Journal of Digital Terminology and Lexicography – ISSN:3103-3601
Vol. 1, No. 2 (2025) DOI: 10.25430/pupj.jdtl.1763979001

Cross-Linguistic Terminology Transfer in Economic and Employment Discourse: Evidence from Greek and French

Angeliki Christopoulou

Greek Public Administration, Executive Structure of the Partnership Agreement, Ministry of Social Cohesion and Family Affairs, Aer. Papanastasiou 34, Athens, 115 27, Greece.

Contributing authors: achristopoulou@mou.gr

Abstract

Specialised language can be analysed through the study of lexical units at the morphological, syntactic, and semantic levels. Within Terminology, diverse theoretical frameworks address these issues from different perspectives. This study employs the Theory of Textual Terminology to analyse the register of Public Administration in Greek and French, with a focus on the domains of Employment and Economy. The research is corpus-driven and relies on a bilingual, comparable, static, and synchronic corpus. Node terms – frequent and semantically salient lexical items – were identified through combined frequency and keyword analyses. Equivalences across the Greek and French corpora were explored using statistical measures of co-occurrence and semi-automatic term extraction. The findings contribute to the development of bilingual terminology resources and underscore the value of digital corpus methodologies in capturing cross-linguistic patterns. Potential applications include enriching digital terminology databases and improving tools for multilingual text processing in the field of Public Administration.

Keywords: LSP, Terminology, Comparative Study, Corpus Linguistics, Public Administration

1 Introduction

Corpus Linguistics occupies a prominent position in the study of language due to its systematic investigation of linguistic phenomena through corpora – comprehensive collections of authentic data processed and interrogated with specialised software tools. Baker (2006) highlights the importance of Corpus Linguistics in its ability to reduce subjective influence through an empirical and objective approach. The application of corpora is intended to counteract subjective bias in the selection of linguistic data. Nevertheless, it is important to acknowledge that methodological decisions – such as which corpus to employ and how to define search parameters – necessarily entail a measure of subjectivity. Even so, Teubert (2004) argues that corpus-based inquiry fosters more systematic and replicable research than approaches relying on subjective data. Indeed, Teubert (2004) underscores this view by asserting that a

word's meaning can only be fully understood when all its corpus occurrences are considered. This holistic approach to meaning forms the central focus of Corpus Linguistics, as it combines quantitative frequency patterns with qualitative interpretation, making it a flexible framework for uncovering functional language use. Essentially, Corpus Linguistics does more than just analyse texts: it brings to light the patterns within the data and uses statistical analysis to understand the nuances of language. It is not simply a rigid methodology, but also a dynamic tool that defines the very terms of its analysis. Pearson (1998) observes that the flexibility of this approach permits ongoing methodological refinement and fosters the exploration of new linguistic avenues, while Condamines and Picton (2022) highlight the concomitant need to develop repositories that both preserve authentic term usage and support efficient, tool-based processing.

The field of Textual Terminology challenges the assumption that concepts exist independently of language (Auger & L'Homme, 1994; Bourigault & Slodzian, 1999; Condamines, 2003, 2005). Instead, it adopts a semantic and discourse-oriented perspective, deriving the meaning of terms and their collocational patterns directly from textual evidence. As L'Homme (2004) asserts, the term is realised within texts,[1] even though its interpretation is anchored in an extralinguistic dimension: the subject field. Lexical items that may initially seem characteristic of general language use can, upon closer examination, function as specialised terms within a particular subject field, bearing domain-specific meanings. The meaning and terminological properties of lexical elements are shaped by the subject field, their register and the communicative context. The comprehension of specialised texts necessitates a contextual approach to language, encompassing the consideration of text production and setting. This article presents a case study for identifying contextual meaning with the aim of becoming a useful and applicable working model. Therefore, terminology extends beyond purely prescriptive, onomasiological approaches. According to L'Homme (ibid.), once terms have been identified as *lexical units* (single-word terms),[2] the emphasis in terminology and terminography shifts towards the study of word combinations in which such terms are found together – designated as lexical collocations. Collocations are defined as word pairs or groups that frequently co-occur and are characterised by three key features: their frequency of co-occurrence, their proximity in the text, and their collocational strength (Brezina et al., 2015; Brezina, 2018). Corpus pattern analysis is used to explore how the surrounding words (syntagmatic criteria) reveal the varied meanings of a term with multiple interpretations (Faber & L'Homme, 2014), or of a word which obtains a specialised meaning in relation to its context and the register.

2 Core Inquiry and Methodological Approach

This study examines the specialised language[3] and terminology found in Public Administration (PA) documents produced in Greece and France between 2010 and 2018, within two thematic areas: (a) Employment, and (b) Economy, domains that reflect the political identity of any administration. This research is primarily motivated by the linguistic challenges posed by documents issued by Greek public authorities. In this context, a comparative and parallel analysis of French administrative language is highly relevant for two main reasons: (a) France did not experience a fiscal crisis of equal severity, and (b) French administrative discourse operates within a regulatory framework, offering a significant point of reference for examining specialised language and its terminology. Furthermore, the decision to conduct comparative analysis was driven by the study's methodological approach. The study addresses three central questions:

a) How are recurrent morphosyntactic patterns and grammatical constructions realised across the two languages?

[1] "[…] le terme se matérialise dans les textes" (L'Homme 2004, p. 81).

[2] As *lexical unit* is considered any lexical item with a specific meaning. Lexical units can be realised as either single-word items or multiword expressions (L'Homme, 2020)

[3] The terms *special language* (Sager, 1990; Cabré, 1998) and *specialised language* (Sinclair & Ball, 1996; Bowker & Pearson, 2002) are used interchangeably. They both refer to language used in specific domains or professions. A more prevalent term in academic circles is *Languages for Specific Purposes* (LSP) (Hoffmann, 1979; Picht, 1987; Beaugrande, 1987; Bergenholtz & Tarp, 1995; Wright & Budin, 1997), which encompasses both *special* and *specialised* languages. LSP is defined as the "natural language used in communication between experts in a domain and characterised by the use of specific linguistic means of expression" (ISO 1087, 2019).

Table 1 Statistics of the corpus

Ad hoc bilingual corpus	Number of PA docs	Number of words
Total	679	4.119.530
GR_subtotal	457	2.323.489
Employment sub corpus	226	1.007.928
Economy sub corpus	231	1.315.561
FR_subtotal	222	1.796.041
Employment sub corpus	113	989.099
Economy sub corpus	109	806.942

b) To what extent is terminological usage consistent, and how is morphological variation managed?
c) How do multi-word terms (MWTs) achieve functional equivalence across Greek and French, and what insights can comparable corpora provide into their structure?

According to ISO 1087 (2019), "a multi-word term is a complex term that consists of more than one word"; Hamon and Nazarenko (2001) argue that complex terms are composed of a head and an expansion.

In Greek, a term may be expanded through one of the following structures:[4]

1) Addition of an adjective as a modifier (Adj+N), ex. *αδήλωτη* ***εργασία*** (undeclared **work**);
2) Addition of a noun as a modifier in genitive case (N+Ngen), ex. ***επίδομα*** *θέσης* (position **allowance**);
3) Addition of another noun in paratactic coordination (N+N), ex. ***κράτος*** *μέλος* (member **state**);
4) Addition of a noun preceded by a preposition (N+PREP+N), ex. *απαλλαγή από το* ***φόρο*** (**tax** exemption).

Gavriilidou and Lampropoulou (2004) explain that language users create three-word terms and multi-word expressions consisting of more than two-term elements by adding new components to the core of a two-word term, thereby expanding its terminological structure by:

a) adding a new head or a new modifier;
b) adding another modifier;
c) combining two two-word terms;
d) linking modifiers or heads through coordination (copulative, disjunctive, or paratactic).

In French, term expansion is realised through one of the following syntactic patterns:[5]

1a) Addition of an adjective as a modifier (N + Adj), ex. ***travail*** *illégal* (illegal **work**);
1b) Adj + N, ex. *nouvelles* ***mesures*** (new **measures**);
2) Addition of a noun preceded by a noun (N + prep. + N) ex. ***impôts*** *en capital* (capital **taxes**);
3) Addition of a verb preceded by a preposition (N + prep. + V), ex. ***allocation*** *à verser* (**benefit** to be paid).

According to Daille et al. (1994), Daille (1994), and Gaussier (2001), two-word terms in French develop and expand into three-word term structures through one of the following processes:

(a) composition,
(b) modification (i.e., the insertion of a modifier), and
(c) coordination.

The compilation of the corpus was guided by the principles of authenticity, representativeness, and careful sampling (Tognini-Bonelli, 2001). The final corpus is a bilingual and comparable resource, characterised as static, complete, and synchronic.[6] It comprises 679 documents produced within the sphere of PA, amounting to 4,119,530 words in total (Table 1). For analytical purposes, the material was

[4] Additional subcategories of two-word term formation can be identified in Greek. The words in bold indicate the terms.

[5] Additional subcategories of two-word term formation can be identified in French. Identically, the words in bold indicate the terms.

[6] Taking into account the variable of temporality, corpora are distinguished as synchronic corpora, that is, those whose period of production and publication is specific and relatively short, and as diachronic or historical corpora, which comprise texts of the same language from different time periods (Kennedy, 1998; Baker, 2006; McEnery & Hardie, 2012; Gabrielatos, 2014) argue that "a synchronic corpus contains texts drawn from a single period –typically the present or the very recent past".

organised into four sub-corpora – two corresponding to each thematic domain – across the two languages examined. Texts in Greek and French were selected based on the following criteria: (i) official documents issued by PA institutions, (ii) accessible either in printed and/or digital form, (iii) specialised in content, (iv) relevant to the thematic areas of Employment and Economy, and (v) issued within the time framework from January 2010 till December 2018, a period of intense political ferment prompted by adverse economic and social conditions, which resulted in rapid transformations within Public Administration and an increased production of administrative documents.

The lower volume of administrative documents in France, compared to Greece, can be attributed to structural differences in legislative practice. In Greece, law-making is characterised by heightened intensity and frequent initiatives presented as "reforms," a term that has gained prominence since 2010. These reforms are often introduced without comprehensive long-term planning, which subsequently generates the need for clarifications and interpretative guidance. This dynamic results in successive modifications, additional circulars, and an overall expansion in administrative texts. Such proliferation, however, does not necessarily lead to greater clarity or more effective regulation. Rather than serving a merely supplementary role, it frequently functions as a mechanism of institutional adaptation, compensating for legal and regulatory lacunae, orienting administrative authorities, and equipping public officials to address the practical complexities of legislative implementation. Concurrently, it facilitates a process of normative evolution by enabling the incorporation of practices not originally envisaged by legislators but subsequently recognised as integral to the effective functioning of the system.

For this study, the widely used and dynamic software Sketch Engine was chosen for its advanced technology and continually updated tools and features.

3 Term Recognition and Retrieval

A semi-automated approach was employed to capture: a) the structure of MWTs and their collocational environments in Greek and French, and b) the most frequent and statistically significant terms and collocations as well as broader clusters, regardless of their frequency. Wordlists were extracted first, followed by keyword lists generated with Sketch Engine tools. The GreekWeb 2014 corpus (elTenTen14) was chosen as the reference corpus for Greek, while the FrenchWeb 2012 corpus (frTenTen12) was selected for French.[7] The study focused on *nouns*, as they are the fundamental units of terminology – encompassing both single-word and multi-word forms – and play a pivotal role in forming collocations and larger lexical clusters. To ensure a comprehensive examination of the wordlist, a keyword list was integrated. Accordingly, the keyword tool was employed with the single-words option selected to identify single-word terms. Subsequent analysis drew upon relative frequency and statistical scores to extract keywords from each sub-corpus. Finally, the most recurrent lexical items were identified and compared across the selected language pairs. The focus of this study is the investigation of a word selected as a *key term*, which is then analyzed within a concordancer. Concordancers' data provide insights both into the meaning of the key term - referred to here as a node term (NT) - and into its patterns of use.

In large corpora, comprising millions or even billions of words, multiple occurrences of the NT can be identified, allowing for a thorough examination of its linguistic behavior and the range of possible meanings. This corpus-driven approach is widely employed by lexicographers in the study of general-language corpora. The study specifically sought to investigate whether semantically and translationally equivalent words – NTs – could be identified: a) within the wordlist, and/or b) within the keyword list of the comparable corpora. Semantic equivalence refers to the degree to which linguistic units (words, phrases, or terms) in different languages convey the same meaning or denote the same concept (Baker, 1992; Cabré, 1998; Temmerman, 2000). Linguistic units that can be substituted for one another in translation and convey the same meaning or function within a given context are considered translationally equivalent. Additionally, the analysis considered the collocational strength of co-occurring

[7] The Greek Web 2019 corpus (elTenTen19), comprising 2,342,091,029 words, was released in Sketch Engine as a reference corpus in 2022. As the research for this study had already been completed, it was not incorporated into the analysis. Similarly, the French Web 2023 corpus (frTenTen23), containing 23,874,070,858 words, became available in Sketch Engine in 2023, but was likewise excluded, since the research for this dissertation had already been concluded.

Table 2 NTs extracted from the Bilingual Corpus

Bilingual corpus	French NT	Greek NT	English Translation of the NTs
Employment sub corpus	travail	εργασία	work
	sécurité/assurance	ασφάλιση	security/insurance
	emploi	απασχόληση	employment
	allocation	επίδομα	allowance/benefit
	cotisation/contribution	εισφορά	contribution/levy
	chômage	ανεργία	unemployment
Economy sub corpus	dépense	δαπάνη	expenditure
	impôt	φόρος	tax
	dette	χρέος	debt
	développement	ανάπτυξη	development
	investissement	επένδυση	investment
	revenu/recette	έσοδα	revenue/income
Employment & Economy sub corpora	réforme	μεταρρύθμιση	reform
	mesure	μέτρο	measure

words that constitute MWTs and collocations, as this metric can serve as a criterion for automatic term extraction. As highlighted by Ananiadou and Zervanou (2004), mutual information quantifies the extent to which a word provides information about other words with which it frequently co-occurs.

After identifying common NTs across the comparable corpora, the study proceeded with a contrastive analysis of the findings. Selection of NTs for further investigation was guided by the following cumulative criteria: a) the NT must be present in the wordlist, indicating its frequency within the corpus, b) it should possess semantic and pragmatic relevance, c) it must be representative of the subject field pertinent to the specific text type, with preference given to NTs also appearing in the keyword list, and d) it should contribute to the formation of MWTs and collocations.

Operating under the assumption that Greek and French terms would correspond, the contrastive analysis employed both quantitative and qualitative methods to assess the extent and nature of correspondences for the selected NTs within the two comparable corpora. Fourteen terms were selected per language: six related to the subject field of Employment in Greek, six to Employment in French, six to Economy in Greek, and six to Economy in French (Table 2). Notably, two NTs occurred consistently across all four subcorpora, appearing in both subject fields and in both languages.

All selected NTs were among the 350 most frequent words in the wordlists. Given that each sub corpus contains over 10,000 entries, the results demonstrate a clear correspondence in occurrence frequency across the four subcorpora, confirming the initial hypothesis (Christopoulou, 2024).

4 Summary of Findings

4.1 Formation of MWTs

Both languages display systematic structures in MWTs, where bi-lexical forms typically consist of a head and a modifier. Three-component terms and broader collocations, which consist of more than two term elements, are formed by adding new term elements to the *core of the bi-component term*, thereby expanding the terminological structure. Valeontīs et al. (1999) argue that a *term element* is defined as any morphological component of a term that carries meaning or possesses a semantically differentiating function. Tri-lexical terms and larger clusters, consisting of more than two-term elements, are typically created by incorporating additional terms into the core of the bi-lexical structure, thereby enriching and extending the terminological unit (Gavriīlidou & Lampropoulou, 2004).

According to L'Homme (2004), MWTs in French are formed according to a limited number of grammatical category patterns. Most multi-word French terms arise through the combination of a noun with an adjective, another noun, or a prepositional phrase. In the corpus under study, no four-word terms were identified; co-occurring word sequences exceeding three lexical items are therefore treated as lexical clusters rather than as MWTs. Beyond co-occurrence frequency, the analysis also considers the collocational strength of term elements, which necessitates restricting the examination of MWTs to tri-lexical units. Prepositions are not regarded as term elements. As noted by Riegel, Pellat, and Rioul (2009), any preposition or prepositional phrase may introduce a nominal complement, whose semantic interpretation depends both on the intrinsic meaning of the preposition and on the contextual relation between the connected constituents. Among French prepositions, *de* is the most frequently employed in the formation of MWTs.

4.2 Structural and Morphological Patterns

The analysis revealed symmetrical tendencies in MWT formation, with Greek bi-lexical terms clustering around N+Ngen and A+N, while French relied heavily on N+PRE+N and N+A. These patterns show structural parallels, particularly where French prepositions correspond to Greek genitive case markers. Both corpora demonstrated a heavy reliance on syntactic devices (genitive or preposition) for expanding MWTs.

4.3 Collocational Strength and Methodological Criteria

To ensure that only stable and meaningful lexical combinations were analysed as MWTs, collocational strength was measured using the log Dice statistics (Rychlý, 2008). For the measurement of the collocational strength the log Dice score was deemed as the most suitable for the analysis of MWTs, as this index allows for the comparison of collocations across corpora, in contrast to other statistical measurements of collocational strength. The maximum value of the log Dice is 14, but practically, collocations with a value equal to or greater than 10 are considered strong. The value of this index is independent of the corpus size, which allows researchers to compare textual data consistently and stably, and for this reason, it is preferred in the qualitative analysis of very large corpora. Usually, the values of statistical indices are dependent on the reference corpus, thus rendering it less practical for comparing collocations from comparable corpora or for data interpretation. The log Dice is based on the frequency of the NT and the collocate, and the frequency of the entire collocation (co-occurrence of the NT and the collocate). LogDice is also used in bilingual term extraction to identify the most probable translation.[8]

4.4 Administrative Discourse and Terminological Behavior

As for the tendencies in administrative discourse, French favours the use of abbreviations as a means of terminological condensation, while Greek displays a preference for named entities as reference anchors, thereby shaping distinct patterns of terminological representation across the two languages.

4.5 Semantic and Contrastive Analysis of MWTs

The quantitative and qualitative analysis of MWTs (based on corpus data) allows for the contrastive examination of the functional equivalence of the selected terms. Subsequently, the study explores the distinctive features of specialised language in administrative documents in Greece and France. Key characteristics of PA language include long complex sentences; frequent use of conjunctions indicative of multiple subordinate clauses, repetition of words instead of pronouns to ensure referential precision, adverbial clauses, nominal groups, recurrent collocation, elaborate syntax, extensive nominalisation, frequent use of the passive voice, and the prevalence of transitive verbs. This language constitutes a register, defined as a functional variety of language associated with a specific context or situational type.

[8] LogDice in https://www.sketchengine.eu/glossary/logdice/ [Accessed 06/11/ 2025]

Table 3 Distribution of most frequent NTs contributing to MWT formation in Greek and French corpora

Greek Corpus			
NT	μέτρο	δαπάνη	
NTs occurrences	105	89	194
% occurrence in the GR_Corpus	13,73	11,63	25,36
% occurrence in the Bilingual Corpus	7,42	6,29	13,71
French Corpus			
NT	dépense	mesure	
NTs occurrences	78	65	143
% occurrence in the FR_Corpus	11,98	9,98	21,96
% occurrence in the Bilingual Corpus	5,51	4,59	10,1

A more detailed analysis of the textual material from the four comparable sub-corpora yielded the following results.

4.5.1 Statistical Data on MWTs of Two and Three Components

In Table 3, term structures from both corpora are presented, along with a visualisation of the NTs that most frequently participate in term formation. The same NTs in both languages – *μέτρο* (measure) and *δαπάνη* (expenditure) from the Greek corpus, and *dépense* (expenditure) and *mesure* (measure) from the French corpus – were found to form the most numerous MWTs.

4.5.2 Frequency, Distribution, and Cross-Linguistic Symmetry

In the Greek corpus, *μέτρο* (measure) and *δαπάνη* (expenditure) together account for more than a quarter of all MWTs (25.36%), with individual frequencies of 13.73% and 11.63%, respectively. The French corpus shows a comparable pattern: *dépense* (expenditure) and *mesure* (measure) collectively represent 21.96% of MWTs, respectively occurring at rates of 11.98% and 9.98%. This concentration indicates that, in both languages, discourse around economic and employment issues relies heavily on a limited set of high-frequency conceptual anchors.

This distribution reveals a striking cross-linguistic symmetry: the same semantically aligned terms (*μέτρο*/*mesure* and *δαπάνη*/*dépense*) dominate both corpora and exhibit comparable proportional weight. Their recurrence suggests that terminology transfer in the domain of Economy and Employment is facilitated by a shared conceptual nucleus, where parallel notions are probable to function as productive anchors for MWT formation. This convergence underscores the universality of key macroeconomic categories, reinforcing the cross-linguistic portability of concepts central to policy, regulation, and financial discourse.

At the same time, subtle divergences are also evident. In the Greek corpus, the slightly stronger prominence of *μέτρο* indicates a discourse more frequently oriented toward policy instruments and regulatory measures. By contrast, in the French corpus, *dépense* assumes greater relative weight, pointing to a discursive emphasis on financial allocation and expenditure management. These nuanced differences suggest that while both languages share a common conceptual architecture, each reflects distinct discursive priorities in framing economic and employment issues.

Table 4 Examples of acronyms and abbreviations lists from the bilingual corpus

Acronym/ abbreviation	Full form	Score	Freq	Ref freq	Rel freq	Rel ref freq
ΠτΚ	Πτωχευτικός Κώδικας	73.690	132	314	84.498	0.160
PACEA	Parcours contractualisé d'accompagnement vers l'emploi et l'autonomie	325.790	394	334	340.681	0.049

Table 5 Examples of named entities lists from the bilingual corpus

NT	Named Entity	Corpus
Ασφάλιση	Ειδική Υπηρεσία Ελέγχου Ασφάλισης (ΕΥΠΕΑ)	GR sub corpus_Employment
Επένδυση	Ανώνυμες Εταιρείες Διαχείρισης Οργανισμών Εναλλακτικών Επενδύσεων (ΑΕΔΟΕΕ)	GR sub corpus_Economy
Assurance	Assurance complémentaire en santé (ACS)	FR sub corpus__Employment
Travail	Plan national de lutte contre le travail illégal (PNLTI)	FR sub corpus__Employment

4.5.3 Implications for Terminology Transfer

The prominence of *μέτρο*/*mesure* highlights regulatory and policy instruments – emphasizing institutional responses and mechanisms, while the occurrences of *δαπάνη*/*dépense* stand for financial allocation and cost structures – foregrounding expenditure and resource management.

Their high frequency not only illustrates their lexical productivity in MWTs formation but also reveals their role as conceptual pivots that facilitate terminological transfer across Greek and French. This suggests that the portability of economic terminology is anchored in the recurrence of a shared set of NTs, enabling parallel formations to emerge across languages while preserving domain-specific precision.

4.6 Lemmatological Lists of MWTs

As mentioned above, the Greek corpus contains 745 MWTs, while the French corpus includes 641 MWTs, resulting in a total of 1,386 MWTs.

4.7 Lemmatological Lists of Acronyms and Abbreviations

Using the keyword lists from each of the four sub-corpora, shortened forms were manually extracted into four distinct lemmatological lists. A notable number of acronyms and abbreviations were identified in the French sub-corpora (Table 4).

4.8 Lemmatological Lists of Named Entities

The list includes only those named entities that are associated with the NTs being analysed (Table 5). It is observed that certain NTs appear frequently in multiple named entities, such as the Greek NT *ασφάλιση* (insurance) and *επένδυση* (investment), and the French NT *assurance* (insurance) and *travail* (work). From the analysis, it follows that the Greek Corpus contains a greater number of named entities compared to the French Corpus. Specifically, 157 named entities from the Greek Corpus and 69 named entities from the French Corpus are included in the lemmatological list from the selected NTs.

Table 6 Snapshot of the bilingual glossary

NT_corpus	Greek NT	French NT	English Translation of the NTs
Εργασία/Travail_ Employment	βελτίωση των συνθηκών εργασίας	amélioration des conditions de travail	improvement of working conditions
	εργασία ιδιωτικού δικαίου	travail de droit privé	employment governed by private law
	εργασία αορίστου χρόνου	travail à durée indéterminée	open-ended employment
Φόρος/Impôt_ Economy	φόρος ακίνητης περιουσίας	impôt sur la fortune immobilière	real estate wealth tax
	φόρος εισοδήματος	impôt sur le revenu	income tax
	φόρος επί των ακινήτων	impôt sur la fortune	wealth tax

4.9 Bilingual Glossary of Term Equivalences

A contrastive analysis of the terms at the pragmatic level led to the compilation of a bilingual glossary of common terms from both corpora with semantic and translational equivalence – or at least pragmatic correspondence (Table 6). The glossary comprises 90 term equivalences.

4.10 Correspondence in the Morphosyntactic Analysis of Terms

Correspondence was observed in the morphosyntactic structure of terms in both Greek and French, with symmetry found in the formation of MWTs from specific parts of speech in both languages.

The comparison of the lexical data reveals that the 447 Greek bi-lexical terms adhere to three syntactic structures, with the most common being:

i) Noun + Noun in genitive (N+Ngen), ex. *εισφορά αλληλεγγύης* (/isforá allilengíis/, in English: *solidarity levy*).
ii) Adjective + Noun (A+N), ex., *αδήλωτη εργασία* (/adíloti ergasía/, in English: *undeclared work*).

In the French corpus, 425 bi-lexical terms were identified, formed using four syntactic structures, with the most frequent structures being:

i) Noun + Preposition + Noun (N+PRE+N), ex. *mesures de protection* (mə.zyʁ də pʁɔ.tɛk.sjɔ̃/, in English: *protective measures*).
ii) Noun + Adjective (N+A), ex. *travail illegal* (/tʁa.vaj i.le.gal/, in English: *illegal work*).

It is observed that the French structure Noun + Preposition + Noun (N+PRE+N) aligns with the most frequent structure of bi-lexical terms in Greek (N+Ngen), as the preposition de in French primarily functions to express the genitive case. This demonstrates perfect symmetry between the two languages in the formation of bi-lexical terms.

The comparative analysis of three-word term structures across the corpora identified 318 Greek and 223 French instances. Overall, fifteen distinct syntactic patterns were observed among the Greek three-word terms, whereas the French terms exhibited ten distinct patterns.

In the Greek corpus, the most frequent three-word terms adhere to the following structures:

i) Noun + Adjective in genitive + Noun in genitive (N+Adjgen+Ngen), ex. *σύστημα κοινωνικής ασφάλισης* (/sístima kinonikís asfálisis/, in English: *social security system*).
ii) Noun + Noun in genitive + Noun in genitive (N+Ngen+Ngen), ex. *Ημερολόγιο Μέτρων Ασφάλειας* (/imerolóyio métron asfálias/, in English: *Safety Measures Log Book*).

In the French corpus, the most frequent syntactic structure for the three-word clauses of the NTs is:

i) Noun + Preposition + Noun + Adjective (N+PRE+N+Adj), ex. *ratio de dette publique* (/ra.si.o də dɛt py.blik/, in English: *public debt ratio*) followed in frequency by:
ii) Noun + Preposition + Noun + Preposition + Noun (N+PRE+N+PRE+N), ex. *rupture du contrat de travail* (/ʁyp.tyʁ dy kɔ̃.tʁa də tʁa.vaj/, in English: *termination of employment contract*).

The analysis of the syntactic structures of three-word terms in the French corpus reveals a notably high frequency of prepositions, particularly the preposition *de*, in the formation of MWTs. Both the quantitative and qualitative analysis of the three-word terms in the corpus demonstrate a clear correspondence and symmetry in the predominant formation patterns across both languages.

5 Further Research Perspectives

The study of the functional equivalence of terms through contrastive comparison provides insights into the ways comparable corpora can inform lexicography, translation, language teaching, and Natural Language Processing (NLP) applications. Term equivalence in the comparable corpus is not always one-to-one, as one MWT may correspond to multiple MWTs in the comparable corpus. Therefore, all terms considered equivalent within the specific context are recorded, based on textual evidence and corroborated through documentation and verification from other sources (parallel texts, legislative texts, authoritative electronic sources and other supporting documentation). The implications of this methodology extend to lexicography, translation, terminology teaching and NLP:

1) In lexicography, the systematic extraction of specialised PA vocabulary contributes to lexicographic practice by identifying recurrent terms, collocations, and MWTs suitable for the creation of dictionaries, thesauri, and digital terminology repositories. The corpus-driven approach ensures that such resources are enriched with contextualised examples, reflecting authentic usage rather than prescriptive lists.
2) Regarding the translation of PA documents, this research supports translators working with administrative discourse. Knowledge of recurring MWTs and their bilingual equivalence enhances accuracy and terminological consistency, particularly in regulatory texts issued by supranational institutions. This approach also clarifies concepts that are often conflated due to legislative plurality or institutional overlap.
3) InTerminology teaching the study of collocations and terms within authentic corpora enables learners to grasp not only lexical meanings but also the communicative functions of terminology in context. The use of concordancers allows learners to explore the full semantic range of a term as it occurs in authentic texts.
4) As for for NLP, a key application of the proposed methodology for extracting terminology in PA is the development of terminological databases that facilitate information retrieval from administrative documents. Consolidating legislation, updating it as amendments are issued, incorporating new terms and concepts, and linking them to laws, circulars, and other regulatory texts – while offering user access through publicly available digital platforms – would constitute a valuable tool for administrative efficiency and transparency. PA institutions could further leverage this potential by developing software systems that integrate language-technology applications, such as domain-specific spell checkers, grammar checkers, and terminology repositories tailored to the linguistic conventions of PA discourse.

Beyond its direct NLP applications, the project also contributes to digital governance by ensuring that the linguistic data produced is not only usable but also sustainably reusable for future research and technological development.

In this regard, the research aligns with the FAIR data principles – Findability, Accessibility, Interoperability, and Reusability – by producing a curated, reusable, and linguistically annotated corpus of PA documents enriched with metadata to support cross-linguistic analysis. Making authentic corpora machine-actionable ensures reproducibility and encourages future empirical research on terminology,

discourse practices, and digital governance. The FAIR framework is designed to enhance the long-term impact of digital research outputs, enabling both humans and machines to easily locate, access, and integrate data in a transparent and sustainable way (Wilkinson et al., 2016; Mons et al., 2017). Its implementation is growing across public-sector organizations and industries, especially as governments adopt open-data policies and prioritise evidence-based decision-making (European Commission, 2019; OECD, 2023).

6 Conclusions

Corpus-based terminology research shows that conventional resources, such as glossaries and dictionaries, are insufficient for capturing the dynamic nature of specialised language. Digital corpora, combined with computational tools, enable precise tracking of terminological behavior and the documentation of emerging terms not yet codified in dictionaries (Frérot & Pecman, 2021; Pearson, 1996).

This study illustrates the value of a corpus-driven approach to PA discourse in Greek and French, focusing on the domains of Employment and Economy. Analysing structural configurations and semantic behavior of MWTs reveals cross-linguistic parallels, functional correspondences as well as asymmetries. The findings highlight how terms and collocations reflect structural, semantic and functional equivalences, demonstrating the complementarity of Textual Terminology and Corpus Linguistics in producing reusable, bilingual resources.

Candidate collocations, identified through statistical measures of word association, represent a central functionality of Corpus Linguistics tools (Rychlý, 2008). Language technology tools further enhance translation practice by enabling: a) deeper understanding of source-text terms, b) enrichment of terminology databases and software, and c) improvement of dictionaries (L'Homme, 2008).

Beyond descriptive analysis, this research contributes to applied fields by providing resources for translators, educators, and lexicographers, and by informing the design of NLP tools tailored to PA language. Its originality lies in combining corpus-driven methods with Textual Terminology to support multilingual communication, knowledge transfer, and terminological standardisation.

The study also underscores the importance of aligning linguistic research with FAIR principles (cf. Vezzani, 2022), ensuring that terminological knowledge is preserved, accessible, and transferable across professional and academic contexts. These findings demonstrate that corpus-based approaches not only document specialised language, but also generate practical, reusable resources that advance both scholarship and professional practice.

References

Ananiadou, S., & Zervanou, K. (2004). Anagnorisī oron. In M. Katsogiannou & E. Euthymiou (Eds.), *Ellīnikī Orologia: Ereuna kai Epharmoges* (pp. 283-311). Ekdoseis Kastaniotī AE.

Auger, P., & L'Homme, M.-C. (1994). La terminologie selon une approche textuelle : une représentation plus adéquate du lexique dans les langues de spécialité. In *Terminologie et linguistique de spécialité. Étude du vocabulaire et textes spécialisés (*ALFA, 7/8, pp. 17-21). Dalhousie University.

Baker, M. (1992). *In other words. A coursebook in translation*. Routledge.

Baker, P. (2006). *Using corpora in discourse analysis*. Continuum.

Beaugrande, R. de (1987). Special purpose language and linguistic theory. *UNESCO ASLED-LSP Newsletter*, *10*(2), 2-11.

Bergenholtz. H., & Tarp, S. (1995). *Manual of specialised lexicography: The preparation of specialised dictionaries*. John Benjamins Publishing Company. https://doi.org/10.1075/btl.12

Bourigault, D., & Slodzian, M. (1999). Pour une terminologie textuelle. *Terminologies nouvelles*, 19, 29-32.

Bowker, L., & Pearson, J. (2002). *Working with specialized language : A practical guide to using corpora*. Routledge.

Brezina, V. (2018). *Statistics in Corpus Linguistics. A Practical Guide*. Cambridge University Press. https://doi.org/10.1017/9781316410899

Brezina, V., McEnery, T., & Wattam, S. (2015). Collocations in context: A new perspective on collocation networks. *International Journal of Corpus Linguistics*, *20*(2), 139-173. https://doi.org/10.1075/ijcl.20.2.01bre

Cabré, M. T. (1998). *La terminologie : Théorie, méthode et applications*. Les Presses de l'Université d'Ottawa / Armand Colin.

Christopoulou, A. (2024). *Eidikī glossa kai orologīa tīs dīmosias dioikīsīs: antiparavolikī meletī ellīnikīs-gallikīs me chrīsī sōmáton keimenōn stous tomeīs tīs apascholīsīs kai tīs oikonomīas* [Doctoral dissertation, University of Athens, Department of French Language and Literature]. National Archive of PhD Theses. http://dx.doi.org/10.12681/eadd/57022

Condamines, A. (2003). *Sémantique et corpus spécialisés : Constitution de bases de connaissances terminologiques* (Doctoral dissertation). Université Toulouse Le Mirail.

Condamines, A. (2005). Linguistique de corpus et terminologie. *Langages*, *157*(1), 36-47. https://doi.org/10.3917/lang.157.0036

Condamines, A., & Picton, A. (2022). Textual terminology: Origins, principles and new challenges. In P. Faber, & M.-C. L'Homme (Eds.), *Theoretical perspectives on terminology: Explaining terms, concepts and specialised knowledge* (pp. 219-236). John Benjamins Publishing Company. https://doi.org/10.1075/tlrp.23.10con

Daille, B. (1994). *Combined approach for terminology extraction: lexical statistics and linguistic filtering*. Université Paris VII, TALANA. Retrieved November 1, 2025, from http://ucrel.lancaster.ac.uk/papers/techpaper/vol5.pdf

Daille, B., Gaussier, E., & Langé, J. (1994). Towards automatic extraction of monolingual and bilingual terminology. In *COLING 1994: The 15th international conference on computational linguistics* (Vol. 1, pp. 515-521). Kyoto.

European Union (2019). Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast). *Official Journal of the European Union*, L 172/56. http://data.europa.eu/eli/dir/2019/1024/oj

Faber P., & L'Homme, M.-C. (2014). Lexical semantic approaches to terminology: An introduction. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, *20*(2), 143-150. https://doi.org/10.1075/term.20.2.01int

Frérot, C., & Pecman, M. (Eds.). (2021). *Des corpus numériques à l'analyse linguistique en langues de spécialité*. UGA Éditions. https://doi.org/10.4000/books.ugaeditions.24195

Gabrielatos, C. (2014). *Corpus Approaches to Discourse Studies: The basics*. Discourse & Communication Studies Research Team. Örebro University.

Gaussier E. (2001). General considerations on bilingual terminology extraction. In D. Bourigault, C.

Jacquemin, & M.-C. L'Homme (Eds.), *Recent advances in computational terminology* (pp. 167-183). John Benjamins Publishing Company. https://doi.org/10.1075/nlp.2.09gau

Gavrilīdou, M., & Lampropoulou, P. (2004). Exagogī oron apo keimena. Mia yvridikī methodos. In M. Katsogiannou, & E. Euthymiou (Eds.), *Ellīnikī Orologīa: Ereuna kai Efarmoges* (pp. 313-326). Athīna: Ekdoseis Kastaniotī AE.

Hamon T., & Nazarenko, A. (2001). Detection of synonymy links between terms: Experiment and results. In D. Bourigault, C. Jacquemin, & M.-C. L'Homme (Eds.), *Recent advances in computational terminology* (pp. 185-208). John Benjamins Publishing Company. https://doi.org/10.1075/nlp.2.10ham

Hoffmann, L. (1979). Towards a theory of LSP. Elements of a methodology of LSP analysis. *Fachsprache*, 1, 12-17.

ISO 1087. (2019). *Terminology work and terminology science - Vocabulary*. International Organization for Standardization. Retrieved from https://www.iso.org/standard/62330.html

Kennedy, G. (1998). *An Introduction to corpus linguistics*. Routledge.

L'Homme, M.-C. (2004). *La terminologie: Principes et techniques*. Presses de l' Université de Montréal.

L'Homme, M.-C. (2020). *Lexical semantics for terminology: An introduction*. John Benjamins Publishing Company. https://doi.org/10.1075/tlrp.20

L'Homme, M.-C. (2008). *Initiation à la traductique* (2e éd.). Linguatech.

McEnery, T., & Hardie, A. (2012). *Corpus Linguistics*. Cambridge University Press. https://doi.org/10.1017/CBO9780511981395

Mons, B., Neylon, C., Velterop, J., Dumontier, M., da Silva Santos, L. O. B., Wilkinson, M. D. (2017). Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. *Information Services and Use*, *37*(1), 49-56. https://doi.org/10.3233/ISU-170824

OECD. (2023). *OURdata Index: Open, useful and re-usable data index*. Organisation for Economic Co-operation and Development. https://www.oecd.org

Pearson, J. (1996). Strategies for identifying terms in specialised texts. *TEANGA: The Irish Yearbook of Applied Linguistics*, *16*, 33-42. Retrieved November 1, 2025, from https://files.eric.ed.gov/fulltext/ED414748.pdf

Pearson, J. (1998). *Terms in context*. John Benjamins Publishing Company. https://doi.org/10.1075/scl.1

Picht, H. (1987). Terms and their LSP environment - LSP phraseology. *Vers l'an 2000. La terminotique, bilan et prospectives*, 32(2), 149-155. https://doi.org/10.7202/003836ar

Riegel, M., Pellat, J.-C., & Rioul, R. (2009). *Grammaire méthodique du français*. PUF.

Rychlý, P. (2008). A Lexicographer-Friendly Association Score. In P. Sojka & A. Horák (Eds.), *Proceedins of recent advances in slavonic natural language procesing, RASLAN* (pp.6-9). Retrieved November 1, 2025, from https://www.sketchengine.eu/wp-content/uploads/2015/03/Lexicographer-Friendly_2008.pdf

Sager, J. C. (1990). *A practical course in terminology processing*. John Benjamins Publishing Company. https://doi.org/10.1075/z.44

Sinclair, J., & Ball, J. (1996). EAGLES *preliminary recommendations on text typology* (Internal working document). Retrieved November 1, 2025, from http://www.ilc.cnr.it/EAGLES96/texttyp/texttyp.html

Temmerman, R. (2000). *Towards new ways of terminology description: The sociocognitive approach*. John Benjamins Publishing Company. https://doi.org/10.1075/tlrp.3

Teubert, W. (2004). *Lexicology and corpus linguistics: An introduction*. In M. Halliday, W. Teubert, C. Yallop, & A. Cermáková (Eds.). Continuum.

Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. John Benjamins Publishing Company. https://doi.org/10.1075/scl.6

Valeontīs, K., Zeritī, K., & Nikolakī, A. (1999). To prosdioristiko syntheto tou ellīnikou symplokou orou kai ī chrīsī tīs genikīs. *ELETO - 2o Synédrio "Ellīnikī Glossa kai Orologīa"*, (pp. 283-316). Athīna. Retrieved September 1, 2025, from https://eleto.gr/download/Conferences/2nd%20Conference/Tomos-2ou_Synedriou.pdf

Vezzani, F. (2022). *Terminologie numérique : conception, représentation et gestion*. Peter Lang. https://doi.org/10.3726/b19407

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018. https://doi.org/10.1038/sdata.2016.18

Wright, S. E., & Budin, G. (1997). *Handbook of terminology management: Basic aspects of terminology management* (Vol. I). John Benjamins Publishing Company. https://doi.org/10.1075/z.htm1