

# ENTRE TERMINOLOGIE, TERMINOGRAPHIE ET HUMANITÉS NUMÉRIQUES

## Le projet DIACOM-fr

PAOLO FRASSI  
UNIVERSITÀ DEGLI STUDI DI VERONA

**Abstract** – The large-scale *Digital Humanities applied to foreign languages and literature* project (2018-2022) has enabled the Department of Foreign Languages and Literatures of the University of Verona to apply a large part of its linguistic, literary and didactic research to the digital humanities. The DIACOM-fr project (terminological database concerning the domain of international trade), which is part of the linguistic axis of this project, has several objectives: 1) to compile a corpus in the field of international trade (and make it available online for public consultation); 2) to develop a digital tool for interrogating this corpus; 3) to extract terminological units; 4) to compile a digital terminology database inspired by Explanatory and Combinatorial Lexicology. The purpose of this contribution is to present the objectives of the project, which has now come to an end.

**Keywords:** terminology; terminography; international trade; terminological database; terminological networks.

## 1. Introduction

La constitution d'une base de données terminologiques présente des enjeux théoriques et méthodologiques majeurs, notamment l'identification des entités terminologiques (simples et complexes), leur affichage, la prise en compte de la synchronie et/ou de la diachronie, la possibilité d'exploiter la base à des fins didactiques ou à d'autres fins, etc. Les décisions prises à ce niveau se reversent sur les modèles informatiques exploités pour la constitution de la base elle-même.

Nous illustrons, dans les lignes qui suivent, l'expérience du projet DIACOM-fr ; il s'agit d'un projet original, car il représente l'une des rares bases de données terminologiques inspirées de la Lexicologie Explicative et Combinatoire (Mel'čuk, Clas, Polguère 1995), affichant les entités terminologiques et leurs propriétés sous forme de *réseau terminologique* et valorisant, en plus de leurs caractéristiques sémantiques et des liens syntagmatiques et paradigmatiques des termes, leurs propriétés diachroniques et leurs propriétés didactiques.

Le projet s'est achevé en décembre 2022 et s'est échelonné sur plusieurs étapes que nous allons énumérer par la suite :

1. La création d'un corpus concernant le domaine du commerce international ;
2. L'exploitation d'un logiciel permettant la consultation du corpus ;
3. L'extraction automatique de termes simples et complexes ;
4. Le classement des termes complexes ;
5. L'identification des propriétés des termes ;
6. La représentation des entités terminologiques et de leurs propriétés dans une base de données terminologiques de type *réseau terminologique*.

Dans cette contribution nous nous proposons de présenter l'essentiel de chacune de ces étapes.

## 2. La constitution du corpus

La constitution d'un corpus représente souvent le point de départ d'un projet de type terminologique ; le fonctionnement des entités terminologiques en contexte est essentiel pour les réflexions théoriques concernant les propriétés des termes ainsi que pour l'activité terminographique.

Tout corpus terminologique se doit d'être représentatif du domaine en question – dans notre cas il s'agit du domaine du commerce international –, d'où l'adoption de deux critères, notamment le critère thématique et le critère textuel. S'ajoute un troisième critère, le critère chronologique, qui fait la spécificité de notre approche de type diachronique.

Par ce dernier critère – le critère chronologique –, nous avons voulu créer un corpus témoignant de l'évolution du domaine en question à travers l'identification de trois périodes-clés dans l'histoire du commerce international : 1) 1850-1914, période de l'industrialisation ; 2) 1945-1970, période du boom économique ; 3) 1985-2020, période de l'éclosion de nouvelles techniques de vente, du développement du marketing et de l'émergence du commerce électronique.

Le critère textuel nous a permis d'obtenir des données provenant de typologies textuelles différentes – textes institutionnels, textes scientifiques ou académiques, textes relevant de la presse spécialisée et textes issus de la documentation d'entreprise –, alors que le critère thématique a garanti la représentativité d'un large éventail de domaines, regroupés en quatre catégories (macroéconomie et économie internationale, produits et services) et quatre types d'activités (management, marketing, logistique, commerce électronique et droit).

Notre corpus se compose, au total, de 586 textes et 264.585 tokens<sup>1</sup>. Afin de garantir une continuité dans le temps des textes choisis, ainsi que plusieurs autres activités les concernant, dont des recherches à partir de paramètres divers, le corpus a été stocké dans la plateforme Alfresco (<https://www.alfresco.com/it/piattaforma>) : il s'agit d'une plateforme open-source, adaptée aux besoins de notre corpus et de sa consultation de la part de notre équipe de techniciens.

Les critères qui ont présidé à l'établissement du corpus se retrouvent dans la plateforme comme autant d'éléments de filtrage : il est possible de consulter des textes relevant de la première, de la deuxième ou de la troisième période, sur l'une des activités identifiées, selon le type de texte ; ou, encore, nous avons décliné dans la plateforme les différents types de produits et de services, en les associant à tel ou tel autre texte ; en outre, les textes sont classés par mots-clés, qui figurent également dans le menu de gauche et qui peuvent être activés en phase de filtrage. Autre aspect activable pour le filtrage : le pays francophone dont il est question dans le texte.

Cette plateforme est à l'usage exclusif des membres de l'équipe DIACOM-fr et ne concerne que des recherches portant sur les textes du corpus : il ne s'agit donc pas de recherches concernant des contextes à partir des termes donnés, qui affectent plutôt la section qui suit (Section 3).

<sup>1</sup> Nous tenons à préciser que la totalité du corpus n'est pas accessible dans sa version publique (voir Section 3) car nous n'avons pas obtenu, pour la totalité des textes, l'autorisation de la part des auteurs. Soulignons, au passage, que tous les textes ont été soumis à un traitement prévoyant, d'abord, le passage du format .PDF au format .TXT et, ensuite, le nettoyage du bruit s'étant glissé dans les documents à l'issue de la conversion.

The screenshot shows the 'DIACOM FR Documents' interface. On the left, there are filters for 'Active' and 'Available' documents, including options for 'Texte intégral public' (No: 455, Yes: 131), 'Mots-clés', 'Type de document', 'Macroéconomie et économie internationale', 'Pays - Publication ou rédaction', 'Produits', 'Services', 'Entreprise', and 'Intervalles chronologiques'. The main table displays the following data:

Titre	Auteur	Texte intégral...	Date de publi...	Date de cons...	Mots-clés	Macroécono...	Produits	Services	Et
Revenus		No	1992	2018	consommat...	Social Issues			
Commerce e...	OMC	No	1999	2018	OMC, comm...	Politique co...			
Rapport sur l...	Organisation ...	No	2016	2018	OMC, PME, ...	Politique co...			
Orienter les e...	Chambre de...	No	2015	2018	Canada, PM...	Politique co...			
Peut-on iden...	Jean-Bernar...	No	2014	2018	DSGE, politi...	Politique co...			
Prévoir la co...	Jean-Cristop...	No	1990	2018	conjoncture, ...				M
Primes de ris...	Patrick Artus	No	1990	2018	prime de risq...	Politique co...			
Promotion d...	Equipe Cana...	No	1998	2018	Canada, co...	Politique co...			
Qualification ...	Patrick Seve...	No	1990	2018	main-d'ouvr...				M
Questions su...	Gilbert Cette ...	No	1990	2018	industrie ma...				M
Rapport d'ac...	Direction du ...	No	2014	2018	France, budg...				M

At the bottom of the table, it indicates 'Visualizzazione di 1-25 su 586' and 'Elementi per pagina 25'.

Figure 1  
Corpus DIACOM-fr stocké sur la plateforme Alfresco.

### 3. Interrogation du corpus

La création du corpus a plusieurs finalités : la première est de rendre ce corpus disponible pour un ensemble relativement vaste d'utilisateurs (chercheurs, terminologues, terminographes, experts du domaine, enseignants, étudiants etc.) pour qu'ils puissent mener des recherches ad hoc concernant les données linguistiques. Pour que cet objectif se réalise, nous nous sommes servis d'un logiciel, lui aussi open source : il s'agit de KonText (voir Machálek 2020) que nous avons adapté à nos besoins. KonText est un moteur de recherche qui se caractérise par « a powerful query language, analytical functions, user-friendly interface, performance scalable to billion tokens corpora and tens or more concurrent users » (Machálek 2020, p. 7003).

Le masque de recherche se présente de manière très simple et donne la possibilité de rechercher les concordances à partir d'un terme donné, comme par exemple *commerce* :

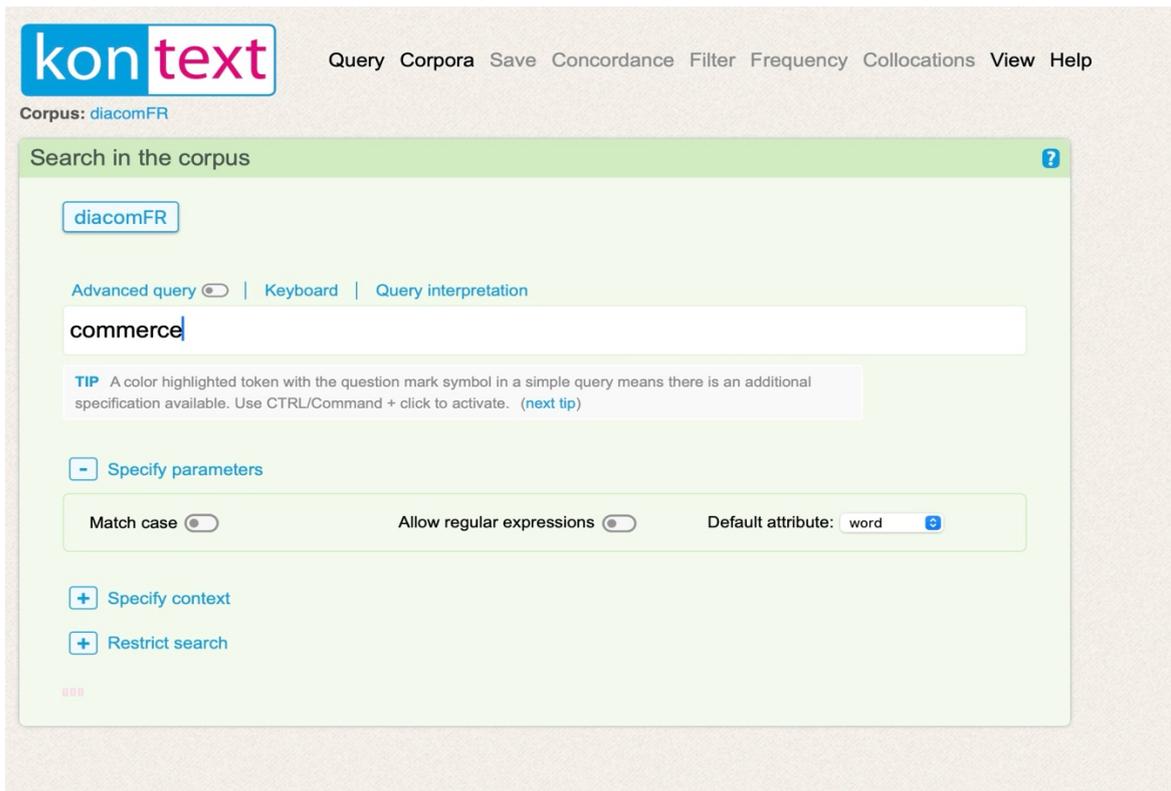


Figure 2  
Masque de saisie à partir de Kontext pour DIACOM-fr.

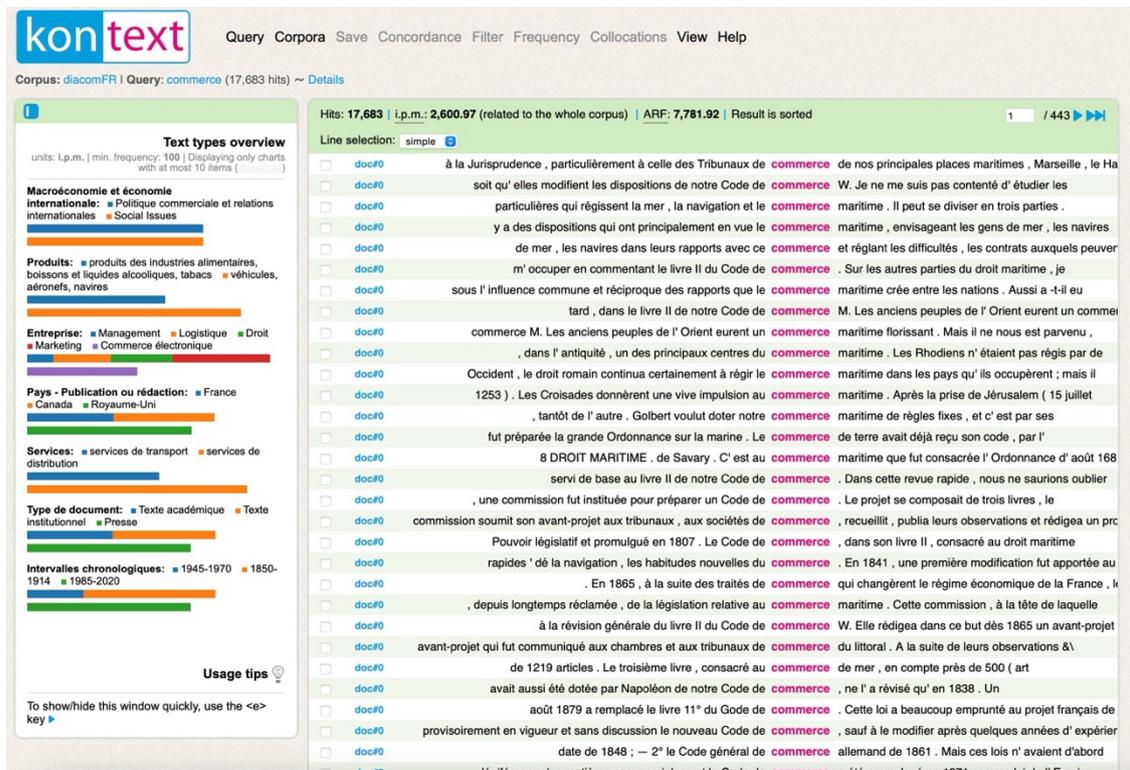


Figure 3  
Concordances de *commerce* dans Kontext pour DIACOM-fr.

Pour ce qui est des concordances, en plus des contextes du terme recherché, KonText offre un ensemble de données statistiques à partir des paramètres adoptés pour la constitution du corpus. En particulier, le masque de gauche « Text types overview » fait état des proportions des sous-domaines principaux, des types de produits et d'entreprise, des pays où les textes ont été publiés, des types de services, des types de documents et des intervalles chronologiques.

Nous précisons aussi que, comme dans la plupart des logiciels de ce type, la longueur des contextes à gauche et à droite du terme recherché est paramétrable.

En outre, les contextes peuvent être modifiés à l'aide des filtres contenus dans « Restrict Search » : ici, il sera possible de restreindre les contextes selon le type de texte, la période historique, le type de produit ou de service etc. – il s'agit par ailleurs des mêmes critères enregistrés dans la plateforme Alfresco (voir Section 2) – comme montré dans la figure qui suit :

The screenshot shows the KonText interface for the 'diacomFR' corpus. The 'Restrict search' section is expanded, showing several filter categories:

- Titre**: Macroéconomie et économie internationale
- Auteur**: Produits
- Date de publication**: Source
- Entreprise**: Pays - Publication ou rédaction
- Services**: Services
- Type de document**:
 

<input type="checkbox"/>	161,908
<input type="checkbox"/> Presse	273,146
<input type="checkbox"/> Presse\Texte académique	20,083
<input type="checkbox"/> Texte académique	1,600,148
<input type="checkbox"/> Texte institutionnel	5,126,948
<input type="checkbox"/> Texte institutionnel\Texte académique	343,455
- Intervalles chronologiques**:
 

<input type="checkbox"/> ===NONE===	2,780
<input type="checkbox"/> 1850-1914	4,353,483
<input type="checkbox"/> 1945-1970	260,157
<input type="checkbox"/> 1985-2020	2,182,192

Figure 4  
Paramètres concernant la « Restrict Search » dans KonText pour DIACOM-fr.

## 4. Extraction des données

Même si la base de données se veut d’abord synchronique et qu’elle se focalise davantage sur les liens syntagmatiques et paradigmatisques qui caractérisent la période actuelle (voir Section 4.2), puisque les liens diachroniques font aussi partie des liens pris en compte, nous avons effectué l’extraction terminologique distinctement pour les trois périodes, en obtenant ainsi trois nomenclatures différentes.

À partir du corpus constitué, nous avons procédé à l’extraction des termes simples et complexes. Pour l’extraction des termes, nous nous sommes appuyés principalement sur le logiciel TermoStat, conçu et élaboré à l’Université de Montréal par Patrick Drouin (voir Drouin 2010).

Ce logiciel est particulièrement performant et s’adapte à la typologie de corpus et de données attendues. La performance de TermoStat a été mesurée par un test effectué sur plusieurs extracteurs et logiciels actuellement disponibles : AntConc, Sketch Engine, TermoStat et TermSuite.

Par rapport aux autres logiciels, TermoStat présente les avantages suivants (voir également Calvi, Drouin, Frassi 2023) :

- Il est relativement fiable pour ce qui est de l’extraction des unités multilexémiques, notamment des collocations. En effet, les termes complexes sont surtout des syntagmes nominaux. Néanmoins, certaines associations, comme les collocations à collocatif verbal, qui se caractérisent souvent par un verbe en première position (ex. : *dresser un bilan*), sont difficilement identifiables par la plupart des extracteurs/logiciels ; or, TermoStat permet l’identification de ces types de termes complexes grâce à la fonction *bigrammes*. Cette fonction n’existe pas dans les autres logiciels : dans AntConc (avec la fonction *Collocate*) et Sketch Engine (avec la fonction *Word sketches*), par exemple, on peut retrouver les collocations à collocatif verbal uniquement en insérant la base dans le masque de saisie ; ces logiciels ne fournissent pas une liste de collocations de ce type à partir d’une extraction automatique. Cet aspect s’avère utile pour nous car nous ne considérons pas les collocations comme des termes à proprement parler, mais comme une propriété de combinatoire des termes (voir à ce propos la Section 4.1).
- L’intérêt de TermoStat réside aussi dans le corpus de référence dont il se sert et de la manière d’identifier automatiquement le statut plus ‘terminologique’, si l’on peut dire, d’un corpus ou d’une liste de mots simples ou complexes. Ce logiciel reçoit un corpus en entrée et, à l’issue d’une comparaison avec un corpus de référence non spécialisé, il propose de manière automatique une liste de candidats-termes simples et/ou complexes. S’appuyant sur l’hypothèse que les termes d’un domaine donné sont bien plus fréquents dans un corpus spécialisé du même domaine que

dans le corpus de langue générale, la mise en opposition de deux corpus joue un rôle d'indéniable importance pour la réduction du bruit.

Ces deux avantages ont été pour nous fondamentaux dans le repérage des termes qui peuplent notre base de données, car ils fournissent des listes sur lesquelles le travail manuel de filtrage – qui est indispensable – a été simplifié. Nous avons intégré nos nomenclatures par les termes simples et complexes figurant dans l'un des ouvrages terminographiques de référence pour les trois périodes. Ces ouvrages résultent d'un dépouillement concernant tous les ouvrages terminographiques du domaine sur les trois périodes, qui nous a donné 5 ouvrages pour la première tranche, 22 pour la deuxième et 79 pour la troisième. Nous avons retenu 2 ouvrages pour chaque période, choisis en raison de la richesse et de l'exhaustivité de leurs nomenclatures : il s'agit de Sacré (1884) et Guyot et Rafalovich (1901) pour la période allant de 1850 à 1914 ; de Kettridge (1949) et Centre international du droit des affaires (1973) pour la période allant de 1945 à 1970 ; de Massabié-François et Poulain (2002) et Morelle (2003) pour la période allant de 1985 à 2020. À l'issue de cette extraction, nous avons retenu 5030 termes (simples et complexes).

## 5. Les propriétés des termes

Une fois les données identifiées, il est nécessaire d'établir le modèle nous permettant de les représenter dans une base de données.

Les bases de données terminologiques existent depuis longtemps ; dans le cadre de notre projet, nous avons choisi de nous appuyer sur un modèle de représentation des données terminologiques assez intuitif et, surtout, respectueux des propriétés paradigmatiques et syntagmatiques des unités terminologiques, ainsi que des propriétés syntactico-sémantiques de chaque entité terminologique. Cela nous a poussé à une réflexion concernant 1) la définition des entités terminologiques et leurs propriétés syntactico-sémantiques ; 2) les propriétés paradigmatiques et syntagmatiques qui les caractérisent, ainsi que d'autres propriétés que le terminographe souhaite valoriser ; 3) le choix d'un modèle dans lequel ces aspects soient représentés de manière intuitive, en vue d'une consultation rapide de la part de tout utilisateur ainsi que dans des objectifs didactiques.

Nous croyons que le modèle plus traditionnel de la base de données – celui qui emprunte à la lexicographie et à la terminographie classiques le caractère linéaire – ne s'adapte pas à nos objectifs ; ainsi, nous avons préféré de nous tourner du côté du modèle des réseaux lexicaux (voir à ce propos Polguère 2014). Ce modèle représente les entités lexicales en respectant leurs propriétés : il se configure comme un graphe lexical dans lequel chaque nœud correspond à une entité lexicale et des liens se tissent entre nœuds, représentant

autant de propriétés paradigmatiques ou syntagmatiques des diverses entités terminologiques.

Ce modèle a été déjà appliqué dans la langue générale : il suffit de penser à WordNet ou FrameNet, pour la langue anglaise, ou, encore, au *Réseau lexical du français* (RL-fr, <https://spiderlex.atilf.fr/fr>) pour ce qui est de la langue française.

Cette dernière expérience lexicographique est notre modèle d'inspiration pour la base de données terminologiques que nous avons constituée.

En amont de la création de cette base il est nécessaire d'identifier :

- 1) Les entités terminologiques ;
- 2) Les types de liens que ces entités nouent entre elles ;
- 3) Les autres propriétés à représenter.

### **5.1. Les entités terminologiques**

Du point de vue linguistique, les unités de base de la terminologie s'appellent *termes* : il s'agit de formes graphiques rattachées à un signifié et appartenant clairement à un domaine de spécialité.

La tradition terminologique et terminographique distingue deux types de termes : les termes simples et les termes complexes (voir par exemple L'Homme 2020 et Cabré 1992).

Nous avons voulu aller plus loin dans l'identification des termes complexes et, à partir de deux études pilotes sur deux sous-corpus différents (les résultats détaillés de ces deux études se trouvent dans Frassi, Calvi, Humbley 2020 ; Frassi 2020, 2023), nous avons été en mesure de démontrer qu'un grand nombre de termes complexes partage les propriétés syntactico-sémantiques qui ont été traditionnellement rattachées aux locutions, notamment par Gross (1996) et Mel'čuk (2013). Plus particulièrement, pour notre base de données nous avons adopté le classement de Mel'čuk (2013), identifiant des cas de locutions fortes, semi-compositionnelles et faibles.

Les locutions ne sont pas, pour autant, les seuls types d'unités multilexémiques apparaissant en terminologie : notre étude pilote a révélé une quantité non négligeable de collocations. Puisque les collocations en terminologie possèdent essentiellement les mêmes propriétés que les collocations dans la langue générale, les mêmes types de liens syntagmatiques se présentent en domaine de spécialité : par exemple, si nous reprenons le répertoire des liens proposés dans le cadre de la Lexicologie Explicative et Combinatoire (voir Mel'čuk, Clas, Polguère 1995 ; Mel'čuk, Polguère 2022 ; voir également la Section 4.2), il est possible d'identifier le sens de 'bon', dans les collocations *rapport fiduciaire* ou *client fidèle*, ou, pour ce qui est des collocations à collocatif verbal, des verbes de réalisation comme dans *réaliser un investissement*. Toutefois, nous ne considérons pas la collocation comme

une entité terminologique ; elle exprime une propriété de combinatoire de la base (terme simple ou terme complexe).

Ainsi, notre base de données va considérer les entités terminologiques correspondant à des termes simples (à savoir des mots graphiques, suites de lettres séparées par des espaces blancs, comme par exemple *commerce*) et à des termes complexes de type *locutions*.

## 5.2. Les liens entre termes

Les graphes lexicaux vont tenir grâce à des liens paradigmatiques et syntagmatiques. Pour l'identification de ces liens, nous nous sommes appuyés sur le répertoire dressé dans le cadre de la Lexicologie Explicative et Combinatoire (Mel'čuk, Clas, Polguère 1995 ; Polguère, Mel'čuk 2022) : ces liens, très généraux et très récurrents dans la langue générale, se trouvent également dans les langues de spécialité. Ainsi, par exemple, le lien de synonymie (*commerce/transaction*) ou le nom typique de l'un des actants (*commerce/vendeur/acheteur/marchandise/prix...*).

## 5.3. Autres propriétés des termes

D'autres aspects vont être pris en compte en tant que propriétés des termes. Il s'agit notamment de la diachronie et des aspects didactiques.

### 5.3.1. La diachronie

Les liens diachroniques permettent de relier, sur l'axe du temps, des unités terminologiques appartenant à des périodes différentes. Ces liens se créent au niveau de la forme et/ou du sens et peuvent être mis au nombre des liens de type paradigmatique.

Pour marquer l'appartenance de tel ou tel autre terme à l'une des trois périodes envisagées, nous avons adopté une notation impliquant l'emploi des trois lettres de l'alphabet grec  $\alpha$  (pour la période 1985-2020),  $\beta$  (pour la période 1945-1970),  $\gamma$  (pour la période 1850-1914). Chaque terme pouvant être potentiellement polysémique, il est aussi accompagné d'un chiffre romain indiquant l'acception. Ainsi, par exemple, *commerce $\alpha$ 1* réfère à la première acception du terme *commerce* dans la période 1985-2020.

Pour revenir à la question des liens diachroniques entre termes, comme nous l'avons déjà souligné dans Frassi (2021), nous avons identifié, avec l'aide de plusieurs études concernant la terminologie diachronique, et notamment Møller (1998), Condamines, Rebeyrolles, Soubeille (2004), Picton (2014) et Dury (2018), les cas de variation les plus récurrents, et plus précisément :

- 1) Expansion (ex. *billet à ordrey1* → *billet à ordrea1* → *billet à ordre international $\alpha$ 1*) : un terme simple devient complexe ou un terme complexe ajoute un élément ultérieur ; cette variation implique généralement une

précision dans le sémantisme du terme à l'issue de cette expansion formelle.

- 2) Ellipse ou réduction (ex. *protection contre les rayonnements* → *radioprotection*<sup>2</sup>): un terme complexe réduit le nombre de ses composantes ; cette variation implique généralement une identité dans le sémantisme des deux termes, l'ellipse se configurant comme une opération de simplification s'appliquant uniquement à la forme.
- 3) Resémantisation (ex. *transbordement*<sub>1</sub> → *transbordement*<sub>1</sub>): un terme demeure tel quel du point de vue de la forme mais subit une variation au niveau du signifié.

### 5.3.2. Les 'propriétés' didactiques<sup>3</sup>

Comme nous l'avons rappelé plus haut, en plus des propriétés traditionnelles, nous poursuivons également l'exploitation des données terminologiques dans un contexte didactique. C'est sur ce dernier aspect que nous focalisons maintenant notre attention.

Chaque terme vedette faisant partie de notre base de données a été associé à un niveau d'apprentissage, et ce dans l'objectif d'aider enseignants et apprenants dans l'exploitation de la base de données elle-même à des fins didactiques. Un champ à part entière, appelé *niveau d'apprentissage*, intègre de ce fait la fiche terminologique du DIACOM-fr. À chaque terme est associé l'un des trois niveaux d'apprentissage identifiés, *élémentaire*, *indépendant* et *expérimenté*, qui correspondent aux trois niveaux généraux de la compétence linguistique établis par le CECR.

Pour attribuer à chaque terme un niveau de compétence, nous avons adopté un critère principalement qualitatif (voir, à ce propos, Frassi et Bonadonna 2022 pour le classement des termes de type *locutions* ; Frassi et Bonadonna 2023 pour le classement des acceptions de la polysémie). En particulier, nous avons associé à l'échelon le plus bas (A2) des termes simples dont les sens pouvaient être considérés comme *de base* (à savoir typiquement associés au domaine du commerce, comme par exemple *commercial* = 'relatif au commerce'). Dans le cas de termes polysémiques, il s'agissait toujours de la première acception, qui était également, du point de vue quantitatif et dans la grande majorité des cas, la plus fréquente. L'échelon intermédiaire, correspondant aux niveaux B1/B2, inclut les termes simples dont l'acception n'est pas considérée comme une acception de base, ainsi que les locutions qui, du point de vue sémantique, sont semi-compositionnelles. Pour ce qui est, finalement, du niveau C1, y sont associés les termes simples dont l'acception,

<sup>2</sup> Nous empruntons cet exemple à Goffin (1989, p. 105).

<sup>3</sup> Plus en général, sur l'exploitation des corpus en didactique du FOS – notamment dans les domaines du commerce et du marketing digital – voir Bonadonna 2023.

puisqu'elle est métaphorique ou métonymique, demande un effort particulier de la part de l'apprenant ; quant aux termes complexes de type locution, il s'agit d'unités multilexémiques non compositionnelles (ex. : *marge arrière*).

Il va de soi que, selon les critères de classement adoptés pour la propriété didactique, la plupart des termes sont des termes de base ou des acceptions de base non métaphoriques ni métonymiques, ou des locutions faibles (compositionnelles du point de vue sémantique).

## 6. La représentation dans la base des données

Dans le présent paragraphe nous faisons état de la méthode choisie pour la représentation des données terminologiques et de leurs propriétés. Ces choix se trouvent déjà dans Frassi et Rospocher (2022) : nous nous bornons ici à rappeler les aspects de base.

Notre approche s'est basée sur l'adoption des technologies du web sémantique via le modèle RDF pour la représentation des données, de leurs propriétés et de leurs liens. Plus spécifiquement, pour la modélisation des entités terminologiques nous nous sommes appuyés sur le modèle Lemon (McCrae *et al.* 2017). Ce modèle a été exploité pour la représentation des entités terminologiques (termes simples ou termes complexes de type *locution*), chacune étant associée à un signifié spécifique et à une partie du discours. Pour la modélisation des liens paradigmatiques et syntagmatiques, ainsi que pour les liens de type diachronique, nous avons adopté le modèle Lexfom (Lexical Functions Ontology Model, Fonseca *et al.* 2016). Celui-ci a été conçu pour la modélisation des fonctions lexicales selon la Théorie Sens-Texte et la Lexicologie Explicative et Combinatoire.

Ces modèles nous permettent ainsi de représenter des entités terminologiques et leurs propriétés sous forme de web sémantique prévoyant une interrogation à partir des entités lexicales qui constituent notre nomenclature.

Les entités terminologiques simples et complexes (celles-ci uniquement de type *locution*) s'affichent à l'écran entourées de leurs propriétés diverses (liens paradigmatiques et syntagmatiques, propriétés diachroniques et propriétés didactiques), comme dans les deux images qui suivent, représentant un terme simple (*transaction*) et un terme complexe de type *locution* (*transaction courante*) :



Figure 5  
Exemple de représentation des entités terminologiques dans DIACOM-fr.

Les propriétés des termes s'affichent sous forme de petits ronds entourant l'entité terminologique elle-même. Ainsi, le nombre de petits ronds dépend du nombre de propriétés valorisées pour chaque entité terminologique : en l'occurrence, *transaction* développe un réseau de liens plus vaste que *transaction courante*. En cliquant sur chacune des propriétés, il est possible d'afficher le type de lien et l'entité vers laquelle ce lien pointe.

Ainsi, par exemple, pour ce qui est de *transaction courante*, en cliquant la propriété diachronique, on aura le réseau suivant, qui atteste la présence de ce même terme dans les deux autres périodes envisagées dans notre diachronie, avec une relation d'identité de sens et de forme :

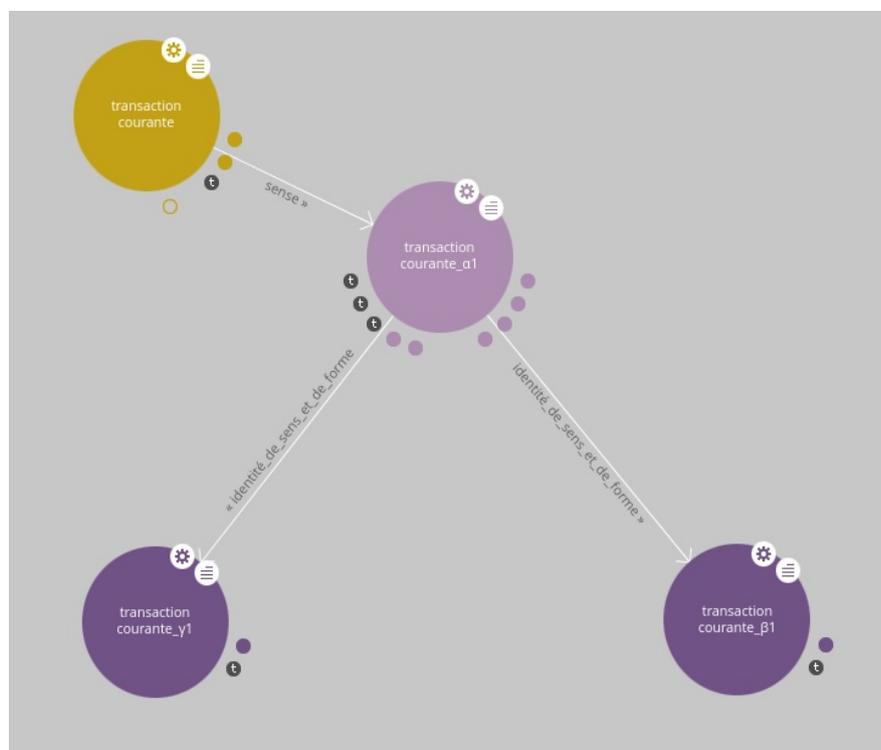


Figure 6  
Les propriétés diachroniques d'un terme dans DIACOM-fr.

Quant aux liens syntagmatiques et paradigmatiques, nous proposons l'exemple du réseau qui se tisse à partir de *transaction* :

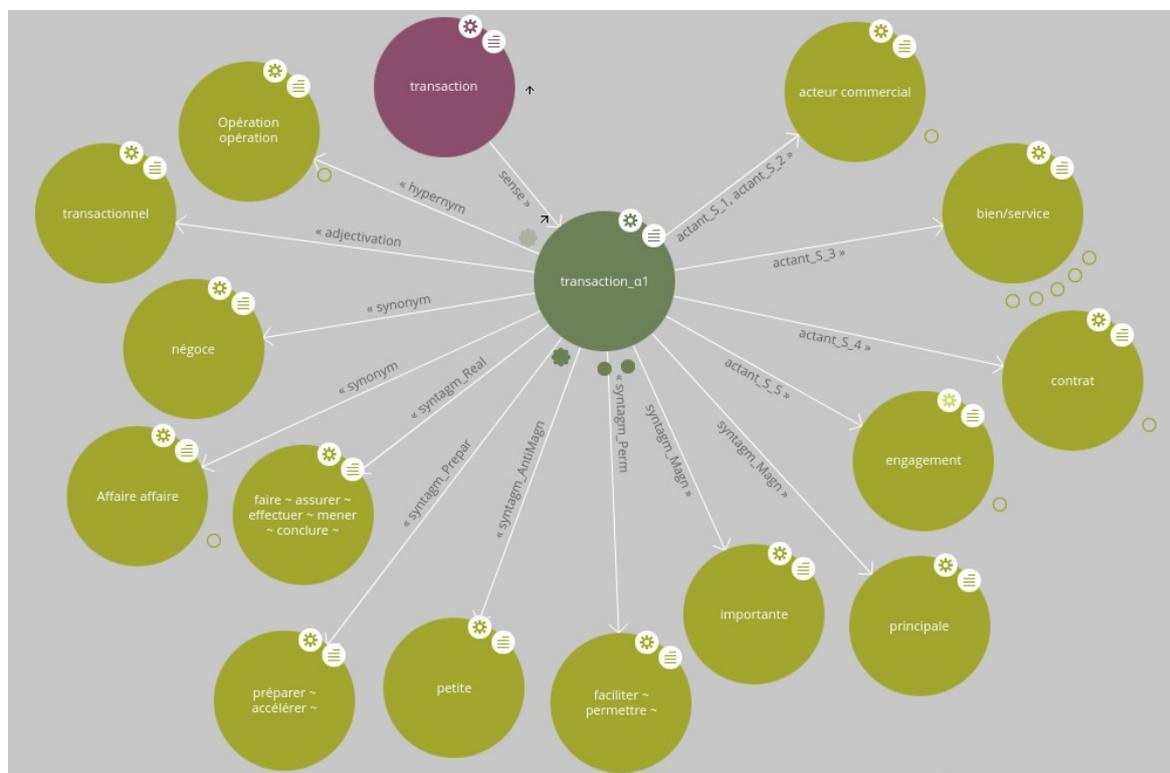


Figure 7

Affichage en réseau des liens paradigmatiques et syntagmatiques (exemple de *transaction*), DIACOM-fr.

Les relations paradigmatiques sont représentées par les liens d'hyperonymie (« hypernym », *opération commerciale*), d'adjectivisation (*transactionnel*) et de noms d'actants (*acteur commercial*, *bien/service*, *contrat* et *engagement*).

Les relations syntagmatiques – précédées de la chaîne de caractères « syntagm\_ » - concernent, pour le cas de *transaction*, des verbes de réalisation (*faire*, *assurer*, *effectuer*, *mener*, *conclure une transaction*), des verbes de préparation (*préparer*, *accélérer une transaction*), des verbes exprimant le sens 'ne pas faire cesser, laisser continuer' (*faciliter*, *permettre une transaction*), les modificateurs d'intensité (*transaction principale*, *transaction importante*) et les modificateurs exprimant le sens opposé à l'idée d'intensité (*petite transaction*).

Les propriétés didactiques ne sont pas encore valorisées dans la base : il s'agit de la dernière étape de notre activité terminographique en phase d'achèvement.

## 7. Conclusion

Nous avons fait le tour d'un projet terminologique ayant plusieurs débouchés et, par conséquent, plusieurs points d'intérêt.

Premièrement, du point de vue de la linguistique de corpus, notre projet permet la consultation d'un corpus récent à partir d'un logiciel – KonText – adapté à cette fin et permettant le paramétrage concernant des critères différents affectant, par exemple, la diachronie, le domaine, le type de texte, le type de produit ou service, etc.

L'aspect le plus novateur est représenté par la méthodologie adoptée pour l'affichage du réseau terminologique lui-même, incluant une réflexion de type linguistique et une réflexion touchant à l'informatique. Au niveau linguistique nous avons identifié non seulement les types principaux d'entités terminologiques – termes simples et termes complexes de type *locution* – mais nous avons également motivé notre choix de ne pas faire figurer, en tant que lemmes dans la base elle-même, les collocations. En outre, nous avons isolé un ensemble de propriétés des termes, dont certaines ne sont pas communément représentées dans les bases de données terminologiques : en plus des liens paradigmatiques et syntagmatiques, en effet, nous nous sommes occupés d'isoler les types de liens intervenant au niveau de la diachronie et de classer les termes selon le niveau de compétence linguistique. Ces derniers aspects sont particulièrement novateurs dans le domaine terminologique. Quant aux premiers, nous nous sommes appuyés sur les types de liens identifiés par la Lexicologie Explicative et Combinatoire, ce qui constitue une tentative certes non-pionnière, mais assez peu répandue.

La méthodologie adoptée pour adapter ces données linguistiques et ces propriétés linguistiques à une base de données informatisée demeure aussi intéressante : nous avons, en effet, réfléchi non seulement à la modélisation des types de propriétés de chaque terme, à travers l'exploitation de Lexfom, mais nous proposons une modélisation, en général, de toutes les données linguistiques à travers de triplets (RDF). Une telle modélisation répond à l'idée général du web sémantique et permet de représenter les données non pas à travers un affichage linéaire et traditionnel, mais à travers un graphe lexical qui permet des avantages de consultation : davantage intuitif, il permet sans doute un meilleur accès aux propriétés des termes et aux liens entre termes.

La base de données DIACOM-fr s'adresse à des types divers d'utilisateurs, dont des linguistes, des terminologues et des terminographes, des enseignants et des apprenants : nous souhaitons que la conformation de notre réseau terminologique puisse leur rendre service.

**Bionote** : Paolo Frassi est Professeur de Linguistique française à l'université de Vérone. Ses intérêts scientifiques concernent la lexicologie et la lexicographie, ainsi que la terminologie,

la terminographie, l'enseignement du FLE. Il a consacré de nombreuses études à la définition de l'adjectif ainsi qu'au statut des phrasèmes entre langue générale et terminologie. Au sein du Département des Langues et Littératures étrangères de l'Université de Vérone, il a dirigé le projet d'envergure 2018-2022 sur « Les humanités numériques appliquées aux langues et littératures étrangères ». Il a participé à plusieurs projets, dont DIACOM-fr (Université de Vérone), RL-fr (ATILF-CNRS, Nancy), FLATIF (PRIN) et collabore avec le Laboratoire ATILF-CNRS de Nancy ainsi qu'avec l'UQAM et l'Université de Montréal.

**Adresse de l'auteur :** [paolo.frassi@univr.it](mailto:paolo.frassi@univr.it)

## Références bibliographiques

- Alfresco <https://www.alfresco.com/it/piattaforma> (25/10/2023).
- Bonadonna, M.F. 2023, *Didactique du lexique et corpus numériques pour le Français L2. Des applications pour le commerce et le marketing digital*, Peter Lang, Bruxelles.
- Cabré M.T. 1992, *La terminologia. La teoria, els mètodes, les aplicacions*, Empúries, Barcelona.
- Calvi S., Drouin P. et Frassi P. 2023, *Acquisition (semi-)automatique des collocations terminologiques*, in Frassi P. (éd.), *Phraséologie et terminologie*, De Gruyter, Berlin, Boston, pp. 239-258.
- Centre International du Droit des Affaires 1973, *Lexique pratique commercial*, Régif, Paris.
- Condamines A., Rebeyrolle J. et Soubeille A. 2004, *Variation de la terminologie dans le temps : une méthode linguistique pour mesurer l'évolution de la connaissance en corpus*, in Williams G. et Vessier S. (éds.), *Proceedings of the 11th EURALEX international Congress, Euralex 2004*, Lorient, 6-10 juillet 2004, Université de Bretagne-Sud, Faculté des Lettres et des Sciences humaines, Lorient, pp. 547-557.
- Drouin P. 2010, *TermoStat. Présentation du logiciel* [http://termostat.ling.umontreal.ca/doc\\_termostat/doc\\_termostat.html](http://termostat.ling.umontreal.ca/doc_termostat/doc_termostat.html) (25/10/2023).
- Dury P. 2018, *La dimension diachronique en anglais de spécialité : une approche terminologique*, Note de synthèse pour l'habilitation à diriger des recherches, Université Paris 7.
- Fonseca A., Sadat F. et Lareau F. 2016, *Lexfom: a lexical functions ontology model*, in *Proceedings of the 5th Workshop on Cognitive Aspects of the Lexicon (CogALex-V)*, pp. 145-155.
- Frassi P., Calvi S. et Humbley J. 2020, *Fouille de textes et repérage d'unités phraséologiques*, in *Terminologie & ontologie : théorie et applications*, Presses Universitaires Savoie Mont Blanc, Chambéry, pp. 321-338.
- Frassi P. 2020, *La force des locutions faibles en domaine de spécialité*, in Zanola M.T. et Conceição M.C. (éds.), *Terminologia e mediação linguística, métodos, práticas e atividades*, UAlg editora, Faro, pp. 43-60.
- Frassi P. 2021, *DIACOM-fr, une base de données terminologiques de type diachronique*, in « Cahiers de Lexicologie » 118 [1], pp. 23-50.
- Frassi P. 2023, *Les unités multilexémiques non libres entre langue générale et langue de spécialité*, in Henrot Sostero G. (éd.), *Alle radici della fraseologia europea*, Peter Lang, Lausanne, pp. 281-304.
- Frassi P. et Bonadonna M.F. 2022, *Les termes complexes de type locution dans l'enseignement du français L2 en langue de spécialité : le cas du domaine du commerce international*, in « LIDIL » 65, pp. 1-19.
- Frassi P. et Bonadonna M.F. 2023, *Termes, polysémie et niveaux d'apprentissage en FLE*, in « Recherches et applications. Le Français dans le Monde » 73, pp. 57-74.
- Frassi P. e Rospocher M. 2022, *Ontologie, locuzioni deboli e legami fra entità terminologiche nell'ambito del progetto DIACOM-fr*, in Chiocchetti E. e Ralli N. (éds.), *Risorse e strumenti per l'elaborazione e la diffusione della terminologia in Italia*, Eurac Research, Bolzano, pp. 138-151.
- Goffin R. 1989, *La terminologie des sciences et des techniques nucléaires. Un cas de diachronie récente*, in Centre de Terminologie de Bruxelles (éd.), *Terminologie diachronique. Actes du colloque organisé à Bruxelles les 25 et 26 mars 1988*, Conseil international de la langue française, Paris, pp. 94-107.

- Gross G. 1996, *Les expressions figées en français : noms composés et autres locutions*, Ophrys, Paris.
- Guyot Y. et Rafalovich A.G. 1901, *Dictionnaire du commerce, de l'industrie et de la banque : Assurance, banque et bourse, comptabilité, droit commercial, terrestre et maritime, économie politique appliquée*, Guillaumin, Paris.
- Kettridge J.O. 1949, *French-English/English-French Dictionary of Commercial and Financial Terms. Phrases and Practice*, Routledge and Kegan Paul Limited, London.
- KonText <http://kontext.korpus.cz/> (25/10/2023).
- L'Homme M.-C. 2020, *La terminologie : principes et techniques*, Les Presses de l'Université de Montréal, Montréal.
- Machálek T. 2020, *KonText: Advanced and Flexible Corpus Query Interface*, in *Proceedings of the Twelfth Language Resources and Evaluation Conference*, European Language Resources Association, Marseille, pp. 7003-7008.
- Massabié-François M. et Poulain E. 2022, *Lexique du commerce international. Les 2 600 mots actuels et pratiques de l'import-export*, Bréal, Rosny-sous-Bois.
- Mel'čuk I. 2013, *Tout ce que nous voulions savoir sur les phrasèmes mais...*, in « Cahiers de lexicologie » 102 [1], pp. 129-149.
- Mel'čuk I., Clas A. et Polguère A. 1995, *Introduction à la lexicologie explicative et combinatoire*, Duculot, Paris.
- Mel'čuk I. et Polguère A. 2021, *Les fonctions lexicales dernier cri*, in Marengo S., *La Théorie Sens-Texte et ses applications. Lexicologie, lexicographie, terminologie, didactique des langues*, L'Harmattan, Paris, pp. 75-155.
- McCrae J.P., Bosque-Gil J., Gracia J., Buitelaar P. and Cimiano P. 2017, *The Ontolex-Lemon Model: Development and Applications*, in *Proceedings of eLex 2017 Conference, 19–21 September 2017, Leiden, Netherlands*, Lexical Computing CZ s.r.o, Leiden, pp. 587-597.
- Møller B. 1998, *À la recherche d'une terminochronie*, in « Meta » 43 [3], pp. 426-438.
- Morelle F. 2003, *Le Dico du commerce international* <https://www.glossaire-international.com/> (25/10/2023).
- Picton A. 2014, *The dynamics of terminology in short-term diachrony: a proposal for a corpus-based methodology to observe knowledge evolution*, in Temmerman R. and Van Campenhout M. (éds.), *The dynamics of culture-bound terminology in monolingual and multilingual communication, Terminology and Lexicography Research and Practice*, John Benjamins, Amsterdam/ Philadelphia, pp. 159-182.
- Polguère A. 2014, *From Writing Dictionaries to Weaving Lexical Networks*, in « International Journal of Lexicography » 27 [4], pp. 396-418.
- Sacré A. 1884, *Dictionnaire de commerce et de droit commercial*, Librairie de A. Marescq, Paris.
- Termostat Web 3.0+ <http://termostat.ling.umontreal.ca> (25/10/2023).