Preliminary Considerations on a Systematic Approach to Semic Analysis: The Case Study of Medical Terminology

[1]Vanessa Bonato, [2]Giorgio Maria Di Nunzio, [3]Federica Vezzani

[1,2,3]University of Padua, Italy

[1]vanessa.bonato@studenti.unipd.it
[2]giorgiomaria.dinunzio@unipd.it
[3]federica.vezzani@unipd.it

Abstract

Semic analysis is a linguistic technique aimed at capturing the essential specificities of terms meaning through the identification of minimum semantic units. This procedure is functional for the achievement of an in-depth comprehension of technical terminology and the acquisition of a specialised conceptual knowledge. In this paper, we focus on semic analysis applied to medical terminology. In particular, we discuss some preliminary considerations in order to establish the starting points for a systematic approach to semic analysis. Firstly, we propose a preliminary experiment to 1) study users' perception of semic analysis and 2) validate the absence of systematicity in its performance. Based on the resulting data, we secondly propose a methodology aiming at increasing the systematic factorisation of semic analysis. Finally, we propose an experimental study to investigate on the potential interrelation in terms of applicability and productivity of Word Embeddings with respect to semic analysis in the framework of the proposed methodological criteria.

L'analisi semica è una tecnica linguistica volta a cogliere le specificità essenziali del significato dei termini attraverso l'individuazione di unità minime di significato. Questa procedura è funzionale al raggiungimento di una comprensione approfondita della terminologia tecnica e all'acquisizione di una conoscenza concettuale specializzata. In questo articolo, ci concentriamo sull'analisi semica applicata alla terminologia medica. In particolare, discutiamo alcune considerazioni preliminari al fine di stabilire i punti di partenza per un approccio sistematico all'analisi semica. In primo luogo, proponiamo un esperimento preliminare per 1) studiare la percezione degli utenti dell'analisi semica e 2) validare l'assenza di sistematicità nelle sue prestazioni. Sulla base dei dati ottenuti, proponiamo in secondo luogo una metodologia volta a favorire la fattorizzazione sistematica dell'analisi semica. Infine, proponiamo uno studio sperimentale volto ad indagare sulla potenziale interrelazione in termini di applicabilità e

produttività dei Word Embeddings rispetto all'analisi semica nel quadro dei criteri metodologici proposti.

1. Introduction

Terminology performs a pivotal function in the context of specialised knowledge. The study of the relation between its linguistic and conceptual dimension is essential to accomplish its objectives, among which term description. The relevance and usefulness of terminology notably emerge when considering the professional contexts in which it is implicated, namely specialised translation and interpreting, and its prominent function in the field of Information Retrieval ([15]). The employment of terminology in different domains and professional activities emphasises the heterogeneous nature of its applications. Therefore, the necessity for different users to productively manage terminology is determined in present-day society. The notion of term, however, manifests to some extents a spectrum of conceptual variations according to the peculiar theoretical paradigms respectively articulated in different theories of terminology ([20]; [21]). According to Gaussier ([11]), "[t]here is no fully operational definition of terms". The metalinguistic implication is the identification of the polysemic nature of the term '*term*' within the terminological context. We assume as a theoretical presupposition that "[a] term is a sign in the Saussurian sense; it is the bound relationship between a word form and *one* meaning of that word form" ([46]). Indeed, "[t]ogether, the signifier and the signified form the sign" ([46]). Moreover, we consider the approach adopted by L'Homme ([21]), which "considers terms as lexical units. This entails that they can be delimited syntagmatically and semantically".

Concerning the lexical structure that terms can assume, it can be exemplified by way of taking as a reference medical terminology. A medical term, for example, can assume the lexical form either of a single element or of a multi-word unit ([8]). However, a lack of consensus in literature can also be devised in relation to the consideration of the lexical structure of terms such as the grammatical category that terms should pertain to. For instance, as L'Homme ([21]) expresses, "[t]here is a larger consensus around *nouns* than around any other parts of speech as far as defining terminological status is concerned". As L'Homme ([21]) states, "[f]or some reason, it is much more difficult to determine the terminological status of verbs and adjectives".

Medical terminology constitutes the terminological framework of the present study. It is inscribed in the macro-structure of special languages, also referred to as specialised languages or LSP (Language for Special Purposes). It is peculiarly characterised by the subsistence of different communicative needs which manifest both at the linguistic and at the conceptual level in intra-specialist contexts and physician-patient interactions. Indeed, physicians are among the specialists that employ this specialised language. However, the medical domain also concerns patients who need to access medical language to communicate with physicians and comprehend information about diagnosis and treatments. Moreover, the accessibility of medical information to patients highly improved due to the Web ([2]). However, patients could necessitate accessing a level of knowledge of medical concepts which differs from the specialised level of conceptualisation possessed by specialists. Furthermore, domain technicalities doubtfully could

be properly understood by non-experts. In this light, a terminological and conceptual asymmetry between different users is determined. Medical terminology inherently features a communicative dimension, which is also considered by León-Araúz ([19]):

> The selection of one term or another depends on different communicative and cognitive factors. For instance, in a doctor-patient situation, the doctor will often use easily understood terms when addressing the patient, whereas in a medical conference he/she will use more specialized standardized medical terminology.

In addition, Dury ([7]) underlines the role played by the connotative aspects in the evolution of the medical language. For instance, according to Dury, the presence of a connotation judged to be insufficiently rewarding, even pejorative, can interfere with the use of a term and foster its obsolescence.

The pragmatic usage of popular terms and bit-by-bit explanations in plain language could be conceived as acts that realise a conceptual negotiation in knowledge exchange in physician-patient interactions. In such way, despite the conceptual gap engendered by the missing information, mutual comprehension is not hampered due to shared conceptual references. Moreover, it can be observed that popular terms frequently present lexical and conceptual forms which are characterised by relational connections such as conceptual generalisations with respect to the corresponding specialised terms. It is the case of the technical terms *carcinoma* and *lymphoma* that can be associated with the same popular term *cancer*.

In this context, semic analysis (see Section 2) is a linguistic technique which can be envisaged as a functional strategy for the achievement of an in-depth comprehension of medical terminology. In this paper, we perform a semic analysis based on the theoretical assumption that the term is a linguistic sign often characterised by a complex meaning, which can be decomposed into minimal units. We adopt two complementary perspectives on the study of medical terminology. Specifically, in the context of semic analysis, we study terms as contextually-independent linguistic signs. Conversely, we study terms as immersed in a contextual dimension when adopting word embeddings algorithms. We consider that:

> […] the semasiological and onomasiological approaches are two complementary terminological methodologies that should be used in the construction of knowledge representation tools. ([37])

We discuss some preliminary considerations to establish the starting points for a systematic approach to semic analysis of medical terminology. In particular, our objectives are

1. the validation of the absence of systematicity in the semic analysis procedure through a study based on the data contained in the terminological records of the TriMED resource ([44]);
2. the proposal of a methodology that aims to define a systematic approach for the performance of semic analysis;

3. the investigation of a potential interrelation in terms of applicability and productivity of Word Embeddings with respect to semic analysis in the framework of the proposed methodological criteria;

4. the investigation of the possibility to automatically or semi-automatically perform semic analysis of medical terminology.

The remainder of this paper is organised as follows. In Section 2 we present an overview of the background studies about medical terminology, semic analysis and word embeddings. Section 3 contains the preliminary experiment on terminological records. In Section 4, we present the methodology for a systematic semic analysis. Section 5 presents an experimental study on Word Embeddings and semic analysis. Finally, we give our conclusions and proposals on possible developments of the current study in Section 6.

2. Background

Generally, the ascription of the properties of unambiguousness, referential precision and monosemy to medical terminology would be desirable. As L'Homme ([21]) affirms, polysemy and synonymy should not ideally occur. It should be, however, considered that, as Gotti underlines ([13]), monoreferentiality is to be considered as restricted to the subject area wherein the term is featured. Notwithstanding this, it can be observed that the idealistic condition of monoreferentiality as restricted to a single domain is partially unfulfilled in the practical usage of medical terminology. For example, the term *cervical* is related to different concepts, as it refers both to the neck and the uterine cervix. Moreover, as Kuzio observes ([17]), polysemy can be featured at the level of both abbreviations and acronyms.

Medical terminology is also peculiarly characterised by the inclusion of both specialised and popular terms. The categorisation that distinguishes specialised terms from popular ones partially corresponds to the distinction between terms that are uniquely characterised by denotation and terms which also integrate connotative traits. Denotation is the essential feature that realises the referential function of terms: in the circumstances in which it is individually employed, the term designates one distinct referent ([3]). Connotation, instead, expresses a subjective qualitative conceptualisation. Moreover, connotative considerations may be influenced by personal former experiences and are individual emotional perceptions that are associated to words ([40]). Particularly, the exploration of connotative traits ascribed to medical terminology could constitute a resource for physicians. Indeed, it could valuably improve physician-patient communication and mutual comprehension.

The study of denotation and connotation of medical terminology can be realised by way of resorting to the linguistic technique of semic analysis. The accurate identification and representation of the concepts that are represented by medical terms can be achieved by way of performing it. Semic analysis is devised in the domain of semantics ([14]). Specifically, it captures the essential particularities of meaning through its methodical factorisation, with the aim of retrieving a collection of minimal traits. As a matter of fact, the systematic fragmentation of terms

meaning into semes, that is to say minimal traits of meaning ([38]), enables to comprehensively express meaning components. By way of exemplification, the concept of the English term *biopsy* can be represented by the following semic analysis ([42]):

/examination/ /tissue removed/ /from a living body/ /discover/ /presence-cause-extent/ /disease/

where each */seme/* is surrounded by the slash "/" character.

The assumption at the root of semic analysis is that meaning should provide distinctiveness to terms. As a matter of fact, meaning derives from oppositions ([43]), which is also required in semic analysis. A reference theory with regard to semic analysis is the interpretative semantics theory, formulated by François Rastier ([34]). Its main topic of study is the textual dimension. Semic analysis is considered as a technique which fosters both the cognitive and linguistic comprehension of linguistic units and the acquisition of semantic knowledge because of its systematic approach. Moreover, it can be efficiently applied with respect to terminology. The assumption of the possibility of performing it with respect to terminology could be derived from Rastier's interpretation of the "concept of concept". A concept is a constructed sememe, whose definition is determined by the norms of a discipline, in such a way its occurrences are identical to its type ([35]). The conventional validity of these disciplinary norms enables for the translation of concepts that consequently elude the variety of languages as well as the difference of contexts ([35]).

The study of semic analysis of medical terminology can also be carried out in the context of Word embeddings, with a view to exploring the possibility to semi-automatically or automatically perform it. Word embeddings are widely employed in the framework of Natural Language Processing and can be defined as the representation of words ([22];[26]) in the form of numerical vectors which incorporate information related both to syntax and semantics as a vector space ([25];[26];[39]). On the basis of the notional principles of word embeddings, the context in which a word is embedded determines the meaning of the linguistic entity itself ([47]). However, it is relevant to mention that, as Lenci states, it is the distributional hypothesis that realises the analogy between the similarity of distributions and the similarity at the meaning level ([18]). Former studies on the interconnection between semic analysis and word embeddings were carried out in the context of the investigation of sememe prediction, as proposed by Xie et al. ([48]). The same issue was examined from a cross-linguistic viewpoint, such as in the study presented by Qi et al. ([31]).

3. A preliminary experiment on medical terminological records

With respect to semic analysis, Pottier claimed that "[w]hat is surprising is the arbitrariness of the choice of semes as compared to the perceptible world" ([29]). Taking as a reference his assertion, in this section we aim to: 1) ascertain that semic analysis is not an objective technique

215

on the basis of a preliminary experiment, and 2) propose a methodology which could increase systematicity in its performance.

Specifically, we perform a preliminary experiment to study users' perception and approach to semic analysis in the context of a study which concerned the compilation of the TriMED model of terminological records ([44]). The terminological records were compiled by two groups of students of the Modern Languages for the International Communication and Cooperation Master's Degree Course of the University of Padua. The analysed datasets concern the academic years 2018/2019 and 2019/2020. The compilation of these records was carried out as a task which was proposed in the context of the Computer-Assisted Translation Tools course. For the objectives of this experiment, the examined subsections of the record concern the analysed term, the language and all information related to the semantic domain. In particular, the focus is placed on the selected definition of the term from a terminological perspective and, primarily, on the semic analysis performed by the students. The articulation of the investigation process contemplates two different approaches:

- Primarily, an intralinguistic approach divided into different phases is adopted. Firstly, we detect some cognitive operations which can be deemed as being involved in the selection of semes. Secondly, we analyse the lexical form of semes focusing on the morphological level. Lastly, we check the correspondence between the semes and the concept represented by the term.
- Contextually, an interlinguistic approach is adopted with respect to three among the languages in which the records are compiled, namely English, Italian and Spanish. The examined semic analysis respectively amount to 458 for the English language, 343 for Italian and 101 for Spanish. The interlinguistic follow-up study is conceived with a view to identifying similar or different phenomena in the performance of semic analysis when considering three different languages.

In detail, students were asked to compile TriMED terminological records of medical terms and, contextually, to fulfil the included section related to semic analysis. They were asked to rely on definitions to identify semes and to express them in the form of lemmas. Moreover, students were asked to adopt the same methodology with respect to all the three languages in which terminological records could be compiled.

At a first stage, the analysis concerns the semes associated to English terms. The perception of the absence of a uniformed criterion in the selection of semes primarily emerges in the circumstance in which, for the same term, different sememes can be individuated. Some correspondences between sememes can be detected because of the presence of shared semes. Such occurrence suggests the hypothesised happenchance that some conceptual units are generally shared in students' perception. Notwithstanding the fact that a term inscribed in the framework of special languages should be widely agreed upon in its conceptual content, we observed that: i) different semantic elements are used to designate a term, ii) different semes are indicated in order to indicate a single conceptual entity. For example, the term *screening* which refers to an examination concerning a group of people, is expressed by the following semes:

/people/, /population/, /group/ and /large number of people/.

A subsequent comparison between the definition of a determined term selected by students and the formulated semic analysis is performed. The comparison results in the ascertainment of the subsistence of a close relation that is established in terms of shared terminology. In these circumstances, it can be inferred that a manual term extraction was carried out by way of extracting terminology from definitions. A potential criterion that can be spotted is the application of the notion of termhood ([16]), which was applied based on the perception of the pertinence of the term in the context of the medical domain. However, the establishment of a total correspondence between the terms contemplated by the definition and the semes included in the sememe is not realised in the totality of the circumstances. Consequently, the implementation of a different criterion in the semic retrieval process is highlighted. As a matter of fact, in numerous cases it is possible to detect the inclusion of semes which present a different linguistic content with respect to the selected definition, hence conflating two approaches to semic analysis. In other cases, neither a total reliance on a definition nor a partial one is adopted as selection criteria. From a morphological viewpoint, the most representative categories are constituted by nouns and verbs. However, even the employment of determiners, prepositions and coordinating conjunctions can be occasionally observed. Acronyms and abbreviations are scarcely considered as semic elements. This circumstance can be potentially attributed to the fact that the selected definitions scarcely include these entities or that a less informative potential can be perceived. Additionally, the application of the notion of transposition ([45]) can be mentioned as a further criterion. This notion derives from translation studies and particularly constitutes a translation technique. Indeed, morphological transformations involving the terminology included in definitions can be observed, the majority of which primarily regard the inflections of nouns and verbs. It can be confirmed, however, that in the context of the performed semic analysis base forms constitute the main carriers of concepts.

Comprehensively, taking as a reference the 2019/2020 dataset, in a total of 198 formulations of semic analysis compiled in English terminological records, 161 of them manifest a close connection with the selected definitions of the terms. Specifically, we consider that a close connection corresponds to a range which oscillates between a total terminological correspondence to a proximity. The proximity is envisaged varying from the majority to a minimum of half of the terms extracted from definitions. With reference to the 2018/2019 dataset, in a total of 260 semic analysis, 226 present the same interconnection. For what concerns the mentioned interlinguistic approach, a shared pattern in terms of adopted strategies can also be observed in semic analysis with respect to Italian and Spanish. In particular, the reference is to the realisation of a terminological extraction from definitions to transpose terminology into semic elements and to morphological modifications. Considering the semic analysis of Italian terms included in the 2018/2019 dataset, 286 include terminology from definitions. Spanish semic analysis comprised in the 2019/2020 dataset feature the same criterion in 78 cases. For what concerns Italian semic analysis, an interesting occurrence is represented by the fact that on some occasions, identical sememes are employed to represent the concept of different terms. For example, a sememe uniquely composed by the seme */malattia contagiosa/*, in English */contagious*

disease/, is used to represent the concepts of *morbillo* and *rosolia*, known in English as *measles* and *rubella*. From the observation of these data, it can be confirmed that, in most cases, students employed definitions as sources of terminological and conceptual knowledge to perform semic analysis in accordance with the guidelines. Moreover, it could be assumed that connotation is apparently not contemplated in semic formulations. We specifically refer to connotation as the subjective perception that is linguistically expressed in semic analysis by way of afferent semes. Therefore, we adhere to the distinction between semes that express denotation and connotation. In relation to this, it could be hypothesised that semes that express connotation could have been regarded as less relevant in the framework of a domain-specific terminology with respect to its referential purpose and less distinctive. In this light, it could be supposed that terms were principally considered from a referential perspective resorting to denotation, excluding individual conceptualisations. A further reason for the exclusion of connotative semes from sememes could be attributed to a heavy reliance on the considered definitions. In fact, both the presence and the absence of connotative semes could potentially be directly proportional to their manifestation in definitions. Moreover, the potential consideration of terms in a specialised context could have exerted an influence on this issue. Indeed, it can be observed that in specialised situations medical terms are not considered from a subjective perspective, hence relying on denotation. The lack of connotative semes could be, therefore, hypothetically related to the fact that the employment of medical terminology in everyday language, in which terms acquire connotation ([3]), was not considered.

Evidence from the analysed records highlights the subsistence of a subjectively biased approach in semic analysis' performance. As a matter of fact, the individuation of a partially shared cognitive mode of operation in the semes' selection is indicative of variability. This circumstance corroborates the impression of the discrepancy that can be observed in its formulations. Indeed, the criterion which emerges as the most considered to perform semic analysis can be itself regarded as a cause of dissimilarities in terms of expression of the conceptual dimension of terminology. In fact, different definitions were respectively selected by different students for the same term, consequentially leading to the integration of different semic elements. A procedural gap is therefore contextually detected, determining the necessity of a systematisation in the realisation of semic analysis.

4. Proposal of a methodology for a systematic semic analysis

Based on this preliminary study, we propose a methodology to increase the objectivity and systematicity of semic analysis in the context of medical terminology. Our objective is to address the semantic discrepancies with reference to definition-based semic analysis. Moreover, we aim to attempt to moderate subjectively biased manifestations in the expression of terminological and conceptual knowledge.

We argue that the focus of semic analysis should not be uniquely placed on the linguistic perspective but equally on the cognitive operations that are performed in the semes' selection

phase. Specifically, for the performance of an objective and systematic semic analysis, we propose four sequential cognitive procedural steps.

1. The primary phase involves the achievement of an in-depth comprehension of the concept. The concept should be considered as a single unit that, at this stage, is not factorised.

2. Secondly, the cognitively elaborated unitary concept is supposed to undergo a further mental elaboration, as the concept should be fragmented into minimal significant constituents. The segmented configuration and the preconceived unitary conceptualisation should manifest a total conceptual correspondence. The selection of semes could be equated with a cognitive retrieval process, as the recall capability should result in an exhaustive collection of semic units.

3. Thirdly, the conceptual elaboration should be transposed in form of a lexical and metalinguistic output, following the formal graphic articulation of semic analysis.

4. Lastly, a conceptual validation should be contemplated. It is substantially conceived as a screening activity aimed at ensuring the conceptual correspondence between the preconceived unitary concept and the originated sememe. Furthermore, the aim is to ensure the referential exactness with respect to the considered terminological entry. This step can be regarded as a follow-up operation which is excluded from the proper performance of semic analysis. As a matter of fact, the main objectives it realises are the verification of its functional effectiveness and, contextually, the ascertainment of the in-depth assimilation of the concept related to the term also with respect to the wider framework of the conceptual map of the domain.

The proposed sequential and complementary processes of conceptual construction and decomposition of the semantic content associated to terms provide for the acquisition of an in-depth comprehension and investigation of concepts and terminology.

Nevertheless, the identification of the exact componential structure of multi-word terms could constitute a challenging operation due to different factors. For example, the lexical structure itself, as terminological particles which are included in the componential structure of these complex entities could be mistakenly regarded as juxtaposed adjectives. It is the case of *superior mesenteric vein*. Moreover, the presence of coordinating conjunctions such as in *basal nuclei and thalamic region* as well as graphic signs as in *positron emission tomography/computed tomography* could elicit the perception of a reference to different terms, hence failing to determine the unitary concept. The question could also be related to the postulated frequent association between words and the concept that is inherently involved in the process of tokenisation. In this framework, we refer to tokenisation to indicate words seen in segmented languages as single orthographic units generally isolated by spaces. Both reasons might be conflated into a main assumption, which is the supposed tendency to being accustomed to focus to a greater extent on the individual elements of a multi-word term and the individual concept they represent in the given succession and less on the unitary concept of the multi-word entity in its compositional entirety. This reasoning hinges on the comparison between two different levels of abstraction which

respectively differentiate the reading of single-word terms from that of multi-word ones and the interpretation of a single concept from that of compositional concepts.

4.1 Domain-specific corpus-based semic analysis

A domain-oriented corpus-based data-driven semic analysis of terminology is proposed as the first criterion for the realisation of an objective and systematic semic analysis. Specifically, the recollection of different definitions for the considered terms to select and identify semes hence constituting a specialised corpus of definitions would precede the applicative phase of semic analysis, which is performed by way of a terminological extraction. This methodological approach was formerly adopted by Elezi ([9]) in an investigation on political and economic terminology. The exploration of definitions as a methodological approach which would contribute identifying the semes linked to determined emotions was adopted by Baider and Constantinou ([1]). In the framework of this study, however, the domain-oriented perspective in the context of terminology is not the only standpoint. As a matter of fact, this methodology is principally targeted as the baseline for the realisation of a systematic recollection of all the semic elements related to a term. The underlying conception is also represented by the consideration of this methodology as a valuable source of knowledge and a strategy to develop greater awareness of the conceptual dimension of terms. Therefore, an accurate usage of terminology would be fostered. A qualitative perspective is adopted with respect to the corpus, rather than a quantitative one. This is because the comprehensive retrieval of all semic units is advocated as an essential function of the adoption of this methodology. Subjectively biased manifestation would not be entirely excluded because a selection concerning the terminological, lexical, and conceptual extraction would still be performed. Notwithstanding this, the data-driven mode of operation would increase the objectivity and systematicity of semic analysis due to the adoption of a criterion for the selection which would expand knowledge.

As a matter of fact, an opportunity which could be derived from the application of this methodology is the possibility to transcend personal knowledge, whose specificity could vary among individuals. In this light, the potentiality for lay people to include terminological units as semic elements hence conveying fine-grained conceptual particularities related to specialised terms which are inherently characterised by advanced conceptualisations could be amplified. Moreover, semic analysis itself could be enhanced in its quality, representing a tool which would provide an improved comprehension of the conceptual contents which are embedded in a specialised domain. Moreover, semic analysis can also be conceived as a learning strategy. Indeed, it would subsequently bridge the gap from a cognitional viewpoint between the previous conceptual knowledge and the advanced conceptual elaboration engrained in technical terms.

The concept of relying on different domain-oriented definitions, hence constituting a specialised corpus of definitions, derives from the following consideration: the reliance on a single definition would fixedly relate the conceptual knowledge represented by semic elements to the circumscribed conceptual or semantic content included in the determined definition. For instance, supposing that three different people would respectively rely on three different definitions and a term extraction would be performed, the resulting extraction could potentially

be dissimilar. The respective exploration of all three definitions performed by a single person, instead, can possibly result in a more accurate extraction.

A further aspect to be considered is that this approach could lead to a more exhaustive and accurate representation of the conceptual content of terminology, including in this respect a variety of conceptual elements. The reasoning further concerns the selection of a suitable type of definition. As Roche stated, while descriptive characteristics are linked to descriptions, essential characteristics are related to definitions ([36]) and are advocated as indispensable for the definition of concepts as well as for the differentiation between different concepts ([36]). From the consideration of these statements, it derives that definitions are further confirmed as fundamental references to convey conceptualisations. In particular, intensional definitions can be proposed and addressed as the appropriate typology of reference lexicographic sources. According to Roche ([36]), "[*i*]*ntensional definition* […] comprises the *superordinate concept* immediately above followed by one or several *delimiting characteristics*". The inclusion of hypernyms can be regarded as valuable to instantiate the terminological unit in the wider conceptual dimension, in this sense hierarchically gradually progressing from the superordinate conceptual elements which are meant to contextualise from a conceptual viewpoint the semes related to the specific term. In detail, in the circumstances in which the proposed methodology is exemplified with reference to the English language, excluding the signalled exceptions, the consulted sources for semes' extraction are the Merriam-Webster Medical Dictionary ([24]) and the TheFreeDictionary's Medical Dictionary ([41]). For what concerns Spanish language, valuable sources that can be indicated are the Diccionario médico from the Clínica Universidad de Navarra (CUN) ([5]) and the Diccionario de cáncer del NCI – Instituto Nacional del Cáncer ([4]). With reference to the Italian language, reliable dictionaries are the Enciclopedia Salute from the Ministero della Salute ([10]) and the Dizionario di Medicina Treccani ([6]).

It should also be considered that the key objective of semic analysis is the accurate and unequivocal determination of the concepts that are represented by terms and, therefore, the adoption of a qualitative perspective. The quantitative minimisation of semic elements is not accounted as an aim contemplated by this technique. As asserted by Rastier ([32]) economy in the number of semes is not ascribed to semic analysis, nonetheless it can be performed and it represents a valuable technique, whose potential is increased when contextually applied.

Moreover, the achievement of a conceptual and terminological systematisation could be fostered by way of contemplating the inclusion of mesogeneric semes. Mesogeneric semes are generic semes that correspond to the domains, that is to say to the "spheres of human activity" ([14]). Their individuation can be considered as strategically fundamental with a view to reducing the occurrence of polysemy. In this sense, semic analysis emerges as a valuable linguistic technique that possesses an expressive quality up to the extent that ambiguity could be significantly or partially minimised. For this reason, provided the fact that the reference terminological domain is medical terminology, the mesogeneric seme /medicine/ is selected.

A strategy that can be proposed to heighten the potentiality offered by semic analysis to distinctively designate a term from a differential viewpoint is the realisation of a parallel terminological analysis. The parallel terminological analysis would be carried out with respect to

221

other terminological elements which could be characterised by a proximity as for their conceptualisations in the interconnected conceptual map of the domain. By way of exemplification, the particularities which characterise the conceptual content represented by the term *rosaceiform dermatitis* could be comprehended by way of recurring to the contemplation of other types of *dermatitis* as well as the concept of *dermatitis*.

4.2 Termhood to address the diastratic and diaphasic dimensions

A second criterion meant to improve the methodology is the application of the property of termhood ([16]) combined with the consideration of the diastratic and the diaphasic dimensions of language. This criterion is oriented to the contemplation of the communicational perspective. Indeed, the aforementioned principles would not properly differentiate specialised terms from popular ones. According to Rastier ([33]), by way of exemplification, when analysing medical discourse, it would seem necessary to define 'patient' and 'sick person', to prevent a synonymic relationship between these sememes. As a matter of fact, they are not found in the same contexts: the first occurs in the words employed by physicians towards their assisted patients; the second, in the words that physicians exchange with each other ([33]). The interconnection of the principles of termhood, diastratic and diaphasic dimensions would improve the specificity which could characterise the representation of terminology from a conceptual viewpoint. Indeed, the domain-contextual knowledge would be captured fostering the adequate achievement of a distinction between medical technical and non-technical terms. In other words, semic analysis of specialised terms would feature specialised terminology while popular terms would be represented by way of adopting non-specialised terms. Consequently, the diverging levels of specificity that can be observed in their contextual usage would be transmitted. Indeed, "popular terms" can be devised as entities that express concepts which stem from the realisation of a semantic adaptation strategy for pragmatic purposes, hence differentiating them from specialised terms. This scheme, aimed at fostering mutual comprehension and simplification, can be equated with a communication-oriented "translation" process, specifically moving from the specialised source concept to the target one.

As it can be observed, the presented methodology for an objective semic analysis focuses on the context-independent inherent semes. Therefore, we delineate a research perspective which could focus on the expression of connotation in semic analysis and its investigation. Moreover, it could focus on the potentiality to objectively capture semantic information conveyed by connotative traits. The cognitive elaboration of the particularities of concepts acquired through semic analysis could prove to be fundamental in the context of specialised translation. Specifically, the former comparison between sememes of term candidates to individuate conceptual variations could constitute the baseline for the selection of the appropriate term for a target text, resulting in a quality translation. Furthermore, semic analysis could also prove to be useful when comparatively applied in the framework of an interlinguistic investigation. As a matter of fact, conceptual discrepancies or similarities could be detected in the case in which an identical concept is supposed to be represented by terms in different languages.

5. Word Embeddings and Semic Analysis: An Experimental Study

In this section, we perform an experimental study to explore the potential subsistence of an interrelation between word embeddings and semic analysis. Specifically, we investigate on the capability of word embeddings to retrieve semic elements in the context of medical terminology hence from a determined terminological perspective. Moreover, we challenge the possibility to perform a (semi-)automatic semic analysis of medical terms as theorised in the methodology. Furthermore, this experiment is conceived as an opportunity to illustrate the implementation of the mentioned criteria to the practical performance of semic analysis.

5.1 Experimental setting

The present experimental study features a peculiar dual analysis. Specifically, the dataset presents two sources of data:

1. the semic analysis of five medical terms – *screening*, *measles*, *asthma*, *malformation* and *dermatitis* – and
2. a set of lists including 50 lexical units each deriving from the application of the technique of word embeddings with relation to the same terms.

We chose these medical terms to compare broader concepts (screening and malformation) with narrower concepts (measles, asthma, dermatitis).

The two reference models for word embeddings are Word2Vec, proposed by Mikolov et al. ([26]), and GloVe, developed by Pennington, Socher and Manning ([27]). We used the pre-trained word embeddings glove-wiki-gigaword-300, for GloVe, based on Wikipedia ([28]), and two Word2Vec models pre-trained on the Google News dataset ([12]) and PubMed ([30]). With reference to the term *dermatitis*, Word2Vec is the only considered model. We used the Gensim open-source Python library for the pretrained models.[1,2]

5.2 A comparative analysis of sememes and word embeddings

The first approach to the investigation features a comparative perspective. The focal issue is individuated in the questioning of the existence of a terminological and conceptual equivalence between the semic units composing the sememe of reference terms and the related lexical units retrieved by way of word embeddings. Contextually, the comparison is also conceived as a means to challenge the recall capability of semic analysis with respect to the computational data. Moreover, from an overturned perspective, it is conceived as a way to assess the potential that word embeddings might possess to capture semic units. The 50 semantically proximal lexical

---

[1]   https://radimrehurek.com/gensim/

[2]   https://github.com/RaRe-Technologies/gensim-data

units are specifically ranked according to correlated numerical values representing vector's proximity computed as a cosine distance among vectors.[3]

An exemplification of the procedural steps involved in this approach can be proposed by taking as a reference the term *screening*. Let us consider the following semic analysis manually performed by a linguistic expert:

/medicine/ /examination/ /exam/ /test/ /testing/ /group/ /individuals/ /population/ /people/ /asymptomatic/ /detect/ /identify/ /likelihood/ /probability/ /disease/ /condition/ /diagnostic test/ /process/ /organ/ /tissue/ /prevention/

The semic elements are subsequently compared with the lexical units which are retrieved by way of the application of word embeddings. We list these units in the Table 1 in the Appendix. Specifically, the following semes feature as terms in the list obtained through the application of the GloVe model:

/testing/, /examination/, /exam/, /test/, and /detect/.

Moreover, the inflected forms of some semes can be detected, as in the case of the terms *tests* and *examinations*. With respect to the list stemming from the training of Word2Vec on PubMed, the seme */testing/* features as a term in the list. Furthermore, also in this circumstance the presence of inflected forms of semes can be observed with specific reference to the terms *exams*, *detecting* and *identifying*. An absence of correspondence between the semic elements and the entities included in the list obtained by way of Word2Vec trained on the Google News dataset can be observed.

We identified some common phenomena which can be detected in this comparative analysis:

- the presence in the lists of semantically-proximal terminology of inflected forms of lexical elements representing semes of reference terms;
- the absence of correspondence between the semes comprised in a determined sememe and the units included in a list generated from the application of a model;
- in two circumstances, the elements included in the lists are multi-word terms whose constituents can be reconnected at the lexical level to the terminological aggregation of terms constituting semes.

An example can be mentioned in relation to the reference term *asthma*, as two of its semes */airway/* and */constriction/* feature in the list obtained from the application of Word2Vec trained on the Google News dataset as the multi-word term *airway constriction*. In several cases, a single lexical unit comprised in the semic analysis constitutes a lexical unit of a semantically-related multi-word term. For example, the seme */defect/* included in the semic analysis of the reference term *malformation* features as a constituent of multi-word terms in the lists such as in *ventricular*

---

[3]   https://link.springer.com/referenceworkentry/10.1007%2F978-1-4614-6170-8_141

septal defect, *birth defect* and *neural tube defect*. In addition, morphological differences between semic elements and the terms comprised in the lists can be in some instances detected. It is the case of the seme */abnormal/* which is employed in the semic analysis of *malformation* and the listed terms *abnormality* and *abnormalities*.

Comprehensively, two considerations can be potentially deducted from the observation of these results. Firstly, it can be validated that semic analysis represents a technique which manifests a conceptual retrieval capability which also manages to comprise elements regarded as contextually semantically-proximal. Therefore, a potential interrelation between semic analysis and word embeddings is established. On the other hand, word embeddings manage in some circumstances to capture some lexical units which represent semic elements in the sememes of reference terms. In this connection, the possibility to selectively consider lexical units deriving from the application of word embeddings as semic elements can be conceived, with a view to further improving the performance of semic analysis. Specifically, the elements characterised by a morphological relationship or related in terms of inflections with respect to the semes included in the sememes can be estimated as suitable for the conceptual representation because of the subsistence of a high semantic proximity.

5.3 Semic analysis of word embeddings

The second step consists in the consideration of a subset of the most semantically-proximal entities with respect to the reference terms, amounting to 10 consecutive terms for each list. This approach is specifically conceived to further challenge and explore the subsistence of an actual conceptual connection with respect to the reference terms. A systematic semic analysis of these units based on the proposed methodology is performed, with a view to investigating this circumstance. Moreover, the aim is to identify the concepts represented by these lexical units to evaluate the possibility for these units to constitute semes of the respective reference terms.

For instance, taking as a reference the term *screening*, a semic analysis of the terms with respect to which a relation of semantic proximity is individuated in the context of word embeddings is carried out. Terminology which belongs to the medical domain is exclusively considered. For exemplification purposes, we propose the semic analysis of three terms which are retrieved by way of the application of word embeddings:

- screened:

/medicine/ /separate/ /undiagnosed/ /disease/ /defect/ /pathologic/ /condition/ /risk/ /tests/ /examinations/ /procedures/ /examine/ /evaluate/ /infection/ /test/ /population/

- mammography:

/medicine/ /x-ray/ /examination/ /breasts/ /detection/ /cancer/ /study/ /test/ /mammogram/ /screening/ /breast/ /diagnostic/ /evaluation/ /abnormalities/ /patients/ /abnormality/ /follow-up/ /breast cancer/ /lumpectomy/ /radiological/ /screen/ /evaluate/ /tumors/ /abnormalities/ /procedure/ /imaging/ /diagnosis/

- diagnostic:

225

/medicine/ /diagnosis/ /identify/ /disease/ /medical/ /symptom/ /technique/ /instrument/ /signs/ /symptoms/ /methods/ /act/

As it emerges from the analysis of the semic formulation of the term *mammography*, the reference term *screening* features among its semic elements. This highlights the subsistence of a conceptual relation between the two terms. Nonetheless, the evaluation of a potential compatibility in the specific form of a conceptual integration is guided by the adoption of a conception of concepts as determined by the comprehensive consideration of all the semic elements included in the sememes. Moreover, provided the selection of the intensional definition for the determination of concepts, hyponyms are consequentially excluded as potential semes. As Löbner ([23]) affirmed,

> an expression A is a hyponym of an expression B iff the meaning of B is part of the meaning of A and A is subordinated of B. In addition to the meaning of B, the meaning of A must contain further specifications, rendering the meaning of A, the hyponym, more specific than the meaning of B.

Provided this standpoint, the hypernym concurs to compose the concept of the hyponym. However, the hyponym includes other conceptual specificities, therefore determining an asymmetry and the subsistence of a related but diverse concept.

Considering this line of reasoning, the terms *screened* and *diagnostic* are regarded as suitable in order to compose the sememe of the reference term *screening*. The term *mammography* is consequentially excluded from the collection of semic elements which represent the conceptual content of the reference term. Furthermore, with reference to the term *screened*, a morphological relationship with respect to the term *screening* can be observed. In this circumstance, the interrelation that is established between the two terms does not uniquely concern a conceptual connection. This signals the subsistence of a synergy which manifests at the linguistic level between semic analysis and word embeddings specifically from a morphological perspective.

The targeted aim of this approach is also identified in the investigation of the possibility to retrieve a superset of semes. The superset of semes is conceived as a collection of semic units which would co-occur in the sememes of both the reference terms and the semantically-proximal terminology, determining a partial recall of conceptual particularities. By way of considering this perspective, the recall capability of word embeddings with respect to the conceptual elements embedded in the concept of reference terms is challenged. Contextually, the potentiality to capture semic elements in the context of semic analysis and the possibility to automatically perform it with respect to medical terminology in the context of the proposed methodology are questioned.

Some considerations can also be proposed as for this different approach to the investigation of the topic:

- Firstly, the subsistence of lexical units which are not related to the medical domain can be detected in the context of the application of the GloVe model. Therefore, this indicates the potential occurrence of noise.

- Secondly, the semic analysis of the semantically-proximal terms respectively feature reference terms as semic elements only in a few circumstances.
- Thirdly, many semantically-proximal terms only manifest a partial semantic interrelation with respect to the reference terms which can be individuated in a shared semantic field.

Provided the limited but not sufficient relatedness of some proximal terms with respect to reference ones in terms of shared semes, a potential outcome can be constituted by the ascertainment of the fact that: 1) semantically-related terms should not be comprised in the sememes of reference terms, 2) the impossibility to detect a superset of semes with the exception of the mesogeneric seme. Consequently, an automatic performance of semic analysis would not enable the retrieval of all the semes which are conversely detected in the context of the application of the proposed methodological criteria. Indeed, many of the semic particles of reference terms cannot be captured by way of the application of word embeddings. Nevertheless, a relevant finding is constituted by the identification of a connection between the two techniques. This connection can also be productively applied in the context of the proposed methodology to improve it in its conceptual recall capability and its lexical exhaustiveness. The application would specifically consist in the introduction in the sememe of additional conceptually related semic elements which would contribute to comprehensively represent the conceptual dimension of terminology. As a matter of fact, determined terms can be potentially deemed as adequate to represent the concepts of the respective reference terms hence constituting semes of the latter. This is due to the establishment of an interrelation realised by way of a conceptual connection and a morphological relationship also in terms of inflections between the reference terms, their semes and the listed lexical entities.

6. Conclusions

In this paper, we proposed a methodology for an objective and systematic semic analysis starting from a preliminary experiment on medical terminological records aimed at the study of users' perception and approach to semic analysis. We focused on the identification of discrepancies both in terms of mode of operations and generated outputs with respect to the performance of semic analysis. On the basis of the observations of the first analysis, we proposed a methodology aimed at reducing the manifestation of subjectively-biased phenomena in the performance of semic analysis. In the context of the proposed methodological criteria for an objective semic analysis, we investigated on the subsistence of a potential interrelation of Word Embeddings with respect to semic analysis of medical terminology. The experimental study led to the successful identification of a connection at the linguistic level between the two techniques. This connection was also envisaged in the framework of the presented methodology as a valuable integration with a view to improving its productivity and conceptual recall capability.

The study also demonstrated that semic analysis as performed in the framework of the theorised methodology effectively manages to capture lexical units which are conceived as semantically-

proximal to reference medical terms. By adopting different approaches to the study of the topic, however, the capability of Word Embeddings to capture semantic information was called into question. The experimental study further provides the foundations to call into question the capability of word embeddings to capture the semantic dimension. As a matter of fact, with respect to medical terminology in the framework of the present investigation, it can be considered as limited. Indeed, the interrelation between word embeddings and semic analysis as performed in the context of the proposed methodology only concerns the morphosyntactic and syntactic dimension. Furthermore, some listed lexical entities do not present a semantic relation with respect to the reference ones, signalling the strong influential faculty exerted by contextual elements on the analysed word embeddings. An additional observation that can be proposed concerns the consideration of the fact that, as semes can be regarded as the fundamental intrinsic components of concepts, the limited semantic similarity which can be observed further calls into question the actual subsistence of a semantic dimension that could be ascribed to word embeddings. Moreover, it should be considered that medical terminology presents a high conceptual specificity which should be accordingly reflected with respect to word embeddings as for the semantic relatedness and proximity. Further investigation on a more extended dataset would however be required to examine the issue. For the aforementioned reasons, we are actually planning a further experimental analysis with a wider set of terms in order to study an interactive threshold on the value of similarity between the term and the embeddings in order to help the expert in the choice of semes during the semic analysis.

As future work, we propose to study a methodology to capture the semantic information expressed by way of connotative semes in the context of the performance of semic analysis. In particular, taking into consideration users' native language, in order to observe if variations in its performance occur according to whether the user is performing it in his/her native language or in a foreign language. We also propose to study the extent to which the proposed methodology improves the semic analysis of terms compared to the results obtained in the preliminary experiment conducted by the students.

Appendix

Glove (Wikipedia)	Word2Vec (Google News)	Word2Vec (PubMed)
screenings	Screening	screenings
screened	screenings	testing
mammography	view Puyol leaped	screen
procedures	screened	diagnostic
testing	Screenings	mammography
checks	prescreening	routine

tests	rescreening	screens
diagnostic	Newborn Hearing	colposcopy
detection	Backtesting portfolio	surveillance
baggage	Puyol leaped	evaluation
routine	screeing	diagnostics
detectors	Doc Soup	assessment
examination	General Cholesterol Glucose	counseling
rigorous	screeners	retesting
diagnosis	Flexi Scope	chlamydia
examinations	RICHFIELD SPRINGS Breast	counselling
undergo	abusers Deron	work-up
evaluation	rectal exams	unscreened
mammogram	screener	point-of-care
screeners	Universal Pictures Bruno	workup
scans	multiphasic blood	identifying
check	design ChemAxon focuses	identification
tsa	PAP smears	detecting
checked	fingerprinting	triage
stringent	detect precancerous growths	colonoscopy
applicants	mammograms pap tests	diagnosis
surveillance	prostate exam	opt-out
procedure	test FOBT	clinic-based
prenatal	colorectal cancer screening	mammogram
screen	Tennessee Outlive	prioritization
mammograms	digital rectal exams	screened
treatment	Hearing Screenings	diagnosing
ultrasound	optical colonoscopy	exams
scanners	USPSTF recommends	confirmation

exam	screening colonoscopy	phenotyping
treatments	Bone density	biennial
guidelines	AFI Film Festival	programme
scanning	PAP smear	programmes
imaging	immunochemical	validation
patients	urine dipstick	high-risk
detecting	colonoscopy	cost-effective
polygraph	colorectal screenings	at-risk
colonoscopy	mammography	checkup
scan	protege Bernard Kerik	check-up
test	mammograms	smear
identification	Nuclear Tipping Point	high-throughput
detect	Dr. Domenico Corrado	evaluations
stricter	mammography screening	detection
monitoring	Samuel A. Bozzette	mammograms
required	AGO Jackman	dipstick

Table 1: Lists of the 50 most semantically-proximal word embeddings to the reference term "screening" with respect to the GloVe model trained on the Wikipedia dataset and the Word2Vec model trained on the Google News and PubMed datasets.

References

[1] Baider, Fabienne, and Maria Constantinou. 2014. "La fureur de gagner, la rage de perdre. Étude contrastive de colere, rage et fureur en français et en grec moderne". *Études romanes de Brno*, 35 (1): 89-104.
https://www.researchgate.net/publication/329511969_La_fureur_de_gagner_la_rage_de_perdre_Etude_contrastive_des_concepts_de_rage_orge_et_de_fureur_lyssa_en_grec_et_en_francais.

[2] Bowman, Julie M., and Marilyn L. Haas. 2010. "When Lung Cancer Consumers Seek Evidence". In *Contemporary Issues in Lung Cancer: A Nursing Perspective*, edited by Marilyn L. Haas, 335-351. 2nd edition. Sudbury, MA: Jones and Bartlett.

[3] Danesi, Marcel. 2016. *Language, Society, and New Media: Sociolinguistics Today*. New York/Abingdon: Routledge.

[4] Diccionario de cáncer del NCI. Instituto Nacional del Cáncer. Accessed March 23, 2021. https://www.cancer.gov/espanol/publicaciones/diccionario.

[5] Diccionario médico. Clínica Universidad de Navarra. 2020. https://www.cun.es/diccionario-medico.

[6] Dizionario di Medicina. Treccani. Accessed March 23, 2021. http://www.treccani.it/enciclopedia/elenco-opere/Dizionario_di_Medicina.

[7] Dury, Pascaline. 2013. « Que montre l'étude de la variation d'une terminologie dans le temps. Quelques pistes de réflexion appliquées au domaine médical ». Debate Terminológico 9 : 2-10.

[8] Džuganová, Božena. 2013. "English medical terminology – different ways of forming medical terms". *JAHR – European Journal of Bioethics*, 4(7): 55-69. https://www.researchgate.net/publication/257622885_English_medical_terminology_-_different_ways_of_forming_medical_terms.

[9] Elezi, Shpëtim. 2018. "Ideological Background of Political and Economic Entries in Explanatory Dictionaries of the Albanian Language". *Journal of Educational and Social Research*, 8(3): 79-86. http://dx.doi.org/10.2478/jesr-2018-0033.

[10] Enciclopedia salute. Ministero della Salute. Accessed March 23, 2021. http://www.salute.gov.it/portale/salute/p1_3.jsp?lingua=italiano&tema=Salute_A_Z.

[11] Gaussier, Eric. 2001. "General Considerations on Bilingual Terminology Extraction". In *Recent Advances in Computational Terminology*, edited by Didier Bourigault, Christian Jacquemin, and Marie-Claude L'Homme, 167-183. Amsterdam/Philadelphia: John Benjamins Publishing Company. https://doi.org/10.1075/nlp.2.09gau.

[12] Google Code Archive. Accessed March 23, 2021. https://code.google.com/archive/p/word2vec/.

[13] Gotti, Maurizio. 2008. *Investigating Specialized Discourse*. 2nd edition. Bern: Peter Lang AG.

[14] Hébert, Louis. 2006. Tools for Text and Image Analysis: An Introduction to Applied Semiotics. *Texto!*. http://www.revue-texto.net/Parutions/Livres-E/Hebert_AS/Hebert_Tools.html.

[15] Introduction to Information Retrieval. Accessed March 23, 2021. https://nlp.stanford.edu/IR-book/information-retrieval-book.html.

[16] Kageura, Kyo, and Bin Umino. 1996. "Methods of Automatic Term Recognition: A Review". *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, 3(2): 259-289. https://doi.org/10.1075/term.3.2.03kag.

[17] Kuzio, Anna. 2019. "Difficulties resulting from language diversity in teaching medical translation and methods to overcome them when teaching medical English to future translators". *Language Value*, 11(1): 23-44. https://doi.org/10.6035/LanguageV.2019.11.3.

[18] Lenci, Alessandro. 2009. "Spazi di parole: metafore e rappresentazioni semantiche". *Paradigmi*, 27: 83-100. doi:10.3280/PARA2009-001007.

[19] León-Araúz, Pilar. 2017. "Term and concept variation in specialised knowledge dynamics". In *Multiple Perspectives on Terminological Variation*, edited by Patrick Drouin, Aline Francœur, John Humbley, and Aurélie Picton, 213-258. Amsterdam & Philadelphia: John Benjamins Publishing Company. https://doi.org/10.1075/tlrp.18.09leo.

[20] Leonardi, Natascia. 2009. "Terminology as a system of knowledge representation: an overview". In *La ricerca nella comunicazione interlinguistica. Modelli teorici e metodologici*, edited by Stefania Cavagnoli, Elena Di Giovanni, and Raffaela Merlini, 37-52. Milano: FrancoAngeli.

[21] L'Homme, Marie-Claude. 2020. *Lexical Semantics for Terminology: An introduction*. Amsterdam/Philadelphia: John Benjamins Publishing Company. https://doi.org/10.1075/tlrp.20.

[22] Li, Yitan, Linli Xu, Fei Tian, Liang Jiang, Xiaowei Zhong, and Enhong Chen. 2015. "Word embedding revisited: a new representation learning and explicit matrix factorization perspective". In *Proceedings of the 24th International Joint Conference on Artificial Intelligence*, edited by Qiang Yang, and Michael Wooldridge, 3650-3656. AAAI Press. https://www.ijcai.org/Proceedings/15/Papers/513.pdf.

[23] Löbner, Sebastian. 2002. *Understanding Semantics*. New York: Oxford University Press.

[24] Merriam-Webster Medical Dictionary. 2021. https://www.merriam-webster.com/medical.

[25] Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg Corrado and Jeffrey Dean. 2013. "Distributed Representations of Words and Phrases and their Compositionality". In *Advances in Neural Information Processing Systems 26*, edited by C.J.C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger, 3111-3119. Curran Associates, Inc. https://proceedings.neurips.cc/paper/2013/file/9aa42b31882ec039965f3c4923ce901b-Paper.pdf.

[26] Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. "*Efficient Estimation of Word Representations in Vector Space*". https://arxiv.org/pdf/1301.3781.pdf.

[27] Pennington, Jeffrey, Richard Socher, and Christopher D. Manning. 2014. "GloVe: Global Vectors for Word Representation". In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, edited by Alessandro Moschitti, Bo Pang, and Walter Daelemans, 1532-1543. Association for Computational Linguistics. http://dx.doi.org/10.3115/v1/D14-1162.

[28] Pennington, Jeffrey, Richard Socher, and Christopher D. Manning. "GloVe: Global Vectors for Word Representation". Accessed March 23, 2021. https://nlp.stanford.edu/projects/glove.

[29] Pottier, Bernard. 1992. "The Components of Communication". In *Ideas, Words, and Things: French Writing in Semiology*, edited by Harjeet Singh Gill, and Bernard Pottier, 113-135. New Delhi: Orient Longman.

[30] PubMed. Accessed March 23, 2021. https://pubmed.ncbi.nlm.nih.gov.

[31] Qi, Fanchao, Yankai Lin, Maosong Sun, Hao Zhu, Ruobing Xie, and Zhiyuan Liu. 2018. "Cross-lingual Lexical Sememe Prediction". In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, edited by Ellen Riloff, David Chiang, Julia Hockenmaier, and Jun'ichi Tsujii, 358-368. Association for Computational Linguistics. http://dx.doi.org/10.18653/v1/D18-1033.

[32] Rastier, François. 1985. "Principes et conditions de la sémantique componentielle". In *Exigences et perspectives de la sémiotique: Recueil d'hommages pour A.J. Greimas. / Aims and Prospects of Semiotics. Essays in honor of A.J. Greimas*, edited by Herman Parret, and Hans-George Ruprecht, 505-527. Amsterdam/Philadelphia: John Benjamins Publishing Company. https://doi.org/10.1075/z.23.43ras.

[33] Rastier, François. June 2005. "La microsémantique". *Texto!* vol. X no.2. http://www.revue-texto.net/Inedits/Rastier/Rastier_Microsemantique.html.

[34] Rastier, François. 2009. *Sémantique Interprétative*. Formes sémiotiques. Paris: Presses Universitaires de France.

[35] Rastier, François. 2010. *Sémantique et recherches cognitives*. Formes sémiotiques. Paris: Presses Universitaires de France.

[36] Roche, Christophe. 2012. "Should Terminology Principles be re-examined?", in *Proceedings of the 10th Terminology and Knowledge Engineering Conference: New frontiers in the constructive symbiosis of terminology and knowledge engineering (TKE 2012)*, edited by Guadalupe Aguado de Cea, Mari Carmen Suárez-Figueroa, Raúl García-Castro, and Elena Montiel-Ponsoda, 17-32. Madrid: Universidad Politécnica de Madrid. https://arxiv.org/ftp/arxiv/papers/1609/1609.05170.pdf.

[37] Santos, Claudia, and Rute Costa. 2015. "Domain specificity". In *Handbook of terminology*, Volume 1, edited by Hendrik J. Kockaert, and Frieda Steurs, 153-179. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/hot.1.dom1.

[38] Schmit, Louis-Marie. 2017. *Les définitions en droit privé*. Thèses de l'IFR. Toulouse: Presses de l'Université Toulouse 1 Capitole, LGDJ – Lextenso Editions. https://doi.org/10.4000/books.putc.2335.

[39] Sfakakis, Michalis, Leonidas Papachristopoulos, Kyriaki Zoutsou, Giannis Tsakonas, and Christos Papatheodorou. 2019. "Automated Subject Indexing of Domain Specific Collections Using Word Embeddings and General Purpose Thesauri". In *Metadata and Semantic Research: 13th International Conference, MTSR 2019, Rome, Italy, October 28-31, 2019, Revised Selected Papers*, edited by Emmanouel Garoufallou, Francesca Fallucchi, and Ernesto William De Luca, 103-114. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-36599-8_9.

[40] Steinberg, Sheila. 2007. *An Introduction to Communication Studies*. Cape Town: Juta & Company Ltd.

[41] TheFreeDictionary's Medical Dictionary. 2003-2021. https://medical-dictionary.thefreedictionary.com/.

[42] TriMED. Accessed March 23, 2021. https://shiny.dei.unipd.it/TriMED/.

[43] Trudel, Éric. 2009. "Éléments de synthèse en sémantique interprétative — Unités thématiques et expressives et approche morphosémantique d'une production sémiotique". *Texto!*, Vol. XIV no. 2. http://www.revue-texto.net/docannexe/file/2284/trudel_synthesesemantique.pdf.

[44] Vezzani, Federica, and Giorgio Maria Di Nunzio. 2020. *"*Methodology for the standardization of terminological resources. Design of TriMED database to support multi-register medical communication*". Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, n. 26(2), 266-298. https://doi.org/10.1075/term.00053.vez.

[45] Vinay, Jean-Paul, and Jean Darbelnet. 1995. *Comparative Stylistics of French and English: A Methodology for Translation*, edited by Juan C. Sager, and Marie-Josée Hamel. Translated by Juan C. Sager, and Marie-Josée Hamel. Amsterdam and Philadelphia: John Benjamins. https://doi.org/10.1075/btl.11.

[46] Warburton, Kara. 2021. *The Corporate Terminologist*. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/tlrp.21.

[47] White, Lyndon, Roberto Togneri, Wei Liu, and Mohammed Bennamoun. 2019. *Neural Representations of Natural Language*. Singapore: Springer Singapore. https://doi.org/10.1007/978-981-13-0062-2.

[48] Xie, Ruobing, Xingchi Yuan, Zhiyuan Liu, and Maosong Sun. 2017. "Lexical Sememe Prediction via Word Embeddings and Matrix Factorization". In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17)*, edited by Carles Sierra, 4200-4206. AAAI Press. https://doi.org/10.24963/ijcai.2017/587.