from the very stage of elaborating the concept system.

Keywords: multilingual terminology, ontoterminology, corpus, experts, linguistic dimension

1. INTRODUCTION

Notre recherche vise à élaborer une ontoterminologie (Roche, 2007b) du domaine de la balance des paiements, lequel regroupe les activités associées à l'établissement et à l'utilisation d'un état statistique qui décrit les relations d'une économie avec le reste du monde. Cette ontoterminologie prendra la forme d'un dictionnaire électronique multilingue, qui présentera de manière structurée les connaissances du

domaine et les termes utilisés pour les décrire en français, en anglais et en allemand, constituant une introduction au domaine à la fois pour les futurs experts, les utilisateurs et les traducteurs de la statistique en question.

L'ontoterminologie s'inscrit dans une approche conceptuelle de la terminologie. Il s'agit de construire une terminologie dont le système de concepts est une ontologie formelle, afin d'en permettre l'opérationnalisation. L'objet de cet article est de définir le rôle que le travail sur le corpus peut jouer dans la modélisation ontoterminologique multilingue.

Si l'on commençait par élaborer le système de concepts exclusivement à partir d'interactions avec les experts (statisticiens), l'affaire serait entendue, et le corpus n'entrerait en scène qu'après, pour jouer un rôle de confirmation et d'instanciation d'informations déjà pour l'essentiel collectées. Or la nature du corpus sur la balance des paiements nous amène à penser qu'il est pertinent de donner sa place à une partie du corpus dès l'étape d'élaboration du système de concepts. Le choix de cette partie du corpus à traiter sera alors déterminé par les différents types de textes qu'elle regroupe, d'où la nécessité d'une typologie.

Cet article est structuré comme suit : (1) Nous présentons d'abord le domaine de la balance des paiements. (2) Nous passons ensuite en revue les caractéristiques du corpus en cours de construction, présentant notamment la typologie qui a émergé lors de sa constitution. (3) Nous en déduisons enfin le rôle que les différentes parties de ce corpus pourront jouer dans l'élaboration et l'enrichissement de l'ontoterminologie du domaine.

2. LA BALANCE DES PAIEMENTS : UN DOMAINE SOCIALEMENT CONSTRUIT DONT LES OBJETS SONT IMMATÉRIELS

2.1. PRINCIPALES CARACTÉRISTIQUES

Publiée à intervalles réguliers par les banques centrales et les offices statistiques des Etats membres du Fonds monétaire international (FMI), la balance des paiements (BP) est un état statistique qui décrit toutes les relations d'une économie¹ avec le reste du monde (FMI, 2009, p. 9). Ces relations sont représentées dans des agrégats statistiques correspondant à des flux, à des stocks et à des revenus, qui sont enregistrés dans des comptes internationaux renvoyant à des objets macroéconomiques (importations, exportations, actifs, passifs, flux financiers, etc.). Les objets² auxquels nous avons

¹ On entend ici par économie une entité nationale ou supranationale régie par des règles homogènes (pays, union économique, telle que l'Union européenne, ou union monétaire, comme la zone euro).

² « Tout ce qui peut être perçu ou conçu » (ISO 1087 : 2019, § 3.1.1).

affaire dans le domaine de la balance des paiements sont donc immatériels (Carsenty, 2020, p. 84). Le domaine de la BP peut être schématiquement représenté par un enchaînement de triangles sémiotiques (voir fig. 1) : un premier triangle correspond aux objets macroéconomiques observés, un deuxième, aux agrégats statistiques calculés par les statisticiens pour représenter et manipuler ces objets, et un troisième symbolise les comptes dans lesquels ces agrégats sont enregistrés.

La BP est une construction sociale (Carsenty, 2020, p. 85). Les statisticiens l'établissent dans le cadre conceptuel défini par le FMI (FMI, 2009, FMI, 2014) et complété par les travaux d'autres organisations internationales (comme Eurostat) et nationales (banques centrales et offices statistiques). Les organisations nationales ou régionales peuvent adopter des règles supplémentaires : il existe ainsi un cadre européen applicable à l'établissement de la BP (cf. UE, 2012 et UE, 2016).

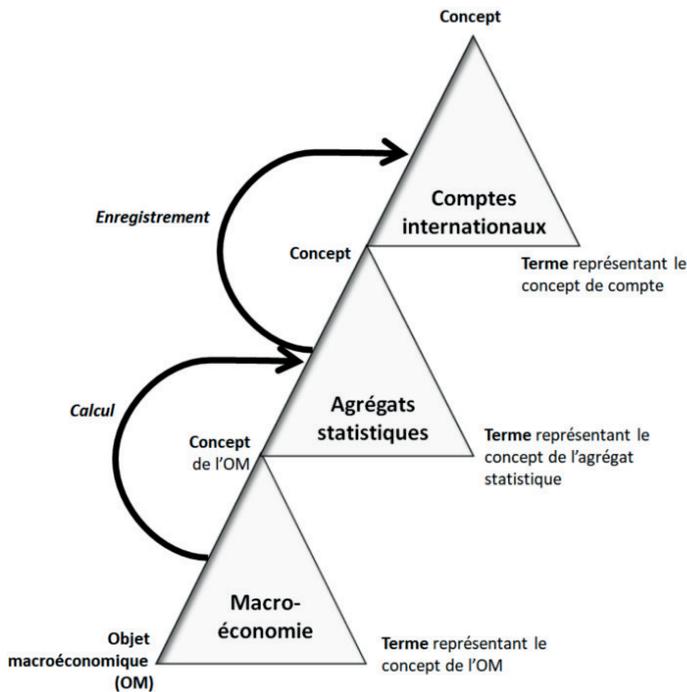


Figure 1. Représentation du domaine de la balance des paiements

Source : représentation propre, d'après Sowa (2000, p. 193).

Les comptes internationaux s'appuient sur le même socle conceptuel que les comptes nationaux : dans les deux cas, on observe des interactions entre des agents économiques et on les agrège, avec les mêmes outils conceptuels, dans des unités institutionnelles, elles-mêmes regroupées dans des secteurs institutionnels, en fonction de leur apparte-

nance à une économie donnée (Banque mondiale *et al.*, 2013, p. 61 ; FMI, 2009, p. 52). Tandis que le système de comptabilité nationale (SCN) « mesure les activités qui ont lieu au sein d'une économie » (Banque mondiale *et al.*, 2013, p. 2), les comptes internationaux accomplissent la même tâche pour les processus en jeu « entre les résidents et les non-résidents » (FMI, 2014, p. 3), c'est-à-dire entre une économie et le reste du monde.

2.2. ÉLABORATION ET UTILISATION

On peut schématiser les activités liées à l'élaboration et à l'utilisation de la BP à l'aide d'un modèle simplifié de communication linéaire (Adler & Rodman, 2006, p. 12) : l'information est « encodée » par l'émetteur et « décodée » par le récepteur. Les statisticiens observent des phénomènes macroéconomiques et collectent des données, qu'ils « encodent » ensuite dans des agrégats et enregistrent dans des comptes. Ce sont des « activités et tâches destinées à convertir des données collectées en informations statistiques » (UNECE, 2019, p. 4 ; notre traduction). Ces informations sont ensuite « décodées » par les utilisateurs de la BP, qui les « convertissent » en connaissances sur des phénomènes et processus macroéconomiques. En fonction des centres d'intérêt de ces utilisateurs, ces connaissances peuvent être utilisées « pour l'analyse économique, la prise de décisions et l'élaboration des politiques » (Banque mondiale *et al.*, 2013, p. 1).

La communication effective est plus complexe : ainsi, la première phase du modèle général des processus applicables en matière de statistiques économiques est destinée à spécifier les besoins des utilisateurs, et donc à sonder ces derniers (UNECE, 2019, p. 11). Mais un modèle simplifié, limité aux seules activités d'encodage et de décodage, constitue une approximation opératoire pour comprendre le domaine.

En résumé, la BP est donc la représentation d'objets macroéconomiques, calculés dans des objets statistiques et présentés dans des comptes. C'est ce domaine dont nous entendons élaborer l'ontoterminologie, c'est-à-dire une terminologie dont le système de concepts est une ontologie formelle (Roche, 2007b).

3. UN CORPUS TRILINGUE SUR LA BALANCE DES PAIEMENTS

3.1. POURQUOI S'APPUYER SUR UN CORPUS

Nous considérons avec Roche (2007b, p. 3), Costa & Silva (2008, p. 6) et Santos & Costa (2015, p. 155) que la terminologie possède une double dimension, à savoir une dimension linguistique et une dimension conceptuelle. Élaborer une ontoterminologie à partir de textes revient donc à considérer qu'il est possible de s'appuyer sur un corpus spécialisé pour construire le système de concepts d'un domaine et d'apparier ce dernier aux termes de ce domaine. Cela correspond à l'idée selon laquelle « c'est

[...] à travers l'acte de la parole, c'est-à-dire le discours, que nous pouvons accéder à la représentation [des] connaissances » (Costa & Silva, 2008, p. 6).

Toutefois, cet accès aux concepts n'est pas aisé dans les textes du fait que les concepts relèvent par définition de l'extralinguistique. Le non-dit est en effet « une des propriétés les plus caractéristiques du texte de spécialité », notamment du fait que « dans toutes les situations de communication, on a tendance à signifier plus que ce que l'on dit » (Costa & Silva, 2008, p. 6), et le matériau lexical observable dans les textes est régi par « l'incomplétude des textes » (Roche, 2007a, p. 163), ce qui signifie que « la compréhension des textes, que le sens des mots, requiert des connaissances extralinguistiques qui par définition ne sont pas incluses dans les textes » (Roche, 2007a, p. 163). L'emploi de figures de rhétorique, de son côté, « suppose que l'auteur et le lecteur partagent la (une) même et préexistante conceptualisation du monde nécessaire à la compréhension de ces expressions » (Roche, 2007a, p. 164). Ainsi, l'ellipse est fréquente dans le domaine de la BP : les comptes dans lesquels sont enregistrés les agrégats statistiques calculés à partir des objets macroéconomiques sont souvent simplement désignés en omettant la mention « compte de » (Carsenty, 2019, p. 32).

Dans ces conditions, pourquoi s'appuyer sur un corpus ? Si un corpus est un ensemble de textes considérés comme représentatifs d'une langue, regroupés à des fins d'analyse linguistique (Tognini-Bonelli, 2001, p. 2), les corpus spécialisés sont plus appropriés pour comprendre la langue employée dans des contextes spécialisés que les corpus de langue générale (Flowerdew, 2004, p. 15). Tognini-Bonelli (2001, p. 2) situe la linguistique de corpus dans le cadre d'une théorie contextuelle et fonctionnelle de la signification. Ce point est important pour qui entend travailler sur des corpus à des fins terminologiques : il amène à poser la question de la distinction à opérer entre le signifié, dont s'occupe la linguistique, et le concept, qui est au centre du travail terminologique. « La manipulation des textes implique des connaissances extralinguistiques du domaine très pointues » (Costa & Silva, 2008, p. 12). Si nous voulons construire une ontoterminologie à partir d'un corpus, il sera donc nécessaire de disposer de moyens d'accès à cet extralinguistique.

3.2. COLLECTE DES TEXTES : AUTHENTICITÉ, ÉCHANTILLONNAGE ET REPRÉSENTATIVITÉ

Si « le texte écrit n'a assurément pas la même valeur dans toutes les communautés professionnelles » (Costa & Silva, 2008, p. 12), la BP est un domaine très normalisé, et cette caractéristique présuppose l'existence de textes dont une fonction est précisément d'établir et de manifester cette normalisation.

Nous constituons un corpus spécialisé de textes sur la BP produits par les organisations suivantes : Banque centrale du Luxembourg, Banque centrale européenne, Banque de France, Banque nationale de Belgique, Banque nationale suisse, Bureau

of Economic Analysis (États-Unis), Banque d'Angleterre, Deutsche Bundesbank (Allemagne), Central Statistics Office (Irlande), Committee on Monetary, Financial and Balance of Payments Statistics (Union européenne), Commission européenne, Eurostat, Fonds monétaire international, Organisation des Nations Unies, Österreichische Nationalbank (Autriche), Office for National Statistics (Royaume-Uni), Statistiques Canada et Statec (Luxembourg). Il s'agit d'organisations (banques centrales, offices statistiques ou autres) d'entités nationales ou supranationales comparables (pays de l'Union européenne et de la zone euro, États-Unis et Canada), ainsi que d'organisations internationales comme le FMI. Ces organisations produisent des textes en français, en anglais et/ou en allemand³.

Tognini-Bonelli (2001, pp. 55-62) mentionne trois aspects importants à prendre en compte lors de la construction et de l'utilisation de corpus : l'authenticité, l'échantillonnage et la représentativité. Nous examinons à présent la constitution du corpus sur la BP sous ces trois angles et détaillons les critères de collecte des textes.

Pour ce qui est de l'*authenticité*, ce critère est d'entrée de jeu rempli ici, puisqu'il n'est pas question de créer le corpus de manière « artificielle » ou « expérimentale » (Tognini-Bonelli, 2001, p. 55) : nous avons collectés ces documents à partir des sites Internet des institutions mentionnées et demandé à des experts chargés d'établir la BP dans ces institutions de nous signaler d'éventuels oublis.

Ce point nous conduit aux critères appliqués à la collecte, notamment à l'*échantillonnage*. Ces critères portent sur les aspects suivants : l'année de parution, le cadre conceptuel, la sélection ou non de fragments et le choix entre corpus statique et corpus dynamique.

Pour la plupart de ces critères, la manière dont se présente le paysage des textes produits sur la BP est déterminante. Le principal manuel de référence actuel est le Manuel de la Balance des paiements et de la position extérieure globale, 6^e édition (FMI, 2009). Depuis la première édition en 1948, « chaque nouvelle édition [...] répond à l'évolution économique et financière, à des changements dans les intérêts des analystes et à l'accumulation d'expérience de la part des statisticiens » (FMI, 2009, § 1.17, p. 3). FMI 2009 explore trois nouveaux champs : « la mondialisation, l'importance croissante de l'analyse de bilan et l'innovation financière » (FMI, 2009, 1.31, p. 4). Il introduit à cette fin de nouveaux concepts et de nouveaux termes pour les désigner. Ce processus entraîne l'extension de l'usage de termes rares ou absents, ainsi que la production de premières définitions, si un consensus se dégage à l'échelle internationale. Nous le voyons avec le terme « entité à vocation spéciale » (EVS), lequel désigne un concept correspondant à un objet dont l'existence n'est possible que dans une économie mondialisée caractérisée par une forte innovation financière. Il apparaît

³ Les institutions qui établissent la balance des paiements publient en effet leurs documents dans leur(s) langue(s) officielle(s), ainsi que très souvent en anglais si celui-ci n'est pas la langue officielle.

dans la 5^e édition du Manuel (1993) – mais sans définition – et demeure introuvable dans les éditions antérieures.

Cet exemple illustre ce que Picht nomme le cycle de vie d'un concept (Picht, 2013, p. 11), c'est-à-dire la période durant laquelle un concept est considéré comme valable et utilisé en tant que tel (2013, p. 14). Dans le cas de la BP, il est assez aisé de déterminer le cycle de vie des concepts et des termes, essentiel pour l'échantillonnage du corpus : les statisticiens appliquent les principes de FMI 2009 depuis 2014. Si les concepts ont changé depuis 2009, aucun consensus stable n'a encore émergé, et la nouvelle édition du Manuel ne paraîtra pas avant 2025.

C'est pourquoi nous adoptons une perspective synchronique dans le cadre conceptuel de FMI 2009. Pour être sélectionné, un document doit donc relever du cadre conceptuel de FMI 2009 ou être compatible avec ce cadre. Concrètement, il s'agit de textes produits à partir de 2009 et se référant à FMI 2009, et de textes produits avant 2009 mais toujours d'actualité (documents réglementaires et certains articles de recherche).

Nous ne retenons que des documents entiers, et non des fragments. Nous estimons en effet avec Bowker (1996, pp. 42-43) que c'est le meilleur moyen de ne pas exclure des segments qui pourraient contenir des informations utiles sur les plans linguistique et terminologique.

Le corpus sur la BP est dynamique : des textes y seront ajoutés au fur et à mesure de leur publication sur les sites web des institutions.

Enfin, pour ce qui concerne la *représentativité*, si Tognini-Bonelli (2001, p. 57) indique qu'il n'y a aucun moyen de la garantir ou de l'évaluer objectivement, notons avec Rastier (2004, en ligne) « [qu']un corpus est adéquat ou non à une tâche en fonction de laquelle on peut déterminer les critères de sa représentativité et de son homogénéité. ». Notre corpus sur la BP est un corpus spécialisé authentique. Il est représentatif du discours employé par les institutions étudiées car les critères d'échantillonnage (perspective synchronique et limitation à un cadre conceptuel bien défini) nous amènent à y regrouper des textes typiques du domaine, comme nous allons le voir. Sa représentativité est donc « d'ordre qualitatif, assise sur la typicalité et la spécialisation » (Doualan, 2018, p. 6, § 17).

3.3. CARACTÉRISTIQUES DU CORPUS EN COURS DE CONSTRUCTION

L'analyse des textes collectés nous amène à distinguer cinq types (voir fig. 2), en fonction des « *situations de communication* » (Maingueneau, 2016, p. 55). Partir des situations de communication permet de tenir compte de « *l'inscription sociale des énoncés* » (Maingueneau, 2016, pp. 56-57). Cette typologie s'intéresse à « *l'identité des locuteurs* », à « *la finalité* » et au « *dispositif* » mis en œuvre (Charaudeau, 2009, p. 50). Elle est réalisée à des fins d'analyse lexicale pour identifier les termes et les concepts que ceux-ci désignent.

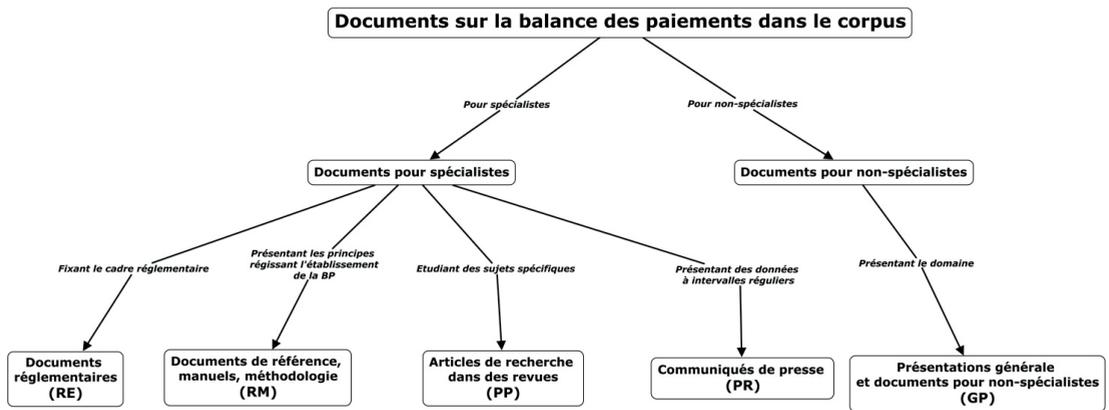


Figure 2. Typologie des documents produits sur la BP

Source : représentation propre.

Le corpus est subdivisé en suivant deux axes : les connaissances nécessaires pour comprendre les textes, et le contenu et les objectifs de ces textes. Nous présentons ce second axe, et les cinq types qui en découlent :

- Textes portant sur des thématiques réglementaires, p. ex. lois stipulant que la BP doit être établie par la banque centrale : nous les regroupons sous le type « Documents réglementaires » (ci-après, **RE**).
- Textes présentant les principes fondamentaux du domaine (échantillonnage des données, calcul des agrégats, conduite d'enquêtes statistiques, etc.) : ces documents sont attribués au type « Documents de référence, manuels et méthodologie » (**RM**).
- Textes portant sur des champs de recherche spécifiques publiés notamment dans des revues spécialisées, que nous classons sous le type « Articles de recherche dans des revues » (**PP**).
- Textes contenant des données sur la BP publiées à intervalles réguliers : nous les attribuons au type « Communiqués de presse » (**PR**).
- Textes qui constituent une présentation générale du domaine de la BP pour les non-spécialistes (**GP**).

Qui dit « *situation de communication* » dit émetteurs et récepteurs. La figure suivante (fig. 3) présente les destinataires de ces types de textes, en fonction de leurs activités respectives : statisticiens, universitaires, acteurs des marchés financiers, autorités (p. ex. gouvernements) et non-spécialistes. Chacun d'eux a des besoins spécifiques lorsqu'il accède aux connaissances portées par ces textes.

Le corpus comprend 504 textes, pour un total d'un peu moins de 36 millions de caractères. L'importance respective des langues est fortement influencée par la présence de documents de référence internationaux en anglais (51 %) et en français (40 %). Si nous réduisons le corpus aux seuls textes produits par des organisations

nationales, la part des textes en allemand passe de 9 % à 19 %, Mais celle des textes en anglais s'accroît également, passant à 58 %.

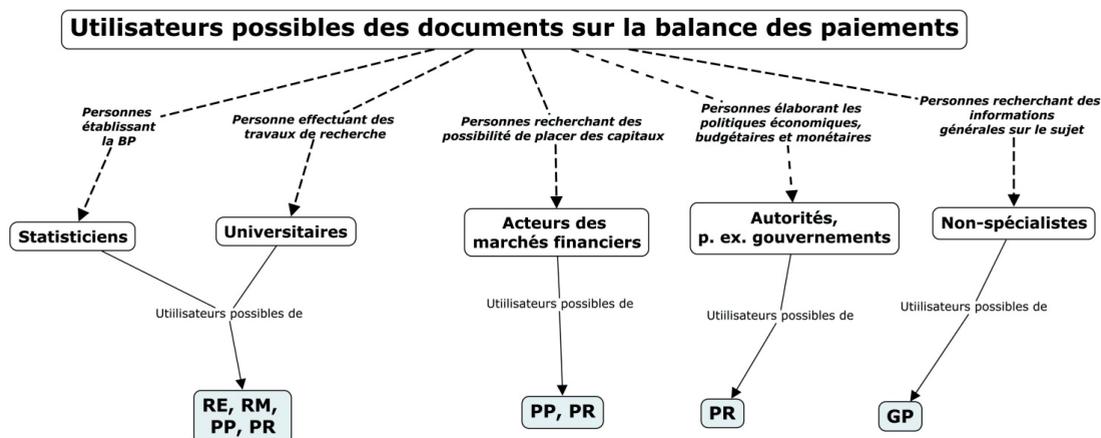


Figure 3. Typologie des utilisateurs possibles des documents sur la BP

Source : représentation propre.

Les deux principaux types sont les documents de référence, manuels et documents sur la méthodologie (RM, 62 textes, près de 22 millions de caractères) et les communiqués de presse (PR, 322 textes, près de 10 millions de caractères). Il apparaît que 99 % des textes ont été publiés depuis 2009. Les raisons en ont été indiquées à la section 3.2. Si nous étudions le nombre de textes, nous observons une prédominance des RM et des RE avant 2014, et une majorité de PR pour ceux produits depuis : les textes produits jusqu'à 2014 visaient à installer le cadre conceptuel et réglementaire de FMI 2009. Depuis, les principes de FMI 2009 sont mis en œuvre et se retrouvent donc dans les communiqués de presse produits par les institutions.

4. RÔLE DU CORPUS DANS LA MODÉLISATION ONTOTERMINOLOGIQUE DE LA BALANCE DES PAIEMENTS

4.1. EXPERTS ET CORPUS

Pour construire le système de concepts de la BP, nous devons recourir aux connaissances des statisticiens. Le corpus, lui, nous permettra de disposer d'un ensemble de candidats termes et d'équivalents pour désigner ces concepts. Peut-il être utilisé dès la modélisation du système de concepts ? La perspective linguistique considère les textes comme de la connaissance en action (Santos & Costa, 2015, p. 169), et la désignation est, en discours, un point d'accès au concept (2015, p. 156).

Condamines (2009, p. 6) estime « que les corpus les plus riches en marqueurs de relations et donc les plus susceptibles d’être utilisés pour construire des ontologie/terminologies sont les corpus de type didactique comme les ‘manuels’. En effet, les auteurs de manuels utilisent fréquemment des contextes définitoires ou du moins des contextes riches en connaissance destinés à faire comprendre un domaine ou une pratique pour des non-experts ». Nous partageons ce point de vue. Une recherche dans AntConc avec la requête « défini ? »⁴ montre que près de 90 % des occurrences (1886/2102) se trouvent dans les RM (voir fig. 4). La fenêtre « Concordance » (voir fig. 5) fournit bel et bien une proportion importante de contextes riches en marqueurs linguistiques dans les RM. Le même phénomène se vérifie dans la partie anglaise du corpus. Or les RM constituent près des deux tiers de notre corpus.

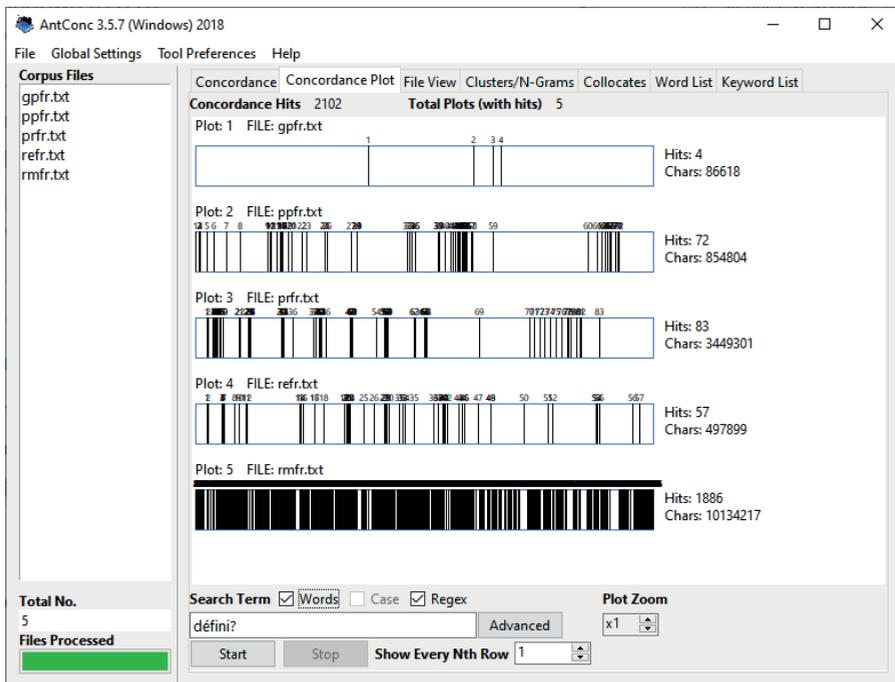


Figure 4. Distribution des résultats de la requête « défini ? » entre les cinq types de documents du corpus

Source : AntConc.

⁴ AntConc est un outil d’analyse de corpus (cf. <https://www.laurenceanthony.net/software/antconcl/>). La requête « défini ? » utilise une expression régulière (case « Regex » cochée) : « ? » est un caractère générique qui remplace un ou plusieurs caractères quelconques. La requête renvoie les occurrences telles que « définition », « défini », « définie », « définit », etc. Si toutes ne seront pas des contextes définitoires, c’est-à-dire des contextes fournissant les éléments de connaissances permettant de forger une définition, certains patrons (tels que « se définit comme ») sont particulièrement productifs. Il est par ailleurs entendu que des contextes définitoires seront également fournis avec d’autres requêtes.

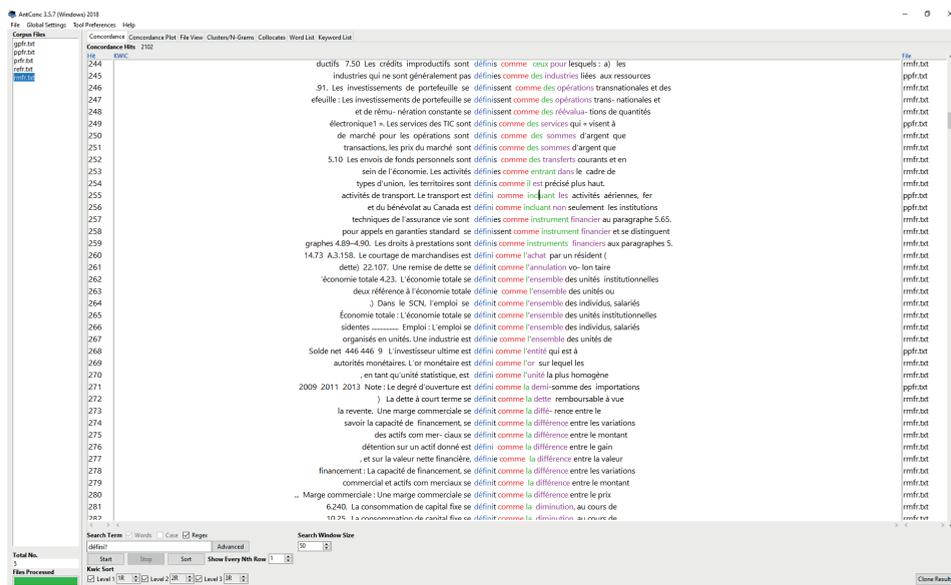


Figure 5. Contextes des marqueurs linguistiques : fenêtre Concordance

Source : AntConc.

4.2. APPROCHE ADOPTÉE

Ces bases posées, retenons le constat fait par Santos et Costa (2015, p. 171), à savoir que les connaissances acquises lors d’interactions avec les experts les aident à inclure dans leur modélisation des relations lexico-sémantiques et des concepts qui n’ont *pas* d’expression linguistique dans un corpus de textes. Autrement dit, en nous appuyant exclusivement sur les textes, nous risquons de ne pas identifier tous les concepts du domaine. Mais à l’inverse, aucun(e) expert(e) ne maîtrise suffisamment son domaine pour être la source unique de la construction du système de concepts.

Ayant la responsabilité de la traduction en français des documents sur la BP à la banque centrale suisse, nous avons acquis une semi-expertise sur le domaine, notamment en interagissant avec les experts. Aussi allons-nous construire le système de concepts en nous appuyant sur ces connaissances *et* sur le corpus, et en priorité sur sa partie constituées par les RM. Ce choix repose sur l’hypothèse suivante : la plupart des termes correspondant aux activités d’encodage (voir 2.2) devraient être présents dans les RM, et la majorité des termes concernant les activités de décodage devraient se trouver dans les PR. Cette hypothèse s’appuie sur les objectifs et groupes cibles respectifs de ces types de textes.

Pour en vérifier la validité, prenons l’exemple d’« unité institutionnelle ». Nous savons que c’est un terme qui désigne un concept dénotant les entités observées par

les statisticiens (en vue de l'activité d'encodage, donc), défini en langue naturelle comme une « entité économique en droit de posséder, de plein droit, des biens et des actifs, capable de prendre des décisions économiques, capable de conclure des contrats et possédant un ensemble complet de comptes ou pouvant en établir un » (Carsenty, 2020, pp. 91-92, d'après FMI, 2009, pp. 54-55). Si nous ignorions cela, nous devrions analyser les différents types de textes. Si le corpus d'analyse était restreint aux seuls communiqués de presse (voir fig. 6), l'unité ne serait pas retenue car elle est absente du nuage de candidats termes affichés dans TermoStat⁵. Par contre, elle apparaît pour les documents de référence et manuels⁶ (voir fig. 7).



Figure 6. Nuage des candidats termes extraits des communiqués de presse en français (corpus >>prfr)

Source : TermoStat.



Figure 7. Nuage des candidats termes extraits des documents de référence et manuels en français (corpus >>rmfr)

Source : TermoStat.

⁵ « Outil d'acquisition automatique de termes qui exploite une méthode de mise en opposition de corpus spécialisés et non spécialisés en vue de l'identification des termes » développé par Patrick Drouin à l'Université de Montréal, Canada (<http://termostat.ling.umontreal.ca/>).

⁶ Certes avec une coquille (« unités institutionnels »), due à un bogue.

L'onglet des contextes (phrases) de TermoStat (voir fig. 8) affiche les contextes riches en connaissance contenus dans les RM, qui permettront de construire une définition en langue naturelle similaire à celle susmentionnée. Le même phénomène apparaît si on répète cette expérience avec les textes en anglais.

comptes d'accumulation du SCN.
Une unité institutionnelle est réputée résidente sur le territoire économique d'un pays donné lorsqu'elle a, sur ce territoire, un centre d'intérêt économique prépondérant, c'est-à-dire lorsqu'elle y exerce ou a l'intention d'y exercer des activités économiques ou d'y effectuer des opérations relativement importantes, soit pour un temps indéfini, soit durant un laps de temps prolongé, notion généralement interprétée comme une année au moins.
En tant qu'agrégé de la production, le PIB d'un pays est égal à la somme des valeurs ajoutées brutes de toutes les unités institutionnelles résidentes qui exercent des activités de production (augmentées des éventuels impôts moins les subventions sur les produits, non inclus dans la valeur de leurs productions).
Une partie de la production d'une unité institutionnelle résidente peut en effet avoir lieu à l'étranger, ce qui est, par exemple, le cas de l'installation de machines ou d'équipements exportés ou bien celui de travaux effectués par une équipe d'experts-conseils envoyée temporairement à l'étranger.
Si la consommation est l'activité par laquelle les unités institutionnelles utilisent des biens ou des services, il en existe deux sortes très différentes.
L'activité de formation brute de capital fixe, à l'instar de la consommation intermédiaire, est le fait d'unités institutionnelles en leur qualité de producteurs.
La séquence des comptes, y compris ceux de patrimoine, dans le SCN pourrait en principe être établie à n'importe quel niveau d'agrégation, même à celui de l'unité institutionnelle individuelle.
Même si chaque unité institutionnelle tient une comptabilité, les concepts nécessaires ou appropriés au niveau microéconomique ne conviennent pas nécessairement au plan macroéconomique.
En règle générale, toute opération entre deux unités institutionnelles requiert systématiquement quatre entrées simultanées et équivalentes dans les comptes du SCN (c'est ce qu'on appelle le principe des quadruples entrées), même si l'opération constitue un transfert et non un échange et même si aucun montant monétaire n'est échangé.
Ces entrées multiples permettent d'enregistrer et d'analyser les interactions économiques entre unités institutionnelles et secteurs.
C'est ainsi que les nomenclatures des unités institutionnelles, des opérations et des actifs peuvent être mises en œuvre avec souplesse en introduisant davantage d'agrégation ou de désagrégation pour tenir compte de la disponibilité des données dans les différents pays et de leurs caractéristiques propres.
Elles sont appelées unités institutionnelles.
En outre, étant juridiquement responsables de leurs actes, les unités institutionnelles sont des centres de décision pour tous les aspects du comportement économique.
En pratique, certaines unités institutionnelles sont contrôlées par d'autres : leur autonomie de décision n'est alors pas complète, et elle peut varier au cours du temps.
Les unités institutionnelles sont regroupées en secteurs institutionnels sur la base de leurs fonctions, de leurs comportements et de leurs objectifs principaux : a. Les sociétés non financières sont des unités institutionnelles dont la fonction principale consiste à produire des biens et des services non financiers marchands.
b. Les sociétés financières sont des unités institutionnelles dont la fonction principale consiste à fournir des services financiers, y compris des services d'intermédiation financière.
d. Les ménages sont des unités institutionnelles composés soit d'un individu, soit d'un groupe d'individus ; toutes les personnes physiques de l'économie doivent appartenir à un seul ménage.
Un sous-secteur comprend des unités institutionnelles complètes, et chaque unité institutionnelle ne peut appartenir qu'à un seul sous-secteur, bien qu'il existe d'autres possibilités de regroupement.
Elle comprend toutes les unités institutionnelles qui résident sur le territoire économique d'un pays.
Une unité institutionnelle est considérée comme étant résidente d'un pays à partir du moment où elle possède un centre d'intérêt économique prépondérant sur le territoire économique de ce pays, c'est-à-dire qu'elle exerce des activités économiques sur ce territoire pendant une période prolongée (une année ou plus constitue une règle d'application pratique).
Dans la structure comptable du SCN, le reste du monde joue donc un rôle analogue à celui d'un secteur institutionnel, même si les unités non résidentes n'en font partie qu'à partir du moment où elles réalisent des opérations avec des unités institutionnelles résidentes.
Les unités institutionnelles remplissent diverses fonctions économiques (production, consommation, épargne, investissement, etc.).
), ont pour effet de créer, de transformer, d'échanger, de transférer ou de faire disparaître de la valeur économique ; elles entraînent des variations du volume, de la composition, ou ? 18 Système de comptabilité nationale de la valeur des actifs ou des passifs des unités institutionnelles.
La plupart des actions économiques sont menées d'un commun accord entre des unités institutionnelles.
Le SCN appelle « opérations » ces actions entreprises d'un commun accord entre deux unités institutionnelles.
Le SCN traite également comme des opérations certaines actions économiques qui ne concernent qu'une seule unité institutionnelle : elles sont appelées opérations internes (ou intra-unités).
Par exemple, la formation de capital fixe pour compte propre est traitée comme une opération entre une unité en sa qualité de producteur et elle-même en sa qualité d'acquéreur de capital fixe ; une telle opération possède une nature identique aux actions entreprises d'un commun accord par deux unités institutionnelles différentes.
C'est ainsi que certaines actions menées unilatéralement par une unité institutionnelle ont des conséquences sur d'autres unités institutionnelles, sans que ces dernières y aient consenti.
Le SCN n'enregistre de telles actions que dans certaines limites ; il le fait essentiellement quand des administrations publiques, ou d'autres unités institutionnelles, prennent possession d'actifs appartenant à d'autres unités institutionnelles, y compris à des unités non résidentes, sans contrepartie intégrale.

Figure 8. Contextes (phrases) avec le candidat terme « unités institutionnelles », extraits des documents de référence et manuels en français

Source : TermoStat.

Ces contextes riches en connaissances permettront de construire dans le logiciel Tedi⁷ le système de concepts de l'ontoterminologie (Carsenty, 2020, pp. 89-91). La figure suivante (fig. 9) présente la construction du concept désigné par « unité institutionnelle ».

⁷ Éditeur d'ontoterminologies multilingues développé par Christophe Roche à l'USMB (<http://ontoterminology.com/tedi>).

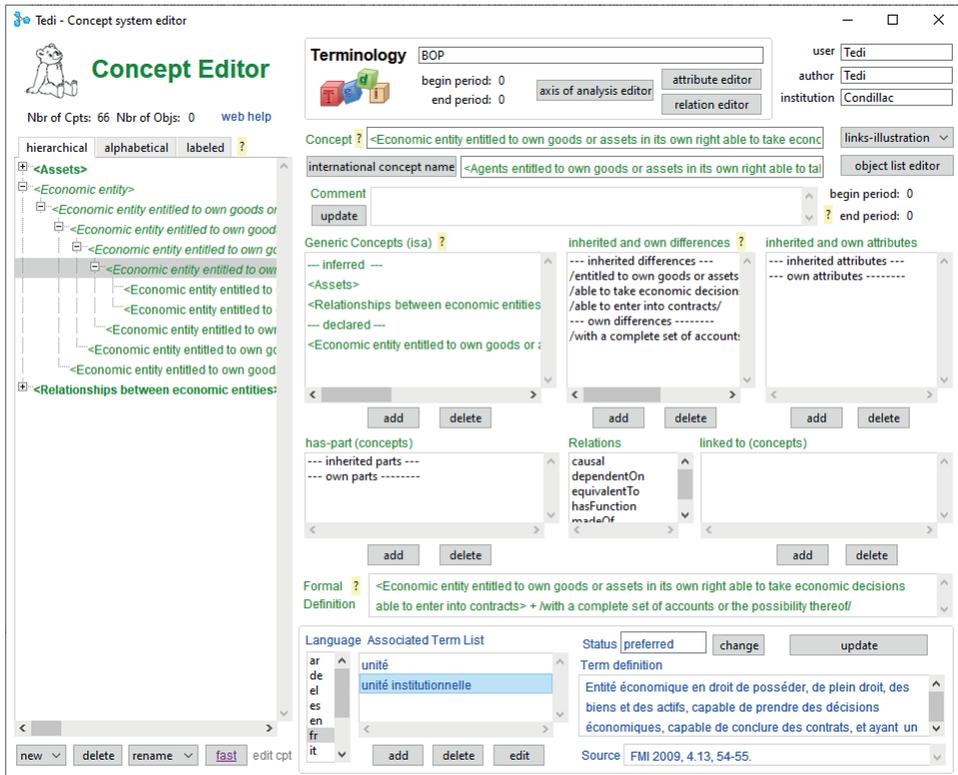


Figure 9. Éditeur de concepts de Tedi avec le concept désigné par le terme « unité institutionnelle »

Source : Tedi.

5. CONCLUSION

La construction d'un corpus en vue d'élaborer l'ontoterminologie de la BP nous amène à établir une typologie des textes collectés. Tous les textes du corpus ne fourniront pas le même type d'information pour l'ontoterminologie car ils reflètent des situations de communication différentes. Or il nous faut à la fois connaître le réel que le matériau linguistique étudié décrit et accéder au système de concepts élaboré par les experts pour penser ce réel. C'est pourquoi il nous semble pertinent d'élaborer cette typologie des textes sur la base des situations de communication présupposées par les textes collectés. Dans le cas des manuels, majoritaires dans le corpus, les destinataires sont les statisticiens qui produisent la BP. Cette typologie dit quelque chose sur l'organisation textuelle du domaine de la BP. Il est probable qu'elle corresponde à une typologie linguistique des textes.

Donner sa place à un corpus spécialisé dans un travail sur l'ontoterminologie, c'est prendre acte du fait que le texte de spécialité est l'« habitat privilégié des termes » (Costa, 2006, p. 80). Reconnaître que la notion de marqueur de relation est « beaucoup plus complexe qu'on pourrait le croire et [...] interroge fondamentalement les possibilités de systématiser les analyses de discours, quelles qu'elles soient » (Condamines, 2005, p. 43), c'est aussi s'interroger sur les moyens d'accéder à l'extralinguistique, essentiel dans tout travail de terminologie. C'est ici que les connaissances des experts sont indispensables, tant pour la modélisation que pour la validation de l'ontoterminologie. Nos connaissances préalables sur le domaine nous conduisent à adopter une approche hybride comprenant des phases sémasiologiques et onomasiologiques. Nous n'abordons pas le corpus sur la BP de manière extérieure puisque nous l'avons construit expressément à des fins d'analyse terminologique et que nous traduisons aussi des textes qui en font partie.

La prochaine étape de nos travaux, à présent que ce corpus existe, sera de modéliser le système de concepts de la BP et les termes correspondants en français, en anglais et en allemand, en nous appuyant sur nos connaissances et sur la partie du corpus constituée par les manuels et documents de référence. Grâce à des statisticiens responsables de la balance des paiements de la Suisse, nous avons noué des contacts avec leurs homologues des autres institutions étudiées. Nous nous réjouissons de leur future contribution à cette recherche.

BIBLIOGRAPHIE

- Adler, R.B. & Rodman, G. (2006). *Understanding Human Communication (Ninth Edition)*. New York : Oxford University Press.
- Bowker, L. (1996). Towards a corpus-based approach to terminography. *Terminology*, 3 (1), 27-52. DOI : 10.1075/term.3.1.03bow.
- Carsenty, S. (2019). *Esquisse des éléments d'une ontoterminologie de la balance des paiements et de la position extérieure*. (Mémoire présenté dans le cadre du Diplôme Universitaire « Fondements et enjeux de la terminologie »). Université Savoie Mont Blanc, Chambéry, France.
- Carsenty, S. (2020). Première approche de la modélisation ontoterminologique de la balance des paiements. *Actes de la conférence TOTH 2020 « Terminologie & Ontologie : Théories et Applications »*, Université Savoie Mont Blanc, Chambéry, France, 26 & 27 novembre 2020 (pp. 81-99). DOI : 10.57726/mhdg-s181.
- Charaudeau, P. (2009). Dis-moi quel est ton corpus, je te dirai quelle est ta problématique. *Corpus*, 8, 37-66. DOI : 10.4000/corpus.1674.
- Condamines, A. (2005). Linguistique de corpus et terminologie. *Langages*, 157, 36-47. DOI : 10.3406/lgge.2005.973.
- Condamines, A. (2009). Comment peut-on construire une ontologie personnelle à partir de textes ? Considérations linguistiques. *Actes de la 8^e conférence internationale Terminologie et Intelligence Artificielle (TIA'09)*, Toulouse, France, 18-20 novembre 2009. <https://hal.archives-ouvertes.fr/hal-00924991/document> [accès : 20.12.2022].
- Costa, R. (2006). Texte, terme et contexte. In D. Blampain, Ph. Thoiron & M. Van Campenhoudt (eds.), *Mots, termes et contextes. Actes des 7^{es} Journées scientifiques du réseau de chercheurs Lexicologie, terminologie, traduction* (pp. 79-87). Paris : Éditions des archives contemporaines.
- Costa, R. & Silva, R. (2008). De la typologie à l'ontologie de textes. *Actes de la conférence TOTH 2008 « Terminologie & Ontologie : Théories et Applications »*, Annecy, France, 5 & 6 juin 2008 (pp. 3-16). http://ontologia.fr/TOTH/Conference/TOTH2008/TOTH2008_actes-website.pdf [accès : 20.12.2022].
- Doualan, G. (2018). De la représentativité à la spécialisation : exemple d'un petit corpus sur la synonymie. *Corpus*, 18. DOI : 10.4000/corpus.3331.
- Flowerdew, L. (2004). The Argument for Using English Specialized Corpora to Understand Academic and Professional Settings. In U. Connor & T. Upton (eds.), *Discourse in the Professions: Perspectives from Corpus Linguistics* (pp. 11-33). Amsterdam : John Benjamins. DOI : 10.1075/scl.16.02flo.
- ISO 1087 (2019). *Travail terminologique et science de la terminologie – Vocabulaire*. Genève : Organisation internationale de normalisation. <https://www.iso.org/fr/standard/62330.html> [accès : 20.12.2022].
- Maingueneau, D. (2016). *Analyser les textes de communication*. Paris : Armand Colin.
- Picht, H. (2013). Concepts as Reflection of Societal Changes. *Terminologija*, 20, 10-23. <https://journals.lki.lt/terminologija/article/view/469> [accès : 20.12.2022].
- Rastier, F. (2004). Enjeux épistémologiques de la linguistique de corpus. *Texte !*. http://www.revue-texto.net/1996-2007/Inedits/Rastier/Rastier_Enjeux.html [accès : 20.12.2022].
- Roche, C. (2007a). Dire n'est pas concevoir. *Ingénierie des Connaissances*, 157-168. <https://hal.archives-ouvertes.fr/hal-00455253> [accès : 20.12.2022].
- Roche, C. (2007b). Le terme et le concept : fondements d'une ontoterminologie. *Actes de la conférence TOTH 2007 « Terminologie & Ontologie : Théories et Applications »*, Annecy, France, 1^{er} juin 2007 (pp. 1-22). <https://hal.archives-ouvertes.fr/hal-00202645> [accès : 20.12.2022].
- Roche, C. (2012). *Terminologie conceptuelle versus Terminologie textuelle*. http://christophe-roche.fr/Bibliographie/2012/Reperes_1_Janvier_2012.pdf [accès : 20.12.2022].

- Santos, C. & Costa, R. (2015). Domain specificity. Semasiological and onomasiological knowledge representation. In H.J. Kockaert & F. Steurs (eds.), *Handbook of Terminology*, 1 (pp. 153-179). Amsterdam : John Benjamins. DOI : 10.1075/hot.1.09dom1.
- Sowa, J. (2000). *Knowledge Representation – Logical, Philosophical, and Computational Foundations*. Pacific Grove : Brooks/Cole.
- Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. Amsterdam : John Benjamins.

BALANCE DES PAIEMENTS (EXTRAIT DU CORPUS)

- Banque mondiale, Commission européenne, Fonds monétaire international, Organisation de coopération et de développement économiques, & Organisation des Nations Unies (2013). *Système de comptabilité nationale 2008 (SCN 2008)*. New York : Nations Unies. <https://unstats.un.org/unsd/nationalaccount/docs/SNA2008FR.pdf> [accès : 20.12.2022].
- European Communities, International Monetary Fund, Organisation for Economic Co-operation and Development, United Nations, & World Bank (2009). *System of National Accounts 2008 (SNA 2008)*. New York : United Nations. <https://unstats.un.org/unsd/nationalaccount/docs/SNA2008.pdf> [accès : 20.12.2022].
- Fonds monétaire international (2009). *Manuel de la balance des paiements et de la position extérieure globale*, 6^e édition. Washington, D.C. DOI : 10.5089/9781589068155.069.
- Fonds monétaire international (2014). Guide pour l'établissement des statistiques selon le MBP6. Washington, D.C. <https://www.imf.org/external/pubs/ft/bop/2017/pdf/Guidefra.pdf> [accès : 20.12.2022].
- International Monetary Fund (2009). *Balance of Payments and International Investment Position Manual*, 6th Edition (BPM6) Washington, D.C. DOI : 10.5089/9781589068124.069.
- International Monetary Fund (2014). *Balance of payments and international investment position compilation guide*. Washington, D.C. DOI : 10.5089/9781484312759.069.
- Union Européenne (2012). Règlement (UE) n° 555/2012 de la Commission du 22 juin 2012 modifiant le règlement (CE) n° 184/2005 du Parlement européen et du Conseil relatif aux statistiques communautaires de la balance des paiements, du commerce international des services et des investissements directs étrangers, en ce qui concerne l'actualisation des exigences relatives aux données et les définitions. Journal officiel de l'Union européenne, L166/22. <https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX%3A32012R0555> [accès : 20.12.2022].
- Union Européenne (2016). Règlement (UE) 2016/1013 du Parlement Européen et du Conseil du 8 juin 2016 modifiant le règlement (CE) n° 184/2005 relatif aux statistiques communautaires de la balance des paiements, du commerce international des services et des investissements directs étrangers. Journal officiel de l'Union européenne, L 171/144. <https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX%3A32016R1013> [accès : 20.12.2022].
- United Nations Economic Commission for Europe (2019). *Generic Statistical Business Process Model V.5.1 (GSBPM)*. <https://statswiki.unecp.org/display/GSBPM/GSBPM+v5.1> [accès : 20.12.2022].

