

From Term Dynamics to Concept Dynamics: Term Variation and Multidimensionality in the Psychiatric Domain

Pilar León-Araúz, Arianne Reimerink
Department of Translation and Interpreting, University of Granada
pleon@ugr.es, arianne@ugr.es

Abstract

Medical terminology is one of the most dynamic terminological domains, and the choice of one term instead of the other is not random, but the result of different perspectives towards reality. VariMed is a research project on medical term variants and its overall objective is to generate a multifunctional resource on the medical domain for linguistic research, translation and technical writing. In this paper, we propose a systematic way of extracting term variants from large corpora within the subdomain of Psychiatry and how to represent them according to cognitive and communicative parameters. Our aim is to discover if different conceptualizations, or different conceptually motivated term variants, of the same concept are preferred in expert or semi-specialized communication. A corpus on Psychiatry was compiled and classified according to user types. A grammar was designed in NooJ (Silberstein, 2003) in order to extract term variants based on the usual lexico-syntactic patterns accompanying synonyms (*also known as; commonly referred to as*, etc.). Corpus analysis results indicate that, from a cognitive perspective, term variants reflect the prototypical dimensions in which psychiatric disorders may be classified. From a communicative perspective, terms and dimensions can also be associated with user-based parameters.

Keywords: term variation; cognitive motivation; communicative motivation

1 Introduction

The medical domain has over 25 centuries of history and involves numerous disciplines which affect all human beings to some extent. Therefore, medical terminology is one of the most dynamic terminological domains, and the use of one term instead of the other implies perceiving and conceptualizing aspects of reality from different perspectives (Prieto Velasco et al 2013: 168). VariMed is a research project on medical term variants and its overall objective is to generate a multifunctional resource on the medical domain for linguistic research, translation and technical writing. In this paper, we propose a systematic way of extracting term variants from large corpora within the subdomain of Psychiatry and how to represent them according to cognitive and communicative parameters. The relationship of specialized communications between a terminological resource and its user implies a

prototypical discursive positioning (Harré and Langenhove 1999), which is reflected in a specialized text sender and receivers with a different background knowledge level. Terminological resources should provide adequate terminological units and an adequate knowledge load (Tarp 2005: 8-9), always according to their potential users' continuum of general-specific language and knowledge purposes (León-Araúz et al 2013: 33).

This is in consonance with the Functional Theory of Lexicography (FTL; Bergenholtz and Tarp 1995, 2003). According to the FTL, there are two main types of lexicographic functions that cover *use situations* and different *user needs* (Wiegand 1989). These functions are cognition and communication-oriented (Bergenholtz and Tarp 2003; Bergenholtz and Nielsen 2006). In cognition-oriented situations, users seek additional information to widen their knowledge about the conceptual structure of a particular subject-field (psychiatry, neurology, oncology, etc.). Bergenholtz and Nielsen (2006: 286) explain that in these situations, the only communicative act taking place is between the terminographer and the users of the resource. The users want knowledge and the lexicographers provide it at a cognitive level, nothing more. The most difficult task is then, for the terminographer to decide *how much* information is to be included and how to represent its underlying structure to make the dictionary suitable to meet users' needs. On the other hand, in communication-oriented situations, two or more persons are engaged in producing or receiving a piece of language. This is the case of a translator who receives and must subsequently produce a text, as well as scientific writers, proofreaders, etc. Here the terminographer acts as a kind of mediator who helps to solve communication problems. We believe any terminological resource should satisfy both (León-Araúz et al 2013: 34).

In section 2, we give a brief overview on term variation. In section 3, we present how term variants are extracted from a specialized corpus on Psychiatry with a pattern-based grammar in NooJ, an NLP application (Silberztein, 2003). In section 4, a selection of the extracted variants is classified according to dimensional features and the results are compared across three subcorpora in order to see if certain dimensions are preferred in one discourse or the other. Finally, section 5 covers the conclusions and further research.

2 Term Variation

Although specialized language initially aspired to having one linguistic designation for each concept for greater precision, it is true that the same concept can often have many different types of linguistic designations. In the same way as in general language, there is terminological variation based on user-based parameters of geographic, temporal or social variation or usage-based parameters of tenor, field, and mode (Gregory and Carroll 1978). However, terminological variation also occurs for reasons that are often considerably more complex and difficult to explain. Freixa (2006: 52) classifies the causes for terminological variation in the following categories: (1) dialectal, caused by different origins of the authors; (2) functional, caused by different communicative registers; (3) discursive, caused by dif-

ferent stylistic and expressive needs of the authors; (4) interlinguistic, caused by contact between languages; (5) cognitive, caused by different conceptualizations and motivations. According to Freixa (2002), certain term variants are not only formally different, but also semantically diverse, as they give a particular vision of the concept. In this sense, Fernández-Silva et al (2011) describe this phenomenon as the linguistic reflection of conceptual *multidimensionality*. Multidimensionality has been defined by many authors (Bowker 1997, Kageura 1997, Wright 1997, Rogers 2004) as the phenomenon in which certain concepts can be classified according to different points of view or conceptual facets. This has important consequences in how domains are categorized and modelled. According to Picht and Draskau (1985, 48 *apud* Rogers 2004, 219), multidimensionality depends on who is the classifier as well as the different knowledge sources that may reflect different criteria when organizing the same domain or knowledge node. For example, botanists would classify roses different from rose growers. However, multidimensionality has also an impact on term variation, since concepts can be designated in more than one way based on the different characteristics that it possesses (Fernández-Silva et al 2011). Thus term variation should not be regarded as a linguistic phenomenon isolated from conceptual representations, since it is one of the manifestations of the dynamicity of categorization and expression of specialized knowledge (Fernández Silva et al in press).

Fernández-Silva (2010: 60-71) classifies the cognitive factors involved in term variation, based on numerous authors, according to two criteria. Firstly, the first category division depends on whether the cognitive factor refers to the conceptual organization or to its usage. Secondly, within the usage category, the categorization of the factor depends on how reality is conceptualized by certain groups of people or individuals or how reality is conceptualized according to the specific context in which the concept is used (Table 1).

Conceptual organization	
	Conceptual system
	Conceptual class
	Multidimensionality of the conceptual system
	Flexibility of the concept
	Linguistic system/culture
Usage	
Different conceptualizations	Knowledge evolution
	Dialects/cultures
	Thematic areas in interdisciplinary contexts
	Schools of knowledge/Ideologies
	Socio-professional groups
	Individual/individual point of view
Adaptation to specific context use	Adaptation to level of expertise of the receiver
	Intention/Aim/Point of view

Table 1: Cognitive factors of term variation (adapted from Fernández-Silva 2010: 61).

As can be inferred by Table 1, all Freixa’s causes for term variation can be approached from a cognitive perspective according to Fernández-Silva. This study combines Freixa’s second (functional) and fifth (cognitive) causes for term variation with Fernández-Silva’s perspective, since it analyses the multidimensionality of the conceptual system and how the different conceptual dimensions correlate with the adaptation to the level of expertise of the receiver. Our aim is to discover if different conceptualizations, or different conceptually motivated term variants, of the same concept, are preferred in expert communication or semi-specialized communication.

3 Extracting Term Variants

A specialized corpus was compiled on the Psychiatry domain, which has more than 10 million tokens, and it was divided according to user and genre types: expert, semi-specialized and encyclopaedic. The expert corpus contains specialized books and journal papers written by experts for experts, such as the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2013). The semi-specialized corpus consists of web pages and brochures written by experts from Medline Plus and the National Institute of Mental Health, which combines basic and clinical research with information for patients, or their relatives, suffering from any kind of mental disorder. Finally, the encyclopaedic corpus consists of a Wikipedia dump which was automatically collected through categories such as *Psychiatry, Syndromes, Disorders*, etc. We considered that Wikipedia should belong to this corpus because, being an encyclopaedic resource, it usually contains metalinguistic information on synonyms and variants that could be useful in our research.

Once the corpus was compiled and classified, a NooJ local grammar was designed in order to extract term variants (Figure 1). The grammar is based on the usual lexico-syntactic patterns accompanying synonyms (*also known as; commonly referred to as*, etc.) combined with specialized terms, namely syndromes, disorders and diseases.

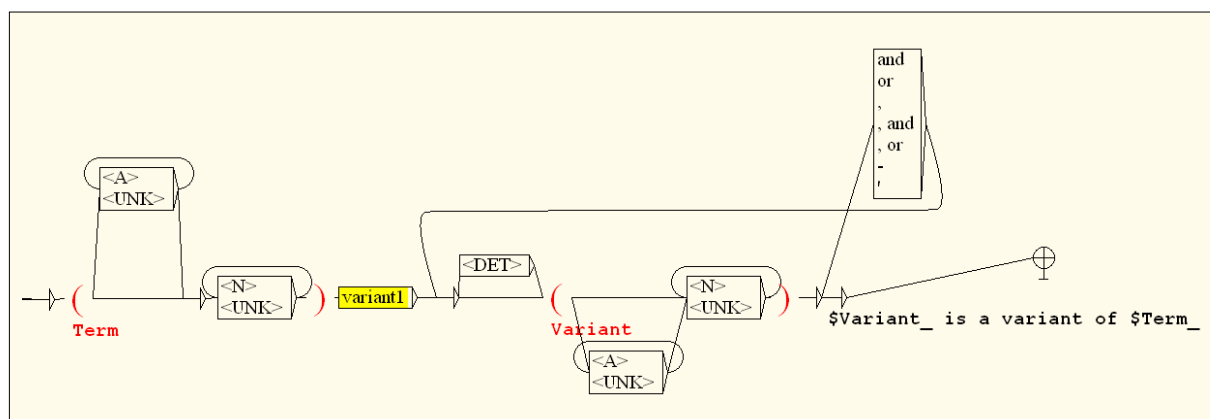


Figure 1: Pattern-based grammar for term variants’ extraction.

de that, not surprisingly, term variance structures are most often lexicalized in the Wikipedia corpus, then in the semi-specialized corpus, and finally in the expert corpus (Figure 4).

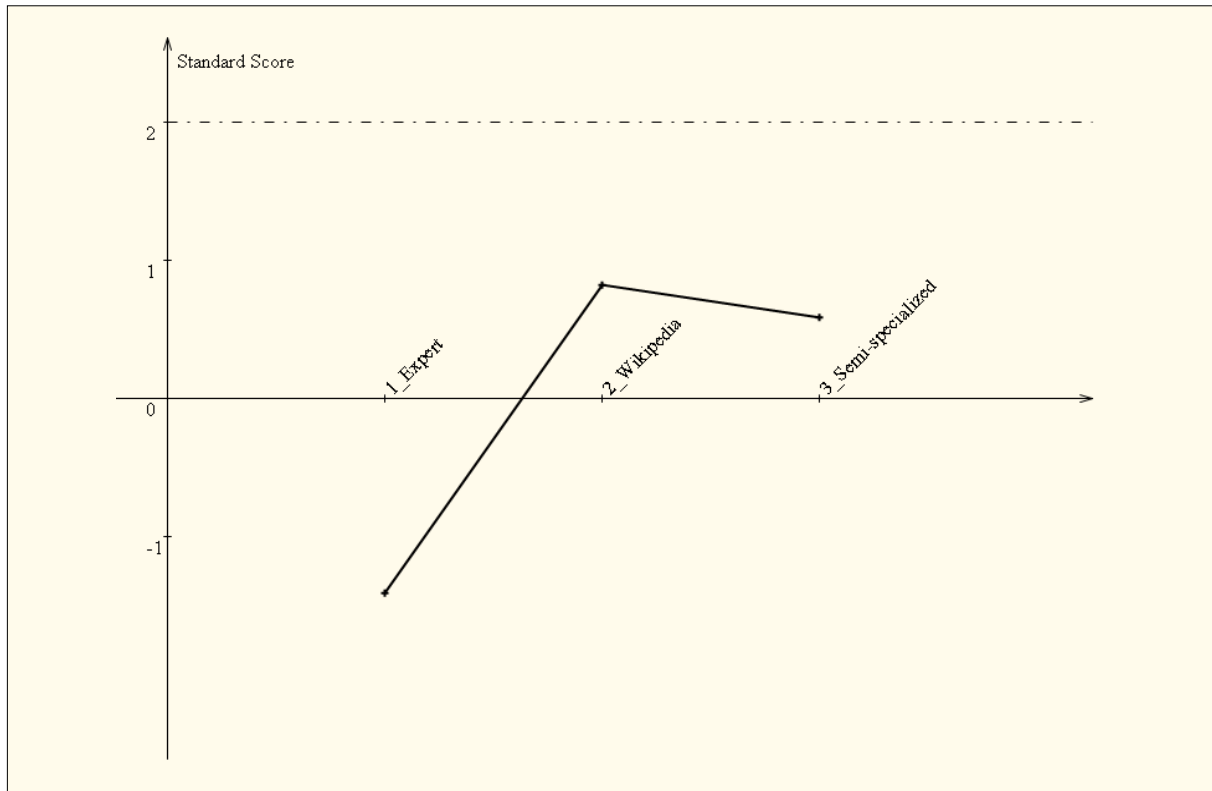


Figure 4: Term variance structures in the three corpora.

4 Representing Term Variants

In terminological resources, users are often confronted with a vast array of variants with no other information on how term variation arises and how their use may be constrained. However, they need to know when to use each of the variants and the conceptual connotations they imply, since this will affect the receiver's interpretation of the message.

In our study, we have found many different types of term variants for the same concept. Some of them were just acronyms, dialectal, orthographic or morphological variants, which, of course, need to be stored in any terminological resource, but their impact on communication is obvious, and their use does not usually need any further explanation. In this paper, however, we focus on dimensional variant types, which need more in-depth study, since they affect both cognitive and communicative situations. Dimensional variants show different conceptualizations of the same concept according to different facets and are usually conveyed by multi-word terms. For instance, *Ganser syndrome*, *nonsense syndrome* and *prison psychosis* are all variants of the same concept, but the first one highlights a DISCO-

VERER dimension (Sigbert Ganser was the first to describe the syndrome), the second one focuses on the SYMPTOM dimension (saying nonsense is one of them) and the third one on the LOCATION dimension (it often takes place in prisons, since it affects inmates). We collected from the corpus all the concepts that showed more than one variant type and classified their corresponding variants according to the dimension conveyed. In Table 2, we show the dimensions we found with an illustrative example.

Dimension	Term variant
+Discoverer	Korsakoff's psychosis
+Symptom	burning-mouth syndrome
+Cause	alcohol-induced amnesic disorder
+Body_part	Broca's aphasia
+Patient	boxer's dementia
+Result	bedwetting
+Intensity	mild cognitive impairment
+Time	short-term insomnia
+Location	prison psychosis

Table 2: Dimensional variant types found in the corpus.

Of course, there are variants that may show different dimensions at the same time, such as *chronic* [+Time] *traumatic* [+Cause] *brain* [+Body_part] *injury associated with boxing* [+Cause] or *alcohol-induced* [+Cause] *amnesic* [+Symptom] *disorder*; and different variants for the same concept that highlight the same dimension, such as *Alice in Wonderland* [+Symptom] *syndrome* and *Lilliputian hallucination* [+Symptom], which refer to the same symptom of the disorder, or *nonsense* [+Symptom] *syndrome* and *.syndrome of approximate answers* [+Symptom], which convey the same dimension but refer two different symptoms.

Terminological resources should add the conceptual dimension conveyed by each variant so that users can make a cognitively sound choice, but as previously stated, term choice also depends on communicative situations, namely, the expert-lay continuum. Therefore, term entries should also add use-related information. When querying the corpus, we see that the variant *dementia pugilistica* is much more often used in the expert corpus as compared to *punchdrunk syndrome* or *boxer's dementia*. In this case, the latin origin of the term points to the usual preference in expert settings. However, this preference is not always as straight forward. In such cases, users should have this information at hand. For instance, in Figure 5 three different variants for the same concept are represented according to usage-based preferences: *postnatal depression* (red), *postpartum depression* (blue), and *baby blues* (black). For this analysis, the three corpora were merged into a single file in order to observe these preferences as a continuum. When a single file is loaded in NooJ, its statistical module automatically splits it into 20 parts. Thus, in Figure 5, the first third of the graph represents the expert corpus, the second the Wi-

ikipedia corpus and the third the semi-specialized corpus. Not surprisingly, *baby blues* is the preferred term in semi-specialized communication, although the other two variants are also commonly found. At first, *postnatal* or *postpartum* would seem interchangeable choices, but corpus analysis indicates that the use of one term or the other imposes a strong constraint on the communicative situation. *Postpartum depression* seems to be a more neutral term most often found in the Wikipedia corpus and *postnatal depression* is the preferred term in expert texts.

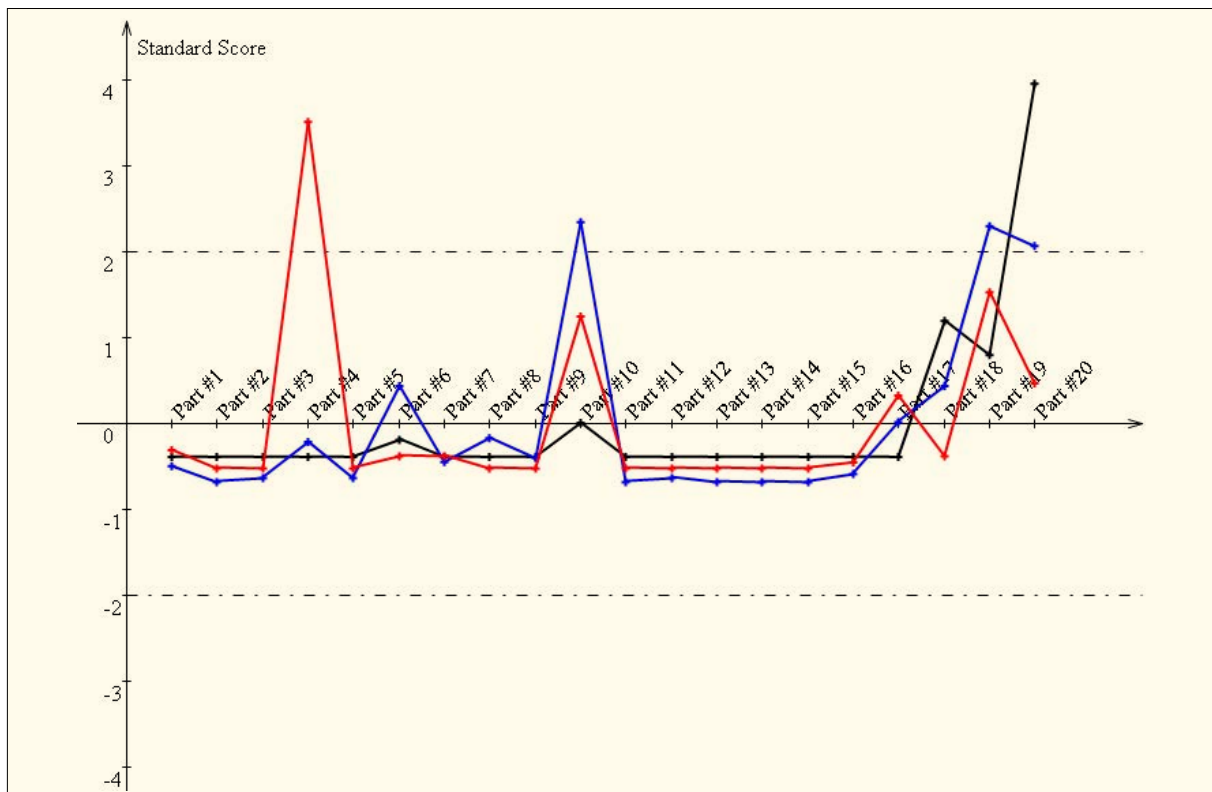


Figure 5: Usage-based preferences for *baby blues* (black); *postpartum depression* (blue); *postnatal depression* (red).

Nevertheless, apart from showing usage information related to individual terms, we believe that more generalized patterns can be found in the comparison of the three corpora with regards to the prevalence of dimensions. For this reason, as a further step, we built a new dictionary with all the dimensional variants related to the same entry in order to automatically annotate them as such in the corpus. So far, the dictionary has 73 concept entries associated with 326 dimensional variants. Figure 6 shows an example of a NooJ dictionary entry, where N means noun, FLX=TABLE refers to the inflectional paradigm, and UNAMB is a special code that lets us deal with multi-word terms as a single unit. Apart from NooJ codes, we can also add any other semantic feature to each entry (e.g. Term, Symptom).


```
ganser syndrome, N+FLX=TABLE+Term+Discoverer+UNAMB  
ganser's syndrome, ganser syndrome, N+FLX=TABL+Term+Discoverer+UNAMB|  
nonsense syndrome, ganser syndrome, N+FLX=TABLE+Term+Symptom+UNAMB  
balderdash syndrome, ganser syndrome, N+FLX=TABLE+Term+Symptom+UNAMB  
syndrome of approximate answers, ganser syndrome, N+FLX=TERMOFTERM+Term+Symptom+UNAMB  
pseudodementia, ganser syndrome, N+FLX=TABLE+Term+Intensity+Old+UNAMB  
prison psychosis, ganser syndrome, N+Term+Location+UNAMB
```

Figure 6: NooJ dictionary entry for *Ganser syndrome* and its variants.

Once we had our own dictionary, we applied it to the main corpus in order to build three new subcorpora, which would only contain sentences where our previously extracted term variants occurred. This was done for a better performance of the system and especially in order to eliminate any bias towards lexical diversity or distribution differences in the three corpora. After that, we performed new queries based on four of the most relevant dimensions found in this domain (Figure 7): +Discoverer (black), +Cause (red), +Symptom (blue), +Body_part (green).

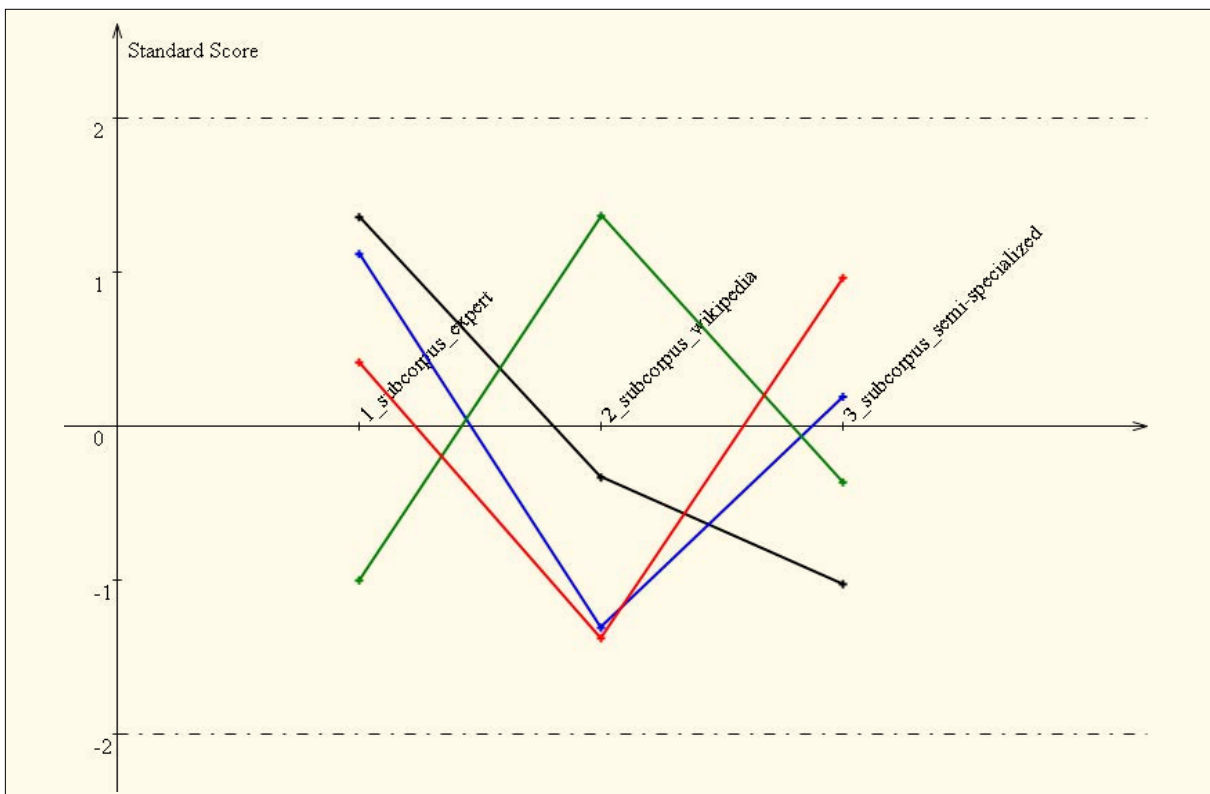


Figure 7: Correlation of dimensional variants with the three corpora.

According to a patient-oriented approach, one might think at first that the +Symptom and +Body_part dimensions could be more semi-specialized than expert-related. However, surprisingly enough, it seems that the +Body_part dimension is only significant in the Wikipedia subcorpus, whereas the +Symptom dimension is most often found in the expert subcorpus. In the expert subcorpus +Discover-

rer is the most prevalent dimension, but +Cause is also significantly represented. As for the semi-specialized subcorpus, +Cause and +Symptom dimensions are most prototypical, but in an inverted way as compared to the expert subcorpus.

This is only a first approach to the study of cognitive and communicative correlation that should be further extended to a higher number of variants and other medical domains, since results can change dramatically when comparing domain-based differences. In this sense, Tsuji and Kageura (1998) observed in a medical corpus that person or virus names were more dominant than other variant types.

5 Conclusion

In this paper we have shown how term variants in the psychiatric domain are cognitive and communicatively motivated. From a cognitive perspective, variants reflect the prototypical dimensions in which psychiatric disorders may be classified. From a communicative perspective, terms and dimensions can also be associated with user-based parameters. However, further studies need to be done with regards to the correlation between the cognitive and communicative factors underlying term variation, especially from a cross-linguistic perspective, since not all cultures conceptualize specialized domains in the same way and nor do they address their audience in the same manner.

6 References

- American Psychiatric Association. 2013. *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.
- Bergenholtz, H. and S. Nielsen. 2006. Subject-Field Components as Integrated Parts of LSP Dictionaries. *Terminology* 12 (2): 281-303.
- Bergenholtz, H. and S. Tarp. 2003. Two Opposing Theories: On H. E. Wiegand's Recent Discovery of Lexicographic Functions. *Hermes: Journal of Linguistics* 31: 171-196.
- Bergenholtz, H. and S. Tarp (eds.) 1995. *Manual of Specialised Lexicography: The Preparation of Specialised Dictionaries*. Amsterdam: John Benjamins.
- Bowker, L. 1997. Multidimensional Classification of Concepts and Terms. In Wright, S. E. and G. Budin (eds.) *Handbook of Terminology Management: Basic Aspects of Terminology Management*. Amsterdam: John Benjamins, 133-143.
- Cruse, D. A. 2002. Hyponymy and Its Varieties. In Green, R., C. A. Bean and S. H. Myang (eds.) *The Semantics of Relationships*. Dordrecht: Kluwer, 3-21.
- Cruse, D. A. 1995. Polysemy and Related Phenomena from a Cognitive Linguistic Viewpoint. In Dizier, P. St. and E. Viegas (eds.) *Computational Lexical Semantics*. Cambridge: Cambridge University Press, 33-39.
- Fernández-Silva, S. 2010. Variación Terminológica y Cognición. Factores Cognitivos en la Denominación del Concepto Especializado. PhD Thesis. Universitat Pompeu Fabra.
- Fernández-Silva, S., J. Freixa and M. T. Cabré. 2011. A Proposed Method for Analysing the Dynamics of Cognition through Term Variation. *Terminology* 17 (1): 49-73.

- Fernández-Silva, S., Freixa, J., Cabré, M.T. (in press). "A proposed method for analysing the dynamics of naming from a intralingual and interlingual perspective". In: Temmerman, R. and Van Campenhoudt, M. (eds.) *The dynamics of culture-bound terminology in monolingual and multilingual communication*. Amsterdam/Philadelphia: John Benjamins.
- Freixa, J. 2006. Causes of Denominative Variation in Terminology: A Typology Proposal. *Terminology* 12 (1): 51-77.
- Freixa, J. 2002. Reflexiones acerca de las causas de la variación denominativa en terminología. In Guerrero, G. and M. F. Pérez (coords.) *Panorama actual de la terminología*. Granada: Editorial Comares-Interlingua, 107-115.
- Gregory, M. and S. Carroll. 1978. *Language and situation: Language varieties and their social contexts*. London: Routledge and Kegan Paul.
- Harré, R. and L. van Langenhove (eds.) 1999. *Positioning Theory: Moral Contexts of Intentional Action*. Malden: Blackwell.
- Kageura, K. 1997. Multifaceted/Multidimensional Concept Systems. In Wright, S. E. and G. Budin (eds.) *Handbook of Terminology Management: Basic Aspects of Terminology Management*. Amsterdam: John Benjamins, 119-132.
- León Araúz, Pilar, Arianne Reimerink, and Alejandro García Aragón. 2013. Dynamism and Context in Specialized Knowledge. *Terminology* 19 (1): 31-61. doi:10.1075/term.19.1.02leo.
- Picht, H. and J. Draskau. 1985. *Terminology: An Introduction*. Guildford: University of Surrey.
- Prieto Velasco, Juan Antonio, Maribel Tercedor, and Clara Inés López Rodríguez. 2013. La multidimensionalidad conceptual en la traducción médica. *Shopos. Revista Internacional de Traducción e Interpretación* 2: 167-183.
- Rogers, M. 2004. Multidimensionality in Concepts Systems: A Bilingual Textual Perspective. *Terminology* 10 (2): 215-240.
- Tarp, S. 2005. The Pedagogical Dimension of the Well-Conceived Specialised Dictionary. *Iberica* 10: 7-21.
- Tsuji, K, Kageura, K. 1998. An analysis of medical synonyms: the word-structure of preferred terms. *Terminology*, 5:2, 229-249
- Wiegand, H. E. (ed.) 1989. *Wörterbücher in der Diskussion* [I] (Lexicographica. Sieres Maior 27). Tübingen: M. Niemeyer.
- Wright, S. E. 1997. Representation of Concept Systems. *Handbook of Terminology Management: Basic Aspects of Terminology Management*. Amsterdam: John Benjamins, 89-97.

Acknowledgements

This study was carried out in the framework of the R&D project VARIMED (FFI2011-23120) financed by the Spanish Ministry of Education, Culture and Sports.