UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Università degli Studi di Padova

Dipartimento di Studi Linguistici e Letterari

Dipartimento di Ingegneria dell'Informazione

Corso di Laurea Magistrale in
Lingue Moderne per la Comunicazione e la Cooperazione Internazionale
Classe LM-38

Tesi di Laurea

A Proposal of a Methodology for an Objective Semic Analysis: A Case Study on Medical Terminology and Word Embeddings.

Relatore
Prof. Giorgio Maria Di Nunzio

Laureando
Vanessa Bonato
n° matr.1202849 / LMLCC

Anno Accademico 2019 / 2020

TABLE OF CONTENTS

INTRODUCTION

The present dissertation sets out to propose a methodology for an objective semic analysis performed on medical terminology and a subsequent investigation on a potential interrelation in terms of contextual interoperable applicability and productivity of word embeddings with respect to the linguistic approach. In particular, the conceptual dimension which is inherently embedded in the context of terminological investigation will be addressed and different techniques and criteria aimed at its accurate expression and definition will be proposed and challenged.

The structural organisation of the thesis will involve an articulation in four different chapters, with a view to progressively delving into the investigation of the topics. In this light, the first chapter will be devoted to a notional overview on terminology with the specific purpose of introducing the reader to the general linguistic framework in which medical terminology is embedded. In this line of argument, the multifarious conceptualisations implicated in the term *terminology* and its relevant applicative engagement in heterogeneous fields of study will be addressed. Furthermore, an insight on the conception of *terms* will be subsequently presented, with a view to offering a dual approach to the topic. In the first place, different theoretical perspectives on the conceptual nature ascribed to terms will be mentioned, with reference to the respective conceptions of related definitions and considerations in the major theories of terminology. Secondarily, the subject of the lexical representation of terms will be considered and, in this connection, reflections on supposed motivations that could be implicated in the perception of a significant complexity in detecting the exact compositional structure of multi-word terms will be proposed.

The second chapter will specifically focus on medical terminology, whose inscription in the enclosing domain of special languages will be in the first section mentioned. In particular, the subsistence of diversified communicative necessities both at the terminological and conceptual level which distinctively manifest in intra-

specialist situations and doctor-patient contexts as well as the potential peculiar cognitive medical knowledge specificity that could be needed on the part of lay people will be highlighted. In this connection, the differentiation between specialised terms and popular ones will be explored by way of exemplifications with a view to addressing the semantic adaptation strategy which could be entailed in the conceptual content of non-technical terms with respect to specialised ones, aimed at considering the pragmatic dimension. Subsequently, an insight on the main characteristics that comprehensively feature medical terms will be offered, specifically problematising the idealistic purposes of monosemy and monoreferentiality as well as emphasising the inclusion of domain-contextualised terminology in the general inherently ambiguous linguistic frame and in daily life experience. Furthermore, the notion of the ascription of denotative and connotative traits to medical terms will be introduced, stressing in this respect the relevance of the identification of the latter in order to foster an emphatic comprehension and communication in the context of doctor-patient interactions. A reflection which will be proposed in the form of an elaboration of a conceptualisation concerning the key function of transparency and accuracy in the communicative dimension in which medical terminology is involved will conclude the chapter.

In the third chapter the effective relevance of semic analysis as conceived in the framework of François Rastier's interpretative semantics theory as a linguistic technique for the accurate identification of the conceptual content of medical terminology and its consequential pivotal function in the enhancement of an in-depth comprehension of terminological entities will be substantiated. In this light, subsequent to a general introduction to the theoretical perspective that will be adopted, considerations on the pertinency of the adoption of semic analysis as an approach for the in-depth terminological investigation will be presented. In the following subchapters, the specific core elements that are comprised in the notional sphere of semic analysis will be mentioned and the postulated sequential cognitive procedural steps which could be assumed to be involved in its performance will be outlined. The approach to the study of semic analysis will then take the configuration of a case study specifically conducted with respect to the English, Italian and Spanish languages and consisting in an

investigation on two different collections of medicine-related terminological records compiled by students of the Modern Languages for the International Communication and Cooperation Master's Degree Course of the University of Padua in the context of the Computer-Assisted Translation Tools course. The research, conceived as a testbench for the identification of the hypothesised inclination towards a subjectively-biased performance of semic analysis, will interestingly simultaneously highlight a partially-shared pattern in terms of mode of operation, hence withal implying to a certain extent a first level of variability, and confirm the mentioned perception as for the discrepancy which can be spotted in the output. As a matter of fact, the individuated majoritarian-applied procedure will be itself addressed as an additional cause for the divergence in the expression of the conceptual content represented by terminology. Taking these premises as a notional baseline and motivation for an in-depth research, a proposal of an objective methodology of semic analysis will be theorised and presented by way of outlining and describing the target function of each different criteria. Particularly, the adopted approach to the performance of semic analysis will also consider the necessity to conceptually and lexically express the aforementioned diverse levels of specificity which can be respectively dichotomously entailed in the knowledge conveyed through the employment of specialised medical terms and popular ones. In this light, the diastratic and the diaphasic dimensions will be purposefully covered.

The fourth chapter will present an additional investigation concerning semic analysis in the framework of the study of the potential subsistence of an interrelation between the linguistic technique and word embeddings. In particular, the potentiality of word embeddings to retrieve semic elements with respect to medical terminology will be explored with a view to challenging the possibility to perform semic analysis in the framework of the proposed objective methodology through the exploitation of an automatic process. In order to offer an overview on word embeddings, a notional introduction will be presented by way of briefly focusing on the distributional hypothesis and the conceptualisations that are respectively at the root of two main models of word embeddings, namely GloVe and Word2Vec. The investigation will be then carried out in the following subchapters in the context of a case study which will

comprise different approaches to the topic. Specifically, in the first place a comparison between the sememes of the terminology antecedently analysed and the lexical entities retrieved by way of the application of word embeddings will be conducted in order to assess the possibility for these lexical outputs to capture semic units. Successively, the analysis will focus on a limited number of entities which are supposed to represent the most semantically-proximal terms with respect to the reference ones and, in this phase, a systematic semic analysis based on the proposed objective methodology of the former will be carried out with a view to both capturing the conceptual content of these units and, consequentially, evaluating their potential to constitute semic elements of the respective reference terms. Lastly, a further step will be conceived in order to challenge the recall capability of word embeddings with respect to elements included in the sememes. In particular, the purpose of detecting a shared superset of semes between the sememes of the reference terms and these of the related ones will be challenged with the aim of investigating a possible interrelation in terms of semic co-occurrences which could result in a partial recall of the conceptual particularities embedded in the concept of the reference terms. From the consideration of these analysis some observations will be then proposed with regard to the detected interrelation in terms of conceptual integration between semic analysis and word embeddings and, to conclude, the attribution to word embeddings of the capability to capture pure semantic information will be specifically addressed and challenged.

CHAPTER 1 Terminology

1.1 Terminology

Terminology performs a prominent function both in written texts and interactive communication within the framework of specialised knowledge. Delineated in substantive terms, it designates and conveys at the linguistic level the conceptual contents that are constitutive of the conceptual network of a language that pertains to a determined domain or, as ISO 1087-1:2000 defines it, a “set of designations belonging to one special language”. Specifically, however, it does not uniquely denote a collection of linguistic domain-oriented outputs, as the word can also refer respectively to a discipline and an interoperable activity that enable for the realisation of terminographic-related analysis (Brenes 2017: 14).

The relevance of the functional role of terminology is particularly perceptible when considering the several objectives it realises, namely the meticulous description of terms and, besides their representative function at the lexical level, the definition of their relation with respect to the conceptual network of specialised fields of study as well as standardisation and the consequential systematic structuration of domain-related terminological knowledge.

First and foremost, terminology is concerned with the description of terms, which is deeply interrelated to the ascription of terms to concepts. Key to the accomplishment of this activity is therefore the study and cogitation of both entities, with respect to whose an interrelation is meant to be established. Furthermore, particular attention is attributed to the modalities and the origination processes by way of which terms designate a specific conceptual reality of a given discipline in the form of a linguistic representation. As a matter of fact, ISO 1087-1:2000 reckons it as the

> science studying the structure, formation, development, usage and management of terminologies in various subject fields.

As for standardisation, in order to thoroughly gain an insight into the nature and evolution of the principle in relation to terminology, it proves to be relevant to briefly consider it from an historical perspective. In particular, the establishment of terminological distinctiveness by way of a univocal synergy between terms and concepts was especially required for ensuring the unambiguous naming of technical specificities, products and procedures in an age characterised by a massive industrial expansion and international commercial exchange (Cabré 1999: 196). In this light, in 1947 the International Organisation for Standardisation (ISO) was officially established, with a view to supporting international cooperation by way of the definition of commonly accepted standards (Cabré 1999: 195) for the transmission of technical knowledge. To date standardisation still upholds its significance with relation to terminology, as it is as a matter of fact implicated in numerous activities that are concerned with the accurate usage of terms. Such activities, for instance, can be mainly identified in the professional contexts of specialised communication and translation, that broadly rely on a faithful transmission of concepts. Practically, the rational behind the realisation of this principle is based on the necessity to create a common agreement on terminology which, as an example, proves to be particularly essential in the case of intra-specialist communication that occurs within the context of a specific domain of the language. In this sense, normalisation exerts an influence in terms of the achievement of harmonisation and mutual intelligibility; moreover, it could also result in the economy of vocabulary, all these aspects related to the fact that it is commonly accepted that a given concept is specifically represented and comprehensively designated by a given term.

The aforementioned descriptive function is also encoded in the processes of compilation and management of terminological records, that can be considered as fundamental terminographic activities aimed at the recollection of global and detailed data concerning concepts. In this respect, ISO 1087-1:2000 also establishes a related definition of terminology work:

> work concerned with the systematic collection, description, processing and presentation of concepts and their designations

The importance of relying on such resources concerns the fact that these databases enable for the acquirement of an in-depth understanding of concepts and, contextually, of the accurate and consistent usage of terms. In particular, terminological records integrate knowledge that touches on various linguistic levels of analysis, such as grammatical, morphosyntactic and semantic sets of information. A fundamental aspect is the systematic organisation of contents, which is a key element of terminological records, that is related both to the user experience and to the automatic management of data. With respect to the user experience, the structured visualisation enhances the user-friendliness hence fostering the acquisition of knowledge. As for the automatic processing, the organised structure is based on the implementation of XML metalanguage which, in broad terms, allows for data exchange and makes it possible for programs to unambiguously interpret information.

1.2 Practical applications

Evidence for the practical incisiveness of the function of terminology chiefly emerges from the consideration of its implication in activities that increasingly captivate attention both with regard to literature and on the part of users. In particular, specialised translation and interpreting largely rely on terminological knowledge in order to realise their specific functions. Additionally, terminology performs a relevant role with relation to Information Retrieval[1] which, in essential terms, can be defined as a discipline as well as an enclosing term that designates all the different activities that allow users to obtain target information by way of searching for documents and sources.

Terminology and the domain of specialised translation are deeply interconnected, principally because of the fact that terms that are inscribed in a specific domain of the language are the units that first and foremost convey the main content of the related specialised texts. In order to acquire knowledge to trace beforehand the

1 https://nlp.stanford.edu/IR-book/information-retrieval-book.html

conceptual network of a particular domain as well as to synchronically manage the translation process, professional translators resort to different terminological resources such as the aforementioned terminological records, specialised dictionaries, glossaries and data extracted from corpora. According to Bowker (2009: 288), the necessity to figure out appropriate ways to address terminological issues including the achievement of translation equivalence may result in the realisation of a kind of translation-oriented terminographic work on the part of translators. Besides, in recent years, the translation activity has been considerably fostered by the functionalities offered by CAT tools, that is to say dedicated software that support translators in the effective management of terminology by way of the implementation of terminological databases and Translation Memories (TM).

With reference to specialised translation, however, the emphasis should also be placed on an additional dimension that extends beyond the practical accomplishment of the translation process and the consequent production of the target text. Indeed, specialised translation implies a significant degree of responsibility in all its variations, that is to say with relation to all the domains that are encoded in the general label of special languages. In particular, the perception of the crucial relevance of this task could be heightened especially when considering medical translation, as the faithful transferral of knowledge and concepts by way of terms is to be regarded of the utmost importance due to its subsequent implications in patients' health. Following the same rationale, this consideration can also generally apply to the interpreting profession and, for the same reason, the awareness of the responsibility of the task could be especially perceived in medical interpretation.

As for its pivotal function with relation to Information Retrieval (IR), it appears relevant to mention it also in order to emphasise the multifaceted nature of its applications and usefulness with relation to disciplines others than terminology itself. As a matter of fact, the information need is linguistically expressed by users in the form of a query, therefore the accurate choice of terms to include in it emerges as fundamental in order to increase the chances to retrieve documents that could be

effectively relevant for the purposes of the research. Indeed, terminology is already relevantly present in query writing (Lindén 2004: 136). Another main aspect that supports its significance is inherently linked to the operating way in which search engines perform their functions.

> Information Retrieval (IR) typically measures the relevance of documents to a query based on word similarity. An important issue in word similarity comparison is the equivalence between terms in a query and terms in a document. (Jing and Tzoukermann 2001: 245)

In this regard, a possible general consideration that could issue is that the necessity in present-day society to productively manage terminology may not be considered of exclusive interest of terminologists, translators and specialists of specific domains. Conversely, in the information era, the profitable usage of terminology in the context of search engines could positively influence the way in which users acquire knowledge. Indeed, well-defined queries can possibly theoretically lead to immediate access to the needed information (Boguski 2009: 253). Moreover, they can led to the nowadays particularly praised economy in time because, as a consequence, the necessity to repeatedly reformulate the query to access the target information could be avoided.

1.3 The notion of term

As previously mentioned, the focal element of investigation of terminology is the term. As a matter of fact, however, on the one hand it can be noticed that to some extents definitions and considerations vary accordingly to the distinctive conceptual paradigms that are implicated in different theories of terminology; on the other, it also emerges a lack of consensus for what concerns terms from a lexical viewpoint. To support this statement as well as to present a frame of reference, it could be thereby useful to provide a concise overview of the major approaches to the conception of the nature of terms and, at a later stage, to examine their lexical constitution.

1.3.1 Term in terminological theories

The Classical or General Theory was developed by the Austrian engineer Eugene Wüster in the context of the Vienna School and it can be traced back to 1930. The interrelation between terms and concepts is conceived in the framework of an onomasiological approach, that postulates the fixation of a univocal connection between a term and the concept it designates; specifically, the baseline is constituted by a single concept that is meant to be represented at the lexical level by a monosemic term (Temmerman 2000: 223). Within this perspective, terms are therefore regarded as key elements whose essential contribution to communication is mutual distinctiveness, which is in line with the uniqueness that concepts inherently feature. The important consequence, provided that the frame of reference is the progressive manifested necessity for standardisation, is that monoreferentiality acquires a central function in precluding the happenchance of linguistic phenomena such as polysemy, homonymy and synonymy. This also implies that the meticulous transmission of knowledge is likely to be effectively fostered and mutual comprehension can be easily accomplished. In other words,

> For Wüster (1985), the conceptual dimension is fundamental. Nevertheless, at its foundations verbal designation plays a somewhat relevant role, not only because designations are indispensable to designate concepts, but also because they are necessary elements for non-ambiguous communication. (Santos and Costa 2015: 155)

Another relevant theoretical account is represented by the Communicative Theory of Terminology, formulated by Maria Teresa Cabré in 1999. In particular, this theory underpins that the actual conceptual difference between terms and general lexical elements precisely manifests itself in the context of the pragmatic and communicative dimension, that is when terms are employed in discourse, as specifically stated by Cabré (1999: 81):

> Terms do not seem to be very different from words when we consider them from the formal or semantic point of view; they differ from words when we consider them as pragmatic and communicative units.

A relevant aspect that issues from the author's perspective consists in the observation that terms can be uniquely labelled and considered as such in the circumstance in which they fulfil a specific function, hence it is the specific function that is attributed to them that acts as a determining factor in the recognition of the status of terms. It could be therefore assumed that, to this end, the phase in which the function is ascribed to the lexical entity should actually precede its categorisation. In this respect, in a comparison with the General Theory, the shift of focus is specifically related to the communicative quality that terms assume, consequentially expanding their functionality with respect to their denotative nature in a broader sense and purposefully assessing them in the context of communicative acts and settings.

Contemporary to the previously mentioned theory, Textual Terminology developed as a new approach. As a matter of fact, in 1999 Didier Bourigault and Monique Slodzian proposed yet another definition of term. The term is a 'construct', that is to say it results from the analysis performed by the terminographer: this analysis takes into account the place that the term occupies in a corpus, a validation by experts and the intended purposes covered by a given terminographic description. (Bourigault and Slodzian 1999: 31)

In this respect the aforementioned role of terminographic work becomes the bases upon which terms can be, figuratively speaking, brought to life. As the denomination of this theory suggests, texts represent the elected specialised linguistic environment for extracting terminological knowledge. Precisely, the analysis focuses on the corpus-based contextualisation of terms with a view to observing their behaviour and instantiation in a textual dimension, as well as on the proper recognition of the status of terms that needs to be beforehand ascertained.

A year later, the Sociocognitive Terminology Theory emerged as an additional theoretical perspective, as it specifically looks at terminology from the viewpoint of cognitive semantics, thereby redefining by implication the consideration of the nature of terms. In the words of Rita Temmerman (2000: 39), who proposed this approach, "the

term is considered the starting point for discovering categorial attribution". The notion of concept is reconsidered, as terms are related to units of understandings that respectively assume the status of categories or concepts, the latter of which are fewer and not prototypical (Kwiatek, 2013: 6). In addition, univocity is regarded as not essentially required, considering the contribution that synonymy and polysemy provide to understanding, thus distancing in that regard from the onomasiology-based setting (Kwiatek, 2013: 6-7).

At a later stage, the Lexico-semantic Model proposed a further conception on the basis of a semasiological approach which implies that the focus is placed on the terms (L'Homme 2020: 26) in the framework of their relation with concepts and the contexts or domains in which they are included. In this sense, terms are conceived as independent lexical units that are potentially polysemic, that is to say they manifest variations both on the basis of the context in which they are employed and also within a single domain. However, a distinctive problem that is strictly related to terminology that in the first place needs to be addressed is represented by the complexity of detecting a connection between terms and the particular specialised domain of interest, as L'Homme (2020: 59) asserts.

To sum up, different or partially different conceptions of terms can be detected in literature, the conceptual shifts of which can be related to the origination of multiple theories that, metaphorically, constitute the lens through which terminology can be observed. Interestingly enough, "[t]here is no fully operational definition of terms" (Gaussier, 2001: 168). In particular, a relevant aspect that is worth highlighting is the emergence, in some cases and for some aspects, of fairly divergent accounts. In this connection, a case in point can be peculiarly found in the dichotomous assumptions that are at the root of the General Theory and the Lexico-semantic Model when confronting the onomasiological approach with the semasiology-based one. In this sense, drawing on these observations it could be argued that, provided the consideration of the issue from a theoretical standpoint, an implication is the detection of the polysemic nature of the term 'term' within the linguistic context, which is even to a great extent differently

conceived. As a matter of fact, as far as the notional question is concerned, the term is either totally or in part variably associated to multifarious concepts. At the same time, however, it is also possible to look at the issue from an upside down point of view and theorise a definition that suggests a uniformed rationale, as proposed by Ahmad, Davies, Fulford and Rogers (1994: 269):

> Various definitions of 'term' are available. The common denominator seems to be that a term is a label – usually lexical – in the special language of a specific domain, designating a particular concept in the knowledge of that domain, and arguably less context-dependent with regard to its sense than a general-language word.

1.3.2 Term in lexicon

From the consideration of the various theories of terminology it emerges that the focal point concerns the notional declinations of concepts and terms, which for the most part appear to be arguably perceived as abstract entities that necessitate definition and qualification. As an outcome, less regard is paid to the lexical structure of terms. In order to propose some notional perspectives about it, in this section the topic will be addressed by way of introducing some examples drawn from medical terminology, upon which the following chapter of the present dissertation is going to focus in more detail on. Indeed, the lexical structure of terms could for its part provide additional points of reflection, starting from the fact that it appears to exist no general agreement on the lexical grammatical category that terms should pertain to.

In this sense, Sager (1996: 58) refers to the hegemonic identification and attestation of terms in form of nouns to the extent that, according to some experts, adjective and verbs are not to be regarded as lexical entities carrying conceptual value. Suspending for a short while the differentiation between words as broadly conceived and terminology, the mentioned conception of adjectives and verbs seems however to be at odds with the assumption that is at the basis of the distinction between functional words and content words. As a matter of fact content words, being the units that pertain to the lexicon that carry meaning as opposed to functional words that are grammatical

entities therefore less informative, generally include these two grammatical categories. Moreover, another element that should be considered is the existence of morphological relationships that are established in language with reference to nouns, adjective, verbs and adverbs. Indeed, a multitude of such cases can be observed: *virus* and *viral*, *hepatitis* and *hepatic*, *muscle* and *muscular, anaemia* and *anaemic*, or even *resectable*, *resectability* and *resect.*

To look more closely, for instance, a medical term can assume the lexical form either of a single graphic element or of a multi-word entity (Džuganová 2013: 57), as it is noticeable when drawing this kind of distinction considering terms such as *neuroblastoma* and *leukocytes*, as compared with *lumbar disk herniation* and *chronic focal pancreatitis*. In this respect, however, terms can be constituted by compounds that present an even higher level of lexical complexity because of the compositional concatenation of several graphic entities. For instance, this interlinked structural pattern characterises terms such as *pancreatic mucinous ductal ectasia*, *endoscopic ultrasound-guided fine-needle aspirations* or even *spinal cord injury without radiological abnormality*.

In this regard, a possible consideration is related to the fact that the exact detection of terms that feature numerous components could presumably constitute a particularly challenging operation for different reasons. To provide some examples in order to propose a first argumentation in support of this statement, the term *superior mesenteric vein* represents a case in point. When considering this term, a lay person that is not familiar with the medical domain could perhaps regard *superior* merely as a juxtaposed adjective, hence failing to identify its actual inclusion in the composition of the multi-word term, or manifest doubts about it. As a matter of fact, it is a specific term employed in anatomy to precisely indicate a particular vein that is situated in the abdominal area; moreover, *superior* is featured in the term abbreviation *SMV*, which accounts for its intrinsic involvement with respect to the conceptual content. Following the same line of argument, another relevant case in point could be the term *basal nuclei and thalamic region*. Because of the presence of the coordinating conjunction, it could

be guessed that a non-specialist could be led to think that two different terms are mentioned. On the contrary, it is conceived as a single concept actually used in nuclear medicine that is also designated by way of the abbreviation *BNTr*. At this proposal, an aspect that is worth highlighting is the modality through which terminology seems to some extent to play in such cases with grammatical conventions. Furthermore, a third term that could possibly lead to the same perception is *positron emission tomography/computed tomography*, which however when presented in this specific graphic modality indicates a unitary concept. Specifically, it designates an imaging technique in the context of nuclear medicine that integrates scans derived from both *positron emission tomography* and *computed tomography*[2], that is also represented by way of an abbreviation, that is *PET/CT*.

Moreover, an additional reflection on the reason why the comprehensive recognition of multi-words terms could be tricky could be proposed. Specifically, from this perspective, the issue could likewise be linked to a postulated frequent association between words and the concept that is at the basis of the process of tokenisation that is to say, in this instance, words seen in segmented languages as single orthographic entities generally isolated by spaces. If that is the case, for what concerns terminology, the tendency would arguably be that of encountering some difficulties in the precise detection of terms that feature such a complex structure, that is to say multi-word terms and, at the same time, in correctly individuating unitary concepts. Following this line of argument, both reasons might be conflated into a single main hypothesis, that is the supposed possibility of generally being accustomed to focus more on the individual terms of a multi-word term and the individual concept they express in the given succession, and less on the one broader concept that multi-word terms convey as a whole. To a certain extent, the focal point of this consideration is therefore based on the comparison between the two different levels of abstraction that, from this perspective, respectively distinguish the reading of terms from that of multi-word terms and the interpretation of a single concept from the one of compositional concepts.

2 https://en.wikipedia.org/wiki/PET-CT

CHAPTER 2 Medical terminology

2.1 Special languages

Medical terminology, intended as a circumscribed domain-related collection of terms, is inscribed in a specific language variety that is medical language, which on its part is included in what could be conceived as a macro-structure named special languages, specialised languages or LSP, that is to say Language for Special Purposes. In order to better contextualise it, the notion of *special languages*, that constitutes the linguistic framework in which terminology is inscribed, will be introduced. Special language is conceived as a functional variety of a natural language, reliant on a specialised field of knowledge or sphere of activities, used, as a whole, by a group of speakers that is more circumscribed than the totality of speakers with respect to whose language the special one is a variety, in order to fulfil the communicative needs (above all the referential ones) of that specialised field (Cortelazzo 1994: 7).

In this respect, two concepts that are regarded as fundamental for the constitutive purpose of a special language emerge, namely the necessity for the realisation of a communicative function and, among the various aspects in it comprised, the acknowledgement of the referential needs as the most relevant ones. A possible implication could be that terminology is in this sense meant to perform a substantial role in conveying specialised knowledge. In this connection, a second presumable consideration concerns the fact that terminology would be therefore immersed in the communicative context, that is to say it would be called to practically effectively and adequately designate and convey concepts that pertain to a given specialised domain. It could be inferred that, as a consequence, terminology would function as a means to express knowledge also in a pragmatic situation in which the specific communication needs that the employment of such special language is required to meet need to be realised. Specialised knowledge, however, is expressively attributed to a particular

category of the population that, on the basis of the acquired experience, possesses the ability to manage its specific vocabulary.

2.2 Medical terminology

Considering the inclusion of medical language in the domain of special languages, it follows that both the cognition and the subsequent employment of medical terminology are to be considered as exclusively accessible to subjects that conceptually and terminologically master the specific field of specialisation. As a matter of fact, physicians and doctors are among the figures that properly use this particular specialised language in intra-specialist communication and, more at large, in professional contexts. Nevertheless the medical field also concerns patients who, in order to communicate with specialists and understand information about their own health status at both levels of diagnosis and treatment, need to access medical language.

A further relevant aspect that should be considered is the highly improved accessibility of medical information to non-expert people, which is featured among the multiple outstanding achievements attributable to the evolution brought by the information era. As a matter of fact, patient-oriented websites as well as television programs, highly popular medical dramas and social media propose and integrate contents that deal with the healthcare environment, also allowing for instance for an increasing awareness of existing illnesses. In this connection, the impact of this kind of indirect contact that lay people establish with medical concepts could be regarded as constructive when considering the promotion of health and the sensitisation with respect to disease prevention, but in a different perspective it could however be conceived as presenting potential unfavourable implications. Indeed, a partial or opaque acquisition of terminology and, notably, concepts, could lead to fallacious perceptions, misunderstandings and, in certain circumstances, even to the inauspicious occurrence of erroneous auto-diagnosis.

Precisely with a view to fostering comprehension and communication, it is clearly indispensable for patients and lay people to be given the opportunity to effectively and adequately comprehend the underlying concepts that terms designate. It could be however guessed that the specificity of the information need that patients may manifest the necessity to satisfy could presumably be different from the level of conceptualisation that specialists possess and exchange in specialised contexts. Furthermore, another key point that should be taken into account is related to the fact that technicalities that are inherently implied in the interrelated conceptual network of the domain doubtfully could be properly understood by subjects that are not professionally involved in the medical setting.

Provided the mentioned terminological and conceptual knowledge asymmetry, a modality through which the cognitional and terminological gap is meant to be at least partially deflected is the context-dependant employment of popular non-technical terms. As a matter of fact, as León-Araúz (2017: 216) asserts, the employment on the part of doctors of terms that entail a lower level of complexity in their cognition usually occurs in doctors-patients conversations; on the contrary, standardised technical terminology will be employed by doctors in settings such as conferences. From our perspective, popular terms can be considered as linguistic entities meant to designate conceptual contents that originate from the realisation of an adaptation strategy for specific pragmatic purposes, hence signalling a peculiar difference with respect to the conceptual and terminological origin of specialised technical terms. In particular, this scheme can be metaphorically equated to a communication-oriented translation process whose direction moves from the source concept, that is to say the specialised one, to the target concept in order to foster simplification and, consequentially, mutual comprehension. With regard to the communicative dimension, León-Araúz (2017: 216) further considers that

> 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.

As a matter of fact, it could be observed that there is no peculiar need for negotiation of understanding to occur in specialised communication, principally due to the fact that knowledge is assumed to be antecedently acquired on the part of experts. On the contrary, both the pragmatic employment of popular terms and bit-by-bit explanations in plain language could be implicitly conceived as acts intended to realise a conceptual negotiation in knowledge exchange in doctor-patient social interactions. To provide an example, the specialised term *dysmenorrhea* designate a largely diffused gynecological condition which occurs in relation with the menstrual cycle that is caused by uterine contractions provoked by lipid substances called prostaglandinis, hence inducing a cramping pain in the lower abdomen[3]. For what concerns both terminology and the underlying concept, on the one hand it can be plausibly inferred that experts would precisely associate the term and the exact health problem as well as the distinct anatomical causal agents and processes involved in the phenomenon. On the other hand, however, patients could presumably be incapable of establishing a connection between the specialised term and the specific designated medical condition. By contrast, this association is likely to be realised in the eventuality in which the popular term *painful period* is mentioned. As it can be noticed, the non-technical term expressly evokes both the perception of the symptom, that is the experienced pain, and the particular physiological condition in which it manifests, as period is a term variant for menstruation. Nonetheless, while the symptomatology and the circumstance in which this gynecological problem surges can be expected to be deductively grasped and therefore the term to be actually instantiated in patients' lexical and conceptual knowledge, the most detailed technical features of the concept could only be partially effectively known. In this respect, in most instances, the existing causal nexus that relates the symptom and its definite anatomical cause, that is the occurrence of uterine contractions due to the presence of prostaglandinis, may not be known to lay people. In short, at one level the missing information that could be ascribed to the lessened level of knowledge specificity that patients possess engenders a conceptual gap, yet at another level the employment of the popular term ensures that mutual comprehension in doctor-patient communication is not hampered, given the shared understanding of the term

3 https://medical-dictionary.thefreedictionary.com/dysmenorrhea

itself and the subsistence of common conceptual references. With reference to the employment of non-specialised terms, however, a relevant consideration that should be considered is that:

> It is important to recognize that in translating medical jargon to lay terms, specificity is very often lost and that the person is likely to have his *own fantasies and ideas* about what the doctor is telling him. (Leigh and Reiser 1985: 31)

Eventually, in this connection, a strategy that could be applied in order to bridge this knowledge gap is the aforementioned explanation in the form of a plain language description of the unfamiliar bit of information. In this sense the accomplishment of this operation postulates the necessity for the concept to be to some extent modulated in a simplified but not overly simplistic way in order for patients to adequately access it.

Additionally, popular term frequently assume lexical and conceptual forms that builds on relational connections such as conceptual generalisations with respect to the corresponding specialised terms. With relation to this, the following exemplifications are extracted from the "Multilingual Glossary of technical and popular medical terms in nine European Languages", commissioned by the European Commission and realised by Heymans Institute of Pharmacology and Mercator School, Department of Applied Linguistics[4]. In particular, it can be noticed that the technical terms *carcinoma* and *lymphoma* are both respectively associated to the popular term *cancer*. Along the same line, *oestrogen* and *progestogen* feature the same popular term, that is to say *female hormone*. A third occurrence of this phenomenon is manifested by the popular term *lung disease* which is respectively associated to the technical terms *emphysema* and *pneumopathy*. Interestingly, it is also possible to notice that the technical term *glaucoma* is associated to the popular term *progressive blindness*, which denotes the potential degenerating development of the disease but does not specifically designate "a group of eye diseases characterized by damage to the optic nerve usually due to excessively high intraocular pressure (IOP)"[5].

4 https://users.ugent.be/~rvdstich/eugloss/welcome.html
5 https://medical-dictionary.thefreedictionary.com/glaucoma

A further lexical form of adaptation that could be observed is the shortening of one-word terms, which could be devised as a technique in order to reduce lexical complexity and foster conciseness of expression. By way of exemplification, considering other terms taken from the aforementioned source, *influenza* is ordinarily contracted as *flu*, which is the term that for instance lay people commonly employ to indicate the specific illness and, as Blake (1998: 122) mentions, is also used in medical transcriptions. A further example is the shortening of the specialised term *implantation*, that is indicated in its popular form by the term *implant*.

In this light, a characteristic that could be inherently related to non-technical terms is the loss of specificity which translates into a partial rendition of the conceptual content that the specialised term designates in the specific conceptual network of the domain. Indeed, as Leigh and Reiser (1985: 31) point out,

> translation of "carcinoma of the cervix" to "cancer of the womb" clearly loses the specificity concerning the type and the exact site of the lesion.

For what concerns in particular multi-word terms, as previously mentioned, it can also be observed that a peculiar characteristic is the manifestation of two strictly related levels of complexity, namely lexical density and, in a consequential relation, conceptual density that is due to the concatenation of elements that implicitly exerts an impact on both levels. In this connection, a linguistic strategy that is widely employed in order to enhance concision is the device of abbreviation. In this respect, some examples can be provided to signal the effective economy both in terms of reduced graphic overcharge in documents and spelling time in communication: *ADHD* (Attention Deficit Hyperactivity Disorder), *ADPKD* (Autosomal Dominant Polycystic Kidney Disease) or *NSAID* (Nonsteroidal Anti-inflammatory Drug).

However, Džuganová (2013: 58-59) stresses the potential inability of lay people to reconnect the abbreviations to the extended representations of terms, as the widespread diffusion of the shortened forms has prevailed over the longer graphic entities; it is for instance the case of *AIDS* (Acquired Immune Deficiency Syndrome) or *SARS* (Severe

Acute Respiratory Syndrome). Generally, these forms effectively allow for quickened knowledge transfer in specialised communication, since it could be fairly assumed that there is no need for experts to access to an explicative integration. Nonetheless, it should additionally be considered that patients could be incapable of identifying the extended term and, consequently, its concept, hence broadening the existing knowledge gap in patient-doctor conversations and intensifying the sensation of conceptual distance that could be perceived with respect to the medical language.

2.2.1 Features of medical specialised and popular terms

Medical terms, being the lexical units whose distinctive function is to designate specific concepts that are inscribed in the highly structured conceptual network of the medical domain, are ordinarily associated to the idea of unambiguousness and referential precision. Indeed, the conceptual value is intrinsically conveyed by way of prefixes, stems and suffixes that comprehensively contribute to the comprehension of the semantic content; in this respect, these terms' minimal lexical components could provide a clue in order to accurately identify concepts in numerous instances (Turgeon 2011: 586). In this sense it could be said that this type of structural decomposition of terms can be indicative of the underlying referents; nevertheless, other leading aspects that are intrinsically linked to the objectives of unambiguousness and referential precision should be considered.

Indeed, for the same purposes, a main characteristic that is interconnected with the idea of the necessity for accuracy as to terminology is monosemy, that is to say a principle related to the assumption that, as L'Homme (2020: 11) underlines, polysemy as well as synonymy should not ideally occur. In this respect, lexical monoreferentiality is addressed as the ideal quality that medical terminology should feature in order to fulfil the claim of absolute accuracy. In this direction, Maglie (2009: 24) refers to monoreferentiality by using the term *semantic univocity*, that is to say the quality that the author ascribes to English for Medical Purposes that entails the circumstance in which a single term is entrusted the task to designate one definite concept. In this light,

it is consequentially excluded for another term the opportunity to represent the same concept, therefore the linguistic device of periphrasis constitutes the only technique through which the same conceptual content can be conveyed (Maglie 2009: 24). However, with regard to this question, Gotti (2008: 34) pinpoints that monoreferentiality is to be considered as restricted to the subject area wherein the term is featured. As a matter of fact, by way of example, the term *terminal* refers to different concepts in different domains; for instance with respect to computer science it designates "an input-output device that communicates with a computer and includes parts such as a keyboard or a printer"[6], whereas with reference to the disciplinary field of medicine it represents "a termination, end, or extremity, especially a nerve ending"[7]. However, it can be noticed that the advocated optimal condition of monoreferentiality within the same domain and the presupposition of an ideal linguistic and conceptual systematisation is, to a certain extent, partially unrealised. To provide an example, the medical term *presentation* can be used to refer both to the manifestation of a condition and the positioning of the fetus at the moment of birth.[8] Following this line of argument, a further occurrence is represented by polysemy that can be featured at the level of both abbreviations and acronyms, as highlighted by Kuzio (2019: 29). For instance, specifically referring to medical terminology, *CT* is employed to designate Computer Tomography, Chemotherapy but also a specific chemotherapy drug named Taxotere and cyclophosphamide. Interestingly, a case in point can also be presented with reference to eponyms. As a matter of fact, various dynamics can intervene in the attribution of denominations to concepts, such as the generation of new terms that feature the inclusion of the name of the expert who discovered or detected an illness or a syndrome. Interestingly, despite the fact that the first impression could be that of a device that confers particularly marked uniqueness to the corresponding term by way of distinctively labelling it, eponyms signpost an important site of challenge for the principle of monoreferentiality, as they can change according to the geographical dimension. In this connection Džuganová (2013: 65) mentions an exemplification presented by Woywodt and Matteson (2007: 424) that concerns the eponym *Morbus*

6 https://medical-dictionary.thefreedictionary.com/terminal
7 *Ibidem*.
8 https://medical-dictionary.thefreedictionary.com/presentation

Basedow which, switching from Germany to English speaking countries, assumes the forms of two different alternatives, that is to say *Grave's disease* and *Morbus Graves*.

In this sense, monoreferentiality could be considered as advisable and as the idealistic feature in order to accomplish lexical and conceptual unequivocal conjunction; nevertheless, in some instances it collides with reality, setting up a dividing line between the theoretical and the practical nature of terminology. Indeed, it should be in any case considered that terminology is inscribed in a language system that is generally acknowledged to be intrinsically ambiguous, therefore manifestations of the complexities of this linguistic stratum can possibly surface.

Another relevant aspect that should be taken into account concerns the fact that medical language experiences a substantial connection with general language, which could also be to a certain extent due to the fact that health represents a topic that recurrently manifests its presence and relevance in everyday life. As previously stated, television programs or TV series as well as digital media contents increasingly incorporate and provide medical information. Moreover, to give an idea of its permeating and ubiquitous nature that could be perceived as even closer to a day-to-day experience, it is enough to consider that each time a patient visits a family doctor or tries to read and interpret a drug leaflet, medical language assumes a key role as it becomes the basis upon which the individual information need can be satisfied to solve a health-related issue.

This partial integration that medical terms experience is also reflected in the light of the necessity of a distinction between terms that uniquely possess denotation and terms that conflate into their conceptualisation both denotation and connotative characterisations. In particular, it can be observed that this differentiation partially corresponds with the aforementioned categorisation that distinguishes specialised terms from popular terms. Nevertheless, it is possible to mention a further phenomenon that characterises medical terminology, that is to say its adoption in common vocabulary and general dictionaries (Balteiro and Calvo-Ferrer 2014: 107) and the subsequent usage of

specialised terms on the part of patients and lay people. Specifically, in this regard, specialised terms acquire also connotative traits, yet maintaining their denotative function also in non-specialised settings.

Denotation is the essential feature that realises the referential function of medical terms, as one specific term is permanently meant to designate one distinct meaning (Danesi 2016: 130), in this respect circumscribing it and differentiating it from others. It could be therefore said that the basic principle is the recollection of the concept of objectivity, as it opposes to connotation due to its neutrality that is not affected by social conditionings and subjective biases. By way of exemplification, the medical term *herpes febrilis* specifically refers to a disease caused by herpesvirus 1 or 2 that is characterised by an inflammatory status that ordinarily visibly manifests in the area adjacent to the lips and nostrils because of a febrile illness or stress[9]. The associated concept is therefore merely considered in terms of a medical condition and for its specific place in the conceptual network of the domain, in this sense conceived from a specialised perspective, ignoring for instance negative social connotations such as the unaesthetic effect.

In this connection, the second distinctive quality that can be ascribed to medical terms, in this case with specific reference to popular used ones is connotation, that is to say the less objective and more qualitative conceptualisation of terms. Connotation is intrinsically linked to a tendency towards subjectivity, due to the individual conception that each person cognitively elaborates of a specific word; moreover, it could be especially asserted that connotative considerations are influenced by personal experiences and individual emotional perceptions and associations (Steinberg 2007: 125). From this perspective, however, the absence of unadulterated objectivity can in some cases be supposed to possibly deviate to a certain extent from the conventional or more diffused mental representations that are attributed to terms, in this way highlighting the evidence of the emergence of different subjective conceptual mind maps.

9 https://medical-dictionary.thefreedictionary.com/herpes+simplex

> Some jargon makes it into the common lexicon, either because a paraphrase or a colloquialism has never been created for it or because it is too cumbersome to do so. Medical terms like *arthritis* and *eczema* have made their way into everyday vocabulary. But in so doing, they acquire connotative meanings – *arthritis* has connotations of aging, while *eczema* in common usage may evoke images of incessant scratching. (Danesi 2016: 130)

Interestingly, a focused analysis on the connotative traits that people ascribe to medical terms could constitute a resource also for doctors, that could in this way gain an in-depth knowledge of the perception that patients manifest by way of observing the choice of lexical and conceptual elements used to describe them. In this sense, mutual understanding and communication would additionally benefit from a comprehension that goes further than the modulated transmission of specialised knowledge, thus embracing a subjective and emotional dimension that is likewise present in patient-physician interactions.

2.2.2 Keyness of transparency and accuracy

From our perspective, the purposes of transparency and accuracy are on an equal level and feature as key elements in medical language, specifically representing essential requirements that medical terminology should possess in order to realise both the communicative function and the denotative one. With reference to terms, Gilreath (1995: 34) refers to transparency from a semantic point of view, that is to say to indicate the extent to which the meaning of a term is straightforwardly intelligibly represented by the latter's components. Differently from his viewpoint, in this interpretation transparency is meant to concern the degree to which the usage of a term manages to convey its concept in a communicative context, that is to say the successful implementation of its communicableness, or, in other words, terminological clearness. As a matter of fact, misunderstanding and opaqueness at the lexical, conceptual and communicative levels, either altogether or singularly considered, would severely hamper and critically compromise the realisation of the main terminology-related dimensions, endangering knowledge transferral in intra-specialist as well as in expert-layperson interactions. As Vitali (1983: 196) suggests, [m]edical language necessitates of a strict set of standardisation rules with regards to terminology in order to get rid of

that Tower of Babel of abbreviations, acronyms and synonyms that overload doctors' publications and verbal productions.

Standardisation, though manifestly necessary and relevant in terms of ontological, morphological as well as lexical systematisation, should not however be considered as the only major objective that should be looked upon in order to obtain transparency and accuracy. In my view, for instance, the eventuality of polysemy does not utterly affect these two principles, as the circumstance in which different terms are employed to intentionally refer to an identical concept does rather impact on consistency. As previously discussed, the tension between what is for the most part considered as ideal, that is absence of polysemy in specialised contexts and the realisation of consistency (Hartley and Paris, 1996: 238), and what is factually present in linguistic use and reality, could be regarded as challenging to resolve. In this sense, it could be proposed a shift in conceptualisation for what concerns the role of terminology, that is to say, a usage of terminology that place the emphasis on the centrality of the transmission of concepts, rather than on the lexical form. In this connection, the denotative function is still considered as the essential prerequisite; nevertheless, the focus is on the accurate and effective usage of terminology intended as the selection and employment of a term that conveys the target concept as opposed to others that do not fit both the communicative and denotative dimensions.

Actually, the underlying cognitive purpose of pragmatically and efficiently modulating the usage of medical language in order to reach a straightforward linguistic and conceptual understanding at least in the determinate contextual setting should be the core of medical communication, intended as both by means of writing and spoken conversation. The idea is therefore to take primarily into account the necessity to reach an effective and purposeful transmission of knowledge, and to transpose this intention in form of a rational and accurate selection of terms for what concerns both standardised terminology in specialised contexts and terminology that is employed in order to discursively explicit the concept, for instance in patient-doctor communication. In this

sense, this argumentation is in line with the observation proposed by Galinski and Nedobity (1988: 8):

> Just like standardization in general, the standardization of terminology is not necessary, sensible, or desirable in all areas of human endeavors. Wherever the safety of human beings in connection with a specialized activity is affected, however, standardization is indispensable in certain areas of nuclear power plants, of medical emergency services, et al.

Generally, however, it could be argued that the purpose of transparency and accuracy needs to precede the effective realisation of referentiality, as from a procedural perspective it represents its propositional baseline. Indeed, the absence of transparency and accuracy would consequentially hinder the comprehension of the conceptual content of terminology or, to a large extent, undermine the possibility to reach a common ground of understanding both in patient-doctor and expert-layperson communication. As a matter of fact, the connection between terminology as the lexical and conceptual representation of reality and terminology as an informative content in reality manifests itself in the moment in which terms are effectively written or pronounced, that is to say in verbal communication, in which pragmatics is inherently featured.

To summarise, medical language with specific reference to medical terminology faces a challenging condition that is defined by a multitude of imprecisions and inconsistencies that, as it emerges from different theoretical abstractions, are deeply interrelated. This absence of absolute correspondence between theory and practice with relation to the use of terminology particularly concerns essential concepts that are profoundly connected with the ideal purposes of monoreferentiality and monosemy. In this connection, both these principles and standardisation are confronted with practical conventions that are deeply engrained in the linguistic and conceptual manifestations of terminology. In this respect, a possible consideration that can be proposed concerns the hypothesis that another ideal that should be looked for and achieved is the balance of necessities, which is to say standardisation could be conceived as primarily essential for what regards specialised terms. From this perspective, the particular reference is for instance to abbreviations, whose characteristic is to inherently feature a less explicit

referential quality at the expressive level due to conciseness. As a matter of fact, the existing occurrences of polysemous acronyms could potentially compromise communication even at the intra-specialist dimension, whose referential precision, accuracy and unambiguousness is manifestly essential also with a view to effectively foster research advancement.

By way of conclusion, as it emerges from this concise overview on terminology and from the previously mentioned considerations about its essential role in medical language, the cardinal notion is that terminology and concepts are clearly inherently interrelated, to the extent that the essential condition that allows for the achievement of an in-depth knowledge of terminology is to comprehend its conceptual content at the maximum level of detail. These theoretical premises on terminology have been conceived as fundamental with a view to progressively delve into the core topic of the present dissertation, that is the investigation on semic analysis performed on medical terminology and the attempt to challenge the subjective quality it features by way of seeking for an improved objective systematisation in its performance. Indeed, semic analysis is a technique that, defined in substantive terms, captures the specific and indispensable particularities of meanings through a methodical and chiselled decomposition aimed at the realisation of a collection of minimal traits. Nevertheless, motivations for its productive application with regards to concepts and, therefore, to terminology, will be presented. In addition, the specific focus on the application of semic analysis with reference to both specialised and popular terms in the context of medical language will help to problematise the topic.

CHAPTER 3 Semic analysis

3.1 Semic analysis

Semic analysis is a linguistic technique devised in the domain of semantics (Hébert 2011: 97) that can be conceived as a structured approach to the definition of the meaning of words. The systematic fragmentation of terms into semes, that is to say basic semantic elements (Schmit 2018: 328), provides for a comprehensive account of the essential significant features that conjointly concur to compose the meaning itself, hence its potential for capturing the core semantic content of words. In this respect, the seme is the element that allows for meaning distinction (Nöth 1995: 317). According to Filipec (1994: 170):

> Semic analysis is based on the nature of the natural language itself, which decomposes continuous phenomena into phenomena which are discrete.

Following the theoretical studies of Bernard Pottier and Algirdas Julien Greimas, a pivotal contribution to the development of this method has been epitomized by a student of theirs, François Rastier, who formulated the interpretative semantics theory (Hébert 2011: 97), which represents the theoretical standpoint that will constitute the baseline for the investigation of the topic. Specifically, the core topic of study of the interpretative semantic theory is represented by the textual dimension (Rastier 2009: I). Nevertheless, for the purposes of the present dissertation, semic analysis will be notionally examined with a view to considering it in the optic of the wide-ranging framework of its possible applications with respect to terms perceived as unitary isolated linguistic elements, excluding therefore the textual perspective. As a matter of fact, semic analysis can notably manifest its theoretical and practical value on different levels of the linguistic system, as it provides for a technique that operatively fosters the cognitive and linguistic comprehension of lexical elements, thereby supporting the acquisition of terminological as well as general semantic knowledge by way of adopting a systematic and methodological analytical prospect.

In this sense, semic analysis could be perceived as a device as well as a proper strategy to access deep understanding of words and terms, which proves to be essential in order to enhance information comprehension, communication, textual decoding up to the realisation of a quality translation. Secondarily, the focus will be particularly shifted on the terminological investigation, with specific reference to medical terminology. However, previous to the application of the principles of semic analysis to terminology both from a theoretical and a practical perspective, a synoptic overview on the effective notional pertinency and applicability of this methodology to the terminological field of study should be contemplated.

3.2 Semic analysis of terminology

As previously mentioned, the essential function of semic analysis is to accurately convey the meaning of words, as the meaning of each word can be fragmented by way of a segmentation in its constitutive elements called semes (Hébert 2011: 98). Hence, it could be inferred that semic analysis shall be exclusively performed with relation to the meaning of words and not to concepts designated by terms. As a matter of fact, from a theoretical standpoint it is argued that meaning and concept cannot be considered as corresponding (Khathchadourian 1967: 34; Wang 2020: 119), albeit no general consensus can definitely be found on the question. Nevertheless, this method productively finds an application also with specific reference to terminology in determining the intrinsic composition of concepts. In this respect, it appears pivotal to provide a definition of concept as formulated by Rastier. 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. (Rastier 2010: 126) The conventional validity of these disciplinary norms enable for the translation of concepts, that consequently elude the variety of languages as well as the difference of contexts. (Rastier 2010: 126)

In this view, from Rastier's idea of concept it could be derived that semic analysis can be effectively performed in order to define the conceptual content of terms, as concepts are uniquely designated by terms and, as broadly understood, sememes can

be considered as outputs that originate from the application of this linguistic method. Another element that is worth highlighting is the reference to norms that are distinctive of a discipline according to which a concept is designated. In a way, this notion could also apply to terminology, as concepts and the corresponding specialised terms are related to the circumscribed conceptual network of a specific discipline (Daille 2017: 68) and, in this sense, interrelationally linked. Moreover, in this connection the aim of standardisation could also be to a certain extent conceived as a linguistic and conceptual normative systematising structure. In this respect, it is related to the organisation of the terminological constitutional taxonomy of a special language and, in this sense, to a particular discipline.

A further reflection that follows this line of argument is specifically related to the distinction between words and terms, which constitutes yet another widely debated as well as not straightforward defined and conceptually determined key issue in the linguistic environment. In this connection, an interesting point is presented by Pérez Hernández (2002: 128), who signalled that the complex nature of the establishment of this demarcation stems from the fact that special languages and general language manifest mutual features as well as a regular flow of entities that relocate from one class to the other and vice versa. Providing this consideration, however, a relevant aspect concerns the inherent linguistic stratum that features terms, as expressed by Marinkovich (2008: 125): [i]n conclusion, it is relevant to remind that terms are composed by words, whose mediating power in the construction process of the meaning of the natural, social and cultural world is unquestionable, thus enabling the communication between people.

The postulated convergence and, on some level, coincidence of words and terms as linguistic entities could be furthermore supported by their possible consideration as a unified topic of study. In particular, the reference is to the detection of the contemporary usage of the expression *specialized lexicography* in a synonymous relationship with *terminology* due to their similarity (Pavel and Nolet 2001: xvii). As a matter of fact, this phenomenon signposts an important point of reflection, as it should be considered that

lexicography is the practical application of lexicology, the latter being the field of study concerned with the general characteristics of lexicon (Dash 2009: 443).

To draw a comprehensive overview, the potential line of thought that could emerge with relation to the univocal connection of the application of semic analysis with respect to the analysis of the semantic content of words can be theoretically challenged by several considerations. From this perspective, key to this assumption could be regarded the aforementioned Rastier's interpretation of the concept of concept; nevertheless, other notional accounts can be presented in order to support this presupposition, specifically concerning the fluid multifaceted contiguity that permeates the conception of words and terms. In this sense, semic analysis could provide for a way to embrace multiple rationales and arguments, hence the necessity to exploit its pivotal function in the definition of the semantic content of linguistic entities in a broader context and, therefore, to broach the subject with reference to terminology. The designated subject of study will be precisely identified in medical terminology, which constitutes an interesting case in point for the investigation of the application of semic analysis as executed with respect to specialised terms as well as popular ones. As a matter of fact, as previously delineated, to this special language multiple connections with general language can be ascribed also with respect to its usage and the diastratic dimension, that is to say both in terms of its social employment in everyday life and with regard to the linguistic expressions that are written and uttered also in order to plainly transmit specialised knowledge. In this connection, given the importance of analysing both specialised and popular terminological linguistic and conceptual structures, possibilities for semic analysis to convey the diastratic and the diaphasic perspectives will be investigated.

3.3 The conceptual map of semes

The most relevant assumption which is inherently ingrained in the theoretical conception at the root of semic analysis is the distinctiveness that meanings and concepts should provide to the delineation of words and terms by way of the selected

semes; as a matter of fact, linguistic entities are meant to be semantically circumscribed in order to precisely impart their referential import within the dimension of knowledge. Provided this aim, to this level of identification an additional one is essentially required, that is to say the determination of a neat meaning distinction by way of opposition (Trudel 2009: 1), through which a word or a term can be univocally distinguished with respect to others. In order to fulfil this purpose, semes can be taxonomically divided into two major groups, that is to say *generic semes* which constitute a *classeme* as opposed to *specific semes* that are comprehensively represented by the term *semanteme* (Hébert 2011: 98). According to Rastier (2009: 29), [b]y definition generic semes cannot distinguish between semes belonging to the same class, such as 'hippopotamus' and 'rhinoceros': their differentiation, when necessary, relates to specific semes.

It therefore derives that within the *sememe*, which encompasses all semes that are considered as relevant (Kurdi 2018: 85), in order to semantically define the lexical entity specific semes are the ones that provide the determinant distinctive traits that for the most part allow for the univocal identification of a word or a term in the widely interrelated knowledge system. With reference to terminology the aforementioned principle of monoreferentiality, though conceptually constellated by the previously delineated issues, could be in this sense intrinsically linked to the achievement of the mutual terminological distinctiveness that can be ascribed at the functional level to the individuation of specific semes. On the contrary,

> Generic semes are inherited from higher classes in the hierarchy (hyperonyms). They make it possible to mark relations of equivalence between the sememes. (Kurdi 2018: 86)

It could be thereby argued that this opportunity to interrelate words and terms by way of relying on a partial common semantic basis constituted by generic semes could prove to be fundamental in order to understand their belonging to a shared taxeme, domain or dimension. Indeed, these linguistic classes respectively designate the minimal classes, the spheres of activities as well as the general classes and are represented by way of *microgeneric*, *mesogeneric* and *macrogeneric* semes, all belonging to the generic semes' category (Hébert 2011: 98). Once again with relation to terminology, this kind of

simultaneously selective as well as potentially shared correspondence of semes between different terms could serve as a means to highlight the greater semantic proximity that could characterise certain terms in the domain-specific conceptual map of a special language with respect to other terminological elements which pertain to the same discipline.

A further major possibility that semic analysis offers is constituted by the opportunity to conflate, in the context of the sememe, both the dimensions of denotation and connotation; in this sense, the distinction that can be made respectively between *inherent semes* and *afferent semes* needs to be introduced. Specifically, as indicated by Rastier (2009: 44), [i]nherent semes belong to the functional system of the language; and afferent semes to other types of codifications: social-based norms and idiolectal norms. This theoretical distinction can also be presented by considering that inherent semes are the ones that are manifested in whichever occasion the sememe occurs; dissimilarly, afferent semes come into play in the moment in which the linguistic context is considered (Malrieu 2002: 121).

3.4 The practical steps of semic analysis

In order to provide a supplementary complement to semic analysis' theoretical overview of the considered fundamental elements and a baseline for the comprehension of the subsequent concepts that will be included in its interoperable application, a selective exploration of the progressive analytical phases that its operational mechanism features as well as the additional mention of some theoretical principles that should be considered in its management could be useful. The consideration that could be regarded as a standpoint is that the focus should be placed not only on the terminological as well as the lexical expressions of semes, that is to say merely on the linguistic perspective, but also on the underlying cognitive operation that is performed in the semes' selection phase. As a matter of fact, notwithstanding these main aspects could be perceived as conceptually distant one from the other, they are deeply interrelated and, in that respect, should be both equally pondered with a particular view to attributing to them the same

relevance. Taking this aspect into account, the ideal circumstance would be represented by the perfect convergence by way of an operative interconnection of these levels of analysis as comprehensively concretised in a lexical, terminological and conceptual correspondence.

The comprehension of a concept, however, can involve a certain amount of intricacy, which can be especially attributed to the particular level of abstraction upon which it is inherently based. Indeed, according to Smyth, Collins, Morris and Levy (1994: 66),

> The process of abstraction is, however, the key activity underlying a concept. "Abstraction" can imply many things. The isolation of a subset of features most relevant to the concept is one aspect of abstraction, as in our attempts to say in what ways sheep and goats are alike and different. […] So the study of concepts is but one way of probing into the structure of our knowledge.

Interestingly enough, this passage includes a concept that, on a different level, could be related to the basic conception that is involved in semic analysis, that is to say the necessity for the delimitation of a collection of selected traits pertaining to a concept with a view to individuating similarities as well as differences with respect to diverse words. Moreover, as mentioned in the previous chapter, a further aspect concerns the possible difficulty that derives from the necessity to conceive a global unitary conceptual entity for what concerns multi-word terms. In this case, however, the reference can be made to terminology as considered in a broader context than medical domain; in other words, this complicatedness could concern terms that pertain to different special languages.

The selection of semes that is required in order to efficiently describe and delineate the conceptual content of terms could be in this sense connected with a cognitive retrieval process, as the basic principle is represented by the manifestation of a recall capability that should result in the proficient and effective constitution of a collection of linguistic elements. Provided this, it could be thought that the first step could be therefore regarded as being constituted by the examination of the term, followed by an investigation on the connection that can be established between the term

itself and the concept it designates. Contrary to this, Rastier proposed a peculiar view on the directionality of the process and advocated for an onomasiological approach, thus declining the application of semasiology (Corriveau 2014: 78). Specifically, a possible observation concerns the fact that for the most part the focus appears thus to be placed on the conceptual dimension in the light of its extrinsic materialisation. In this sense, the cognitive elaboration of the concept stands as definitely essential, indeed representing the prerequisite for the definition of the link between the two entities. As it emerges from the directionality of the process, semic analysis could be regarded as a methodology that postulates an onomasiological approach, as the conceptualisation goes from the conceptual structure to the linguistic manifestation of a given term (Corriveau 2014: 78).

In this respect, within the context of the performance of semic analysis an additional layer of relative complexity could be detected, since the cognitively shaped unitary concept could be conceived as supposed to undergo a second mental elaboration in order for the concept to be decomposed into minimal substantial elements. In the framework of this transitional phase, the key objective should be thereby identified in the establishment of an exact and rigorously accomplished equivalence relation between the preconceived unitary conceptualisation of the semantic content related to the term and its segmented configuration. In this connection, a possible implication is that the level of elaboration that this methodic analysis entails can result in a deeply accurate comprehension of concepts and, therefore, terminology. Indeed, knowledge is in this sense enhanced specifically through the processes of both the construction and the subsequent decomposition of the semantic structure as meanings and concepts could be observed by way of adopting two different perspectives that transmute their dichotomous nature into a complementary double-faced sequential level of understanding.

Subsequently, the cognitive process that implies the selection of semes should be concretised in a lexical output, that is to say in the effective expression of the designated minimal elements. In this sense, this phase entails the consideration of their outward

significant structure both from the metalinguistic and the linguistic viewpoint, about which Rastier provided some guidelines. The metalinguistic usage is subjected to specific conditions, such as the use of demarcation signs. (Rastier 2009: 36) It can be further agreed that semes are represented by way of lexemes, regardless of their morphological category as well of the "categorial meaning" they are attributed to: /vertical/ and /verticality/ are thus equivalent denominations (Rastier 2009: 36). In Rastier's interpretation, the baseline for the linguistic expression of semes is therefore constituted by the lexeme, which is a linguistic entity specifically defined by Hamawand (2011: 267) as "[a] unit of lexical meaning which can take a set of inflectional endings". Provided this definition, it can be asserted that semes could be generally composed at the lexical level by different parts of speech, for instance nouns, verbs, adjectives, adverbs as well as determiners and pronouns. Furthermore, with relation to the elements that exerts an influence on the inherent definition of semes and, in a broader sense, to the characterisation of the semantic dimension, interestingly Rastier referred to the impact that the situational framework and the social dimension present in this specific respect. The seme is certainly defined by the relations between sememes, but these relations themselves are determined by the linguistic and situational context. (Rastier 2009: 36) In other words, the semantic description cannot isolate the functional system of the language from the other social norms that regulate the (too) conveniently called pragmatic conditions. (Rastier 2009: 36). As it can be evinced, all these aspects should be regarded as relevant because of the close link that they manifest with reference to the semantic sphere from the perspective of its cognitive as well as expressed manifestation. Consequentially, it derives that they should be comprehensively taken into consideration in order to accurately conduct a reasoned and exhaustive choice of semes and, in a broader sense, to form an idea about their graphical, lexical and relational nature.

A last step that could be potentially added with reference to the analysis could be a conceptual validation, that is to say a feedback phase that involves the check of the substantial existence of an equivalent correspondence between the semantic content conveyed by the different semes included in the sememe and the concept as unitarily

conceived, that is to say as a unified abstraction and, finally, their referential correctness with reference to the given term. This stage could be considered as simultaneously detached from the proper performance of the analysis as well as a sort of follow-up activity, as it concerns the verification of its productive functional effectiveness, which could be regarded as the substantive objective of this process. As it emerges from the directionality of this particular additional frame, also in this case semic analysis could be perceived as contemplating an onomasiological approach, that proceeds from the comprehensive conceptual designation of a word or a term to investigate the validity of the specific connection that has been established between the particular concept or meaning and the given linguistic entity. Hence, the aforementioned theoretical line of argument followed by Rastier would in this instance likewise be adopted, ensuring the continuity of the directionality of the conceptualisation in the methodological frame. In my view, provided this proposed additional phase, this latter supplementary stage of analysis could constitute a valuable yardstick in order to assess the actual assimilation of the general semantic content of a term. In particular, this screening action would support the evaluation of the cognitive systematisation of conceptual knowledge both with reference to the specific terminological entity but also in a broader sense with respect to the comprehension of the place it occupies in the conceptual map of knowledge as similarly or distinctively compared to other domain-related elements of terminology.

With a view to applicatively illustrating the several postulated steps which are supposed to be ingrained in the practical as well as cognitive process at the basis of the performance of semic analysis, a functional example will be offered. In the first place, as previously indicated, the in-depth cognition of the comprehensive concept designated by a term should be achieved. For this purpose, as will be more extensively substantiated in section 3.6 of this chapter, reliable dictionaries prove to constitute an eligible and adequate source of information. For instance, in order to capture and cognitively elaborate the concept that is represented by the term *fibromyalgia*, a lexicographical research should be carried out. As a matter of fact, the primarily adopted perspective should be oriented to the keen understanding of the concept envisaged as a

single definite unit which, at this stage, is not fragmented. By way of exemplification, the following definition could be considered as a reference: "*a chronic disorder characterized by widespread pain, tenderness, and stiffness of muscles and associated connective tissue structures that is typically accompanied by fatigue, headache, and sleep disturbances*"[10].

Subsequent to this primary step, which constitutes the essential groundwork for the following phase, the concept should be cognitively factorised in its basic constituent. With reference to the mentioned definition, the concept could be thought as formed by the following particles: chronic disorder, widespread pain, tenderness, stiffness, muscles, associated connective tissue structures, fatigue, headache and sleep disturbances.

At this point, the transposition of the conceptual elaboration in form of a lexical and metalinguistic output would represent the succeeding stage involved in the process. As a matter of fact, particular attention should also be paid to the metalinguistic system that characterises the formal articulation of semic analysis, as each seme is supposed to be expressively delimited hence distinguished from the others by way of forward slashes. For instance, the following semic analysis could be hypothesised:

> /chronic disorder/ /widespread pain/ /tenderness/ /stiffness/ /muscles/ /associated connective tissue structures/ /fatigue/ /headache/ /sleep disturbances/

Lastly, a validation phase should be contemplated, as the semic composition of the sememe should conceptually coherently match the unitary concept as cognitively conceived in the first step of the process.

10 The definition was taken from https://www.merriam-webster.com/dictionary/fibromyalgia#medicalDictionary

3.5 The performance of semic analysis: a case study

As previously delineated, the primary purpose of the present dissertation is to propose a potential improvement in the establishment of an objective methodology with the aim to refine and harmonise along this line the performance of semic analysis. In this respect, in order to test the subsistence of the preliminarily hypothesised subjectivity and the subsequent factual ascertainment of the necessity to circumscribe and possibly moderate the influence of the effectively existing range of subjectively-biased manifestations, an investigation on semic analysis performed on medical terminology in the context of terminological records has been carried out.

For the present study, the examined data were extracted from terminological records respectively 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. Specifically, the realisation of these records was carried out in the form of a task in the context of the Computer-Assisted Translation Tools course. In particular, the considered reference datasets concern the academic years 2018/2019 and 2019/2020 and, in both cases, the configuration of terminological records was based on the TriMED model (Vezzani, Di Nunzio and Henrot 2018). In details, while the 2018/2019 collection of records was assembled into one single file, for the more recent mentioned academic year various files have been analysed, as each of them contained indications about different terms.

For what regards the structural content organisation of the datasets and, as a consequence, of the contained terminological records, they presented a different range of features from a quantitative point of view. Specifically, the dataset for the year 2018/2019 included in the form of the Excel structure the following sections:

- the term;
- a univocal identification code;
- the indication of the language;

- the etymology of the term as conceived in its derivation and composition;
- a selected definition;
- the performed semic analysis;
- enlisted collocation as well as phraseology which could be reconnected to the given term.

The 2019/2020 dataset was also structured in an Excel file; however, the structural composition of the information presented some differences from a logical perspective, as features were conceived as conceptually grouped on the basis of their categorial pertinence. Accordingly, the four major logical containers were named as following:

- formal features;
- semantics;
- corpus;
- references.

Specifically, the macro-section dedicated to the formal features included the possibility to fill in information concerning:

- the univocal id;
- the term;
- the language;
- the morphologic genre;
- the transcription of the tonic accent;
- indications about orthography;
- indications about etymology;
- orthographic variants of the term;
- acronyms, related expansions or forms.

The semantics section provided the possibility to insert additional data concerning a definition of the term, the semic analysis, the phraseology comprehensive of collocations and colligations and, lastly, semantic variants that could be expressed in the form of synonyms, hyponyms and hyperonyms. Under the categorial label of corpus, the proposed features comprised the specialized context that presupposed the integration of a sentence in which the term was contextually employed, the field and the detected register. The last subdivision was specifically dedicated to the references, in the context of which credits in the form of the specific URLs of the websites from which data were extracted could be mentioned. Moreover, all four macro-sections were characterised by the repetition of the indication of the analysed term, its univocal id and the reference to the language to which the term pertained to.

Precisely, for the purposes of this study, the subsections that were considered for the present investigation were the ones concerning the given term, the language and, comprehensively, all information specifically related to the semantic sphere. More in details, particular attention was attributed to the selected definition of the term, however preponderately from a terminological point of view and, reasonably, the primary focus was on the performed semic analysis.

The process of investigation followed different lines. From a procedural point of view, the first phase consisted in the consideration of the possibility to detect some of the cognitive operations that could potentially be manifestly noticeable in the selection of semes through their transposition at the lexical level of expression which, as previously delineated, constitutes the output of the antecedent elements' choice. A second phase concerned the analysis of semes from a purely lexical perspective, that is to say with peculiar reference to the morphological dimension. It should be furthermore considered that, provided the fact that semic analysis has been performed by students, the present analysis included a third step that, in this instance, assumed the form of a sort of extrinsic check of the conceptual correspondence between the selected semes and the given terminological entry, about which some considerations could be proposed. The analysis of these distinct but simultaneously interrelated and consequential phases has

been applied with respect to three among the different languages in which terminological records were compiled, that is to say English, Italian and Spanish. Precisely, from a quantitative point of view, the distinct formulations with regard to semic analysis that have been considered comprehensively amounted to 460 for what concerns English language, 343 for Italian and 101 for Spanish. This interlinguistic follow-up study has been specifically carried out in order to achieve an insight about two main research perspectives, namely to ascertain the possible presence of similar phenomena in the context of the performance of semic analysis or whether evident differences could be spotted when considering three distinct languages and, moreover, to evaluate whether the hypothesis of the detection of a lack of a single common semes' selection criterion would be generally consolidated.

In order to better contextualise exemplifications, a translation into English of the different terminological data with reference to Italian and Spanish language has been proposed. With respect to this, however, the premise that should be taken into account is that the introduced translation was not performed by the students who compiled the terminological records. Moreover, the translation process could imply a certain degree of complexity, as in some cases it could entail a slight shift in the terminology as translation equivalence could be occasionally unachievable and, furthermore, choices could be needed to be made as for the potential translation candidates and the subsequent selection of the target translated term for the source one. Nevertheless, in this connection the main objective is to provide a translation as accurate and faithful as possible to the original data, with a particular view to avoid alterations in the conceptual content of terms which, with specific reference to the context of semic analysis, remarkably constitutes a key element.

3.5.1 Investigation on English terminological records

In order to identify the manifestation of a postulated subjective attitude towards the conceptual consideration of the principles of semic analysis and, consequentially, towards its applicative performance, the attention has been focused on a first analysis of

the semes' collections in their respective connection with the corresponding English terms.

First of all, evidence for the absence of a uniformed criterion in the selection of semes emerged in the moment in which, with reference to one single term, different collections of semes and, consequentially, sememes, were detected. By way of exemplification, many variants of the clusters of semes for the term *diagnosis* could be tracked:

- */identification/ /disease/ /patient/ /examination/ /analysis/ /symptom/ /sign/ /investigation/*
- */identification/ /disease/ /examination/ /symptoms/*
- */identification/ /disease/ /examination/ /symptoms/ /signs/*
- */identification/ /diseases/ /examination/ /symptoms/ /signs/ /investigations/*
- */to identify/ /diseases/ /medical examination/*
- */judgment/ /resulting/ /examination/ /illness/*
- */recognize/ /exact/ /character/ /of/ /disease/ /by/ /examing/*
- */identification/ /illness/ /living being/ /exam/ /symptom/*
- */discovery/ /naming/ /ill/ /working/ /properly/*
- */statement/ /disease/ /examining/*
- */process/ /identification/ /nature/ /cause/ /disease/*
- */medicine/ /evaluation/*
- */act/ /process/ /identify/ /determine/ /nature/ /cause/ /disease/ /injury/ /evaluation/ /patient history/ /examination/ /review/ /laboratory data/ /opinion/ / medical/ /human being/*

Focusing on a first examination concerning the semes' selection, it emerges that in some instances the same seme could be found; it is the case for example of the semes /identification/, /examination/ and /disease/, which are expressed in various formulations of the analysis. In this sense it could be stated that for the same term some

particular conceptual elements are to a certain extent commonly shared in students' perception. Considering these exemplifications as a whole, in this respect it could be noticed that, despite the fact that as previously mentioned the sememe is the unit which represents a concept and that the conceptual content designated by a term in the context of a special language should be well-defined and potentially semantically widely agreed upon, nevertheless different semantic traits are generally employed in order to designate it.

Following the same line of argument, a variation in the performance of semic analysis can also be found with reference to the term *screening*:

- */examination/ /testing/ /group/ /individuals/*
- */examination/ /test/ /population/ /people/ /asymptomatic people/ /high probability/ /desease/ /diagnostic test/*
- */process/ /exam/ /disease/ /organ/ /tissue/ /living being/*
- */test/ /large number/ /people/ /human beings/ /prevention/ /disease/*
- */test/ /large number of people/ /likelihood/ /disease development/*
- */testing/ /presence of a disease or other condition/*

In this case, the semes that can be recurrently individuated are /test/ and /testing/, whereas the reference to the fact that the given term concerns an examination of a group of people is differently expressed by way of semes through the usage of both various lexical and conceptual forms, such as /people/, /population/, /group/, as well as /large number of people/. These examples could be therefore considered as a first piece of tangible evidence of the fact that the presence of subjectivity in the semes' selection effectively permeates the practical application of semic analysis, as it can be argued that concepts whose semantic nature should not be nuanced or affected by substantial variations seem to recall to the mind different conceptual elements. With reference to the same case, it can be for instance noticed that the seme /prevention/ appears only

once, and the same frequency of occurrence can be likewise observed with regard to both /organ/ and /tissue/.

The second step that was conceived as potentially useful in order to spot other cognitive operations that feature in the selection of semes is represented by the comparison between the definition of the term and the performed semic analysis. In that respect, a previously postulated underlying assumption was confirmed, that is to say the subsistence of a close connection that is often established between the terminology included in the selected definition and the semes which compose the sememe. A case in point is represented by the previously mentioned semic analysis of the term *diagnosis* as, interestingly, the first three provided examples of semic analysis can be reconnected to the same definition "*The identification of diseases by the examination of symptoms and signs and by other investigations*"[11] and the majority of the semes are shared. Another representative case that manifestly indicates the existence of this link can be found in the semic analysis of the term *meningitis*, whose selected definition is "*Inflammation of the membranes that surround the brain or spinal cord, caused by infection.*"[12]. As for the semes, the absolute correspondence with respect to the extracted terminology can be spotted:

/inflammation/ /membranes/ /surrounding/ /brain/ /spinal cord/ /infection/

To provide another significant example, for the terminological entry *FRA-HS*, which has been correlated to the definition "*score for prediction of risk of osteoporotic fractures*"[13], the following semic analysis has been performed:

/score/ /prediction/ /risk/ /osteoporotic fractures/

11 A reference to the definition was not present in the corresponding dataset. The source of the definition was therefore retrieved by way of a web search query: https://www.dictionary.com/browse/diagnosis

12 The definition was taken from https://www.collinsdictionary.com/it/dizionario/inglese/meningitis

13 The definition was taken from https://europepmc.org/article/MED/28160026

Provided these exemplifications, it could be hence inferred that a manual term extraction has been carried out on the part of students by way of extracting terminological units from the definition of the term. In the light of this consideration, it could be consequently assumed that these elements were conceived as the bearers of the conceptual content of the term and, for this reason, they were also regarded as suitable for the composition of the sememe. With regard to term extraction, according to Wong, Liu and Bennamoun (2008: 501):

> The main aim in term extraction is to determine whether a word or a phrase is a term which characterises the target domain. This key question can be further decomposed to reveal two critical notions in this area, namely, *unithood* and *termhood*. Unithood concerns with whether or not sequences of words should be combined to form more stable lexical units, while termhood is the extent to which these stable units are relevant to some domains. […] While the former is only relevant to complex terms, the latter concerns both simple terms and complex terms.

Specifically, in the context of this investigation, termhood can be considered as evaluated on the basis of the personal perception of the belonging of a term to the medical domain, which then potentially translates into the consequential selection of the terms which conceptually concur to designate the concept. It derives that, in this sense, this notion potentially manifests a relation with respect to the principles that could be considered as significant for the performance of semic analysis.

In addition to this, however, the application of another criterion in the selection of semes could be recognised, as in some cases a perfect correspondence between the semes which compose the sememe and the terminology included in the description of the term has not been established. In this sense, the activity related to terms' retrieval is in numerous cases implemented by way of the inclusion of semes that present a different terminological and conceptual content, thus conflating two different approaches to semic analysis; namely the mentioned term extraction and a selection that does not completely hinge on one selected definition. A case in point can be represented by the semic analysis of the term *asthma*, taking as a reference the selected definition "*A generally chronic disorder characterized by weezing, coughing, difficulty in*

breathing, wheezing and a sense of constriction in the chest"[14], whose sememe comprises:

/Chronic/ /disorder/ /coughing/ /wheezing/ /difficulty/ /breathing/ /sense/ /constriction/ /chest/ /respiratory illness/

As it emerges from a comparison between the selected semes and the definition, it can be observed that /respiratory illness/ does not match, while the other semes exhibit a total terminological as well as lexical compatibility with the terminological units that are employed in the definition. Another case in which the same criterion could be detected concerns the term *vaccination*, defined as "*The process of giving a person or animal, usually by injection, a substance containing a harmless form of a disease, to prevent them from getting that disease.*"[15] and analysed in the form of semic analysis as follows:

/process/ /living being/ /injection/ /substance/ /prevent/ /disease/

A further observation that can be proposed is related to the numerous cases in which semic analysis could not be associated with neither of the two mentioned criteria. In other words, the close link between the description of the term comprised in the record and the related semes did not manifest in the totality of the circumstances, hence it could be inferred that a more loosely knit connection was also perceived. By way of exemplification, for the term *rosaceiform dermatitis* the definition which has been proposed is "*Invasion of the sebaceous follicles by Demodex folliculorum and other Demodex species leads to spongiotic changes in the follicular epithelium*"[16]. The related semic analysis features the following semes, which, as it can be noticed, broadly

14 The definition was taken from https://www.collinsdictionary.com/it/dizionario/inglese/asthma
15 The definition was taken from https://www.collinsdictionary.com/dictionary/english/vaccination
16 The definition was taken from https://books.google.it/books?id=4Xn1CAAAQBAJ&pg=PA348&lpg=PA348&dq=A+chronic-recurrent+inflammatory+skin+disease+with+marked+pruritus+and+variability+in+lesion+morphology&source=bl&ots=3raERAAvjZ&sig=ACfU3U0KUaTyif0INpJ7M-7igMJflLV6Dw&hl=it&sa=X&ved=2ahUKEwj-od-X-7zmAhUoxKYKHWRLCLwQ6AEwCXoECAsQAQ#v=onepage&q=A%20chronic-recurrent%20inflammatory%20skin%20disease%20with%20marked%20pruritus%20and%20variability%20in%20lesion%20morphology&f=false

distance themselves from the terminological content of the definition in order to provide other conceptual elements:

/skin/ /inflammation/ /disorder/ /disease/ /medicine/ /follicle/

From a lexical perspective, various observations can be proposed. First of all, it can be generally asserted that, from the investigation of terminological records, it appears that nouns and verbs constitute the two most representative morphological categories related to the performance of semic analysis. Nevertheless, the presence of determiners as well as prepositions and coordinators, even though in a limited number of cases, can be noticed. In this respect, it could be emphasised the fact that this occurrence seems to be at odds with the formerly quoted definition of lexeme, on the basis of which it could be claimed that they should not feature as constitutive semic elements. However, some examples of their employment in the context of the following semic analysis can be provided respectively with relation to the terms *platelet*, *saline*, *fungal*, *gestational* and *spine*:

- */blood cell/ /stop the bleeding/*
- */isotonic solution/ /solution of sodium chloride/ /distilled water/*
- */of or caused by/ /fungus/*
- */relative to/ /the process or period/ /developing/ /inside the womb/ /between conception and birth/*
- */series of vertebrae/ /extending from the skull to the small of the back/ /enclose/ /spinal cord/ /provide support/ /torax and abdomen/ /backbone/*

In addition, in some instances also the presence of two negative markers, that is to say the no-negation and the not-negation could be observed, as in the cases of the semic analysis of the terms *asymptomatic*, *mass*, *serum* and *non-falciparum*. As it can be observed, interestingly the negator ‘no’ recurrently features as part of the semes that include other lexical unities, whereas ‘not’ has been considered as a minimal element as detached from the others.

- */producing or showing/ /no symptoms/*
- */large body/ /matter/ /no definite shape/*
- */portion/ /animal fluid/ /plant fluid/ /blood serum/ /straw-coloured/ /liquid/ /plasma/ /no fibrinogen/ /no blood cells/ /clotting/*
- */malarian/ /infection/ /not/ /plasmodium/ /falciparum/*

Moreover, it can be noticed that acronyms and abbreviations are scarcely represented by semes. In this sense, two different considerations could be hypothesised as possible reasons: at one level, the analysed definitions scarcely employed such linguistic entities in order to provide a semantic contextualisation of terms, yet at another level acronyms and abbreviations could be potentially perceived as less explicitly informative and, in this sense, more representative of a lexical variant of the form of the term itself than of the conceptual content it designates.

Furthermore, another criterion that can be detected with reference to semes' composition from a purely morphological point of view can be associated with the concept of transposition, which is a notion deriving from translation studies that specifically constitutes a translation technique. In particular, Shuttleworth and Cowie (2014: 190) referred to transposition by resorting to the definition proposed by Vinay & Darbelnet (1958/1995: 36), for whom it implied "replacing one word class with another without changing the meaning of the message". The application of this procedure is precisely manifested in the following examples, in which once again it can be emphasised the reliance on the term's definition in order to select semes. A first case can be for instance detected by comparing the chosen definition for the term *injection*, that is to say "*if you have an injection, a doctor or nurse puts a medicine into your body using a device with a needle called a syringe*"[17] and the following semic analysis, in the context of which this technique has been applied with respect to the shift from the noun 'injection' to the verb 'to inject':

/medicine/ /needle/ /syringe/ /to inject/ /doctor/ /nurse/

17 The definition was taken from https://www.collinsdictionary.com/dictionary/english/injection

To provide another example that concerns the term *antiviral*, it can be noticed that the adjective 'capable' included in the definition "*capable of checking the growth or effect of a virus*"[18], has been transposed into the noun 'ability':

/ability/ /growth/ /virus/

The majority of the morphological transformations, however, primarily concerned changes in the inflections of nouns as well as verbs. In particular, with reference to nouns, an exemplification is provided by the semic analysis of the term *virology*, whose definition is "*branch of medicine concerned with the study of viruses and the diseases they cause*"[19]:

/branch/ /medicine/ /study/ /virus/ /cause/ /disease/

In this case the inflectional affix -s, which indicates the plural form of the considered nouns, has been removed in order to employ the singular form. Another instance that can be considered concerns verbs, as it emerges from the comparison between the semic analysis of the term *epidemic* and its definition, "*Outbreak of disease that spreads quickly and affects many individuals at the same time*"[20]:

/outbreak/ /disease/ /spread/ /quick/ /multitude/ /people/ /simultaneity/

As it has been highlighted the verb, which is conjugated in the third person singular in the definition, can be detected in its base form in the performed semic analysis. Comprehensively, a possible reason that could allegedly be hypothesised for these morphological choices could be related to the fact that in this particular context base forms could have been considered on the part of some students as the main carriers of concepts and, in this view, it could be supposed that inflections were contextually considered as not particularly informative from a conceptual viewpoint.

18 The definition was taken from https://www.collinsdictionary.com/it/dizionario/inglese/antiviral
19 The definition was taken from https://www.collinsdictionary.com/it/dizionario/inglese/virology
20 The definition was taken from https://www.merriam-webster.com/dictionary/epidemic

3.5.2 Investigation on Italian terminological records

With reference to Italian medical terminology, the semes' selection strategy represented by the correlation between definitions and semic elements can be to a great extent observed. This circumstance could be emphasised by way of exemplifications such as the correspondence that can be noticed when comparing the definition of *morbo di Alzheimer*, "*Patologia degenerativa del sistema nervoso caratterizzata da un quadro di demenza presenile o senile*"[21] and the selected semes to convey the concept:

/patologia/ /degenerativa/ /sistema nervoso/ /demenza/ /presenile/ /senile/

The English translation of the term *morbo di Alzheimer* is *Alzheimer's disease*. Its definition can be translated as "Degenerative pathology of the nervous system characterised by a condition of presenile or senile dementia"[22]. The related semic analysis would therefore result in:

/pathology/ /degenerative/ /nervous system/ /dementia/ /presenile/ /senile/

Moreover, cases of partial reliance on the definition of a term could be identified, as in the comparison that can be made between the chosen definition of the term *D-dimero* and the performed semic analysis. Provided the definition, "*Il D-Dimero misura la concentrazione plasmatica di un prodotto di degradazione della fibrina, la principale responsabile della formazione di coaguli (trombi) nei vasi sanguigni*"[23] the selected terms that constitutes the semic analysis have been identified in:

/d-dimero/ /misurare/ /concentrazione/ /degradazione/ /fibrina/ /coaguli/ /vasi/ /sanguigni/

21 A reference to the definition was not present in the corresponding dataset. The source of the definition was therefore retrieved by way of a web search query: http://www.treccani.it/enciclopedia/morbo-di-alzheimer/

22 My translation of the definition provided in the corresponding terminological record.

23 A reference to the definition was not present in the corresponding dataset. The source of the definition was therefore retrieved by way of a web search query: https://www.corriere.it/salute/esami-sangue/cuore-metabolismo/d-dimero_e73d9ce0-dc4a-11df-be1f-00144f02aabc.shtml

The term *D-dimero* can be translated as *D-dimer*, whose previously mentioned definition can be translated in English as follows: "The D-dimer measures the plasma concentration of a fibrin degradation product, which is the main responsible for the formation of blood clots (thrombi) in the blood vessels"[24]. The related semic analysis would therefore result in:

/d-dimer/ /measure/ /concentration/ /degradation/ /fibrin/ /blood clots/ /vessels/ /blood/

As it emerges, semic analysis does not include a seme that presents the information about the specific reference to the particular concentration that is measured, that is to say *plasma concentration*. Besides, the medical term *thrombi* is present in the definition of the term, but it is not contemplated in the semic analysis. In this case, it could be inferred that uniquely some terms have been considered as essential for the conceptual definition of the term; however, it could be potentially argued that reference to *plasma* should have been included in order to specifically designate the concept in its essential elements.

For what concerns morphological transformations, it can be asserted that transposition is also occasionally applied with reference to semic analysis in Italian terminological records. For instance, the application of this technique can be noticed as integrated in the semic analysis of the term *diagnostica per immagini*, whose definition is "*La definizione diagnostica per immagini indica la metodica che permette di conoscere, attraverso la formazione di immagini radiologiche nel vivente, la presenza di uno stato patologico.*"[25]:

/diagnostica/ /immagine/ /metodica/ /conoscenza/ /formazione/ /radiologico/ /vivente/ /presenza/ /stato/ /patologia/

24 My translation of the definition provided in the corresponding terminological record.

25 A reference to the definition was not present in the corresponding dataset. The source of the definition was therefore retrieved by way of a web search query: http://www.treccani.it/enciclopedia/diagnostica-per-immagini_%28Enciclopedia-Italiana%29/

This exemplification entails various kind of morphological shifts which, in order, are represented by a plural form of a noun transformed into its singular form, a verb that is transposed into the corresponding noun, an adjective in its plural feminine form rendered as an adjective in the singular masculine one and an adjective transposed into a noun. The majority of these morphological choices can be also exemplified by way of an English translation of the same example, that is to say providing the analysis of the term *diagnostic imaging*, whose definition is "Diagnostic imaging indicates the methodology that makes it possible to know, through the formation of radiological images in the living being, the presence of a pathological state."[26]:

/diagnostic/ /image/ /methodology/ /knowledge/ /formation/ /radiological/ /living being/ /presence/ /state/ /pathology/

Furthermore, an interesting occurrence that could be noticed by way of analysing various semic analysis is the employment, for different terms, of the same sememe. For instance the terms *epidemia* and *contagio*, in English *epidemic* and *infection*, have been paired with two identical semes which, additionally, constitutes the only elements employed to designate both concepts, in this case /fenomeno/ and /malattia/, translatable in English as /phenomenon/ and /disease/. Following the same line of argument, the seme /malattia contagiosa/, in its English translation /contagious disease/, has been used to refer to two different terms, that is to say *morbillo* and *rosolia*, known in English as *measles* and *rubella*. In this respect it could be noticed that an underlying fundamental principle of semic analysis is not applied, that is to say the distinctiveness that the selection of semes should confer to terminology in order to univoquely distinguish between different terms. As a matter of fact, in the light of the foregoing notions, a single concept should be represented by a particular sememe, that should in this view ideally exclude excessive generalisation that could convert sememes into potential umbrella cluster of semes for different terms. In this sense, it could be implied that in this cases, in order to enhance the possibilities for semic analysis to efficiently circumscribe a concept, more semes should be included.

26 My translation of the definition provided in the corresponding terminological record.

3.5.3 Investigation on Spanish terminological records

For what concerns Spanish terminology's semic analysis, the operational procedure which consists on the fact that semes are selected on the basis of terminology extracted from a definition can be regarded once again as a highly detectable phenomenon. A case in point can be represented by the semic analysis of the term *efecto secundario*, whose provided definition is "*Consecuencia indirecta y generalmente adversa del uso de un medicamento o terapia*"[27] and whose selected semes are:

/consecuencia/ /indirecta/ /adversa/ /uso/ /medicamento/ /terapia/

The English translation of the considered term is *side effect*. Its definition can be translated as "Indirect and generally adverse consequence deriving from the use of a medicine or a therapy"[28]. The related semic analysis would therefore result in:

/consequence/ /indirect/ /advers/ /use/ /medicine/ /therapy/

From a morphological perspective, the aforementioned phenomenon of inflexional transformation can be detected, as in the case of the term *circulación*, whose chosen definition is "*Movimiento regular de la sangre que, partiendo del corazón, se distribuye por las arterias a todo el cuerpo y vuelve a través de las venas*"[29], and whose corresponding semic analysis is:

/movimiento/ /regular/ /sangre/ /parte/ /corazón/ /distribuye/ /arterias/ /cuerpo/ /vuelve/ /venas/

As it can be evinced, from the definition of the term *circulación*, which is translated in English as *circulation*, "Regular movement of blood that, departing from the heart, is distributed through the arteries throughout the body and returns through the veins"[30] all

27 The definition was taken from https://dle.rae.es/efecto#MqUF2ND
28 My translation of the definition provided in the corresponding terminological record.
29 The definition was taken from https://dle.rae.es/circulaci%C3%B3n?m=form
30 My translation of the definition provided in the corresponding terminological record.

elements were extracted, thus entailing a total terminological extraction in order to provide the conceptual delineation of the term, and a morphological transformation that concerns a verb can be noticed. Specifically, the morphological shift is in this case represented by the change in the inflexion of the verb.

/movement/ /regular/ /blood/ /depart/ /heart/ /distributed/ /arteries/ /body/ /returns/ /veins/

3.5.4 Resulting data and observations

Comprehensively it could be asserted that, with reference to the 2019/2020 dataset, in a total of 198 analysed semic analysis' cases relatable to English terminological records, two-thirds of them presents a close connection with the selected definition of the term, in a range that oscillates from the existence of a total correspondence on the basis of terminology between the two semantic expressions to their proximity, that is to say from the presence within the semes of the majority to a minimum of half the terms extracted from the correlated definitions. For what concerns the 2018/2019 dataset, in a total of 262 considered semic analysis more than a half could be perceived as featuring the same interconnection.

Generally, from a comparative interlinguistic perspective, it emerges that similarities in the strategies adopted for the performance of semic analysis could be detected with respect to English, Italian and Spanish. Specifically, the particular reference is to the reliance on terminology extraction from definitions in order to individuate semes, as well as the application of modifications from a morphological point of view. In particular, for what concerns terminological records of Italian terms comprised in the 2018/2019 dataset, the amount of analysis that featured terminology extracted from definitions rises up to two-thirds, whereas with respect to Spanish, whose records were included in the 2019/2020 dataset, the same selection's criterion occurs in more than 50% of the cases. The consideration that emerges from these data is

therefore the tendency to regard definitions as fundamental sources of knowledge and information about terminology and concepts:

> Empirically the terminological definition […] is a compromise solution between the lexicographical definition and encyclopaedic description, for the purpose of improving the usage of nouns so that they can function as terms, and also for reflecting – not reproducing or representing – the method of constituting classes of beings and the operation of the conceptual system. (Rey 1995: 42)

Moreover, from a comparison between the semic analysis of terms and the included semantic variants, that ranged from synonyms to hyponyms and hyperonyms, it could be observed that in a high number of cases the selected semes did not comprise the majority of the terminology extracted from that particular section.

Another relevant aspect that could be emphasised is related to the apparent absence of connotation in the selection of the semes constitutive of sememes which can be detected in terminological records independently from the considered language. As a matter of fact, terms were apparently mainly considered in their referential role by relying on denotation, that is to say excluding personal perceptions or conceptualisations about specific medical conditions or, more in general, medical concepts. This approach could constitute an interesting point of reflection with respect to semic analysis performed on medical terminology, as it could be arguably inferred that connotation could be potentially perceived as less relevant with respect to the referential function that terms should fulfill within the particular domain. A second interpretation could be represented by the fact that connotation could have possibly be deemed as less objectively distinctive considering the differential aim with regard to terminological distinctiveness that semic analysis could be thought as supposed to underpin. Furthermore, in this connection, a heavy reliance on the term's definition could exert a tangible impact on this circumstance, as the presence or the absence of connotative traits within sememes could be directly proportional to the occurrence of the expression of connotative features in the definition itself. In this respect, it could be also inferred that terms could have been analysed by way of focusing on their specialised contextualisation, that is to say without reference to the conceptual

dimension of their employment in general language or settings that do not involve intra-specialist communication in which terminology is not regarded from a purely denotative and subjectively-unbiased perspective.

3.6 Proposals for an improved methodology for semic analysis

As previously delineated, evidence from analysed terminological records highlights the subsistence of a wide range of manifestations that could be ascribed to the trend represented by the application of a subjective approach to the performance of semic analysis. Along the same line, Pottier (1992: 117) claimed that "[w]hat is surprising is the arbitrariness of the choice of semes as compared to the perceptible world". In this respect, this tendency can be regarded as the basic reason for the consideration of the potential necessity to objectively systematise its methodology and to propose some criteria that could be in this view applied. With reference to the terminological leitmotif of the present dissertation, the framework that will constitute the baseline for the application of the notional proposals of this potential implementation will be medical terminology.

As mentioned in the previous chapter, medical terminology can be conceived as a special language that, principally due to the communicative and situational dimensions it needs to embrace, includes both specialised terms and popular terms. Referring once again to underlying cognitive operations, this particular categorial distinction could be perceived as an additional layer of complexity with respect to semic analysis because of the fact that, for instance, the different level of specificity could entail conceptual differences that should be accordingly accurately conveyed. Moreover, another consideration that could represent a potential difficulty is related to the cases of lexical correspondence that can feature a determined specialised term and its popular counterpart as, consequentially, in the circumstance in which terms are analysed as detached from a contextualised linguistic setting, the issue about the level of specificity that should be expressed by way of semes could constitute a tricky point.

A third observation that stems from the previously presented case study concerns the fact that, in some instances, for the same term the definitions that were selected by students differed one from the others. In a comparative perspective it can be asserted that, in many cases, the consequential result was the presence of different semes that culminated in a discrepancy in terms of conceptual elements. Considering the necessity of reducing and harmonising such occurrences in the performance of semic analysis, this aspect should be therefore preliminarily considered and the objective of achieving a potential normalisation with respect to this partial semantic distance especially in definition-based semic analysis should be addressed.

The first proposal for a methodological improvement is represented by the possibility to conceive a domain-oriented corpus-based semic analysis, thus with specific reference to special languages, that is to say to adopt a methodological perspective which assumes as a baseline for the selection of semes an earlier recollection and exploration of various definitions for an analysed term that would precede the applicative phase of semic analysis, which will be realised by way of a terminological extraction. This methodological approach based on lexicographical resources was also respectively adopted by Baider and Constantinou (2014)[31] in the context of a study which focused on the retrieval of the semes that could represent the concepts linked to emotions and by Elezi (2018)[32] in relation to an investigation on political and economic terminology through the analysis of semic elements in definitions included in specific dictionaries of the Albanian language.

Within the framework of the present dissertation, however, this methodology is not solely applied from a domain-oriented perspective within the field of terminology; therefore considering domain-specific definitions composing a vertical corpus, as it is principally and purposely advocated as well as targeted as the groundwork for the

31 https://www.researchgate.net/publication/329511969_La_fureur_de_gagner_la_rage_de_perdre_Etude_contrastive_des_concepts_de_rage_org e_et_de_fureur_lyssa_en_grec_et_en_francais

32 https://www.researchgate.net/publication/327862833_Ideological_Background_of_Political_and_Economic_Entries_in_Explanatory_Dictionar ies_of_the_Albanian_Language

objective recollection by way of a term extraction of all the semes which convey a specific concept related to a determined term. Moreover, it is conceived as a valuable source of information and a strategy to obtain greater awareness with respect to the conceptual content of terms and, consequentially, to foster the accurate usage of terminology. In the words of Tribble (1997: 112),

> If one wishes to investigate the lexis of a particular current domain (e.g. health) a specialist micro-corpus can often be more useful than a much larger general corpus.

In this particular context, it could be postulated that the corpus would be mostly considered from a qualitative perspective than from a quantitative one, as the pivotal function attributed to its employment would be aimed at the comprehensive retrieval of all the elements that should constitute the sememes in order to exhaustively circumscribe and represent concepts, therefore excluding numerical considerations based on frequencies.

This approach could be perceived as a data-driven procedure which, however, would effectively not entirely exclude subjective choices from the spectrum of possibilities, as a selective terminological, lexical and consequential conceptual extraction would still be performed. Nevertheless, it would possibly contribute to an increased level of objectivity to semic analysis due to the fact that semes' selection would be linked to a criterion that would allow for a more expanded knowledge. As a matter of fact, the potentiality that this methodology could provide is to transcend the subjective knowledge that individuals could possess, whose level of detail about the specificity of a concept could differ also to a great extent. In this connection, the application of this approach to semic analysis of terminology could even amplify the opportunity for lay people to capture and introduce terminological entities by way of semes, thus expressing fine-grained conceptual particularities with specific reference to specialised terms involving specialised hence advanced conceptualisations. In this sense, the quality of semic analysis would be enhanced as it could be considered as a tool that allows for an improved understanding and cognition of the concepts pertaining to a specialised field. From this perspective, semic analysis could be therefore

conceived as a methodology that at one level, from a semantic viewpoint, provides the opportunity to capture the conceptual content and the links between concepts and terms, yet at another level it represents a learning strategy that allows for the acquisition of knowledge in multifarious ways. Consequentially, a domain-oriented corpus-based approach could possess the potentiality to partially fill the cognitional gap between the previously possessed knowledge of a concept and the specialised conceptual elaboration that is inherently involved in specialised terminology. Moreover, it would provide for an enhanced possibility to fully conceptually define the term. As a matter of fact, the reliance on one single definition would statically and univocally link the semantic knowledge expressed by way of semes to the delimited expression of the meaning of a word or of the concept of a term which that particular definition presents. In other words, by way of exemplification, in the happenchance that three different definitions are respectively considered by three different people, the output of – a full or a partial – term extraction could accordingly differ also to a large extent. On the contrary, the consideration of all three definitions on the part of three different people could result in a term extraction whose output comprises a larger number of shared terminological and conceptual elements. In this sense, a more detailed amount of information that, in essence, could be asserted to be attributable to a delimited concept, could potentially represent an increased opportunity to comprehensively and more accurately express the concept of a term and to include various perspectives and conceptual particles with regard to its semantic content.

Subsequent to the consideration of definitions as a valuable source in order both to gain access to conceptual knowledge and to constitute a first criteria for the accomplishment of the purpose of theorising an objective methodology for the performance of semic analysis, however, the line of reasoning implicated in the selection of a suitable type of definition should be more accurately problematised and defined. In this connection, an interesting point to ponder which could be conceived as a starting point for the focus on the topic is represented by the assertions that while descriptive characteristics are linked to descriptions, essential characteristics are related to definitions (Roche 2012: 18-19) and are advocated as indispensable for the definition

of concepts as well as for the differentiation between different concepts (Roche 2012: 21). Recalling the aforementioned notional principle of semic analysis, which is based on the necessity to confer distinctiveness to linguistic entities, definitions could be thought as being therefore confirmed as an essential reference work in order to convey conceptualisations. Understanding a conceptual system, however, could possibly entail a challenging and demanding cognitive activity whose level of complexity could occasionally hamper the accurate comprehension of concepts and, consequentially, compromise the semes' retrieval process. With a view to fostering the conceptual cognition with particular reference to specialised terminology involving advanced knowledge, the employment of intensional definitions could be proposed. Specifically, in the words of Roche (2012: 26), "*intensional definition* […] comprises the *superordinate concept* immediately above followed by one or several *delimiting characteristics*". In this sense, the inclusion of hypernyms could prove to be particularly profitable in order to inscribe the term in a wider conceptual sphere, therefore hierarchically and progressively moving from the superordinate conceptual particles in order to introduce and conceptually contextualise the semic ones that are related to the specific term itself. Considering the particular terminological context of the dissertation, the thoughtful selection of the kind of definition that could be proposed in order to delineate the lexicographic baseline for the performance of an objective semic analysis should however be implemented by the indication of reliable medical dictionaries. Specifically, in the circumstances in which semic analysis as elaborated on the basis of the proposed methodology will be instantiated, hence with reference to the English language, the sources from which definitions have been retrieved in order to implement semic elements are the Merriam-Webster Medical Dictionary[33] and the TheFreeDictionary's Medical Dictionary[34]. In details, exceptions will be represented by the semic analysis of the term *rosaceiform dermatitis* as, due to the impossibility up to date to retrieve a definition in the mentioned dictionaries, a different source that will be accordingly indicated will be employed in order to convey its conceptual content, and in the specifically indicated circumstances in which different references, such as semic elements extracted from the related section in the terminological records, will be

33 https://www.merriam-webster.com/medical
34 https://medical-dictionary.thefreedictionary.com/

considered. Generally, however, the employment of different lexicographic sources in order to emphasise the potential of relying on various definitions and elements for a more exhaustive recall of conceptual particles has proven to be valuable. Moreover, given the consideration of the Spanish and Italian languages in the context of the aforementioned investigation, an indication concerning eligible dictionaries with a view to determining concepts in the medical domain could be regarded as functional. In particular, for what concerns the Spanish language, reliable sources could be the Diccionario Médico from the Clínica Universidad de Navarra (CUN)[35] and the Diccionario de cáncer del NCI – Instituto Nacional del Cáncer[36]. With respect to the Italian language, instead, the Enciclopedia Salute from the Ministero della Salute[37] and the Dizionario di Medicina Treccani[38] could be for this purpose mentioned.

An exemplification of the potentiality of this theoretical proposal from an applicative perspective can be offered by taking into consideration in greater detail a previously mentioned term, *screening*, for which six different semes' combinations were identified. The consideration that emerged was that some semes were recurrently present in various formulations of semic analysis, that is to say /test/ and /testing/ as well as /examination/ and /disease/; furthermore, several semes indicated in different lexical ways the reference to population as the target of this examination process, such as /people/ /group/ /individuals/ and /population/. By way of resorting to a corpus-based analysis built on the recollection of various definitions, it could be possible to retrieve these conceptual elements, hence fostering the recall capability with respect to the possible formulation of the same conceptual particle, including semes whose occurrence in the considered semic analysis amounted to one case, such as /prevention/, /organ/, /tissue/ and /process/. Consequentially, the proposed semic analysis would encompass all these terms, establishing an increased conceptual specificity and a potential homogeneity of choices:

35 https://www.cun.es/diccionario-medico
36 https://www.cancer.gov/espanol/publicaciones/diccionario
37 http://www.salute.gov.it/portale/salute/p1_3.jsp?lingua=italiano&tema=Salute_A_Z
38 http://www.treccani.it/enciclopedia/elenco-opere/Dizionario_di_Medicina

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

As it can be observed, the number of semes included in the sememe has significantly expanded as compared with the semic analysis of the term which has been taken into account as its benchmark, whose maximum number of semes amounted to eight different ones. However, the minimisation from the perspective of the quantitative dimension related to semes is not accounted as the principal aim of semic analysis, as the qualitative aspect and the potentiality to accurately and unambiguously determine concepts with relation to terms emerges as the focus of the adoption of this technique. In this direction, according to Rastier (1985: 519) 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.

Furthermore, a potential way to favour the conceptual terminologically-oriented systematisation can be constituted by the consideration of the essential necessity to recur to *mesogeneric semes*. As matter of fact, this element could be regarded as particularly pertinent in order to enhance the contextualisation of a term in a given field, especially provided the aforementioned possibility for terms to be characterised, as Gotti (2008: 34) affirmed, by monoreferentiality only in a specific subject area. The individuation of the mesogeneric seme can therefore be regarded in this sense as a key strategy to specifically minimise terminological polysemy, thus conferring to semic analysis the potentiality to represent a significant technique that, by way of a segmentation of meaning or concept, still is expressive up to the extent that linguistic ambiguity could possibly be marginalised or, at the very least, partially reduced. For instance, the inclusion of /medicine/ in all semic analysis of medical terms could facilitate the delineation of the exact domain in which the terms and the related concepts are specifically contextually inscribed. In this sense, semic analysis of the term *screening* could be further improved:

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

Following the same line of argument, another opportunity that semic analysis can offer is to individuate instances of terminological plesionymy, described by Arnold (2015: 3) as "near-synonymy", the potentiality of which has also been emphasised by Manset (2019: 23).

Another strategy that could generally improve the potential of semic analysis to univocally designate a term in order to distinguish it from others is represented by the possibility to carry out a parallel terminological analysis of other terms that could present a conceptual proximity in the interrelated map of concepts pertaining to a determined semantic field. For instance, in order to comprehend the specificities of the concept that is terminologically represented by the term *rosaceiform dermatitis*, it could be useful to take into account the various existing types of *dermatitis* and, broadly, what *dermatitis* generally entails from a conceptual perspective, in order to precisely grasp the particularities of the specific concept with respect to other relatable or less specific ones. In this view, provided that the focus of semic analysis is in this case one term, it could be nonetheless useful to contextualise it from a more general taxonomical viewpoint also in order to individuate more accurately information related to generic semes, which are equally relevant to inscribe concepts in the semantic map of the domain.

Taking as an example the term *rosaceiform dermatitis*, this term constitutes a specific type of dermatitis and, consequently, it conceptually differs from *seborrhoeic dermatitis*, *atopic dermatitis* and *contact dermatitis* for the specificity and the terminological specification of the causative agent or circumstance of its appearance. With reference to the 2019/2020 dataset, these terms respectively feature a semic analysis:

- *rosaceiform dermatitis: /skin/ /inflammation/ /disorder/ /disease/ /medicine/ /follicle/*
- *atopic dermatitis: /chronic/ /skin/ /disorder/ /inflammation/ /disease/ /medicine/ /pruritus/*
- *contact dermatitis: /skin/ /inflammation/ /disorder/ /disease/ /medicine/ /allergic/ /hypersensitivity/*
- *seborrheic dermatitis: /chronic/ /skin/ /disorder/ /inflammation/ /disease/ /medicine/ /seborrhea/ /scalp/*

However, a corpus-based semic analysis can be performed with respect to these terms in order to comprehend and delineate in greater detail the specificity that these types of dermatitis entail:

- *rosaceiform dermatitis*: /medicine/ /skin/ /inflammation/ /disorder/ /disease/ /follicle/ /papules/ /pustules/ /calcineurin inhibitors/ /erupting/ /facial erythema/ /onset/[39]
- *atopic dermatitis*: /medicine/ /chronic/ /skin/ /disorder/ /inflammation/ /disease/ /pruritus/ /atopic eczema/ /dermatopathy/ /allergic rhinitis/ /stress/ /asthma/ /reactions/ /allergy/ /eczematous/ /condition/ /xerosis/ /hypersensitivity/ /atopy/ /itching/ /dermatitis/
- *contact dermatitis*: /medicine/ /skin/ /inflammation/ /disease/ /disorder/ /allergic/ /hypersensitivity/ /irritant/ /allergic/ /acute/ /chronic/ /contact/ /substance/ /irritating/ /substance/ /dermatitis/ /redness/ /itching/ /allergen/ /immune response/ /inflammatory rash/ /eczema/ /allergic reaction/ /dermatitis/
- *seborrheic dermatitis:* /medicine/ /chronic/ /skin/ /disorder/ /inflammation/ /disease/ /seborrhea/ /scalp/ /dermatitis/ /face/ /chest/ /itchy/ /inflammatory/ /scaly/ /hairline/ /macular/ /dandruff/ /eruption/ /lesions/ /sebaceous glands/ /oily/ /scale/

39 For the semic analysis of the term, in addition to the semes that featured in the aforementioned sememe of the term *rosaceiform dermatitis* retrieved from the collection of terminological records, a related definition included in the following source has been employed: Powell F. C., *Rosacea: Diagnosis and Management*, New York: CRC Press, 2008, p.73.

As it can be noticed, although the number of semes has significantly increased, the possibility to comprehensively retrieve more conceptual elements has been proportionally heightened as well as the accuracy with regard to their related symptomatologies as well as the causes of the different conditions.

This methodological proposal, however, would not prove to be sufficient in order to distinguish between specialised terms and popular ones, therefore a notional integration would be needed. In particular, Rastier (2005) provided an insight with reference to the difference of terminological communication that is established at the intra-specialist level as opposed to the doctor-patient's one. By way of exemplification, when analysing medical discourse, it would seem necessary to define 'patient' and 'sick person', in order to prevent a synonymic relationship between these sememes. (Rastier 2005) As a matter of fact, they are not found in the same contexts: the first occurs in the words employed by doctors towards their assisted patients; the second, in the words that doctors exchange with each other. (Rastier 2005)

Provided this communicational standpoint, a second main potential improvement that could be proposed concerns the relational application of the principle of termhood as combined with the consideration of the diastratic as well as the diaphasic perspective in the context of semic analysis. As a matter of fact, as it emerges from a quote by Gadet (2007: 24), diastratic and diaphasic variations can be regarded as interrelated. It would be expected that the division of variation into different types would cause a discontinuity, nevertheless, diatopic, diastratic and diaphasic variations constantly interact: the lower the socio-cultural status of speakers is and the more familiar the situation is, the more they will rely on regional forms. (Gadet 2007: 24) As a consequence, the diastratic spectrum is wider at the lower stratum of the social scale. (Gadet 2007: 24) In this view, the interconnection at an interoperable level of the notions of termhood, diastratic and diaphasic variation could provide an improved level of specificity to the conceptual designation of terminology, specifically capturing the domain-contextualised informational content and, simultaneously, allowing for an enhanced hence less generic distinction between specialised medical terms and popular

ones. In this respect, semes could be thought as conveying specialised terminological characterisation with reference to specialised terms, while featuring non-specialised terms in the case of semic analysis of popular terms. A possible contrastive argumentation that could be however proposed is that the hypothesised corpus-based approach as a productive methodological procedure to confine subjectivity in the performance of semic analysis could seem to be at odds with the principle of semic analysis itself considering that, according to Baider and Constantinou (2014: 91), [f]rom the perspective of semic analysis, semes are not uniquely inscribed in a lexicographic definition, but are defined according to the contexts and the discourses in which they appear. Nonetheless, it could also be argued that definitions seem to represent a potential relevant source of information and semantic knowledge from a methodological and a functional perspective with respect to terms also in this specific setting, especially given that the specific focus on the diastratic dimension could constitute a way to contextualise terms, by way of distinguishing between specialised terms and popular ones.

The practical application of this principle could be exemplified by way of resorting to the semic analysis of the term *measles*, that has been retrieved from the 2019/2020 reference dataset:

> */disease/ /virus/ /contagious/ /children/ /fever/ /mucus/ /nose/ /conjunctivitis/ /red spot/*

In this case, terms such as *fever* and *red spot* could be considered as popular terms, while the following semic analysis can be elaborated in order to convey knowledge by way of resorting to specialised terminology that could be more commonly employed in intra-specialist communication, such as *pyrexia* and *exanthem*:

> /medicine/ /disease/ /virus/ /contagious/ /children/ /pyrexia/ /mucus/ /nose/ /conjunctivitis/ /exanthem/

The composition of the sememe, however, could be improved in its conceptual specificity by way of the application of the first criteria that has been indicated, that is to say a systematic corpus-based semes' recollection. As a result, the following semic analysis of the term could be proposed:

/medicine/ /disease/ /virus/ /contagious/ /children/ /pyrexia/ /mucus/ /nose/ /conjunctivitis/ /exanthem/ /acute/ /rubeola/ /eruption/ /skin/ /rush/ /vaccination/ /5-day measles/ /hard measles/

Following the same line of argument, the terminological specialisation that is involved in the performance of semic analysis could heavily be interrelated with the selected definitions of the specific term. To provide an example, the 2019/2020 dataset features the semic analysis of the term *asthma*:

/chronic/ /disorder/ /coughing/ /wheezing/ /difficulty/ /breathing/ /sense/ /constriction/ /chest/ /respiratory illness/

The selection of semes has been based on the terminology that was present in the definition that has been included: "*A generally chronic disorder characterized by wheezing, coughing, difficulty in breathing, wheezing and a sense of constriction in the chest*"[40]. In this respect, the selection of definitions that feature specialised terminology in the conceptual delineation of the term could exert a major impact on semic analysis from a diastratic and a diaphasic perspective. To highlight this circumstance, the following corpus-based semic analysis performed on the basis of definitions that mostly comprised specialised terms can be proposed:

/medicine/ /condition/ /dyspnea/ /airway/ /inflammation/ /wheezing/ /spasmodic/ /constriction/ /bronchi/ /reversible/ /obstruction/ /syndrome/ /chronic/ /tracheobronchial tree/ /bronchospasm/ /hyperreactivity/ /lung disorder/

40 The definition was taken from https://www.collinsdictionary.com/it/dizionario/inglese/asthma

From a terminological as well as a conceptual viewpoint, it particularly emerges that the terms that constitute the semes in the first compositional structure of the sememe of the term *asthma* could be potentially employed in doctor-patient communication in order to explicate the specific manifestations of the disease and, generally, to refer to the concept by way of resorting to popular terms. By way of adopting this lexical choice, understanding would not be hampered, as the patient could easily access the explanation provided by the physician without detecting particular difficulties in the comprehension process. On the contrary, the employment of terminology such as *dyspnea*, *obstruction* and *tracheobronchial tree* could represent a terminological and conceptual obstacle as, for instance, specialised knowledge could be needed in order to precisely cognitively determine the concepts that are related to these linguistic entities, and to reconnect them to the specific term.

Following this line of argument, the distinction between the specialised and the popular usage of the same term can be noticeable when comparing the semic analysis provided in the dataset of the term *malformation* with an additional one that can be proposed which features specialised and domain-targeted terminology:

- */condition/ /shape/ /living being/*
- /medicine/ /structural/ /defect/ /deformity/ /development/ /localised/ /morphogenesis/ /embryonic/ /interaction/ /anomalous/ /abnormal/ /formation/ /structure/

An issue that could however be considered as left open is the expression of connotative traits and the way in which semic analysis could potentially be objectively performed also with reference to this particular aspect of meaning. As a matter of fact, as previously considered, connotation is inherently linked to the personal cognitive sphere, that is to say the perception or the feeling that a person expresses towards a particular term and the related concept. In this respect, a potential argument that could be mentioned is the fact that some traits of meaning that can be ascribed to a term could

be commonly shared and conceived in the social setting. Nonetheless, symptomatology or the idea attached to a particular term could be experienced differently, which reflects on the lexical as well as the conceptual references that people employ in order to define the connotative meaning. As a consequence, the expression of these semantic elements of meaning could explicitly lexically manifest in a wider range of possibilities, thus significantly complicating in this respect the intrinsic accomplishment of an objective performance of semic analysis.

To sum up, it could be argued that semic analysis constitutes a significant technique that could perform a pivotal function with relation to terminology and terminological research in order to gain an in-depth knowledge of terms and, more specifically, of the conceptual contents that represent the baseline for the general understanding of the semantic map related to a special language. In particular, semic analysis emerges as profitable in order to address the diastratic dimension, with a view to linguistically formulate and determine it from a terminological conceptual perspective. As a matter of fact, its employment with reference to both specialised terms and popular ones can serve as an expressive tool to the extent that the communicative setting is precisely instantiated, therefore demarcating the semantic and pragmatic level of contextual adaptability that could be in this connection ascribed to medical terminology. In this respect, the usage of this technique to pinpoint this dimension pertaining to language could lead to the adoption of a particularly detailed perspective on the comprehension of the different conceptual shifts and terminological specificity of the compositional elements that distinguish the two main variations in the specialisation of medical special language.

Precisely, in this direction, semic analysis could especially be conceived as a fundamental strategy within the terminological field in order to effectively comprehend and investigate terminology. Recalling the phases of its performance, the consideration of its methodological procedural application as conceived in the decomposition of unitary concepts could prove to be useful because of the multifaceted cognitive elaborations that it implies, and it could further implement the functionalities offered by

definitions in conveying semantic and conceptual contents. Indeed, definitions convey by the same means references to specific semantic fields as well as the identification of meanings and concepts; however, semic analysis could be perceived as a method to potentially enhance their elaboration and the related acquisition of information. Furthermore, in this connection, the cognitive elaboration of conceptual specificities obtained through the performance of semic analysis could prove to be for instance essential at an interoperable level with particular reference to specialised translation. As a matter of fact, the selection of an adequate term to be employed in a determined target text could be for instance guided by way of relying on a previous comparative contemplation of the sememes respectively associated to each different term candidate in order to discern conceptual nuances and, consequentially, carry out a precise and accurate translation. In this connection, the performance of semic analysis could prove to be useful in the context of its application in a comparative perspective also at an interlinguistic level with respect to the individuation of potential conceptual differences or actual similarities in the circumstance in which terms in different languages supposedly convey the same concept.

In the light of the foregoing theoretical notions and reflections, the proposed simultaneous corpus-based hence data-driven approach combined with the reference to the principle of termhood and the consideration of both the diastratic and the diaphasic dimensions could further intensify the productivity of semic analysis in the context of the investigation of the conceptual constitutive structure of medical terminology, by way of highlighting specific linguistic necessities and terminological particularities that this special language manifests.

Subsequent to the postulation of a proposal of an objective methodology for the performance of semic analysis, the contribution of the following chapter will be oriented to the presentation of an additional study on the productivity of word embeddings with relation to this linguistic approach from a specific terminological-oriented perspective by way of employing as well as further exemplify the application of the proposed theoretical criteria. Precisely, the potentiality for data obtained through

the technique of word embeddings to constitute a relevant basis for the semes' retrieval procedure will be explored and, in this connection, a pivotal underlying implication whose effective realisation will be challenged is the possibility for an automatic performance of semic analysis of medical terminology. As a matter of fact, the ascertainment of the profitable applicability of this data-driven prospect in the context of this linguistic technique would represent a valuable quality to be ascribed to semic analysis' procedural dynamics. In this respect, the proposed research would once again provide for an in-depth investigation concerning a specific objective of the linguistic technique, that is to say the exhaustive recollection of the elements constituting concepts by way of semes. With a view to questioning the subsistence of such a possible synergy between the two approaches to the study of linguistic entities, the notional focus will be consequentially placed not only on the expression of the conceptual content of reference terms conceived as isolated linguistic elements but also on the consideration of the contextual environment in which terminology is inscribed, that is to say with reference to terms that manifest semantic proximity with respect to a determined reference term. Previous to the presentation of the case study, which will be articulated in a systematic analysis consisting in the adoption of different perspectives to the investigation of the topic, a brief notional introduction to word embeddings correlated to a concise overview on two distinct approaches implicated in its application and an overview on the distributional hypothesis will be provided.

CHAPTER 4 Semic analysis and word embeddings: an investigation on a potential interrelation

4.1 A notional introduction to word embeddings

Word embeddings are widely employed in the framework of Natural Language Processing and can be defined as the representation of words in the form of vectors which incorporate information related both to syntax and semantics whose nature is numerical (Sfakakis, Papachristopoulos, Zoutsou, Tsakonas and Papatheodorou 2019: 107). By the same token, the textual dimension is also supposed to feature a different type of representation which implies a numerical vector mapping rather than a lexical configuration, as it is instantiated in the vector space. In particular, the conceptualisation that is at the root of the core notional principle of word embeddings is that the context in which a word is embedded determines the meaning of the linguistic entity itself (White, Togneri, Liu and Bennamoun 2019: 83). The inference of the kind of relatedness that subsists between distributions as manifested in the vector space and meaning is however defined in the context of the distributional hypothesis, as Lenci (2009) highlights: [i]n itself, the vector space uniquely registers positions dominated by the syntagmatic relations that words manifest in the linguistic contexts. The cosine (or the Euclidean distance) measures the similarity of the distributional schemes of words: two words that are proximal in the vector space are simply words that feature similar semantic distributions in the linguistic contexts. (Lenci 2009). It is the distributional hypothesis that realises the analogy between the similarity of the distributions and the similarity at the meaning level. (Lenci 2009).

An exemplification that could be for instance mentioned is the one proposed by Harris (1954: 157) with reference to *oculist* and *eye-doctor* which, due to the fact that they occur in the same contexts unless for the most part in occasions in which they are both employed in the framework of the same sentences, are synonymous. In this connection, the ascription of an influential faculty to the contextual dimension in order

to define the meaning of a word was also theorised by John Firth (1957: 7), whose essential deduction is that "the complete meaning of a word is always contextual" (Skelac and Jandrić 2020: 42). In particular, the distributional hypothesis is to be considered as the notional groundwork for two methods, Word2Vec and Glove (Yang, Luo, Chueng, Ling and Chin 2019: 91), which can be defined as "the two most popular semantic embeddings having the capability to capture the semantics of the language" (Kabir, Alam and Islam 2020: 307).

Precisely, Word2Vec represents a renowned approach that relies on neural networks with a view to learning word embedding by way of transforming words into vectors whose similarity at the semantic level should translate into their proximity in an N-dimensional space of vectors, hence implying a similarity that manifests both in terms of meaning and contexts (Saidani, Adi and Allili 2020: 75). In particular, it relies on large datasets in order to apprehend semantic and syntactic relationships between words by way of resorting to an approach that is based on probabilistic prediction (Ashi, Siddiqui and Nadeem 2019: 246). Word2Vec, developed by Mikolov et al. (Mikolov, Chen, Corrado and Dean 2013), implements two neural models that are characterised by different architectures, namely the Continuous Bag-of-Words (CBOW) and Skip-Gram. Specifically, while the Continuous Bag-of-Words relies on the words that are present in the context to predict the target word, the Skip-Gram conversely predicts the contextual words that surround the target one from the word itself (Choudhari and Veenadhari 2020: 72).

For what concerns the GloVe model (Pennington, Socher and Manning 2014), the subsistence of the co-occurrence of terms is considered, hence word embeddings are learned from the related matrix implying an activity that does not involve prediction (Khattak, Jeblee, Pou-Prom, Abdalla, Meaney, Rudzicz 2019: 2). From a comparative perspective, specifically,

> Word vectors estimated using GloVe are conceptually similar to those derived from word2vec but uses an underlying count-based model, rather than word2vec's prediction-based model. Because GloVe typically computes statistics over larger context windows than word2vec, it permits

> capturing longer-term dependencies, although the *order* of those dependencies will be lost. (Khattak, Jeblee, Pou-Prom, Abdalla, Meaney, Rudzicz 2019: 2)

Recalling the essential conception that is at the basis of the aforementioned distributional approach,

> Distributional methods take literally the principle that words are defined by the contexts in which they appear, which lead to representations that are quite different from those of existing human dictionaries. (Col, De Angelis and Poibeau 2017: 44)

Taking into account the fact that one of the proposed criteria for the theorisation of an objective methodology of semic analysis involves the employment of dictionaries as sources for conceptual knowledge, the introduction of such a diverse approach could be thought as a discrepant, even dichotomous conception. On the contrary, the adoption of both perspectives should be hypothesised and therefore investigated in order to challenge a possible juxtaposition, respectively concretised in the postulation of an objective methodology and the possibility to evaluate by way of its application the opportunity to automatically perform it.

4.2 Semic analysis and word embeddings: a case study

As previously mentioned, the investigation of the subsistence of a potential for the terminological outputs generated by way of the application of the technique of word embeddings to represent or to concur to the composition of elements that constitute the conceptual content conveyed by sememes will be in this section proposed. In particular, studies on the interrelation between semic analysis and word embeddings were formerly carried out in order to investigate sememe prediction, as proposed by Xie, Yuan, Liu and Sun (2017)[41]. Moreover, the topic was also examined from a cross-linguistic perspective, as in the study presented by Qi, Lin, Sun, Zhu, Xie and Liu (2018)[42]. In the framework of this dissertation, however, a different approach based on a dual analysis featuring the application of the proposed methodology for an objective semic analysis of

41 https://www.researchgate.net/publication/318829690_Lexical_Sememe_Prediction_via_Word_Embeddings_and_Matrix_Factorization

42 https://www.aclweb.org/anthology/D18-1033.pdf

medical terminology will be presented. Specifically, the dataset will be constituted by two different sources of data, that is to say the semic analysis' formulations of five medical terms, four of them antecedently proposed in the third chapter, and a set of lists featuring 50 terminological elements each stemming from the application of the technique of word embeddings[43] with respect to the same medical terms. In particular, the semic analysis of the terms *screening*, *measles*, *asthma*, *malformation* and *dermatitis* will be considered and, by the same token, data regarding word embeddings will be analysed with respect to the same reference terminology. In details, word embeddings were trained on the basis of two different models, namely the GloVe model glove-wiki-gigaword-300 based on Wikipedia[44] and Word2Vec trained on two different datasets: the Google News dataset[45] and PubMed[46]. Taking as a reference the resulting data, the first step of the investigation will be based on a comparison between the semes included in the sememes and the totality of the semantically-proximal terminology retrieved through word embeddings with relation to each reference term. This process will be specifically aimed at the assessment of terminological as well as conceptual correspondences in order to challenge the recall capability of semic analysis with respect to the computationally obtained data and, vice versa, the potential that word embeddings might possess to capture semic elements. At a later stage, a second type of systematic analysis will be carried out. Subsequent to the determination of a circumscribed number of terms that are comprised in the lists whose semantic proximity will be effectively investigated, individuated in a range of 10 terms for each list, a semic analysis of the selected terminological entities will be carried out following the proposed methodological criteria for its objective performance. As a matter of fact, this procedure will prove to be essential in order to detect the concepts that the various terms represent as well as the eventual conceptual connections that these terms manifest with respect to the reference medical terminology and, consequentially, to evaluate their potential to constitute semic elements which could be considered as effectively suitable in order to define the conceptual content of the reference terms. In addition, considering

43 As for the collections of terminological records, the Excel file containing the mentioned data concerning word embeddings was kindly provided by Prof. Giorgio Maria Di Nunzio.
44 https://nlp.stanford.edu/projects/glove
45 https://code.google.com/archive/p/word2vec/
46 https://pubmed.ncbi.nlm.nih.gov

the provided sets of semic analysis of related terms, a further step of the study will be specifically aimed at the investigation of the possibility to identify a superset of semes conceived as a group of semic elements originally present in the sememes of the reference terms which would hence partially convey their conceptual content. In particular, the overturned perspective that will be proposed in the framework of this step with respect to the previous one was envisaged as functional with a view to further questioning the potentiality of word embeddings to capture semantic elements within the context of semic analysis and, in a broader sense, to efficaciously individuate semantic features both in terms of similarities and relatedness.

4.2.1 A comparative analysis of sememes and word embeddings

As previously delineated, the first phase of the present case study concerns the investigation on the potentiality for the detection of a point of intersection between semic analysis and word embeddings from a comparative perspective. At this stage, indeed, the main point at issue is to question the subsistence of a terminological hence conceptual equivalence between the elements included in the sememes' composition of each reference term and the related terminological entities that were retrieved by way of word embeddings. In order to facilitate the perception of the potential recall capability that semic analysis performs with respect to word embeddings' data and vice versa, the semic analysis of the reference terms as formulated in the third chapter will be re-proposed, followed by considerations on the individuated correspondences between the semes and the terminological units featured in the three different aforementioned lists. In particular, the potentiality of semic analysis to retrieve terminological and conceptual particles that are also featured as semantically-proximal terms will additionally be tested by way of resorting to indicative numerical values. As a matter of fact, the 50 listed terms that were comprised in each set were ranked in descending order from the most proximal term with respect to the reference one to the less semantically-related entity and correlated by a numerical indication which quantitatively represents vectors' proximity. Specifically, the range of the values could be comprised between 1 and 1, respectively signalling a perfect similarity and the opposite condition. Within the

context of the provided data, however, the 50 most proximal vectors to the ones of the reference terms were considered. With reference to the multi-word terms composed of two lexical entities that were comprised in the lists, specifically, the character underscore which was added by the software in order to create bigrams has been omitted; nevertheless, in the circumstances in which multi-word terms are mentioned, a specification will be included.

The first reference term with respect to which the comparative investigation will be performed is *screening*:

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

The following table includes the aforementioned 50 listed terms that were retrieved by way of the application of word embeddings which are hence considered as semantically-proximal to the reference one with respect to the GloVe and Word2Vec models trained on the indicated datasets. The same table will be presented for all the analysed terms in order to provide a reference to the subsequently mentioned terminological entities.

GloVe (Wikipedia)	Word2Vec (Google News)	Word2Vec (PubMed)
screenings	Screenings	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_screening	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

Taking as a reference the results obtained through the application of the GloVe model, by way of comparing the semes with the terms comprised in the related list it could be noticed that the seme *testing* could be found as the 5th term, the seme *examination* as the

13th term, the seme *exam* as the 35th, the seme *test* as the 45th and the seme *detect* as the 47th. A consideration that could stem from a contrastive observation of this reduced set of semes with respect to the other listed terminological entities is that the terms could also be found in some cases in their inflected forms. By way of exemplification, the term *tests* was instantiated in the 7th position, while the plural form of *examination* in the 16th. With reference to the Word2Vec model trained on the Google News dataset, it could be observed the absence of correspondence with the semic elements composing the semic analysis. Considering the list of terms originating from the training of Word2Vec on PubMed, instead, the seme *testing* constituted the 2nd result. Also in this circumstance, terms featuring inflected forms with respect to several semes could be spotted, such as the presence of the terms *exams*, *detecting* and *identifying*.

The same process could be applied with respect to the term *measles* in order to detect potential correspondences by taking as a reference the following semic analysis:

/medicine/ /disease/ /virus/ /contagious/ /children/ /pyrexia/ /mucus/ /nose/ /conjunctivitis/ /exanthem/ /acute/ /rubeola/ /eruption/ /skin/ /rush/ /vaccination/ /5-day measles/ /hard measles/

GloVe (Wikipedia)	Word2Vec (Google News)	Word2Vec (PubMed)
polio	whooping_cough	Mumps
vaccination	chickenpox	rubella
tetanus	pertussis	rabies
diphtheria	polio	varicella
immunization	Measles	polio
malaria	meningococcal_disease	influenza
vaccinations	meningitis	rotavirus
mumps	mumps	smallpox
rubella	vaccine_preventable_diseases	dengue
typhoid	measles_outbreak	vaccination
vaccinated	diphtheria	poliovirus
smallpox	paralytic_polio	poliomyelitis
rabies	rubella	chikungunya
meningitis	bacterial_meningitis	vaccine
cholera	influenza	vaccinations

hepatitis	malaria	laboratory-confirmed
poliomyelitis	measles_vaccination	vaccinees
vaccines	pneumococcal_meningitis	pandemics
tuberculosis	H#N#	encephalitis
influenza	pertussis_whooping_cough	vaccinia
chickenpox	Hib_disease	virus
vaccine	measles_outbreaks	flavivirus
epidemics	chicken_pox	parainfluenza
pertussis	vaccination	vaccines
dysentery	swine_flu	immunisation
epidemic	measles_vaccinations	epidemics
immunisation	tuberculosis	vaccinated
diarrhea	measles_epidemic	reassortant
diarrhoea	mosquito_borne_illness	leptospirosis
dengue	meningococcal	pneumococcal
immunized	Hib_meningitis	meningococcal
malnutrition	mosquito_borne_disease	tetanus
inoculation	polio_virus	flu
immunizations	immunization	enterovirus
outbreaks	Whooping_Cough	tick-borne
encephalitis	Hib	unvaccinated
pneumonia	flu	pox
infections	whooping_cough_pertussis	outbreaks
flu	Whooping_cough	seroconversion
whooping	meningococcal_meningitis	norovirus
pox	typhoid_fever	hantavirus
diseases	vaccinated	vaccinating
h1n1	rotavirus_diarrhea	flaviviruses
preventable	poliomyelitis	parvovirus
vaccinate	tuberculosis_TB	outbreak
pandemic	cholera	immunization
inoculations	Rotavirus	seroconverted
avian	chikungunya_fever	poxvirus
outbreak	rotavirus	booster
infection	swine_flue	tularemia

From the comparative analysis which was performed with respect to the GloVe model, it emerged that only the seme *vaccination* was featured, precisely as the 2nd term. Moreover, the seme *disease* could be retrieved in its inflected plural form as the 42nd terminological entity. With respect to Word2Vec based on the Google News dataset, *vaccination* occupied the 24th position. As for the same model trained on PubMed, the same seme could be detected as the 10th result, while its inflected form as the 15th.

The comparative approach can also be proposed with reference to the term *asthma*, whose sememe is composed by the following semes:

/medicine/ /condition/ /dyspnea/ /airway/ /inflammation/ /wheezing/ /spasmodic/ /constriction/ /bronchi/ /reversible/ /obstruction/ /syndrome/ /chronic/ /tracheobronchial tree/ /bronchospasm/ /hyperreactivity/ /lung disorder/

GloVe (Wikipedia)	Word2Vec (Google News)	Word2Vec (PubMed)
Allergies	Asthmatic	asthmatic
diabetes	Asthma	rhinitis
allergy	allergies	allergic
allergic	asthmatics	asthmatics
hypertension	allergic_asthma	atopy
arthritis	COPD	wheezing
emphysema	allergy	exacerbations
obesity	asthmatic_symptoms	wheeze
bronchitis	asthma_sufferers	eczema
sufferers	eczema	atopic
respiratory	induced_asthma	exacerbation
chronic	asthma_bronchitis	hyperresponsiveness
migraine	inhaled_steroids	allergies
eczema	hay_fever	bronchitis
medication	hypertension	hyperreactivity
asthmatic	respiratory_symptoms	allergy
medications	lung_disease	asthma-related
aliments	lung_diseases	rhinosinusitis
diseases	asthma_flare_ups	bronchospasm
epilepsy	asthma_exacerbations	urticaria
rheumatoid	asthma_eczema	airway
symptoms	recurrent_wheezing	bronchiectasis
illnesses	airway_constriction	omalizumab
migraines	eczema_hay_fever	bronchoconstriction
disorders	angina	dermatitis
schizophrenia	allergic_diseases	moderate-to-severe
fibrosis	aggravate_asthma	cough
cardiovascular	chronic_lung	colds
osteoporosis	respiratory	airways
inflammation	persistent_wheezing	bronchial
wheezing	asthmas	bronchodilator
lung	lung_function	eosinophilia
pneumonia	wheezing	anaphylaxis
inhaler	atopic_eczema	psoriasis

disease	allergy_symptoms	bronchiolitis
psoriasis	bronchitis_asthma	bronchodilators
acne	asthma_hay_fever	inhaled
bronchial	seasonal_allergies	conjunctivitis
alzheimer	asthma_exacerbation	emphysema
cancers	sinusitis	sensitisation
cough	migraine	illnesses
tuberculosis	diabetes	allergen
insomnia	bronchiolitis	montelukast
pulmonary	migraines	gout
gout	chronic_bronchitis	airflow
diabetic	steroid_inhaler	alveolitis
congestive	exacerbations	silicosis
cystic	bronchial_asthma	rheumatic
cancer	asthma_chronic_bronchitis	moderate-severe
sinusitis	respiratory_illness	mild-to-moderate

Referring to the outcome of the application of the GloVe model, correspondences could be observed with reference to the semes *chronic*, which constituted the 12th listed term, *inflammation*, which represented the 30th result and *wheezing*, which featured at position 31. Considering the terminological entities originating from Word2Vec on Google News dataset, the presence of the seme *wheezing* could be spotted as the 33rd result, whereas in the case of Word2Vec on PubMed the following semes could be retrieved: *wheezing* as the 6th listed term, *hyperreactivity* as the 15th, *bronchospasm* as the 19th and *airway* as the 21st.

The elements comprised in the sememe of *malformation* were also susceptible of the same type of comparison with respect to the related terminological entities stemming from the application of word embeddings. For this purpose, the following semes were therefore considered:

/medicine/ /structural/ /defect/ /deformity/ /development/ /localised/ /morphogenesis/ /embryonic/ /interaction/ /anomalous/ /abnormal/ /formation/ /structure/

GloVe (Wikipedia)	Word2Vec (Google News)	Word2Vec (PubMed)
congenital	malformations	malformations
malformations	atresia	anomaly
abnormality	congenital_disorder	anomalies
arteriovenous	hypoplasia	hypoplasia
foetus	congenital_anomaly	agenesis
deformity	brain_malformation	congenital
fetal	diaphragmatic_hernia	craniosynostosis
abnormalities	hemangioma	hamartoma
malformed	coarctation	fistulae
stenosis	pyloric_stenosis	fistula
stillbirth	abnormalities	polydactyly
hydrocephalus	deformity	hemangioma
miscarriage	hydrops	ventriculomegaly
uterine	birth_defect	dural
dysplasia	hypoplastic	microcephaly
atrophy	Joubert_syndrome	fistulas
hypothyroidism	vascular_malformation	atresia
lesion	cystic_hygroma	hydrocephalus
deformities	abnormality	arteriovenous
intracranial	ventricular_septal_defect	hypoplastic
fetus	holoprosencephaly	cyst
lymphatic	congenital_malformation	cysts
foetal	myelomeningocele	malrotation
endometrial	pulmonary_stenosis	shunts
hypoglycemia	brachial_plexus_palsy	infantile
mucinous	agenesis	dermoid
arrhythmia	cysts	abnormality
life-threatening	fibroma	nonsyndromic
chiari	genetic_abnormality	hemangiomas
cardiomyopathy	neural_tube_defect	dwarfism
calcification	congenital_abnormalities	synostosis
hyperplasia	congenital_anomalies	clefts
atresia	hydronephrosis	obliteration
hypoplasia	osteopetrosis	laceration
insufficiency	Wilms_tumor	cavernous
lesions	congenital_abnormality	lipoma
aneurysm	developmental_abnormalities	bifid
ischemia	genetic_defect	chylothorax
adrenal	Apert_syndrome	dysplasias
infarction	esophageal_atresia	malformed
ulceration	congenital_disorders	dysplasia
chromosomal	anencephaly	non-syndromic
syringomyelia	congenital_malformations	lipomas

hemangioma	Spina_bifida	stenosis
iatrogenic	congenital_diaphragmatic_hernia	fenestration
defects	deformities	meningioma
atherosclerosis	volvulus	facies
hypoxia	teratoma	arachnoid
hemorrhage	craniosynostostosis	aplasia
ectopic	congenital_deformity	dysgenesis

Taking as a reference the terminological entities indicated as proximal in the context of the application of the GloVe model, the seme *deformity* could be found as the 6th listed term. Also in this circumstance, the presence of an inflected form of a seme could be observed, specifically the term *deformities* which occupied the 19th position. Moreover, the seme *defect* was instantiated in its plural form, that is to say *defects*, which represented the 46th result. In the framework of the application of Word2Vec trained on the Google News dataset, the same seme constituted the 12th result.

The concluding comparative analysis will be carried out by way of considering the sememe of the single term *dermatitis*. An alteration with respect to the proposed reiterated analysis will be however represented by the consideration of results exclusively deriving from the adoption of one model, Word2Vec.

/medicine/ /inflammation/ /skin/ /inflammatory/ /rash/ /itching/ /redness/ /disorders/ /blister/ /formation/ /swelling/ /crusting/ /weeping/

Word2Vec (Google News)	Word2Vec (PubMed)
contact_dermatitis	eczema
eczema	atopic
skin_irritations	urticaria
allergic_dermatitis	allergy
skin_irritation	rhinitis
pruritic	allergic
folliculitis	rosacea
atopic_dermatitis	conjunctivitis
allergic_reactions	psoriasis
atopic_eczema	eruptions

rashes	irritant
vaginal_yeast_infections	allergies
skin_rashes	blistering
lichen_planus	scabies
cystic_acne	rashes
itchy_scalp	pustular
otitis_externa	pruritic
melasma	atopy
pyoderma	anaphylaxis
nail_fungus	rash
mouth_ulcers	pruritus
papular	vitiligo
dermatosis	eosinophilia
Demodex	photosensitivity
keratitis	asthma
Allergic_reactions	sensitisation
flea_bites	erythema
Urticaria	hypersensitivity
diaper_dermatitis	mucocutaneous
vaginitis	blisters
cutaneous_lupus	acne
latex_allergies	mastocytosis
cystitis	hyperpigmentation
asthma_rhinitis	bullous
Atopic_dermatitis	maculopapular
allergies	skin
canker_sores	depigmentation
sunburn	allergens
seborrhoeic_dermatitis	angioedema
itchy_skin	itching
Keratosis	nodosa
itchy_rash	alopecia
seborrheic_dermatitis	cutaneous
itchy_rashes	tinea
scalp_psoriasis	irritation
keratosis	itch
Dermatitis	dermatological
poison_ivy_rash	burns
skin_lesions	diathesis
Seborrheic_dermatitis	allergen

Specifically, an absence of correspondences could be detected with regard to the list related to the application of Word2Vec trained on the Google News dataset. On the

contrary, linguistic entities which constitute semic elements could be found with reference to the same model trained on PubMed, such as the term *skin* as the 36th element and *itching* as the 40th.

Comprehensively, from the consideration of the presented analysis, further observations could be proposed with respect to the comparison between semes comprised in the sememes representing the conceptual dimension of reference terms and the indicated semantically-proximal terms. As a matter of fact, it could be observed that the presence of inflected forms of different lexical units representing semes is not the only phenomenon that could be detected. Indeed, an interesting finding was represented by the fact that, in two circumstances, the terminological entities comprised in a list obtained by way of the technique of word embeddings were multi-word terms whose constituents could be reconnected to the terminological aggregation of different terms constituting semes at a lexical level. By way of exemplification, the sememe of the term *asthma* comprised two distinguished semes, that is to say /airway/ and /constriction/, which in the list obtained by the application of the Word2Vec method trained on the Google News dataset featured in form of linguistic components of the multi-word term *airway constriction*. In addition, the terminological elements which were conceived as semes in the sememe, *chronic* and *lung*, were also conflated in the multi-word term *chronic lung*. Furthermore it could be noticed that, in several cases, a single lexical unit included in the semic analysis for its conceptual value constituted a terminological element of a semantically-related multi-word term. For instance, considering the seme /defect/ which was included in the sememe of *malformation*, its presence as a linguistic entity could be spotted in different multi-word terms such as *ventricular septal defect*, *birth defect* and *neural tube defect*, each indicating a specific type of defect. In this connection, a further example can be proposed with reference to the seme /skin/ which concurred to represent the conceptual content of *dermatitis*. Considering the reference term *dermatitis*, the terminological entity *skin* also featured as a component of different multi-word terms in the list obtained by way of Word2Vec on the Google News dataset, such as in the following instances: *skin irritation*, *skin irritations*, *skin rashes*, *itchy skin* and *skin lesions*. Moreover, a third phenomenon could

be identified with respect to the morphological dissimilarities which, in some instances, constituted the linguistic and at a certain extent the conceptual differential traits between the semes and the terms comprised in the lists. An exemplification of this circumstance can be proposed by taking as a reference the seme /abnormal/, employed among other semic entities to conceptually define the term *deformation*. As a matter of fact, the examination of the related lists obtained through the application of the models GloVe and Word2Vec on the Google News dataset highlighted the presence of both *abnormality* and *abnormalities* as outcomes. A second instance that can be mentioned concerns the seme /detect/ in the sememe of *screening*, with respect to which a semantically-related term that was actually instantiated was *detection*.

In this respect, two essential assumptions could be inferred by the consideration of the outcomes of the present comparative analysis. In the first place, a noticeable circumstance is represented by the ascertainment that semic analysis can be effectively confirmed as a valuable technique whose conceptual retrieval capability additionally succeeds in incorporating entities that are considered as contextually semantically-related, thus substantiating in this sense a possible interrelation between semic analysis and word embeddings. Conversely, however, it is also possible to consider the issue from an overturned perspective, that is to say by way of highlighting the potentiality of word embeddings to capture some semic elements. On the basis of the mentioned observations, a hypothesised possibility to selectively conceive terminological entities originating from the application of word embeddings as semes in the formulation of the sememes for the accomplishment and the improvement in the performance of semic analysis could be regarded as conceivable. Precisely, in this connection, the semes manifesting morphological relationships or inflections with respect to the ones instantiated in the sememes could be regarded as suitable for the conceptual representation, due to the noticeable semantic proximity that can be ascribed to them in relation to the respective reference terms.

4.2.2 Semic analysis of word embeddings

In order to discuss and explore in greater detail the potential interrelation at the performative level between semic analysis and word embeddings, a second procedural step in the framework of the present case study was envisaged. In particular, the conceptual substrate of the terminology that was retrieved by way of the technique of word embeddings will constitute the focus of the analysis, therefore implying a perspective shift from a comparative approach between semic elements and these terminological entities to the selective investigation of the conceptual specificities that can be attributed to the latter. Moreover, the possibility to retrieve a superset of semes by way of observing both semes contemplated in the sememes of reference terms and related ones will be investigated. As a matter of fact, the finding of the analysis and the knowledge thus gained will contribute to an in-depth as well as reasoned interpretation of the kind of interrelation which can be potentially ascribed to the two techniques. In this connection, the possibility to perform an automatic performance of semic analysis of medical terminology will be questioned and problematised.

For what concerns the sample of the study, the analysis will focus on a subset of data individuated in the number of ten terminological entities for each of the lists obtained from the application of the two models of word embeddings on the three aforementioned different datasets. In particular, the criterion which was exploited in order to elect them was represented by the parameter of maximum semantic proximity with respect to the reference terms, hence considering the ten consecutive most semantically-related ones. With regard to the procedural steps, the present analysis will be specifically articulated in two different phases. Firstly, a selection of the terms that will be actually considered for the performance of the semic analysis will be carried out in order to prevent redundancies, such as in the circumstances in which a term is both featured in its singular form and in the plural one, with respect to whom the semic analysis was provided taking as a reference the singular form. Furthermore, a single semic analysis was also proposed in the case in which the same term was included in more than one list or the variant with the capital letter was conceived as a distinct

terminological unit, as in the latter instance the conceptual content of such entities proved to be unaffected by the orthographic dissimilarity. Additional choices were constituted by the omission of the semic analysis of the reference term in the circumstance in which it was featured as a terminological element in the lists due to the consideration that it was already antecedently formulated in the previous sub-section, and in the occurrence in which enlisted terms were analysed in relation with other reference terms. Secondly, a semic analysis of the designated terminology will be performed. Specifically, the methodological criteria elucidated in section 3.6 of the previous chapter will be applied, thus representing an additional opportunity to further exemplify the proposed approach. Moreover, an aspect that is relevant to mention is related to the procedure that was selected as suitable with respect to the performance of the semic analysis of multi-word terms. As a matter of fact, in some instances, the retrieval of definitions of determined multi-word terms by way of a query in the framework of the indicated lexicographic resources did not generate results. In order to perform the analysis of terms with relation to which an entry was not instantiated, an alternative methodology was therefore conceived. In particular, in such cases, the sememes will respectively feature the semes related to each terminological element constituting the multi-word terms, hence composing a unitary semic analysis. To conclude, observations and considerations will be offered with a view to assessing the conceptual pertinence of the analysed terms and questioning their potential conception as semes with respect to the reference ones.

The following table incorporates the terminology with respect to whose a relation of semantic proximity was individuated with respect to the term *screening* by way of the application of the indicated models trained on the signalled datasets. In this particular case, a supplementary exclusion in matter of analysis concerned the lexical elements *view Puyol leaped*, *backtesting portfolio* and *Puyol leaped* which are not inscribed in the considered medical domain, thus implying their impossibility to compose in this context the sememe of the reference term. With relation to this, however, specific observations will be subsequently proposed.

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

Taking into consideration the outlined criteria, the following formulations of semic analysis could be proposed with respect to the selected terminology:

- *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/
- *procedures*: /medicine/ /act/ /procedure/ /diagnosis/ /treatment/ /operation/
- *testing*: /medicine/ /analysis/ /tests/ /findings/
- *checks*: /medicine/ /examine/ /test/ /checking/
- *tests*: /medicine/ /examination/ /trial/ /substance/ /reagents/ /disease/ /fluids/ /tissues/ /excretions/ /body/ /psychological/ /behavioral/ /trait/ /detect/ /identify/ /bodily/ /fluids/ /observation/ /evaluation/ /procedure/ /reaction/ /constituent/ /diagnostic/ /procedure/ /condition/ /testing/ /analysis/ /diagnosis/ /method/
- *diagnostic*: /medicine/ /diagnosis/ /identify/ /disease/ /medical/ /symptom/ /technique/ /instrument/ /signs/ /symptoms/ /methods/ /act/
- *detection*: /medicine/ act/ /discovery/

- *baggage*: /medicine/ /mental/ issues/
- *newborn hearing*: /medicine/ /newborn/ /born/ /infant/ /neonatal/ /ability/ /perceive/ /sounds/
- *screen*: /medicine/ /separate/ /population/ /undiagnosed/ /disease/ /defect/ /pathologic/ /condition/ /risk/ /tests/ /examinations/ /procedures/ /measuring/ /pap smear/ /identify/ /examine/ /evaluate/ /process/ /select/
- *routine*: /medicine/ /behavioral/ /sequence/
- *colposcopy*: /medicine/ /procedure/ /physician/ /cervix/ /vagina/ /instrument/ /colposcope/ /check/ /precancerous/ /abnormal /areas/ /magnify/ /photographs/ /examination/ /endoscope/ /vaginal/ /cervical/ /epithelia/ /evaluate/ /assist/ /biopsying/ /lesions/ /uterine cervix/ /upper vagina/ /identified/ /pap smear/ /optical/ /neck/ /womb/
- *surveillance*: /medicine/ /monitoring/ /procedure/ /control/ /spread/ /infectious disease/ /supervision/ /incubation period/ /scrutiny/ /methods/ /observation/ /health/ /populations/ /disease/ /prevalence/ /analysis/ /collection/ /health data/ /occurrence/ /testing/
- *evaluation*: /medicine/ /appraisal/ /assessment/ /measurement/ /value/ /effectiveness/ /progress/ /evaluation/ /health care/ /structure/ /process/ /outcome/ /criteria/ /standards/ /diagnostic/ /classifications/

In the second table, the terminology related to the term *measles* that was retrieved by way of word embeddings is presented, subsequently followed by the semic analysis of the listed terms.

GloVe (Wikipedia)	Word2Vec (Google News)	Word2Vec (PubMed)
polio	whooping cough	mumps
vaccination	chickenpox	rubella
tetanus	pertussis	rabies
diphtheria	polio	varicella
immunization	Measles	polio
malaria	meningococcal disease	influenza
vaccinations	meningitis	rotavirus
mumps	mumps	smallpox

rubella typhoid	vaccine preventable diseases measles outbreak	dengue vaccination

- *polio*: /medicine/ /poliomyelitis/ /polio/ /infantile paralysis/ /infectious/ /viral/ /disease/ /central nervous system/ /symptoms/ /nonparalytic/ /infection/ /total paralysis/ /acute/ /infectuous disease/ /epidemics/ /virus/ /poliovirus/ /coxsackievirus/ /echovirus/
- *vaccination*: /medicine/ /vaccines/ /prevent/ /diseases/ /introduction/ /vaccine/ /body/ /produce/ /immunity/ /inoculation/ /protect/ /disease/ /injection/ /bacteria/ /virus/ /antigen/ /stimulate/ /immune system/ /antibodies/ /microbe/ /preventing/ /microorganisms/ /treated/ /harmless/ /inducing/ /development/ /scar/
- *tetanus*: /medicine/ /fatal/ /disease/ /affects/ /central nervous system/ /painful/ /muscular/ /contractions/ /bacteria/ /body/ /wound/ /cut/ /exposed/ /contaminated/ /vaccination/ /neurotropic/ /toxin/ /tetanospasmin/ /clostridium tetani/ /acute/ /contraction/ /voluntary muscles/ /neck/ /jaw/ /toxin/ /bacterium/ /infects/ /lockjaw/ /infectious/ /tonic spasm/ /muscles/ /exotoxin/ /C. tetani/ /Clostridium/
- *diphtheria*: /medicine/ /fatal/ /contagious disease/ /nose/ /throat/ /skin/ /formation/ /membrane/ /tonsils/ /throat/ /infectious disease/ /bacterium/ /Corynebacterium diphteriae/ /toxin/ /inflammation/ /membranous/ /fibrinous/ /exudate/ /mucous membrane/ /pharynx/ /tracheobronchial tree/ /degeneration/ /peripheral nerves/ /heart muscle/ /tissues/ /acute/ /false membrane/ /obstruct/ /breathing/ /bloodstream/ /systemic/ /damage/ /heart/ /nervous system/ /febrile/
- *immunization*: /medicine/ /protection/ /patients/ /susceptible/ /communicable diseases/ /administration/ /suspension/ /organisms/ /protein/ /heterologous/ /inactivated/ /toxin/ /process/ /inducing/ /immunity/ /antigen/ /immune system/ /prevent/ /infection/ /illness/ /pathogen/ /infection/ /vaccine/ /immune response/ /immune/ /inoculation/ /vaccination/ /disease/ /treatment/ /organism/

- *malaria*: /medicine/ /infectious disease/ /recurrent/ /symptoms/ /chills/ /fever/ /enlarged spleen/ /disease/ /treated/ /medication/ /recurs/ /endemic/ /outbreaks/ /sweating/ /protozoan/ /Plasmodium/ /red blood cells/ /transmitted/ /bite/ /infected/ /sporozoan/ /erythrocyte phase/ /infection/ /exoerythrocytic cycle/ /liver/ /parenchyma cells/ /erythrocytic schizogenous cycles/ /gametocytes/ /gametes/ /prostration/ /death/ /immunologically/ /mediated/ /sequelae/ /acute/ /chronic disease/ /attacks/ /destruction/ /blood cells/ /release/ /toxic substances/ /genus/ /anopheles/
- *mumps*: /medicine/ /viral/ /infection/ /salivary glands/ /painful/ /swelling/ /cheek/ /parotid glands/ /inflammation/ /epidemic parotitis/ /acute/ /infectious/ /contagious/ /disease/ /mumps virus/ /Rubulavirus/ /fever/ /testis/ /ovary/ /pancreas/ /meninges/ /inflammatory/ /paramyxovirus/ /parotids/ /ovaries/ /testes/ /vaccination/
- *rubella*: /medicine/ /acute/ /exanthematous/ /disease/ /rubella virus/ /Rubivirus/ / enlargement/ /lymph nodes/ /fever/ /constitutional/ /reaction/ /birth defects/ /incidence/ /maternal/ /infection/ /first trimester/ /fetal/ /congenital rubella syndrome/ /contagious/ /eruptive/ /virus/ /producing/ /congenital defects/ /infants/ /born/ /infected/ /pregnancy/ /German measles/
- *typhoid*: /medicine/ /typhus/ /typhoid fever/ /salmonella typhi/ /infection/
- *whooping cough*: /medicine/ /pertussis/ /contagious disease/ /spasms/ /paroxysms/ /coughing/ /intake/ /air/ /respiratory system/ /bacterium/ /Bordetella pertussis/ /spasms/ /inspirations/ /infectious disease/ /genus/ /B. pertussis/ /convulsive/ /spasmodic/ /cough/ /crowing/ /breath/
- *chickenpox*: /medicine/ /acute/ /contagious disease/ /varicella-zoster virus/ /skin/ /eruptions/ /fever/ /malaise/ /varicella/ /formation/ /vesicles/ /herpesvirus/ /genus/ /Varicellovirus/ /species/ /human herpesvirus 3/
- *pertussis*: /medicine/ /acute/ /infectious/ /inflammation/ /larynx/ /trachea/ /bronchi/ /Bordetella pertussis/ /bouts/ /spasmodic/ /coughing/ /breath/ /recurrent/ /inspiratory stridor/ /laryngeal spasm/
- *meningococcal disease*: /medicine/ /Neisseria meningitidis/ /meningococcus/ /bloodstream/ /bacteria/ /infectious diseases/ /fatal/

- *meningitis*: /medicine/ /inflammation/ /meninges/ /membranous/ /brain/ /spinal cord/ /meningitis/ /infection/ /bacteria/ /virus/ /fungi/ /bleeding/ /cancer/ /diseases/ /immune system/ /inflammatory response/ /chemotherapy/ /chemical agents/ /fatal/ /condition/ /bacterial/ /viral infection/ /fever/ /vomiting/ /headache/ /stiff neck/ /Neisseria meningitidis/ /streptococcus pneumoniae/ /viruses/ /mumps virus/ /pia mater/ /arachnoid/ /disease/ /illness/ /coxsackieviruses/ /meningococcus/ /serotype/ /haemophilus influenzae/
- *vaccine preventable diseases*: /medicine/ /suspension/ /attenuated/ /killed microorganisms/ /viruses/ /bacteria/ /rickettsiae/ /administered/ /prevention/ /treatment/ /infectious/ /diseases/ /preparation/ /pathogen/ /bacterium/ /virus/ /individual/ /antibody/ /production/ /cellular/ /immunity/ /virulent/ /organisms/ /infection/ /antibodies/ /protection/ /pathological/ /process/ /body/ /affect/ /disorder/ /abnormal/ /condition/ /inflammation/ /defect/ /signs/ /symptoms/ /smallpox/ /impairment/ /response/ /agents/
- *measles outbreak*: /medicine/ /disease/ /virus/ /contagious/ /pyrexia/ /mucus/ /nose/ /conjunctivitis/ /exanthem/ /acute/ /rubeola/ /eruption/ /skin rush/ /vaccination/ /increase/ /incidence/ /condition/
- *rabies*: /medicine/ /acute/ /viral disease/ /central nervous system/ /zoonosis/ /saliva/ /bite/ /infected/ /disease/ /hydrophobia/ /symptom/ /fever/ /depression/ /confusion/ /painful/ /muscle spasms/ /sensitivity/ /thirst/ /painful/ /swallowing/ /excessive salivation/ /loss/ /muscle tone/ /immunization/ /fatal/ /infectious/ /transmitted/ /neuroptic lyssavirus/ /Lyssavirus/ /replicates/ /salivary glands/ /infected/ /saliva/ /rabid/ /fatigue/ /headache/ /increased/ /salivation/ /excitability/ /muscle weakness/ /spasms/ /hallucinations/ /abnormal/ /behaviour/ /paralysis/ /coma/ /death/
- *varicella*: /medicine/ /acute/ /contagious/ /disease/ /Varicella-Zoster virus/ /Varicellovirus/ /eruption/ /papules/ /vesicles/ /pustules/ /stages/ /symptoms/ /incubation period/ /chickenpox/
- *influenza*: /medicine/ /flu/ /grippe/ /infectious respiratory disease/ /disease/ /influenza virus/ /acute/ /viral infection/ /respiratory tract/ /contagious/ /viral infection/ /inflammation/ /fever/ /chills/ /muscular pain/ /prostration/ /influenza

viruses/ /attacks/ /respiratory epithelial cells/ /catarrhal//onset/ /headache/ /muscle aches/ /cough/ /epidemics/ /pandemics/ /contagious/ /respiratory disease/ /orthomyxoviruses/ /respiratory/ /infection/

- *rotavirus*: /medicine/ /reoviruses/ /transmitted/ /fecal/ /oral/ /acute gastroenteritis/ /diarrhea/ /antigenic/ /RNA viruses/ /gastroenteritis/ /viruses/ /infect/ /gastroenteritis/ /genus/ /agents/ /epidemics/ /severe/ /fatal/ /infants/ /infant diarrhea/ /virus/
- *smallpox*: /medicine/ /infection/ /variola virus/ /poxvirus/ /disease/ /contagious/ /fatal/ /symptom/ /blisters/ /pustules/ /skin/ /epidemics/ /variola/ /vaccination/ /acute/ /viral/ /illness/ /febrile disease/ /Orthopoxvirus/ /skin/ /eruption/ /sloughing/ /scar/ /formation/
- *dengue*: /medicine/ /tropical disease/ /virus/ /transmitted/ /acute disease/ /fever/ /prostration/ /headache/ /aches/ /bones/ /joints/ /muscles/ /enlargement/ / lymph nodes/ /stage/ /skin rash/ /body/ /breakbone fever/ /infectious disease/ /acute/ /joint pain/ /rash/ /RNA-virus/ /Dengue virus/ /infectious/ /tropical disease/ /flaviviruses/ /bleeding/ /disease/ /dengue hemorrhagic fever/ /gums/ /gastrointestinal tract/ /shock/ /death/ /dandy fever/ /dengue fever/

The third table includes the semantically-proximal terms with reference to the term *asthma*, subsequently correlated to the respective semic analysis. A clarification is however required with respect to the abbreviation *COPD*, with regard to which a definition was not present. In this circumstance, the query was reformulated and the extended lexical form *chronic obstructive pulmonary disease* was therefore searched for in order to retrieve the semic elements.

GloVe (Wikipedia)	Word2Vec (Google News)	Word2Vec (PubMed)
allergies	Asthmatic	asthmatic
diabetes	Asthma	rhinitis
allergy	allergies	allergic
allergic	asthmatics	asthmatics
hypertension	allergic asthma	atopy
arthritis	COPD	wheezing
emphysema	allergy	exacerbations

obesity bronchitis sufferers	asthmatic symptoms asthma sufferers eczema	wheeze eczema atopic

- *allergies*: /medicine/ /abnormal/ /reactions/ /immune system/ /substances/ /altered/ /bodily/ /reactivity/ /hypersensitivity/ /antigen/ /exposure/ /pathological/ /reaction/ /sneezing/ /respiratory/ /itching/ /skin/ /rashes/
- *diabetes*: /medicine/ /disorders/ /urination/ /polyuria/ /diabetes mellitus/ /diabetes insipidus/ /diseases/ /symptoms/ /weight loss/ /glucosuria/ /condition/ /increase/ /pituitary gland/ /urinary/ /defect/ /insulin/ /hyperglycemia/ /inducing/ /abnormal/ /conditions/ /secretion/ /excretion/ /urine/
- *allergic*: /medicine/ /allergy/ /response/ /allergen/ /caused/ /affected/ /stimulated/ subject/ /allergic reaction/
- *hypertension*: /medicine/ /high blood pressure/ /abnormally/ /arterial blood pressure/ /arterial disease/ /chronic high blood pressure/ /pressure/ /tension/ /body fluid/ /intraocular/ /cerebrospinal fluids/ /high arterial blood pressure/ /systolic blood/ /diastolic blood pressure/ /renal/ /endocrine/ /disorder/ /thickening/ /inelasticity/ /arterial walls/ /hypertrophy/ /left heart ventricle/ /risk factor/ /pathological/ /conditions/ /systemic/ /condition/ /symptomless/ /dizziness/ /palpitations/ /fainting/ /headache/
- *arthritis*: /medicine/ /inflammation/ /joint/ /disease/ /cause/ /pain/ /stiffness/ /musculoskeletal system/ /group/ /diseases/ /limit/ /movement/ /state/ /joints/ /swelling/ /inflamed/ /joints/ /infectious/ /metabolic/ /constitutional/ /causes/
- *emphysema*: /medicine/ /chronic respiratory disease/ /over-inflation/ /air sacs/ / alveoli/ /lungs/ /decrease/ /lung/ /function/ /breathlessness/ /pathological/ /condition/ /abnormal/ /air spaces/ /breathing/ /infection/ /susceptibility/ /irreversible/ /expansion/ /destruction/ /alveolar walls/ /distension/ /body tissues/ /retention/ /air/ /lung disease/ /condition/ /expansions/ /interstitial/ /subcutaneous tissues/ /lung/ /rupture/ /loss/ /pulmonary/ /elasticity/ /shortness/ /impairment/ /heart/

- *obesity*: /medicine/ /abnormal/ /accumulation/ /body fat/ /obese/ /condition/ /body weight/ /fat/
- *bronchitis*: /medicine/ /inflammation/ /nose/ /lungs/ /windpipe/ /trachea/ /air tubes/ /lung/ /bronchi/ /acute/ /chronic/ /viral infection/ /bacterial infection/ /lung disease/ /mucous membrane/ /bronchial tubes/ /lining/ /bronchus/ /cough/ /sputum/ /fever/ /breathlessness/ /wheezing/ /pain/ /chest/ /sputum/ /production/ /infective/ /flare-ups/ /chronic inflammation/
- *sufferers*: /medicine/ /subjected/ /pain/ /distress/ /injury/ /loss/ /damages/
- *asthmatic*: /medicine/ /suffering/ /asthma/
- *allergic asthma*: /medicine/ /condition/ /recurrent/ /attacks/ /dyspnea/ /airway/ /inflammation/ /wheezing/ /spasmodic/ /constriction/ /bronchi/ /bronchial asthma/ /cause/ /suffocation/ /acute/ /attack/ /status asthmaticus/ /medical emergency/ /fatal/ /bronchoconstriction/ /shortness/ /breath/
- *COPD*: /medicine/ /chronic obstructive pulmonary disease/ /chronic lung diseases/ /chronic bronchitis/ /emphysema/ /asthma/ /pulmonary/ /disease/ /chronic/ /irreversible/ /airway/ /obstruction/ /slowed/ /exhalation/
- *asthmatic symptoms*: /medicine/ /suffering/ /asthma/ /disease/ /perceived/ /patient/ /phenomenon/ /perception/ /bodily/ /signs/
- *asthma sufferers*: /medicine/ /medicine/ /condition/ /dyspnea/ /airway/ /inflammation/ /wheezing/ /spasmodic/ /constriction/ /bronchi/ /reversible/ /obstruction/ /syndrome/ /chronic/ /tracheobronchial tree/ /allergens/ /bronchospasm/ /hyperreactivity/ /lung disorder/ /subjected/ /pain/ /distress/ /injury/ /loss/ /damages/
- *eczema*: /medicine/ /inflammatory/ /conditions/ /skin/ /vesiculation/ /acute/ /stage/ /erythematous/ /edematous/ /papular/ /crusting/ /lichenification/ /scaling/ /duskiness/ /erythema/ /hyperpigmentation/ /itching/ /burning/ /vesicles/ /intraepidermal/ /spongiosis/ /hereditary/ /allergic rhinitis/ /asthma/ /noncontagious/ /inflammation/ /redness/ /outbreak/ /lesions/ /discharge/ /serous/ /encrusted/ /scaly/ /patches/ /blisters/ /burst/ /serum/ /crusty/ /releasing/ /condition/ /redness/ /oozing/ /vesicular/ /crusted/

- *rhinitis*: /medicine/ /inflammation/ /mucous lining/ /nose/ /nasal mucous membrane/ /nasal mucous membranes/ /nasal mucosa/ /rhinorrhea/ /nasal congestion/ /itching/ /sneezing/
- *atopy*: /medicine/ /clinical/ /ipersensitivity/ /allergy/ /hereditary/ /predisposition/ /develop/ /inherited/ /antibody/ /reagin/ /atopic/ /spontaneous/ /genetically/ /sensitivity/ /state/ /antigens/ /IgE/ /susceptibility/ /allergic reaction/ /diseases/ /antigen/
- *wheezing*: /medicine/ /labored/ /breathing/ /hissing/ /breath/ /asthma/ /infants/ /transient/ /viral/ /respiratory/ /infection/
- *exacerbations*: /medicine/ /increase/ /severity/ /disease/ /symptoms/ /signs/ /cause/ /severe/
- *atopic*: /medicine/ /atopy/ /allergic/ /hereditary/ /predisposition/ /developing/ /allergic reactions/ /exposure/ /antigens/

The following table integrates the terms that were detected as semantically similar to *malformation*, with respect to whose a systematic semic analysis was performed.

GloVe (Wikipedia)	Word2Vec (Google News)	Word2Vec (PubMed)
congenital	malformations	malformations
malformations	atresia	anomaly
abnormality	congenital disorder	anomalies
arteriovenous	hypoplasia	hypoplasia
foetus	congenital anomaly	agenesis
deformity	brain malformation	congenital
fetal	diaphragmatic hernia	craniosynostosis
abnormalities	hemangioma	hamartoma
malformed	coarctation	fistulae
stenosis	pyloric stenosis	fistula

- *congenital*: /medicine/ /birth/ /conditions/ /present/ /mental/ /physical/ /traits/ /anomalies/ /malformations/ /diseases/ /hereditary/ /gestation/ /condition/ /congenital/ /disorder/ /fetal/ /acquired/ /uterus/ /defects/ /environmental/

- *abnormality*: /medicine/ /state/ /condition/ /malformation/ /abnormal/ /anomaly/ /deformity/ /impairment/ /dysfunction/ /defect/ /structure/ /function/ /behaviour/ /phenomenon/
- *arteriovenous*: /medicine/ /arterial/ /venous/ /artery/ /vein/ /arteries/ /veins/
- *foetus*: /medicine/ /product/ /conception/ /embryo/ /formation/ /amnion/ /pregnancy/ /gestation/
- *deformity*: /medicine/ /distortion/ /part/ /body/ /malformation/ /permanent/ /structural/ /deviation/ /disfigurement/ /congenital/ /acquired/ /deformed/ /misshapen/ /distorted/ /physical/ /blemish/
- *fetal*: /medicine/ /fetus/ /development/ /utero/
- *malformed*: /medicine/ /abnormally/ /faultily/ /formed/ /malformation/
- *stenosis*: /medicine/ /abnormal/ /narrowing/ /contraction/ /body/ /passage/ /opening/ /arctation/ /coarctation/ /stricture/ /canal/ /orifice/ /constriction/ /narrowing/ /duct/ /cardiac valves/ /tubular/ /organ/ /intestinal canal/ /blood /vessel/ /diameter/ /bodily/
- *atresia*: /medicine/ /congenital/ /absence/ /closure/ /body/ /opening/ /tubular/ /structure/ /obstruction/ /patent lumen/ /orifice/ /passage/ /anus/ /intestine/ /external ear canal/ /degeneration/ /resorption/ /ovarian follicles/ /developmental/ /formation/ /lumen/ /organ/ /duct/ /channel/ /disappearance/ /anatomical/ /part/ /ovarian follicle/
- *congenital disorder*: /medicine/ /birth/ /conditions/ /present/ /mental/ /physical/ /traits/ /anomalies/ /malformations/ /diseases/ /gestation/ /condition/ /congenital/ /disorder/ /fetal/ /acquired/ /uterus/ /defects/ /environmental/ /derangement/ /abnormality/ /function/ /state/ /disturbance/ /structure/ /genetic/ /embryonic/ /failure/ /development/ /exogenous/ /trauma/ /disease/ /abnormal/
- *hypoplasia*: /medicine/ /incomplete/ /development/ /underdevelopment/ /organ/ /tissue/ /deficiency/ /cells/ /atrophy/ /destruction/ /part/ /growth/ /body/ /structure/ /condition/ /decrease/ /immature/
- *congenital anomaly*: /medicine/ /structural/ /abnormality/ /birth/

- *brain malformation*: /medicine/ /part/ /central nervous system/ /forebrain/ /prosencephalon/ /midbrain/ /mesencephalon/ /hindbrain/ /rhombencephalon/ /nerve/ /tissue/ /cranium/ /gray matter/ /white matter/ /organ/ /skull/ /neurons/ /structures/ /glia/ /structural/ /defect/ /deformity/ /development/ /localised/ /morphogenesis/ /embryonic/ /interaction/ /anomalous/ /abnormal/ /formation/ /structure/
- *diaphragmatic hernia*: /medicine/ /protrusion/ /abdomen/ /opening/ /diaphragm/ /chest cavity/ /condition/ /congenital/ /acquired/ /chest/ /weakness/ /respiratory/ /diaphragm/ /thoracic/
- *hemangioma*: /medicine/ /congenital/ /vascular/ /malformation/ /benign/ /tumor/ /formed/ /blood vessels/ /clustered/ /birth/ /parts/ /body/ /liver/ /bones/ /capillaries/ /surface/ /skin/ /birthmark/ /vascular tumor/ /proliferation/ /mass/ /subcutaneous/ /tissues/ /regression/ /benign/ /skin lesion/ /dense/ /masses/ /dilated/ /anomaly/
- *coarctation*: /medicine/ /constriction/ /stricture/ /stenosis/ /narrowing/ /constricting/ /aorta/ /blood vessel/ /contraction/ /canal/ /vessel/
- *pyloric stenosis*: /medicine/ /narrowing/ /passage/ /stomach/ /small intestine/ /condition/ /infants/ /gastric pylorus/ /congenital/ /muscular/ /hypertrophy/ /scarring/ /pyloric opening/ /congenital/ /malformation/ /contraction/ /scar tissue/
- *anomaly*: /medicine/ /deviation/ /birth defect/ /structural/ /abnormality/ /structurally/ /irregular/ /congenital defect/ /birth/ /hereditary/ /defect/ /bodily/ / part/
- *agenesis*: /medicine/ /absence/ /organ/ /primordium/ /embryo/ /form/ /part/ /incomplete/ /development/ /body/ /tissue/ /formation/ /failure/
- *craniosynostosis*: /medicine/ /premature/ /closure/ /cranial/ /sutures/ /skull/ /deformities/ /ossification/ /cranium/ /obliteration/ /malformed/ /fusion/ /suture joints/ /birth/ /infancy/ /fibrous/ /joints/ /coronal/ /suture/ /sagittal/
- *hamartoma*: /medicine/ /benign/ /nodule/ /overgrowth/ /mature/ /cells/ /tissues/ /focal/ /malformation/ /abnormal/ /tissue/ /develop/ /mass/ /anomalous/ /development/ /part/ /organ/

- *fistulae*: /medicine/ /permanent/ /abnormal/ /passageway/ /organs/ /body/ /passage/ /epithelial/ /surface/ /duct/ /injury/ /disease/ /congenital/ /disorder/ /abscess/ /cavity/ /hollow organ/ /surgically/ /vascular/ /access/ /conduit/ /internal/ /organ/ /skin/ /epithelialized / /part/ /fluids/ /secretions/

The last table comprises the terminological entities related to the reference term *dermatitis* with respect to which the semic analysis was carried out.

Word2Vec (Google News)	Word2Vec (PubMed)
contact dermatitis	eczema
eczema	atopic
skin irritations	urticaria
allergic dermatitis	allergy
skin irritation	rhinitis
pruritic	allergic
folliculitis	rosacea
atopic dermatitis	conjunctivitis
allergic reactions	psoriasis
atopic eczema	eruptions

- *contact dermatitis*: /medicine/ /skin/ /inflammation/ /disease/ /disorder/ /allergic/ /hypersensitivity/ /irritant/ /allergic/ /acute/ /chronic/ /contact/ /substance/ /irritating/ /substance/ /dermatitis/ /redness/ /itching/ /allergen/ /immune response/ /inflammatory/ /rash/ /eczema/ /allergic reaction/ /dermatitis/
- *skin irritations*: /medicine/ /organ/ /body/ /membranous/ /protective/ /covering/ /epidermis/ /dermis/ /corium/ /ectodermal/ /epidermis/ /sweat/ /sebaceous/ /glands/ /mesodermal/ /connective tissue/ /blood vessels/ /nerves/ /inflammatory/ /reaction/ /tissues/ /injury/ /response/ /stimulus/ /irritability/ /soreness/ /roughness/ /inflammation/ /bodily/ /part/
- *allergic dermatitis*: /medicine/ /inflammation/ /skin/ /contact/ /substance/ /allergic/
- *pruritic*: /medicine/ /pruritus/ /itching/

- *folliculitis*: /medicine/ /inflammation/ /infection/ /hair follicles/ /skin/ /hair/ /hair follicle/ /lesion/ /papule/ /pustule/ /follicle/ /boils/ /pimples/ /staphylococcal infections/ /follicles/
- *atopic dermatitis*: /medicine/ /chronic/ /skin/ /disorder/ /inflammation/ /disease/ /pruritus/ /atopic eczema/ /dermatopathy/ /allergic rhinitis/ /stress/ /asthma/ /reactions/ /allergy/ /eczematous/ /condition/ /xerosis/ /hypersensitivity/ /atopy/ /itching/ /dermatitis/
- *allergic reactions*: /medicine/ /allergy/ /local/ /general/ /reaction/ /organism/ /contact/ /allergen/ /exposed/ /sensitized/ /immunologic/ /inflammation/ /tissue damage/
- *atopic eczema*: /medicine/ /dermatitis/ /allergy/ /allergen/
- *urticaria*: /medicine/ /vascular/ /reaction/ /skin/ /transient/ /patches/ /wheals/ /skin/ /itching/ /hives/ /eruption/ /systemic/ /hypersensitivity/ /condition/ /pruritic/ /welts/ /dermographism/ /allergic reactions/ /histamine/ /allergic/ /condition/ /itchy/ /appearance/
- *rosacea*: /medicine/ /disease/ /redness/ /erythema/ /face/ /flushing/ /skin/ /pimples/ /papules/ /pustules/ /veins/ /telangiectasias/ /stages/ /disease/ /swell/ /nose/ /rhinophyma/ /chronic/ /vascular/ /follicular/ /dilation/ /cheeks/ /hyperplasia/ /sebaceous glands/ /erythematous/ /chronic dermatitis/ /capillaries/ /acne rosacea/ /idiopathic/ /disorder/ /inflammation/ /chin/ /forehead/ /eyelids/ /eruptions/ /tissue/ /oedema/ /blood vessels/ /inflammatory/ /telangiectasia/ /inflamed/ /lesions/ /burning/ /irritation/ /tissues/
- *conjunctivitis*: /medicine/ /inflammation/ /redness/ /lining/ /eye/ /eyelid/ /conjunctiva/ /infection/ /allergic reaction/ /physical/ /infrared/ /ultraviolet light/ /bacteria/ /virus/ /chemical/ /infectious/ /bacterial/ /viral/ /contagious/ /discharge/ /allergy/ /chemical/ /irritation/ /radiation/ /sclera/ /itching/ /burning/ / tearing/ /pathogenic/ /viruses/ /allergens/ /irritants/
- *psoriasis*: /medicine/ /chronic/ /non-contagious/ /disease/ /inflamed/ /lesions/ /scabs/ /skin/ /recurrent/ /macules/ /papules/ /patches/ /scales/ /guttate/ /erythrodermic/ /exfoliative/ /pustular/ /inherited/ /condition/ /eruption/ /maculopapules/ /lesions/ /elbows/ /knees/ /scalp/ /trunk/

- *eruptions*: /medicine/ /lesions/ /skin/ /disease/ /redness/ /prominence/ /rash/ /efflorescent/ /developing/ /dermatosis/ /mucous membranes/ /local/ /manifestation/ /exanthemata/ /lesion/ /blemish/ /rash/ /process/ /erupting/ /condition/ /mucous membrane/ /pustule/ /appearance/

Taking as a reference the performance of the second step of the conceived dual analysis, some considerations could be offered in order to further investigate the potential subsistence of an interrelation between semic analysis and word embeddings with a view to eventually individuating the nature of the implicated points of intersection. A first observation that could be proposed consists in the aforementioned ascertainment of the subsistence of terminological elements which are not related to the specific medical domain that can be noticed in the table containing the semantically-related terms with respect to the reference one *screening*, hence signalling the potential verification of rumour. Specifically, this circumstance could be detected in the context of the application of the GloVe model, while in the case of Word2Vec trained on Google News and Pubmed, the latter of which can be considered as the most relevant collection of scientific literature, it did not occur. In connection to this, it particularly emerged the presence of the semantically proximal term *baggage* which is inscribed in the subdomain of psychology and is defined as "*A popular term for a constellation of mental issues based on memories and experiences that any person has when entering into an intimate relationship*"[47]. Nevertheless, it could also be supposedly considered as a revelatory element of the polysemic nature of the reference term *screening*, which can for instance additionally refer to the examinations that are conducted in airports. A second observation concerns the ascertainment of the fact that the presence of reference terms as semic elements in the formulations of the semic analysis of semantically-related terms could be merely detected in a few circumstances. In this respect, on first impression, this evidence could be deemed as indicative of the unrelatedness at the semic level of the considered linguistic entities, hence suggesting a conceptual inconsistency. Conversely, the line of thought that was conceived in order to discern a possible correlation consisted in the conception of the conceptual dimension as

47 https://medical-dictionary.thefreedictionary.com/baggage

determined by the comprehensive consideration of all the semes comprised in the sememes, therefore evaluating in this sense a potential compatibility in terms of conceptual integration. As a matter of fact, in this perspective, the concept should be regarded as the only essential discriminating element, in this connection implying that the sememe should be perceived at the notional level as the touchstone for the determination of a conceptual relatedness up to its transposition in form of semic elements in the composition of the sememe of the related reference term. Furthermore, provided the selection of the intensional definition as the type of lexicographic source that should be employed to represent the conceptual content of terms, this approach was accordingly adopted, hence excluding hyponyms as potential semes. Considering the postulated viewpoint, hyponyms are effectively related to the respective hypernyms; nevertheless, they do not concur to precisely define and convey the more general concept of the former, as an additional level of specification is entailed, hence signalling a difference at the semic level. Precisely, this standpoint could be substantiated by way of resorting to the assertions proposed by Löbner (2013: 83):

> an expression A is a hyponym of an expression B if 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.

From the consideration of the directionality of the relationship, the hypernym concurs to compose the conceptual content of the hyponym, however, the hyponym entails the inclusion of additional conceptual particles hence producing an asymmetry and, consequentially, determining the subsistence of a related but different concept whose essence is comprehensively dissimilar. As a matter of fact, to the sememe of the hyponym, more elements should be added in order to convey, consequentially, a diverse concept, which is deductively supposed for this reason to entail the presence of partially different semes. For instance, taking into account the relationships that can be established between the terms *inflammation*, *dermatitis* and *contact dermatitis*, in the circumstance in which the semic analysis of the term *dermatitis* should be performed, *dermatitis* could constitute the hyponym of *inflammation*, therefore *inflammation* can be employed in order to partially convey the semantic meaning of the reference term.

Conversely, *dermatitis* represents the hypernym of *contact dermatitis*. In this connection, by way of exploiting the logical connection established in the quotation, being the former the entity B and the latter A, it can be inferred that the concept of *dermatitis* should be included in that of *contact dermatitis*. In contrast, the conceptual entity that is lexically represented by the term *contact dermatitis* should not be incorporated in the sememe of *dermatitis* as it would add elements that do not strictly pertain to its concept which is supposed to be more generic.

A further consideration that could emerge from the analysis of the sememes with respect to the reference terms consists in the fact that many semantically-related terms manifest a semantic connection which, however, could be regarded as partial due to a relatedness that can be ascribed to their belonging to the same determined semantic field. For instance, considering the reference term *measles*, it could be detected that terms such as *varicella*, *mumps* and *rubella* share some identical conceptual elements, namely represented by the semes /acute/, /contagious/ and /disease/, and are generically conceptually related to a causative agent that can be ascribed to a specific virus. Furthermore, taking in consideration that *chickenpox* is the popular term for *varicella*, by extension the identical semes are shared with respect to *measles*. By the same token, this occurrence can also be noticed with reference to the terms that are semantically-related to *asthma*, such as *bronchitis* and *rhinitis*, which represent different types of inflammations. Comprehensively, given the partial hence limited possible relatedness realised by way of some semic elements which are embedded in the conceptual nature of semantically proximal terms and, principally, on the grounds that they do not concur to define the concept of the reference terms, a first possible resulting observation that could be offered is that these terms should not be included in the sememes of the reference terms. As a result, it also emerged as unfeasible to detect a superset of semes which could potentially partially convey the conceptual content of the reference terms due to the fact that, for what concerns the relation between the reference terms and the semantically-related terminology, occasionally the semes with regard to which a mutual correspondence was individuated in the sememes circumstantially differed according to the determined semantically-similar term under consideration or, in other cases, no

semic co-occurrence was identified in the respective sememes with the only exception of the mesogeneric seme /medicine/. By way of exemplification, the subsistence of correspondences between the semes representing the concept of *malformation* and the ones conveying the conceptual content of some related terms could be variably observed. As a matter of fact, it could be for instance noticed that the semes /abnormal/, /deformity/, /defect/ and /structure/, comprised in the sememe of the related term *abnormality* and in that of the reference one *malformation*, are not included in that of *deformity*, which for its part shares with the latter the seme /structural/. Conversely, for what concerns related terms and multi-word terms among which *congenital*, *arteriovenous*, *foetus*, *diaphragmatic hernia* and *pyloric stenosis* can be mentioned, none of the semic elements corresponded.

On the basis of the outlined reasons, the terms which could be potentially considered as suitable in order to represent the conceptual essence of the reference ones can be identified in the following lists:

- *screening*: screenings, screened, testing, checks, tests, diagnostic, detection, Screening, Screenings, screen, screens.
- *measles*: vaccination, immunization, vaccinations, Measles, vaccine preventable diseases.
- *asthma*: Asthmatic, Asthma, asthmatics, COPD, asthmatic, wheezing, wheeze.
- *malformation*: congenital, malformations, abnormality, deformity, abnormalities, malformed, congenital disorder, congenital anomaly, anomaly, anomalies.
- *dermatitis*: pruritic.

Interestingly, as it can be evinced, many of the chosen terms manifest not only a conceptual connection with the reference terms up to the extent that they could be conceived as semes of the latter, but it can be additionally spotted a morphological relationship with the reference terms themselves also with respect to inflections. From the two comparative analysis contemplated in the case study, it can be therefore ascertained that many of the semes that constitute the conceptual content of the

reference terms which were retrieved by way of performing the semic analysis based on the proposed methodology could not be captured by way of the application of word embeddings. This circumstance can be hence indicative of the fact that an automatic performance of semic analysis would not culminate in the retrieval of the totality of the semes that were individuated by way of applying the contemplated criteria as outlined in section 3.6 of the previous chapter and, consequentially, could not be contextually hypothesised. Nevertheless, an interrelation between the two techniques could be identified in the detection of the close connection at the morphological and inflectional level that in many instances is established between the reference terms, the terminological entities originating from word embeddings and the semes included in the sememes of the reference terms. The observation of this phenomenon could constitute a potential valuable factor to consider in the performance of semic analysis as antecedently illustrated which, in the light of the analysis of the resulting data, could consequentially be enhanced in its potential in terms of lexical exhaustiveness by way of including other conceptually-related semic particles.

By way of conclusion, the reasoning upon which the quest for an interrelation between semic analysis and word embeddings was grounded also prodded the formulation of observations that are strictly related to the rationale that is at the notional root of word embeddings, with particular reference to the circumstantial elaboration of the semantic dimension. As a matter of fact, contextual co-occurrence is the type of relation that can fundamentally be detected and established between vectors; however, from the purely semantic viewpoint, in the framework of the present study performed on medical terminology in many circumstances examined terms could not be considered as fully semantically similar or highly related in terms of semic elements. In this connection, the consideration of the postulation of semes as essential and intrinsically engrained components of concepts and evidence from the analysed data call into question the actual subsistence of the capability of word embeddings to consistently capture semantic information. This perspective could also be substantiated by the assessment of a high conceptual pertinency which was mostly detected with respect to terms that feature a morphological or grammatical connection, which can be linked to

the syntactic and morphosyntactic levels. In this sense, the ascription of a purely semantic dimension to the geometrical representation of words could be challenged. Moreover, the detection of circumstances in which terms retrieved by way of word embeddings did not manifest any type of semantic relation with respect to a reference term could further corroborate the claim. The reference, in this case, is specifically to lexical entities such as *view Puyol leaped* and *Puyol leaped* which are manifestly not related to medical terminology and, especially, figure among the closest vectors to the reference one in the vector space. Interestingly enough, a query on Google Search formulated as *"view Puyol leaped"* resulted in the retrieval of multiple documents consisting in articles related to football, in which the term *screening* was present as a neighbour term in the context of the same sentence and Puyol represented the surname of a footballer[48]. A focused research which was subsequently carried out led to the ascertainment of the inscription of the term *screening* in the jargon of the football sector; nonetheless, the resulting strings in their entirety conceived as a unitary concept do not manifest a semantic connection with respect to the reference term. This occurrence could be hence considered as indicative of the strong influence that contextual elements exerts on word embeddings, hence potentially undermining in our perception the principle that word embeddings could actually capture pure semantic information. Notwithstanding the factual semantic relatedness which can be frequently detected in the outputs originating from word embeddings also in terms of shared semantic field, our perspective could be validated by way of contemplating the state-of-the-art limitations that determine the impossibleness of envisaging an automatic semic analysis in the context of the application of the proposed objective methodology. As a matter of fact, a selection would be needed in order to extract suitable linguistic and conceptual elements whose prerequisite for the inclusion as semes of a reference term would be that of manifesting a strict conceptual and semantic pertinency in relation to it. Furthermore it should be considered that, given the fact that the domain of medical terminology is inherently characterised by a high level of conceptual specificity, the output of word embeddings should accordingly reflect the same level of specificity in

48 The mentioned sentence can be retrieved in different sources among which, by way of exemplification, an article on the New York Post: https://nypost.com/2010/07/08/spains-puyol-heads-home-winner-over-germany/

terms of semantic similarity or relatedness. In this light, it can be evinced that in the context of the presented analysis word embeddings do not perform perfectly the task of retrieving terms that actually manifest in most cases a semantic similarity in the form of semic elements or semic correspondences. The topic, however, requires further investigation that could be for instance performed on a more extended dataset.

CONCLUSIONS

In the framework of the present dissertation, it could be evinced that semic analysis proves to constitute a pivotal linguistic technique for the acquisition of the specialised conceptual knowledge which is ingrained in the semantic dimension of medical terminology. The perception of its strategical relevance for the in-depth investigation of terminological entities particularly emerged from the consideration of its procedural application, which could be potentially conceived as presupposing different steps involving diversified but complementary approaches to the comprehension and identification of concepts. Specifically, four main cognitive elaborations were in this regard individuated, respectively consisting in the primary acquisition of the concept as a single knowledge unit, the deconstruction of its conceptualisation into basic constituents instantiated as *semes*, the formal articulation of the latter both at the lexical and metalinguistic level in the form of a *sememe* and, lastly, the validation of the subsistence of a rigorous correspondence between the elements comprised in the sememe and the unitary concept.

A procedural gap was however individuated in the state-of-the-art performance of semic analysis by way of an investigation on the application of the technique in the context of terminological records and, with a view to objectively systematising its management, the challenge was specifically addressed by way of proposing potential criteria whose primary aim is the achievement of a non-subjective terminological and semantic expression of the concepts that, in the terminological domain of the dissertation, are inscribed in the medical conceptual network. The theorisation of an objective methodology of semic analysis was hence conceived with a view to systematising its performance by way of limiting individual manifestations but also, hopefully, to offering an additional perspective and lines of reasoning. In particular, the focus was placed on the inherent semes, which are context-independent therefore unconditioned by social conceptualisations. Specifically, the outlined methodology for an objective semic analysis was conceived as a corpus-based hence data-driven

approach supported by the employment of domain-related lexicographic resources and realised by way of terminological extraction additionally combined with the application of the principle of termhood oriented to the expression of both the diastratic and the diaphasic dimensions at the lexical and conceptual level. Precisely, the line of reasoning at the root of these criteria consisted in the individuated necessity to rely on different valuable domain-oriented sources to acquire an in-depth conceptual and lexical knowledge of a concept for the purpose of enabling its determination through a selective extraction and recollection of terms that are both specifically instantiated in the medical terminological domain by way of filtering out terms that do not convey contextually-related conceptual particularities, and capable of transmitting the determined level of specificity that is implied in the elective circumstantial utilisation of specialised terms and popular ones. In this connection, a challenging issue which was raised that could be deemed to be of particular interest to the study of the topic was the selection of the intensional definition as an adequate source in order to accurately circumscribe and confer distinctiveness to concepts. Comprehensively, these principles proved to be effective in the determination of the conceptualisations that are inscribed in the conceptual network of the medical domain and potentially suitable in order to delimit subjectively-biased manifestations in terms of terminological and notional selection of semic elements. Along this line of argument, the access to a thorough and extensive knowledge that this methodological outlook on semic analysis enables could lead to a keen understanding of the conceptual sphere of terminology from a rather broader perspective. As a matter of fact, the potentiality offered by the proposed objective methodology of semic analysis to closely explore a terminological and conceptual domain is not to be considered as merely confined to the study of medical terminology, but rather on the investigation and acquisition of concepts that are enclosed in other domains ranging from economics and law to physics and tourism.

The subsequent investigation on the possibility to individuate a potential interrelation between semic analysis and word embeddings constituted a further occasion to explore the conceptual dimension of terminology and exemplify the proposed methodology. The analysis, based on a case study on terminological entities

originating from the application of the technique of word embeddings, suggested that a connection could be identified and interestingly exploited with a view to potentiating the suggested methodology. As a matter of fact, a close relation was individuated in the subsistence of morphological relationships and connections in form of inflections between the reference terms, the related semes and word embeddings. In this respect, even though an automatic semic analysis could not be considered as a suitable approach with a view to retrieving all the semic elements which could be collected by way of application of the outlined objective methodology, the detection of an interrelation between semic analysis and word embeddings could be envisaged as a relevant finding which could provide for an improved recall capability of conceptual particles. A further consideration that could be proposed concerns the possibility to challenge the attributed potentiality of word embeddings to capture semantic information with relation to medical terminology as, in the context of the conducted case study and with specific reference to semic analysis as performed on the basis of the proposed objective methodology, the capability of word embeddings to actually capture semic elements and conceptual specificities could be regarded as limited. As underlined, however, further research should be conducted in order to challenge the semantic dimension ascribed to word embeddings.

By way of conclusion, the present dissertation also lays the groundwork for further investigation. For instance, further exploration would be required in order to assess the possibility to systematise and objectively capture the semantic knowledge which is inscribed in the connotative traits that are equally contemplated in the performance of semic analysis in the form of afferent semes. As a matter of fact, their inherent subjectively-biased nature represents a relevant challenge with respect to its rationalisation that could be deemed as essential to face with a view to exhaustively recollecting all the semic elements which are representative of the conceptual content of terminology. As mentioned, a focused research on this dimension of the conceptual sphere could potentially constitute a valuable improvement in doctor-patient interactions in terms of communication and mutual comprehension, additionally highlight the pivotal function of terminology and its study in nowadays society. In this

sense, this improvement could be attained through a study of semic analysis in the context of a terminological investigation which could also rely on field research.

REFERENCES

Bibliography

Ahmad, K., Davies, A., Fulford, H., Rogers, M.., "What is a term? The semi-automatic extraction of terms from text" in Snell-Hornby, M., Pöchhacker, F., Kaindl, K. (eds.), *Translation Studies: An Interdiscipline: Selected papers from the Translation Studies Congress, Vienna, 9-12 September 1992*, Benjamins translation library v. 2, Amsterdam/Philadelphia: John Benjamins Publishing Company, 1994, pp. 267-278.

Arnold, C., *Synonymy and Register. A Corpus-based Study in American English*, GRIN Verlag, 2015.

Ashi, M. M., Siddiqui, M. A., Nadeem, F., "Pre-trained Word Embeddings for Arabic Aspect-Based Sentiment Analysis of Airline Tweets", in Hassanien, A. E., Tolba, M. F., Shaalan, K., Azar, A. T. (eds.), *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2018*, Vol. 845 of Advances in Intelligent Systems and Computing, Cham: Springer International Publishing, 2019, pp. 241-251.

Baider, F., Constantinou, M., "La fureur de gagner, la rage de perdre. Étude contrastive de colere, rage et futeur en français et en grec moderne", *Études romanes de Brno*, 35(1), 2014, pp. 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, Last accessed on 20 September 2020.

Balteiro, I., Calvo-Ferrer, J. R., "General Remarks on the Challenges of Integrating Scientific and Technical Words in General Dictionaries", in Calvo-Ferrer, J. R., Campos Pardillos, M. Á. (eds.), *Investigating Lexis: Vocabulary Teaching, ESP, Lexicography and Lexical Innovation*, Newcastle upon Tyne: Cambridge Scholars Publishing, 2014, pp. 103-121.

Blake, R. S., *Delmar's Medical Transcription Handbook*, 2nd edition, Clifton Park NY: Cengage Learning, 1998.

Boguski, M.S., "Online Health Information Retrieval by Consumers and the Challenge of Personal Genomics" in Willard, H. F., Ginsburg G. S. (eds.), *Genomic and Personalized Medicine: V1-2*, Vol. 2, 1st edition, San Diego: Academic Press - Elsevier, 2009, pp. 252-257.

Bourigault, D., Slodzian, M. "Pour une terminologie textuelle", *Terminologie Nouvelles*, n. 19, 1999, pp. 29-32.

Bowker, L. "Terminology" in Baker, M., Saldanha, G. (eds.), *Routledge Encyclopedia of Translation Studies*, 2nd edn., London and New York: Routledge, 2009, pp. 286-290.

Brenes, P., *Terminology for beginners*, In My Own Terms, 2017, available at http://inmyownterms.com/wp-content/uploads/2017/06/TERMINOLOGY-eBOOK.pdf, Lst accessed on 17 September 2020.

Cabré, M. T., Sager, J.C. (ed.) *Terminology: Theory, Methods, and Applications*, Terminology and Lexicography Research and Practice 1, Amsterdam/Philadelphia: John Benjamins Publishing Company, 1999.

Choudhari, P, Veenadhari, S., "Sentiment Classification of Online Mobile Reviews Using Combination of Word2vec and Bag-of-Centroids", in Swain, D., Pattnaik, P. K., Gupta, P. K. (eds.), *Machine Learning and Information Processing: Proceedings of ICMLIP 2019*, Vol. 1101 of Advances in Intelligent Systems and Computing, Singapore: Springer Singapore, 2020, pp. 69-80.

Col, G., De Angelis, R., Poibeau, T., "Continuity in the Interactions Between Linguistic Units", in La Mantia, F., Licata, I., Perconti, P. (eds.), *Language in Complexity: The*

Emerging Meaning, Lecture Notes in Morphogenesis, Cham: Springer International Publishing, 2017, pp. 29-48.

Corriveau, J-P., *Time-costrained Memory: A Reader-based Approach To Text Comprehension*, New York, London, Hove: Psychology Press, 2014.

Cortelazzo, M., *Lingue speciali. La dimensione verticale*, seconda edizione, Padova: Unipress, 1994.

Daille, B., *Term Variation in Specialised Corpora: Characterisation, automatic discovery and applications*, Vol. 19 of Terminology and Lexicography Research and Practice, Amsterdam/Philadelphia: John Benjamins Publishing Company, 2017.

Danesi, M., *Language, Society, and New Media: Sociolinguistics Today*, New York/Abingdon: Routledge, 2016.

Dash, N. S., "The Art of Lexicography", in Muhvić-Dimanovski, V., Sočanac, L. (eds.), *Linguistics*, Enclyclopedia of Life Support Systems, Oxford: Eolss Publishers Co. Ltd., 2009, pp. 437-474.

Džuganová, B., "English medical terminology – different ways of forming medical terms", *JAHR – European Journal of Bioethics*, 4(7), 2013, pp. 55-69.

Elezi, S., "Ideological Background of Political and Economic Entries in Explanatory Dictionaries of the Albanian Language", *Journal of Educational and Social Research*, 8(3), 2018, pp. 79-86, https://www.researchgate.net/publication/327862833_Ideological_Background_of_Political_and_Economic_Entries_in_Explanatory_Dictionaries_of_the_Albanian_Language, Last accessed on 20 September 2020.

Filipec, J., “Lexicology and Lexicography: Development and State of the Research”, in Luelsdorff, P. A. (ed.), *The Prague School of Structural and Functional Linguistics: A Short Introduction*, Vol. 41 of Linguistic and Literary Studies in Eastern Europe, Amsterdam/Philadelphia: John Benjamins Publishing Company, 1994, pp. 163-183.

Firth, J.R., *Papers in Linguistics, 1934-1951*, London: Oxford University Press, 1957.

Gadet, F., *La variation sociale en français*, Paris: Éditions Ophrys, 2007.

Galinski, C., Nedobity, W., "Special Languages, Terminology Planning and Standardization" in Strehlow, R. A. (ed.), *Standardization of Techical Terminology: Principles and Practices*, STP 991, Vol. 2. West Conshohocken, PA: ASTM International 1988, pp. 4-13.

Gaussier, E., “General Considerations on Bilingual Term Extraction” in Bourigault, D., Jacquemin, C., L’Homme, M.-C. (eds.), *Recent Advances in Computational Terminology*, Natural Language Processing vol. 2, Amsterdam/Philadelphia: John Benjamins Publishing Company, 2001, pp. 167-183.

Gilreath, C. T., “Resolving Term Disputes with Weighted Onometrics”, in Wright, S. E., Strehlow, R. A. (eds.), *Standardizing and Harmonizing Terminology: Theory and Practice*, Vol. 1223 of ASTM special technical publication.West Conshohocken, PA: ASTM International, 1995, pp. 25-52.

Gotti, M., *Investigating Specialized Discourse*, 2nd edition, Bern: Peter Lang AG, 2008.

Hamawand, Z., *Morphology in English: Word Formation in Cognitive Grammar*, London and New York: Continuum International Publishing Group, 2011.

Harris, Z. S., “Distributional Structure”, *WORD*, 10:2-3 1954, pp. 146-162, available at https://doi.org/10.1080/00437956.1954.11659520, Last accessed 17 September 2020.

Hartley, A. F., Paris, C., “Automatic text generation for software development and use” in Somers H. (ed.) *Terminology, LSP and Translation: Studies in language engineering in honour of Juan C. Sager*, Vol. 18 of Benjamins Translation Library, Amsterdam/Philadelphia: John Benjamins Publishing Company, 1996, pp. 221-242.

Hébert, L., *Tools for Text and Image Analysis: An Introduction to Applied Semiotics*, Texto!, 2011, available at http://www.revue-texto.net/Parutions/Livres-E/Hebert_AS/10.Semic-Analysis.pdf, Last accessed 17 September 2020.

International Organization for Standardization, *Terminology work – Vocabulary – Part 1: Theory and Application*, BS ISO 1087-1:2000.

Jing, H., Tzoukermann, E., “Determining semantic equivalence of terms in information retrieval: An approach based on context distance and morphology”, in Bourigault, D., Jacquemin C., L’Homme, M.-C. (eds.), *Recent Advances in Computational Terminology*, Natural Language Processing vol. 2, Amsterdam/Philadelphia: John Benjamins Publishing, 2001, pp. 245-260.

Kabir, K. L., Alam, F. F., Islam, A. B., "Word Embeddings for Semantic Resemblance of Substantial Text Data: A Comparative Study", in Somani, A. K., Shekhawat, R. S., Mundra, A., Srivastava, S., Verma, V. K. (eds.), *Smart Systems and IoT: Innovations in Computing: Proceeding of SSIC 2019*, Vol. 141 of Smart Innovation, Systems and Technologies, Singapore: Springer Singapore, 2020, pp. 303-312.

Khathchadourian, H., *A Critical Study in Method.* Springer, Dordrecht. 1967.

Khattak, F. K., Jeblee, S., Pou-Prom, C., Abdalla, M., Meaney, C., Rudzicz, F., "A survey of word embeddings for clinical text", *Journal of Biomedical Informatics: X*, Vol. 4, Elsevier Inc., 2019.

Kurdi, M. Z., *Natural Language Processing and Computational Linguistics 2: Semantics, Discourse and Applications*, Cognitive Science Series, Hoboken-London: Wiley-ISTE, 2018.

Kuzio, A., "Difficulties resulting from language diversity in teaching medical translation and methods to overcome them when teaching medical English to future translators", *Language Value*, 11(7), 2019, pp. 23-44.

Kwiatek, E., *Contrastive Analysis of English and Polish Surveying Terminology*, Newcastle upon Tyne: Cambridge Scholars Publishing, 2013.

Leigh, H., Reiser, M. F., *The Patient: Biological, Psychological, and Social Dimensions of Medical Practice*, 2nd ed., New York: Plenum Medical Book Company, 1985.

Lenci, A., "Spazi di parole: metafore e rappresentazioni semantiche", *Paradigmi*, 27, 2009, pp. 83-100.

León-Araúz, P., "Term and concept variation in specialised knowledge dynamics" in Drouin, P., Francœur, A., Humbley, J., Picton, A. (eds.), *Multiple Perspectives on Terminological Variation*, Vol. 18 of Terminology and Lexicography Research and Practice, Amsterdam & Philadelphia: John Benjamins Publishing Company, 2017, pp. 213-258.

L'Homme, M.-C., *Lexical Semantics for Terminology: An introduction*, Vol 20 of Terminology and Lexicography Research and Practice, Amsterdam/Philadelphia: John Benjamins Publishing Company, 2020.

Lindén, K., "Finding Cross-Lingual Spelling Variants", in Apostolico, A., Melucci, M. (eds.), *String Processing and Information Retrieval: 11th International Conference, SPIRE 2004, Padova, Italy, October 5-8, 2004 Proceedings*, Volume 11, Volume 3246

of Lecture Notes in Computer Science, Berlin/Heidelberg: Springer-Verlag Berlin Heidelberg, 2004, pp. 136-137.

Löbner, S., *Understanding Semantics*, second edition, London and New York: Routledge, 2013.

Maglie, R. B., *Understanding the Language of Medicine*, Roma: ARACNE editrice S.r.l., 2009.

Malrieu, J. P., *Evaluative Semantics: Cognition, Language and Ideology*, Routledge Frontiers of Cognitive Science, London and New York: Routledge. Taylor & Francis e-Library, 2002.

Manset, L., "*Existe-t-il une synonymie parfaite? Étude au moyen des notions de sens dénotatif, de connotation socio-pragmatique et de variation linguistique Approches théorique et pratique*", Liège: Université de Liège, 2019.

Marinkovich, J., "Palabra y término: ¿Diferenciación o complementación?", *Revista Signos*, 41(67), 2008, pp. 119-126.

Mikolov, T., Chen, K., Corrado, G., Dean, J., "Efficient Estimation of Word Representations in Vector Space", 2013, available at https://arxiv.org/pdf/1301.3781.pdf (last accessed 17 September 2020)

Nöth, W., *Handbook of Semiotics*, Bloomington and Indianapolis: Indiana University Press, 1995.

Pavel, S., Nolet, D., *Handbook of Terminology*, Québec: Minister of Public Works and Government Services Canada, 2001.

Pennington, J., Socher, R., Manning, C., "GloVe: Global Vectors for Word Representation", in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, October 25-29, 2014, Doha, Qatar, Association for Computational Linguistics, pp. 1532-1543.

Pérez Hernández, C., "Explotación de los corpora textuales informatizados para la creación de bases de datos terminológicas basadas en en conocimiento", *Estudios de Lingüística Española (EliEs)*, 18, available at http://elies.rediris.es/elies18/ (last accessed 24 August 2020).

Pottier, B., "The Components of Communication", in Gill, H. S., Pottier, B. (eds.), *Ideas, Words, and Things: French Writing in Semiology*, New Delhi: Orient Longman, 1992, pp. 113-135.

Powell, F. C., *Rosacea: Diagnosis and Management*, New York: CRC Press, 2008.

Qi, F., Lin, Y., Sun, M., Zhu, H., Xie, R., Liu, Z., "Cross-lingual Lexical Sememe Prediction", in Riloff, E., Chiang, D., Hockenmaier, J., Tsujii, J. (eds.), *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Brussels, Belgium, Association for Computational Linguistics, pp. 358-368, https://www.aclweb.org/anthology/D18-1033.pdf, Last accessed on 20 September 2020.

Rastier, F., "Principes et conditions de la sémantique componentielle" in Parret, H., Ruprecht, H.-G. (eds.), *Aims and Prospects of Semiotics: Essays in honor of Algirdas Julien Greimas*, Amsterdam/Philadelphia: John Benjamins Publishing Company, 1985, pp. 505-528.

Rastier, F., "La microsémantique" in *Texto!*, vol. X n.2, June 2005, available at http://www.revue-texto.net/Inedits/Rastier/Rastier_Microsemantique.html (last accessed 02 June 2020)

Rastier, F., *Sémantique Interprétative*. Formes sémiotiques, Paris: Presses Universitaires de France, 2009.

Rastier, F., *Sémantique et recherches cognitives*. Formes sémiotiques, Paris: Presses Universitaires de France, 2010.

Rey, A., Sager, J. C. (ed.), *Essays on Terminology*, Benjamins translation library v.9, Amsterdam/Philadelphia: John Benjamins Publishing Company, 1995.

Roche, C., "Should Terminology Principles be re-examined?", in Aguado de Cea, G., Suárez-Figueroa, M. C., García-Castro, R., Montiel-Ponsoda, E. (eds.), *Proceedings of the 10th Terminology and Knowledge Engineering Conference: New frontiers in the constructive symbiosis of terminology and knowledge engineering* (TKE 2012), 19-22 June 2012, Madrid: Universidad Politécnica de Madrid, 2012, pp. 17-32.

Sager, J. C., *A Practical Course in Terminology Processing*, Amsterdam/Philadephia: John Benhamins Publishing, 1996 [1990].

Saidani, N., Adi, K., Allili, M. S., "Semantic Representation Based on Deep Learning for Spam Detection", in Benzekri, A., Barbeau, M., Gong, G., Laborde, R., Garcia-Alfaro, J. (eds.), *Foundations and Practice of Security: 12th International Symposium, FPS 2019, Toulouse, France, November 5-7, 2019, Revised Selected Paper*, Vol. 12056 of Lecture Notes in Computer Science, Cham: Springer International Publishing. 2020, pp. 72-81.

Santos, C. & Costa, R., "Domain specificity: Semasiological and onomasiological knowledge representation" in Kockaert, H. J., Steurs, F. (eds), *Handbook of Terminology, Volume 1*, Amsterdam/Philadelphia: John Benjamins Publishing Company, 2015, pp. 118-124.

Schmit, L.-M., *Les définitions en droit privé*, Thèses de l'IFR, Toulouse: Presses de l'Université Toulouse 1 Capitole, LGDJ – Lextenso Editions, 2018.

Sfakakis, M., Papachristopoulos, L., Zoutsou, K., Tsakonas, G., Papatheodorou, C., "Automated Subject Indexing of Domain Specific Collections Using Word Embeddings and General Purpose Thesauri", in Garoufallou, E., Fallucchi, F., De Luca, E. W. (eds.), *Metadata and Semantic Research: 13th International Conference, MTSR 2019, Rome, Italy, October 28-31, 2019, Revised Selected Papers*, Vol. 1057 of Communications in Computer and Information Science, Cham: Springer International Publishing, 2019, pp. 103-114.

Shuttleworth, M., Cowie, M., *Dictionary of Translation Studies*, Abingdon and New York: Routledge, 2014.

Skelac, I., Jandrić, A., "Meaning as Use: From Wittgenstein to Google's Word2vec", in Skansi S. (ed.), *Guide to Deep Learning Basics: Logical, Historical and Philosophical Perspectives*, Cham: Springer International Publishing, 2020, pp. 41-53.

Smyth, M. M., Collins, A. F., Morris, P. E., Levy, P., *Cognition in Action*, 2nd edition, Hove and Hillsdale: Lawrence Erlbaum Associates, Publishers, 1994.

Steinberg, S., *An Introduction to Communication Studies*, Cape Town: Juta & Company Ltd, 2007.

Temmerman, R., *Towards New Ways of Terminology Description: The Sociocognitive Approach*, Amsterdam/Philadelphia: John Benjamins Publishing, 2000.

Tribble, C., "Improvising corpora for ELT: Quick-and-dirty ways of developing corpora for language teaching", in Melia, J., Lewandowska-Tomaszczyk, B. (eds.) *PALC '97 Proceedings, Practical Applications in Language Corpora*, Lodz: Lodz University Press, 1997, pp. 106-117.

Trudel, E., "Éléments de synthèse en sémantique interprétative — Unités thématiques et expressives et approche morphosémantique d'une production sémiotique", in *Texto!*, Volume XIV n. 2, 2009, available at http://www.revue-texto.net/docannexe/file/2284/trudel_synthesesemantique.pdf (last accessed 24 August 2020).

Turgeon, M. L., *Linné & Ringsrud's Clinical Laboratory Science – E-Book: The Basics and Routine Techniques*, 6th ed., Maryland Heights: Elsevier Mosby, 2011.

Vezzani, F., Di Nunzio, G. M., Henrot, G., "TriMED: banca dati terminologica multilingue", in Spampinato D. (ed.), *AIUCD 2018 – Book of Abstracts*, Bologna, 2018, pp. 237-241.

Vinay, J.-P., Darbelnet, J., *Stylistique comparée du français et de l'anglais : méthode de traduction*. Paris : Didier, 1958.

Vinay, J.-P., Darbelnet, J.. Sager, J. C. and M.-J. Hamel, (eds.) *Comparative stylistics of French and English: a methodology for translation*, Amsterdam/Philadelphia: John Benjamins Publishing Company, 1995.

Vitali, E. D., "Il linguaggio delle scienze biomediche" in *Atti del Convegno; Il linguaggio della divulgazione*, Selezione dal Reader's Digest, Milano, 1983, pp. 185-198.

Wang, M., *Linguistic Semiotics*, Vol. 3 of Peking University Linguistics Research, Singapore: Springer Singapore, 2020.

White, L., Togneri, R., Liu, W., Bennamoun, M., *Neural Representations of Natural Language*, Vol. 783 of Studies in Computational Intelligence, Singapore: Springer Singapore, 2019.

Wong, W., Liu, W., Bennamoun, M., “Determination of Unithood and Termhood for Term Extraction”, in Song, M., Brook Wu, Y.-f. (eds.), *Handbook of Research on Text and Web Mining Technologies*, Hershey: IGI Global, 2008, pp. 500-529.

Woywodt, A., Matteson, E., “Should eponyms be abandoned? Yes”, *British Medical Journal*, 335(7617):424, 2007, pp. 424-425.

Xie, R., Yuan, X., Liu, Z., Sun, M., “Lexical Sememe Prediction via Word Embeddings and Matrix Factorization” in Sierra, C. (ed.) *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17)*, Melbourne, Australia 19-25 August 2017, International Joint Conferences on Artificial Intelligence, pp. 4200-4206, https://www.researchgate.net/publication/318829690_Lexical_Sememe_Prediction_via_Word_Embeddings_and_Matrix_Factorization, Last accessed on 20 September 2020.

Yang, H., Luo L., Chueng, L. P., Ling, D., Chin, F., "Deep Learning and Its Applications to Natural Language Processing" in Huang, K., Hussain, A., Wang, Q.-F., Zhang, R. (eds.), *Deep Learning: Fundamentals, Theory and Applications*, Cognitive Computation Trends 2, Cham: Springer International Publishing, 2019, pp. 89-109.

Sitography

Braun-Falco O., Plewig G., Wolff H. H., Winkelmann R. K., *Dermatology*, Springer-Verlag Berlin Heidelberg, 1991, https://books.google.it/books?id=4Xn1CAAAQBAJ&pg=PA348&lpg=PA348&dq=A+chronicrecurrent+inflammatory+skin+disease+with+marked+pruritus+and+variability+in+lesion+morpholog
y&source=bl&ots=3raERAAvjZ&sig=ACfU3U0KUaTyif0INpJ7M-7igMJflLV6Dw&hl=it&sa=X&ved=2ahUKEwj-od-X-

7zmAhUoxKYKHWRLCLwQ6AEwCXoECAsQAQ#v=onepage&q=A%20chronic-recurrent
%20inflammatory%20skin%20disease%20with%20marked%20pruritus%20and
%20variability%20in
%20lesion%20morphology&f=false, Last accessed on 20 September 2020.

Dictionary.com. Dictionary.com, LLC., 2020, *Diagnosis*, https://www.dictionary.com/browse/diagnosis, Last accessed on 20 September 2020.

Collins Dictionary. Collins, 2020, *Antiviral*, https://www.collinsdictionary.com/it/dizionario/inglese/antiviral, Last accessed on 20 September 2020.

Collins Dictionary. Collins, 2020, *Asthma*, https://www.collinsdictionary.com/it/dizionario/inglese/asthma, Last accessed on 20 September 2020.

Collins Dictionary. Collins, 2020, *Injection*, https://www.collinsdictionary.com/dictionary/english/injection, Last accessed on 20 September 2020.

Collins Dictionary. Collins, 2020, *Meningitis*, https://www.collinsdictionary.com/it/dizionario/inglese/meningitis, Last accessed on 20 September 2020.

Collins Dictionary. Collins, 2020, *Vaccination*, https://www.collinsdictionary.com/dictionary/english/vaccination, Last accessed on 20 September 2020.

Collins Dictionary. Collins, 2020, *Virology*, https://www.collinsdictionary.com/it/dizionario/inglese/virology, Last accessed on 20 September 2020.

Corriere della Sera, 2020, RCS Mediagroup S.p.a., *D-Dimero*, https://www.corriere.it/salute/esami-sangue/cuore-metabolismo/d-dimero_e73d9ce0-dc4a-11df-be1f-00144f02aabc.shtml, Last accessed on 20 September 2020.

Diccionario de cáncer del NCI, https://www.cancer.gov/espanol/publicaciones/diccionario, Last accessed on 20 September 2020.

Diccionario de la lengua española. Real Academia Española, 2020, *circulación*, https://dle.rae.es/circulaci%C3%B3n?m=form, Last accessed on 20 September 2020.

Diccionario de la lengua española. Real Academia Española, 2020, *efecto*, https://dle.rae.es/efecto#MqUF2ND, Last accessed on 20 September 2020.

Diccionario médico. Clínica Universidad de Navarra. 2020, https://www.cun.es/diccionario-medico, Last accessed on 20 September 2020.

Dizionario di Medicina, Treccani, http://www.treccani.it/enciclopedia/elenco-opere/Dizionario_di_Medicina, Last accessed on 20 September 2020.

Enciclopedia salute, Ministero della Salute, http://www.salute.gov.it/portale/salute/p1_3.jsp?lingua=italiano&tema=Salute_A_Z, Last accessed on 20 September 2020.

Google Code Archive, https://code.google.com/archive/p/word2vec/, Last accessed 20 September 2020.

Heymans Institute of Pharmacology and Mercator School, Department of Applied Linguistics, *Multilingual Glossary of technical and popular medical terms in nine European Languages*, Last updated on 3 June 2000,

https://users.ugent.be/~rvdstich/eugloss/welcome.html, Last accessed on 20 September 2020.

Introduction to Information Retrieval, https://nlp.stanford.edu/IR-book/information-retrieval-book.html, Last accessed on 21 September 2020.

Lapi, F., Bianchini, E., Michieli, R., Pasqua, A., Cricelli, I., Mazzaglia, G., Frediani, B., Prieto-Alhambra, D., Brandi, M. L., Cricelli, C., "Assessing Risk of Osteoporotic Fractures in Primary Care: Development and Validation of the FRA-HS Algorithm", *Calcified Tissue International*, 100(6):537-549, 2017, Available at https://europepmc.org/article/MED/28160026, Last accessed on 20 September 2020.

Merriam-Webster Dictionary. Merriam-Webster, Incorporated. 2020, *Epidemic*, https://www.merriam-webster.com/dictionary/epidemic, Last accessed on 20 September 2020.

Merriam-Webster Medical Dictionary. Merriam-Webster, Incorporated. 2020, https://www.merriam-webster.com/medical, Last accessed on 20 September 2020.

Merriam-Webster Medical Dictionary. Merriam-Webster, Incorporated. 2020, *Fibromyalgia*, https://www.merriam-webster.com/dictionary/fibromyalgia#medicalDictionary, Last accessed on 20 September 2020.

New York Post, "Spain's Puyol heads home winner over Germany", Associated Press, 2010, https://nypost.com/2010/07/08/spains-puyol-heads-home-winner-over-germany/ Last accessed on 20 September 2020.

Pennington, J., Socher, R., Manning, C. D., GloVe: Global Vectors for Word Representation, https://nlp.stanford.edu/projects/glove, Last accessed on 20 September 2020.

PubMed. National Library of Medicine, National Center for Biotechnology Information, https://pubmed.ncbi.nlm.nih.gov, Last accessed on 20 September 2020.

TheFreeDictionary's Medical Dictionary, Farlex, 2003-2020, https://medical-dictionary.thefreedictionary.com/, Last accessed on 20 September 2020.

TheFreeDictionary's Medical Dictionary, Farlex, 2003-2020, *Baggage*, https://medical-dictionary.thefreedictionary.com/baggage, Last accessed on 20 September 2020.

TheFreeDictionary's Medical Dictionary, Farlex, 2003-2020, *Dysmenorrhea*, https://medical-dictionary.thefreedictionary.com/dysmenorrhea, Last accessed on 20 September 2020.

TheFreeDictionary's Medical Dictionary, Farlex, 2003-2020, *Glaucoma*, https://medical-dictionary.thefreedictionary.com/glaucoma, Last accessed on 20 September 2020.

TheFreeDictionary's Medical Dictionary, Farlex, 2003-2020, *Herpes simplex*, https://medical-dictionary.thefreedictionary.com/herpes+simplex, Last accessed on 20 September 2020.

TheFreeDictionary's Medical Dictionary, Farlex, 2003-2020, *Presentation*, https://medical-dictionary.thefreedictionary.com/presentation, Last accessed on 20 September 2020.

TheFreeDictionary's Medical Dictionary, Farlex, 2003-2020, *Terminal*, https://medical-dictionary.thefreedictionary.com/terminal, Last accessed on 20 September 2020.

Treccani, 2020, "Alzheimer, morbo di" in *Treccani Enciclopedia on line*, http://www.treccani.it/enciclopedia/morbo-di-alzheimer/, Last accessed on 20 September 2020.

Treccani, 2020, "Diagnostica per immagini", in Passariello R., Alger E., Marin D., Napoli A., *Enciclopedia Italiana* – VII Appendice 2006, http://www.treccani.it/enciclopedia/diagnostica-per-immagini_%28Enciclopedia-Italiana%29/, Last accessed on 20 September 2020.

Wikipedia 2020, *PET-CT*, last updated on 20 February 2019 https://en.wikipedia.org/wiki/PET-CT, Last accessed on 20 September 2020.

RIASSUNTO IN ITALIANO

Il presente lavoro di tesi trova fondamento nell'obiettivo di teorizzare e proporre una metodologia che consenta un approccio oggettivo all'applicazione dell'analisi semica, definibile come una tecnica linguistica inscritta nell'ambito della semantica composizionale la cui funzione principale è delineare univocamente ed esaustivamente il contenuto concettuale espresso a livello lessicale da una determinata entità linguistica. Conseguentemente alla proposta di una prospettiva metodologica, lo studio si focalizza sulla ricerca di una possibile sussistente interrelazione tra analisi semica e word embeddings, i quali possono essere considerati come una rappresentazione vettoriale di unità lessicali, consentendo contestualmente un approfondimento sul versante applicativo dell'analisi stessa. Per quanto concerne la dimensione prettamente terminologica, il filo conduttore è costituito dallo studio della terminologia medica, la quale si inscrive nell'ambito della presente tesi come dominio linguistico di riferimento e oggetto di relative considerazioni e riflessioni.

Il primo capitolo, concepito come introduzione nozionale al contesto terminologico in cui si situa il presente elaborato, presenta una panoramica concettuale sulla natura teorica della *terminologia* evidenziando in primo luogo le differenti concettualizzazioni che sottendono all'impiego polisemico del termine nell'ambito linguistico e ponendo in rilievo la funzione preponderante che essa svolge anche rispetto a contesti ed ambiti di studio eterogenei quali la traduzione specializzata, l'interpretariato e il settore del reperimento dell'informazione. Successivamente, la considerazione di varie concezioni relative alla definizione di *termine* alla luce delle maggiori correnti teoriche terminologiche si integra ad un approfondimento sulle diverse rappresentazioni lessicali che il termine può assumere. In particolar modo, la riflessione proposta si focalizza sull'individuazione delle possibili difficoltà che possono sorgere nel processo di identificazione dei termini complessi nella completezza strutturale della determinata composizione lessicale. In relazione a ciò, una prima circostanza rispetto alla quale è ipotizzabile il riscontro della menzionata complessità è

la tendenza nel dominio della terminologia medica all'inclusione di elementi linguistici o segni grafici che possono indurre ad un errato riconoscimento della concatenazione terminologica caratteristica delle entità complesse. Una seconda contingenza che potrebbe essere implicata nel riscontro di difficoltà nell'individuazione della struttura complessiva della suddetta tipologia lessicale e dunque del concetto come unica entità è l'ipotizzabile frequente associazione tra il concetto di parola ed il concetto insito nella tokenizzazione, inteso come la concezione per cui nell'ambito delle lingue che prevedono un processo di segmentazione, le parole sono distinte a livello ortografico mediante spazi. Le due prospettive possono inoltre convergere in una singola ipotesi, definibile come la tendenza abitudinaria a focalizzare progressivamente l'attenzione sui singoli termini di cui un termine complesso si compone e in misura minore sul concetto generale da esso rappresentato. In questo senso, diversi livelli di astrazione possono essere percepiti nella comprensione di termini semplici rispetto a termini complessi e, conseguentemente, nell'interpretazione di concetti correlati a termini composti da un solo elemento rispetto a concetti che, trovando rappresentazione in termini complessi, prevedono una natura composizionale.

Il secondo capitolo propone un approfondimento relativo alla terminologia medica intesa nell'accezione di congiunto di entità terminologiche di cui si compone il linguaggio medico. In primo luogo, un'introduzione nozionale relativa all'inclusione dello stesso nell'ambito delle lingue speciali viene presentata con il proposito di fornire uno sguardo d'insieme sul contesto linguistico nel quale si inscrive. Successivamente, la considerazione della varietà dei contesti in cui è impiegata, specificamente la comunicazione tra esperti del settore e l'interazione tra medico e paziente, conduce ad osservazioni sulle dissimili necessità comunicative e sul differente livello di specificità che è insito nella dimensione concettuale di termini specializzati e termini popolari. In particolare una riflessione viene proposta in relazione ai termini popolari, ai quali viene ascritto un contenuto concettuale derivante da una strategia di adattamento a fini pragmatici che marca una differenza rispetto alla natura terminologica e concettuale che caratterizza i termini specializzati. La suddetta strategia di adattamento può inoltre essere metaforicamente rappresentata come una traduzione orientata alla comunicazione

che si muove dal concetto specializzato di partenza al concetto meno tecnico di arrivo, consentendo dunque una semplificazione. Un elemento che può complicare la comprensione è l'uso di abbreviazioni che, pur apportando maggiore concisione rispetto all'impiego di termini complessi nella forma estesa, può ampliare il divario conoscitivo del dominio previamente sussistente tra medico e paziente e, generalmente, suscitare la percezione di distanza concettuale del paziente rispetto al linguaggio medico. Tra le caratteristiche ordinariamente ascritte alla terminologia medica rientrano i concetti di assenza di ambiguità, precisione referenziale e monosemia; tuttavia, il principio ideale di monoreferenzialità spesso non si realizza nell'ambito del medesimo dominio di riferimento a livello pratico, considerando anche l'inscrizione della terminologia nel più ampio sistema linguistico che presenta intrinsecamente fenomeni di ambiguità. L'uso di termini medici avviene inoltre in contesti d'uso quotidiani, maturando dunque una connessione con il linguaggio comune e affermando la propria funzione essenziale nel consentire la soddisfazione di esigenze informative relative a questioni mediche. In relazione a ciò si pone la necessità di attuare una distinzione tra termini ai quali è unicamente associata la denotazione e termini che presentano denotazione e connotazione, tra i quali figurano termini specializzati utilizzati in contesti non specializzati. Generalmente è possibile ipotizzare la necessità di raggiungere un equilibrio a livello di esigenze comunicative, per cui la standardizzazione può essere considerata essenziale nell'ambito della terminologia specializzata, in particolare per quanto concerne le abbreviazioni. Il concetto di trasparenza inoltre può potenzialmente assumere una sfumatura concettuale differente, concepibile come chiarezza nell'uso della terminologia in prospettiva comunicativa.

Il terzo capitolo presenta un approfondimento relativo all'analisi semica, il cui approccio teorico di riferimento nell'ambito della presente tesi è rappresentato dalla semantica interpretativa concepita da François Rastier[49]. Specificamente, l'analisi semica viene contestualmente adottata come tecnica linguistica per pervenire ad uno studio dettagliato della dimensione concettuale insita nella terminologia. La

49Rastier F., *Sémantique Interprétative*. Formes sémiotiques, Paris, Presses universitaires de France, 2009.

presentazione di una mappa concettuale relativa alle differenti tipologie di semi implicate nella tecnica linguistica prelude alla postulazione di differenti passaggi che, a livello cognitivo e applicativo, si possono individuare nello svolgimento dell'analisi stessa. In primo luogo, la comprensione approfondita del concetto unitariamente concepito rappresentato a livello lessicale da un termine dovrebbe essere acquisita mediante la consultazione di risorse lessicografiche quali dizionari. Successivamente, il discernimento a livello cognitivo degli elementi concettuali che compongono il concetto dovrebbe essere attuato. La suddetta elaborazione dovrebbe dunque essere trasposta in forma lessicale ed espressa considerando inoltre la dimensione metalinguistica propria dell'articolazione formale dell'analisi semica, costituendo a questo punto un congiunto di semi. In ultimo luogo, una fase di validazione dovrebbe essere contemplata la quale, pur costituendo un passaggio addizionale rispetto allo svolgimento dell'analisi il cui prodotto è generato nella fase precedente, consente di valutare la sussistente coerenza concettuale che dovrebbe essere instaurata tra il concetto unitario e l'espressione dello stesso nel semema originato. L'ipotesi relativa alla sussistenza di una consistente componente soggettiva nello svolgimento dell'analisi semica è stata confermata nel contesto di un caso studio condotto su dati estratti da schede terminologiche basate sul modello TriMED[50] compilate da due diversi gruppi di studenti nell'ambito del corso Tecnologie per la Traduzione, offerto nel Corso di Studio Magistrale in Lingue Moderne per la Comunicazione e la Cooperazione Internazionale presso l'Università degli Studi di Padova. Lo studio ha interessato tre differenti lingue caratterizzandosi dunque per la propria natura interlinguistica; specificamente l'analisi è stata condotta considerando 460 formulazioni di analisi semica svolte in lingua inglese, 343 in italiano e 101 in spagnolo. In particolare, è stato possibile osservare l'esigua inclusione di varianti semantiche tra i semi selezionati, l'adozione di una prospettiva prettamente denotativa, l'esclusione della dimensione connotativa e in particolar modo la similarità di strategie adottate per lo svolgimento dell'analisi. Il suddetto caso studio ha infatti contestualmente permesso di osservare una forte tendenza all'adozione di un approccio consistente nell'estrazione terminologica di entità lessicali e concettuali comprese nelle

50 Vezzani F., Di Nunzio G. M., Henrot G., "TriMED: banca dati terminologica multilingue", in Spampinato D. (ed.), *AIUCD 2018 – Book of Abstracts*, Bologna, 2018, pp. 237-241.

definizioni presenti in dizionari relative al termine osservato al fine di determinare i semi relazionati al concetto di riferimento. Pur costituendo un criterio di selezione considerevolmente utilizzato, è stato possibile notare che l'impiego spesso riscontrato di una sola definizione ha originato l'inclusione in forma di semi di entità lessicali e concettuali strettamente dipendenti dalla selezione e dalla formulazione della definizione stessa, generando dunque differenze nell'espressione del concetto nel semema relativo ad uno stesso termine di riferimento. In relazione a ciò, è possibile ipotizzare che la presenza o assenza di elementi connotativi nel semema possa essere collegata all'inclusione o meno della dimensione connotativa nella definizione selezionata. Considerando le diverse manifestazioni di soggettività osservate, comprendenti inoltre scelte divergenti a livello morfologico, il capitolo presenta una proposta relativa ad una metodologia oggettiva per lo svolgimento dell'analisi semica volta ad apportare maggiore sistematicità nell'espressione concettuale e terminologica di concetti. In particolare, al fine di limitare la suddetta soggettività anche relazionata alla conoscenza individuale di un determinato concetto, si propone un approccio basato su dati mediante la costruzione di un corpus composto da definizioni tratte da dizionari specializzati relativi ad un determinato dominio, dalle quali attraverso una procedura di estrazione terminologica e l'applicazione del concetto di termicità, è possibile selezionare entità lessicali e concettuali appartenenti esclusivamente al dominio terminologico di riferimento trasposti successivamente in forma di semi. A tale scopo, la tipologia di definizione che viene proposta per lo svolgimento dell'analisi semica è la definizione intensionale. Generalmente, le potenzialità che la metodologia proposta presenta sono molteplici, tra le quali la sussistenza della possibilità di derivare una conoscenza articolata dei concetti che compongono il dominio medico e trascendere dunque la conoscenza individuale che potrebbe costituire un fattore sostanziale di soggettività. In questo senso, l'analisi semica può essere concepita come una strategia di apprendimento che incrementa la comprensione della terminologia e della rete concettuale insita in un determinato dominio. Oltre a ciò, la considerazione di varie definizioni tratte da risorse lessicografiche diversificate può potenzialmente permettere di ottenere una rappresentazione di un determinato concetto indipendente dalla conoscenza contenuta in una singola fonte e, contestualmente, inglobare nel semema

entità concettuali e terminologiche derivanti da molteplici risorse consentendo una collezione esaustiva di semi. Inoltre, l'adozione di tale approccio può potenzialmente incrementare la possibilità di corrispondenze a livello lessicale e concettuale tra sememi formulati da differenti individui, consentendo quindi il raggiungimento di maggiore uniformità nell'espressione di un determinato concetto. La potenzialità offerta dall'analisi semica di rappresentare in modo univoco un concetto specifico può inoltre essere consolidata mediante lo svolgimento di analisi terminologiche parallele di termini che manifestano prossimità concettuale ad esso, consentendo quindi di collocare precisamente il termine di riferimento nell'ambito semantico e catturarne le specificità rispetto ad altre entità terminologiche. Un esempio di ciò può essere costituito dalla considerazione del fatto che, al fine di comprendere le peculiarità concettuali del termine *dermatite rosaceiforme*, può essere utile effettuare una comparazione rispetto a termini come *dermatite seborroica*, *dermatite atopica* e *dermatite da contatto*. Tale procedimento può consentire anche di pervenire ad una maggiore comprensione e acquisizione di informazioni che possono essere espresse in forma di semi generici, con il proposito di raggiungere un'adeguata conoscenza concettuale che si pone come ulteriore fine della suddetta metodologia. Inoltre, la metodologia presentata pone l'obiettivo di considerare a livello concettuale e lessicale le dimensioni diastratica e diafasica mediante l'inclusione selettiva nel semema di entità specializzate per esprimere concetti relativi a termini specializzati o, in alternativa, di unità che presentano un livello di tecnicità meno elevato come ad esempio i termini popolari, inglobando dunque nella concettualizzazione adottata la considerazione della dimensione comunicativa e pragmatica nel contesto del dominio medico.

Il quarto capitolo propone la ricerca di una possibile interrelazione e sinergia tra analisi semica e word embeddings, con l'obiettivo primario di verificare la sussistenza della possibilità di condurre l'analisi semica con riferimento alla terminologia medica in modo automatico. In primo luogo, il concetto di word embeddings viene presentato mediante un'introduzione nozionale che si focalizza inoltre sulla concettualizzazione che sottende all'ipotesi distribuzionale e, successivamente, sulla descrizione di due modelli principali di word embeddings, Word2Vec e GloVe. Il seguente sottocapitolo

presenta un caso studio condotto in relazione ad alcuni termini di riferimento analizzati nel capitolo precedente e i relativi sememi, e termini ad essi relazionati, considerati nell'ambito dell'applicazione dei word embeddings semanticamente vicini ai termini di riferimento. I dati rispetto ai quali è stata condotto il caso studio, così come le schede terminologiche analizzate nel terzo capitolo, sono stati resi disponibili nello specifico per lo svolgimento della presente tesi dal Prof. Giorgio Maria Di Nunzio. L'analisi condotta consiste nell'adozione di differenti approcci allo studio di una possibile relazione tra le due differenti tecniche, in particolare è possibile individuare principalmente due prospettive. In primo luogo è stata condotta un'analisi comparativa consistente in un confronto tra i semi contenuti nei sememi dei termini di riferimento formulati sulla base della metodologia proposta e le rispettive 50 entità linguistiche che, nell'ambito dell'applicazione dei word embeddings, sono state rilevate quali maggiormente prossime ai vettori dei termini di riferimento. Alcune delle osservazioni tratte dalla suddetta analisi riguardano la constatazione della presenza di forme flesse di diverse entità lessicali espresse sotto forma di semi nelle differenti liste di termini considerati semanticamente vicini, la sussistenza di casi in cui i termini complessi presenti nelle liste di unità lessicali semanticamente vicine presentano un'aggregazione di termini espressi da semi e l'inclusione di termini compresi nei sememi quali elementi costitutivi dei termini complessi contenuti nelle menzionate liste generate mediante word embeddings. Un'ulteriore manifestazione è costituita dalla presenza, nelle suddette liste, di varianti morfologiche di termini coinvolti nella formulazione di sememi dei termini di riferimento. Sulla base di queste considerazioni, è stato possibile confermare la sussistenza della potenzialità ascrivibile all'analisi semica di intercettare anche termini che sono considerati semanticamente relazionati ai termini analizzati e, contestualmente, la capacità della tecnica dei word embeddings di catturare alcune entità comprese nei sememi di riferimento. Questa prima fase contemplata nell'ambito del caso studio ha permesso di pervenire all'ipotesi di considerare selettivamente alcune entità originate dall'applicazione dei word embeddings, al fine di includere ulteriori particelle concettuali nella formulazione dell'analisi semica; nello specifico, la considerazione è rivolta a termini che presentano una correlazione a livello morfologico o forme flesse dei termini rappresentati in forma di semi compresi nei sememi. In

secondo luogo, l'analisi si è concentrata sulle prime dieci entità linguistiche che risultano essere più simili a livello semantico rispetto ai termini di riferimento; in particolare, lo studio di tali entità è stato condotto mediante lo svolgimento dell'analisi semica adottando la metodologia precedentemente esplicata. In seguito ad una valutazione del contenuto concettuale rappresentato dalla terminologia analizzata, è stato possibile constatare che la presenza di un termine segnalato come semanticamente vicino può anche essere indicativa della natura polisemica del termine di riferimento e che questi ultimi sono inclusi nel semema delle entità analizzate in poche circostanze. Oltre a ciò, è stato possibile osservare che molti termini analizzati presentano una connessione con i termini di riferimento a livello di comune appartenenza ad un determinato campo semantico; tuttavia, la suddetta tipologia di relazione non viene considerata un fattore sufficiente ai fini dell'inclusione dei termini nel semema dei termini di riferimento. Una precisazione che è stata altresì proposta riguarda la scelta di includere nell'analisi semica semi corrispondenti ad iperonimi piuttosto che ad iponimi sulla base di una considerazione relativa al fatto che gli iponimi apportano elementi concettuali che esulano dalla natura più marcatamente generica dei termini di riferimento. Sulla base delle formulazioni di analisi semica generate in questa fase è stata inoltre condotta una ulteriore analisi al fine di verificare la sussistenza della possibilità di individuare un super-insieme di semi comune al termine di riferimento e ai termini ad esso relazionati. Tale approccio ha però condotto alla constatazione dell'impossibilità di costituire tale super-insieme a causa della presenza di corrispondenze parziali tra elementi la cui occorrenza si manifesta solamente in relazione ad alcuni dei termini considerati ad eccezione del solo sema ricorrente in ogni analisi semica rappresentato dal sema mesogenerico. Infine, è stato possibile notare che i termini selezionati tra le entità considerate semanticamente vicine ai termini di riferimento che possono essere inclusi in forma di semi nei rispettivi sememi manifestano nella maggioranza dei casi una connessione a livello morfologico rispetto ai termini di riferimento o rappresentano forme flesse degli stessi. Sulla base di questa osservazione, si ritiene che l'analisi semica della terminologia medica svolta mediante l'applicazione dei criteri individuati nella metodologia proposta non possa essere condotta in modo automatico, poiché l'adozione di un approccio automatico non

consentirebbe di includere la totalità dei semi che rappresentano i concetti. Nonostante ciò, come precedentemente menzionato, una potenziale interrelazione tra analisi semica e word embeddings è stata individuata precisamente nella possibilità di includere elementi che presentano legami morfologici oppure forme flesse come semi all'interno dei sememi di riferimento. Il caso studio proposto ha infine permesso di elaborare alcune considerazioni rispetto all'effettiva capacità dei word embeddings di catturare la dimensione semantica insita nelle entità linguistiche. Tale potenzialità, nel contesto della presente tesi e sulla base dei dati analizzati relativi al dominio medico, viene infatti messa in discussione data l'esigua vicinanza a livello semantico e concettuale manifestata dai termini relazionati ai termini di riferimento, la quale si riscontra quasi unicamente a livello morfologico o grammaticale dunque morfosintattico o sintattico. Inoltre, è stata riscontrata la presenza di elementi che non manifestano alcuna relazione o similarità semantica con il termine di riferimento concepito all'interno del dominio medico i quali, piuttosto, presentano unicamente una vicinanza contestuale nell'ambito della medesima frase. In relazione a ciò, si potrebbe ritenere dunque che la sussistenza di tali circostanze evidenzia l'irrealizzabilità di un approccio automatico data la necessità di operare una selezione rispetto alle entità linguistiche generate mediante l'applicazione dei word embeddings. La constatazione di tale inapplicabilità riscontrata nel contesto dell'analisi presentata deriva infine da un'ulteriore riflessione concernente il dominio terminologico considerato, caratterizzato da una significativa specificità concettuale che dovrebbe quindi manifestarsi in una vicinanza o similarità semantica piuttosto accurata rispetto alle entità linguistiche catturate mediante word embeddings. Sebbene i word embeddings consentano spesso di individuare elementi lessicali che presentano una connessione a livello semantico rispetto ai termini di riferimento, l'assenza di tale livello di accuratezza, la forte influenza della dimensione contestuale e il fatto che si possa ritenere che una quantità esigua di entità possano concorrere a rappresentare semi nel semema di tali termini costituiscono ulteriori fattori che portano quindi a mettere in discussione la potenzialità dei word embeddings di catturare efficacemente ed effettivamente informazioni semantiche. A tale riguardo, si sottolinea la necessità di condurre ulteriori studi per valutare questo aspetto.

In conclusione, il presente elaborato evidenzia la rilevanza dell'analisi semica quale tecnica linguistica volta all'acquisizione di una conoscenza approfondita di concetti specializzati e della terminologia inclusi in un determinato dominio. Le differenti fasi che possono essere potenzialmente individuate nello svolgimento dell'analisi, caratterizzate da differenti approcci allo studio della dimensione concettuale, consentono infatti una comprensione accurata e minuziosa dei concetti e dei termini ad essi associati. La metodologia individuata al fine di consentire la sussistenza di un'analisi semica oggettiva volta alla limitazione di manifestazioni soggettive nell'espressione dei concetti relativi al dominio medico si pone in questo contesto anche come strumento per raggiungere una cognizione esaustiva dei concetti in esso contemplati. Inoltre, l'adozione di tale metodologia può consentire una maggiore sistematizzazione nella selezione degli elementi compresi nell'analisi semica rispetto ad altri domini inclusi nell'ambito delle lingue speciali, sottolineando la rilevanza dell'analisi semica per lo studio di concetti in un'ampia prospettiva. I casi studio sviluppati pongono le basi per ulteriori ricerche, sia rispetto allo studio dell'analisi semica che in relazione all'effettiva sussistenza della dimensione semantica che viene generalmente ascritta ai word embeddings. Con particolare riferimento all'analisi semica, considerando il fatto che la metodologia proposta si concentra sulla possibilità di sistematizzare la conoscenza relazionata alla denotazione, emerge la necessità di approfondire lo studio dei semi afferenti che esprimono la connotazione, in modo tale da consentire un'esplorazione comprensiva degli elementi contemplati nell'architettura concettuale dell'analisi semica. Tale prospettiva di ricerca potrebbe altresì assumere rilevanza anche nel contesto comunicativo, contribuendo a favorire e potenziare la mutua comprensione nelle interazioni tra medico e paziente.