# French Specialised Medical Constructions:

# Lexicographic Treatment and Corpus Coverage in General and Specialised Dictionaries

**Ornella Wandji Tchami[1][2], Ulrich Heid[2], Natalia Grabar[1]**
STL[1] UMR 8163 CNRS, University of Lille3; IWIST[2], University of Hildesheim
e-mail: ornwandji@yahoo.fr, heidul@uni-hildesheim.de, natalia.grabar@univ-lille3.fr

## Abstract

The aim of this paper is to examine four French dictionaries : the *on-line Larousse*, the *Petit Robert* (2009), the *Larousse Médical*, and the *TLFi*, with regard to the way specialised verbal expressions and syntactico-semantic constructions (extracted from two medical subcorpora differentiated according to the level of expertise of their author and readership) are introduced and described within and across the resources. According to our results, specialised verbal expressions and constructions are not given an important place in the queried resources, even in the medical dictionary. Very few of the studied expressions and constructions are found in the analysed dictionaries, and those which are presented are not given a homogeneous and consistent description.

**Keywords:** verb argument structure; specialised verb usage; lexicography; terminography; corpus analysis; phraseology; preferred co-occurrence and collocation

## 1 Context

Standard medical language contains specific terminology and specialised phraseology which are hard to understand for non-expert users (McCray 2005; Zeng-Treiler et al. 2007), and which can therefore render the communication difficult between medical doctors and patients (Jucks & Bromme 2007; Tran et al. 2009). Studies have been conducted in different domains in order to find ways to improve communication between doctors and patients (Kharrazi 2009; Chy et al. 2012; Kokkinakis & Toporowska Gronostaj 2006; Smith and Wicks 2008; Zeng-Treiler & Tse 2006; Chmielik & Grabar 2011). Some of these studies suggested the simplification of the medical doctors' vocabulary. In NLP, text simplification refers to the process of reducing the linguistic complexity of a text, while still retaining the original information and meaning (Siddharthan 2014). The simplification can concern syntax (Brouwers et al. 2014), or the lexicon (Elhadad 2006; Leroy et al. 2012), or simply focus on surface characteristics of the text, i.e., the number of characters and syllables per word, capitalization, punctuation and ellipses, etc. (Tapas & Orr 2009). Several researchers have investigated the use of text simplification for facilitating access to medical texts, by simplifying terminology (Elhadad 2006; Grabar & Hamon 2014). Any simplification requires resources, which in turn presupposes the description of specificities of both the experts language and that of the non-experts. Such description can be done through the comparison of corpora representing each of the language varieties.

In a previous study (Wandji et al. 2015), we performed a comparative analysis of four French medical subcorpora whose author and intended readership have different levels of expertise. The comparison was based on the arguments and collocates of verbs, which were forehand labelled using the semantic categories of the Snomed international terminology. The study resulted in the extraction of syntactico-semantic constructions and preferred co-occurrences, which are specific to our medical subcorpora and among which 38 were selected for further analysis.

The aim of this paper is to examine and compare four French dictionaries : the on-line *Larousse*[1], the *Petit Robert* (2009), the *Larousse Médical*[2], and the *TLFi*[3], focusing on these 38 syntactico-

[1] http://www.larousse.fr/dictionnaires/francais/

[2] http://www.larousse.fr/archives/medical

semantic constructions and preferred co-occurrences. We are going to investigate whether the above mentioned structures appear in the nomenclature or in the micro-structure of the studied dictionaries, and analyse the techniques used to describe them in each dictionary. We expect the *Larousse Médical* to focus on expert medical language, and the *TLFi* to broadly cover all language varieties, while the *Petit Robert* and *on-line Larousse* may focus more on the general language.

# 2 Material

## 2.1 Corpus

The analysed data are extracted from two French subcorpora of medical texts. The first, called *expert subcorpus*, gathers texts intended for medical experts and comes from the *CISMeF*[4] portal, which indexes medical texts according to three different categories: texts for medical experts, texts for medical students, texts for patients or non-experts. The second subcorpus is composed of forum texts i,e. discussions between patients and/or persons participating in a platform called *Doctissimo, Hypertension, Problèmes cardiaques*[5]. These subcorpora are further described in table 1.

| Corpus | Size (#words occ.) | Description |
|---|---|---|
| Expert | 1,285,665 | Scientific publications and brochures |
| Forum | 1,588,697 | Messages from participants in a forum |

Table 1: Description of the corpus.

## 2.2 Snomed International Terminology

We use the Snomed International Terminology (Cote 1996), one of the largest medical terminologies freely available for French, as a source of linguistic information for the semantic annotation of our texts. This terminology groups medical terms into eleven semantic categories, of which nine are considered in this study[6]. These categories (described in table 2) are used for labeling the verb arguments with semantic information.

| Categories | Examples |
|---|---|
| Topography or anatomical locations | *Heart, hand, vessel, etc.* |
| Social status | *Husband, child, former smoker, patient, etc.* |
| Procedures | *Caesarean, surgery, radiography, x-ray, etc.* |
| Living organism | *Bacteria, viruses and animals* |
| Professional occupations | *Doctor, anaesthesiologist, ambulance team, etc.* |
| Functions and dysfunctions of the organism | *arterial pressure, proteinuria, etc.* |
| Disorders and pathologies | *Cancer, diabetes, arterial hypertension, etc.* |
| Chemical products | *sodium, heparin, etc.* |
| Physical agents and artefacts | *catheter, prosthesis, tube, etc.* |

Table 2: Description of the Snomed categories used.

[3] http://atilf.atilf.fr/
[4] http://www.cismef.org/
[5] http://forum.doctissimo.fr/sante/hypertension-problemes-cardiaques/liste_sujet-1.htm
[6] The semantic class containing modifiers is not taken into consideration in this study.

## 2.3 Data

The analysed expressions and constructions (see *Appendix*) are extracted from the subcorpora presented above. The main selectional criterion is their frequency in the whole corpus. Each selected item should have at least 5 occurrences[7] in the corpus.

## 2.4 Dictionaries

Four French dictionaries are analysed in the scope of our study : the *on-line Larousse*, the 2009 version of *Petit Robert*, as well as *TLFi*, which is the computerised version of the *TLF* (*Trésor de la Langue Française*), a dictionary of the French language of the nineteenth and twentieth centuries, printed in six volumes. This dictionary principally covers the general language but also involves various specialised domains like medicine, law, sport, etc. The last resource is a medical dictionary : *Larousse Médical*, which we assume will give a clue on how dictionaries of specialised language deal with the 38 constructions retrieved from the corpus. We have selected different types of resources because together, they constitute a representative sample of French lexicography for general and specialised language and they provide us with enough data for the comparison. The queried dictionaries are connected to our corpus in the sense that they all deal to a certain extent with specialised languages, in this case the medical language. Indeed, though the *on-line Larousse* and the *Petit Robert* are known as general language dictionaries, each of them tend to provide domain-specific meanings of the entries, and Medicine is among the proposed domains.

# 3 Method

## 3.1 Corpus Preprocessing and Annotation

The two subcorpora are downloaded, converted into plain text and recoded in UTF-8 format. A syntactic analysis is performed with the Cordial dependency parser (Laurent 2009). The syntactically annotated sentences are then processed with Perl programs that perform the semantic annotation by projecting the Snomed international terminology onto the nominal arguments of the lemmatised sentences. Each time a nominal phrase matches an entry of the Snomed terminology, the semantic category of this entry is associated to the nominal phrase. The categories of the terminology add semantic information to the verb arguments. Hence, at the end of this stage, each verb argument appearing in the terminology is labelled with a semantic category, in addition to its syntactic function, as in the following example :
(1) Le patient présente un cancer (the patient has a cancer) => le patient_s/S/ présente un cancer_do/D/.

## 3.2 Selection of Syntactico-semantic Constructions and Preferred Co-occurrences

From the annotated sentences, we extract argument structures with terms carrying the Snomed categories (see table 2) :
(2) Le patient présente un cancer (the patient has a cancer) => have: patient_s/S/, cancer_do/D/.
For each verbal argument structure (*V+s/Scat+do/Scat, V+s/Scat+do/Scat+io/Scat*[8]), and for each verb/noun pair (*V+s/Scat, V+do/Scat and V+io/Scat*), the frequency in each subcorpus is computed. The identified constructions are shown in the first column of the table in the Appendix.
These constructions are then searched individually in the nomenclature and in the micro-structure of the four dictionaries. A particular attention is paid to the way each verbal expression or construction is introduced and described in the dictionaries articles : is it through a definition ? An example ? An

[7] As exception, some constructions with less than 5 occurrences were kept because they seemed to hide a specialised meaning.
[8] V=verb, s=sujet, do=, direct Object, io=indirect Object, Scat=Snomed category.

expression? Or otherwise?

## 4 Results and Discussion

The table in the Appendix provides the results of our study. We have investigated the presence of the studied syntactico-semantic structures and preferred co-occurrences in the dictionaries. Several questions were asked: does the searched item appear in the article of one of its constituent words (*En*) ? If yes, is it under the verb lemma entry (*V*), its past participle (*P*), or its noun argument (*N*) ? Or the item appears under another lemma entry which is not part of the analysed construction (*O*) ? How is the construction introduced in the article: through an example (*Ex*) which can match the searched medical meaning (√) or not (*x*) ? Through a special section dedicated to specialised usages of the entry (in our case medicine (*Us*)) ? Or the searched construction is simply mentioned somewhere in the article (*Ar*) or in other articles of the dictionary (*Dic*[9]) ? Does the article propose a medical meaning of the entry ? Is it compatible with the searched construction (*Se*)?

None of the studied expressions and constructions figures in the headwords list of the four dictionaries. This remark is somehow obvious because traditionally, dictionary entries consist of single words rather than constructions or multi-word expressions. Consequently, the analysed constructions were principally searched under the entries of the verbs and in the articles describing their noun arguments. As expected, all the syntactic patterns in the active form (*subject +verb+object*(s), etc.) were found under the verb entries of the non-specialised dictionaries ; the passive voice with an omitted agent (frequently used in the expert corpus) is only found in the *Petit Robert*, s.v. *indiqué*, *recommandé* and *conseillé*. None of the dictionaries provides comments or remarks that highlight the function and importance of this passive form in medical care texts. Surprisingly, in the analysed dictionaries, particularly in the *TLFi*, several notes are given on the passive voice, but none is related to medical texts or evokes the reason of the frequent omission of the verb's agent in expert medical texts sentences.

Based on the results of the table in the Appendix, our first observation concerns the coverage of the dictionaries, with regard to the studied expressions and constructions. None of the dictionaries individually covers more than 44% of the 38 analysed constructions. However, the sum of all the constructions appearing in the four dictionaries corresponds to 61% of the list, versus 52% for the constructions which are actually given a description. Out of the 24 constructions found in the analysed dictionaries, only 11 are presented in more than one.

Moreover, we have tried to investigate to which extent the most frequent constructions[10] (a total of 14 constructions with more than 10 occurrences each) extracted from the expert subcorpus are covered in the *Larousse Médical* and the *TLFi*, which are expected to contain such specialised verbal structures. The same experiment was carried out between the most frequent constructions in the forum subcorpus and the two general language dictionaries (*Larousse* and *Petit Robert*), which are expected to deal with more general or popular expressions. We found out that among the 14 most frequent constructions in the expert corpus, 9 are presented in the *Larousse Médical*, which is an average score, while only 7 are found in the *TLFi*.

In the same way, out of the 12 most frequent constructions[11] extracted from the forum corpus, 8 are found in the *TLFi*, 6 in the *Petit Robert,* and only 3 in the *on-line Larousse*, which means that only one of the general dictionaries reaches the average number of the 12 most frequent forum constructions. This clearly indicates that the general language dictionaries do not properly cover the forum expressions.

Another remark concerns the lexicographic treatment of the studied items in the dictionaries. The specialised constructions are not always given a homogeneous, consistent or accurate description in the dictionaries. Apart from the *TLFi* which in 70% of cases provides medical usages of verbs

[9] This last option only concerns the *Larousse Médicall*.

[10] Their numbers of occurrences are coloured brown in the table (see Appendix).

[11] Their numbers of occurrences are coloured blue in the table (see Appendix).

through sections dedicated to specialised domains (usually introduced by markers), the description in the other dictionaries is less homogeneous. No emphasis is laid on specialised usages of the verb entry, the level of specialisation is scarcely highlighted and domain markers are not always used.

Most of the time, the constructions are introduced in the articles through examples. This trend is particularly noticeable in the *Petit Robert*, where 12 out of 14 constructions (see *Appendix*) are presented via examples. These examples are generally followed by a glose or a little explanatory text, which somehow brings these examples close to definition. For instance, under reading 3, s.v. *développer* (*faire croître*), the example *personne qui développe une maladie* is glosed with *chez qui cette maladie s'installe et progresse*, and the whole is introduced by a phraseological marker.

This technique can be questionable for at least two reasons. First, it combines two important yet different types of lexicographic information in a single example. Indeed, examples are usually given to illustrate a particular meaning (of the headword) described in a given section. By providing a specialised usage of the entry as example, the lexicographer does not only illustrate a general meaning of the entry by a specialised one, but he also omits to lay emphasis on the specialised character of the entry usage given as example, which might result in a loss of information. Secondly, this technique can be confusing for the reader, especially when the concerned article already has a section dedicated to medical language, but which does not contain the medical expression headed by the entry. For example, in the *TLFi*, s.v. *prescrire*, there is a section dedicated to medicine but the collocation *prescrire un médicament* appears elsewhere in the article, without any explanation.

However, we have noticed some few exceptions to this practice. In the *Petit Robert*, s.v. *indiquée*, reading 3, there is a specialised example (illustrating a medical meaning of the entry) that is provided in the body of the article, on the same level with the other general meanings of the entry. The specialised definition is : *signalé comme étant le meilleur (médicament, traitement)*, it is followed by *traitement indiqué dans telle ou telle affection*, which is syntactically and semantically compatible with the searched construction (*traitement+indiqué+pp(D)*).

The studied constructions are sometimes simply introduced in a list of items (phrases) in the article. This practice is frequent in the *TLFi* and the *on-line Larousse*. For instance, in the *Larousse*, the construction *S développer D* is found in a section entitled “expressions”, and is followed by an explanatory phrase as follows : *Développer une maladie, en être effectivement atteint* . The lexicographer presents this structure as an expression used in the general language, and no indication is given with regard to its specialised character.

Some of the constructions are not found under the lemma entries of the items they are composed of, but instead in other articles. For example, in the *TLFi*, *diminuer le risque (de+D)* was found neither under the entry *diminuer*, nor under *risque*. Instead, s.v. *incarner*, we have the phrase *diminuer le risque d'ongle incarné*.

## 5 Conclusion

In this paper, we have conducted a comparative analysis of the content of four dictionaries (three general and one specialised) with regard to the way specialised verbal constructions extracted from two medical subcorpora (different according to the level of expertise of their author and readership) are described. The obtained results show that only few of the studied constructions are found in the dictionaries. Those which are presented are not given a homogeneous and consistent description within and across the dictionaries. As far as the methodology is concerned, we have observed that in most cases, the lexicographer does not really draw the reader's attention on the specialised nature of the described verb usage.

This work shows that corpus analysis is essential for the constitution of adequate resources for text simplification. For future work, we are planning to extend our study to more verbs, in order to gather enough material for the constitution of a text simplification resource where medical experts expressions and their non-experts equivalent are aligned.

## 6 References

Brouwers, L., Delphine, B., Anne-laure, L. & Thomas, F. (2014). Syntactic sentence simplification for French. *In Proceedings of the 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations (PITR) EACL*, pp. 47–56.

Chmielik, J. & Grabar, N. (2011). Détection de la spécialisation scientifique et technique des documents biomédicaux grâce aux informations morphologiques. *TAL*, 51(2), pp.151–179.

CISMef portal. Accessed at :http://www.cismef.org/[03/02/2014]

Côté, R. (1996). Répertoire d'anatomopathologie de la SNOMED internationale, v3.4. Université de Sherbrooke, Québec.

*Doctissimo, hypertension problèmes cardiaques (Forum)*. Accessed at : http://forum.doctissimo.fr/sante/hypertension-problemes-cardiaques/liste_sujet-1.htm [03/02/2014]

Dominique, L., Nègre, S., & Séguéla, P. (2009). L'analyseur syntaxique Cordial dans Passage. *In Proceedings of the TALN, 9*.

Elhadad, N. (2006). Comprehending technical texts: Predicting and defining unfamiliar terms. *In American Medical Informatics Association.*, pp. 239–243.

Grabar, N. & Hamon, T. (2014). Automatic extraction of layman names for technical medical terms. *In ICHI 2014,* Pavia, Italy.

Jucks, R. & Bromme, R. (2007). Choice of words in doctor-patient communication: an analysis of health-related internet sites. *Health Commun*, 21(3), pp. 267–77.

Kanungo, T. & Orr., D. (2009). Predicting the readability of short web summaries. *In Proceedings of the Second ACM International Conference on Web Search and Data Mining (WSDM '09)*. ACM, New York, NY, USA. pp. 202–211.

Kharrazi, H. (2009). Improving healthy behaviors in type 1 diabetic patients by interactive frameworks. *In Proceedings of the American Medical Informatics Association*, pp. 322–326.

Kokkinakis, D. & Toporowska, M.G. (2006). Comparing lay and professional language in cardiovascular disorders corpora. *In James Cook University Pham T., editor, WSEAS Transactions on Biology and Biomedicine*, pp. 429–437.

*Larousse Médical*. Accessed at : http://www.larousse.fr/archives/medical[13/02/2016]

*Le Trésor de la Langue Française informatisé* (dictionary). Accessed at : http://atilf. atilf.fr/ [13/02/2016]

Leroy, G., Endicott J., Mouradi, O., Kauchak, D. & Just, M. (2012). Improving perceived and actual text difficulty for health information consumers using semi-automated methods. *In Proceedings of the American Medical Informatics Association.*

McCray, A. (2005). Promoting health literacy. *Journal of the American Medical Informatics Association.*, 12, pp.152–163.

Siddharthan, A. (2014). A survey of research on text simplification. *In International Journal of Applied Linguistics*, 165 (2), pages: 259-298.

Smith, C., & Wicks, P. J. (2008). Patients Like Me: Consumer health vocabulary as a folksonomy. *In Proceedings of the American Med. Informatics Association. 2008 Symposium*, pages 682–686.

Tran, T.M., Chekroud, H., Thiery, P., & Julienne, A. (2009). Internet et soins: un tiers invisible dans la relation médecine/patient? *Ethica Clinica*, 53, pp. 34–43.

Treitler, Q.Z., Tse, T., Divita, G., Keselman, A., Crowell, J., & Browne, A.C. (2006). Exploring lexical forms: first-generation consumer health vocabularies. *In Proceedings of the American Medical Informatics Association.* 2006, pp.1155.

Treitler, Q.Z., Kim, H., et al. (2007). Text characteristics of clinical reports and their implications for the readability of personal health records. *Studies in Health Technology & Informatics 129 (Pt 2)*: 1117-21.

Wandji, T.O., Grabar, N., Heid, U. (2015). Syntagmatic behaviors of verbs in medical texts: Expert communication vs. forums of patients. In: *Proceedings of the 11th International Conference on Terminology and Artificial Intelligence*. pp. 99–106. Universidad de Granada, Granada, Spain.

## Appendix

| Constructions and/or preferred co-occurrences/ collocations | Frequency | | | | *TLFi* | | | | | *Larousse* | | | | | *Petit Robert* | | | | | *Larousse Médical* | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Pro freq occ** | | **For freq occ** | | En | Ex | Us | Ar | Se | En | Ex | Us | Ar | Se | En | Ex | Us | Ar | Se | En | Ex | Ar | Dic |
| **administrer médicament** | 92 | **12** | 18 | 3 | N | - | - | √ | - | V | - | - | - | √ | V | x | - | | √ | - | - | - | 2 |
| **J administrer P** | | 9 | | 2 | x | - | - | - | - | x | - | - | - | - | x | - | - | - | - | - | - | - | - |
| **augmenter + risque (de)** | 239 | **52** | 281 | **15** | x | - | - | - | - | x | - | - | √ | - | x | - | - | - | - | - | | - | 2 |
| **augmenter+tension (+n)** | | 1 | | **26** | O | x | - | - | - | x | - | - | - | | x | - | - | - | - | - | | - | - |
| **améliorer+état/ santé/état de santé** | | 2 | | **19** | V | √ | - | - | √ | V | - | - | - | - | x | - | - | - | - | - | - | - | - |
| **J/S appliquer P** | 64 | **15** | 41 | 8 | x | - | - | - | - | x | - | - | - | - | x | - | - | - | - | - | - | - | 1 |
| **C appliqué sur T** | | 1 | | 0 | x | - | - | - | - | x | - | - | - | - | x | - | - | - | - | - | - | - | 1 |
| **P appliqué à D** | | 5 | | 0 | x | - | - | - | - | x | - | - | - | - | x | √ | - | - | - | - | - | - | - |
| **appliquer+méthode** | | 2 | | 5 | VO | x | √ | - | - | x | - | - | - | - | x | √ | - | - | - | - | - | - | - |
| **appliquer recommandation** | | 6 | | 6 | x | - | - | - | - | x | - | - | - | - | x | - | - | - | - | - | - | - | - |
| **baisser+tension** | 11 | 1 | 183 | **46** | x | - | √√ | - | - | x | - | - | - | - | x | - | - | - | - | - | - | - | - |
| **C/P est conseillé+ pp** | 92 | **13** | 657 | **25** | x | √ | - | - | | x | - | - | - | - | P | √p | - | - | √ | - | - | - | - |
| **J/S découvrir D chez S** | 19 | 2 | 231 | **30** | x | - | - | - | | x | - | - | - | - | V | √ | - | - | √ | - | - | - | - |
| **S développer D** | 68 | **54** | 77 | 1 | x | - | - | - | | x | - | - | √ | - | x | √ | - | - | √ | - | - | - | 2 |
| **J diagnostiquer S** | 31 | 5 | 148 | 0 | x | - | - | - | | x | - | - | - | - | x | - | - | - | - | - | - | - | - |
| **diminuer+ risque (de D ou F)** | | **34** | | 6 | O | √ | √√ | - | - | x | - | - | - | - | x | - | - | - | - | - | - | - | - |
| **S exposé à C** | 83 | **23** | 27 | 0 | x | - | - | - | x | x | | - | x | x | x | - | - | - | - | - | - | - | - |
| **exposer+à un risque (de)** | 83 | **25** | 27 | 3 | N<br>O | x | √ | - | - | x | - | - | - | - | N<br>V | - | - | √ | - | - | - | - | - |
| **Exposer à+n de médicament** | | **23** | | 0 | x | - | - | - | - | x | - | - | - | - | x | - | - | - | - | - | - | - | - |
| **J évaluer S** | 212 | 7 | 7 | 0 | x | - | - | - | - | x | - | - | - | - | x | - | - | - | - | - | - | - | - |
| **hospitaliser + victime (de+N)** | 31 | 5 | 59 | 0 | x | - | - | - | - | x | - | - | - | - | x | - | - | - | - | - | - | - | - |
| **hospitaliser +patient** | | 5 | | 0 | x | - | - | - | - | x | - | - | - | - | x | - | - | - | - | - | - | - | - |
| **identifier décès** | 129 | 2 | 13 | 0 | x | - | - | - | | x | - | - | - | - | x | - | - | - | - | - | - | - | - |
| **P impliquer P** | 77 | 10 | 33 | 2 | x | - | - | - | | x | - | - | - | - | x | - | - | - | - | - | - | - | - |
| **P/C est indiqué+ pp** | 439 | **100** | 194 | **17** | x | √ | √√<br>ac | - | √ | x | - | - | - | - | P | √p | - | - | √ | - | - | - | 2 |
| **indiquer+traitement** | 439 | **17** | 194 | 3 | V | - | √ | - | √ | x | - | - | - | - | NP | - | - | √ | √ | - | - | - | 1 |

| | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **J/S observer D + pp** | 162 | **33** | 30 | 2 | V | √ | - | - | | x | - | - | - | - | x | - | - | - | - | - | - | - | 4 |
| **Prescrire médicament** | 73 | 0 | 433 | **26** | V | - | √ | √ | √ | x | - | - | - | - | N | √ | - | - | - | - | - | - | 6 |
| **Prescrire traitement** | | 3 | | 7 | VN | -<br>- | √<br>√√ | √<br>- | √<br>- | x | - | - | - | - | N | √ | - | - | √ | - | - | - | - |
| **S présenter D** | 423 | **42** | 194 | **10** | x | - | - | - | | x | - | - | - | - | x | - | - | - | - | - | - | - | 4 |
| **P/C est recommandé+ pp** | 341 | **37** | 87 | 5 | x | - | - | - | | x | - | - | - | - | P | √p | - | - | √ | - | - | - | 1 |
| **subir ablation** | | 1 | | **44** | x | - | - | - | √ | x | - | - | - | √ | x | - | - | - | √ | - | - | - | 4 |
| **subir+intervention (chirurgicale)** | | 5 | | **30** | V | - | √√ | - | - | V | √ | - | - | √ | x | - | - | - | √ | - | - | - | 2 |
| **subir+AVC** | | 0 | | **12** | V | - | √ | - | - | x | - | - | - | - | x | - | - | - | - | - | - | - | - |
| **S subir D** | 63 | 4 | 378 | **22** | x | - | - | - | - | x | - | - | - | - | x | - | - | - | - | - | - | - | - |
| **C subir P** | 63 | 1 | 378 | 0 | x | - | - | - | - | x | - | - | - | - | x | - | - | - | - | - | - | - | - |
| **J suivre S** | 272 | 6 | 502 | 0 | V | - | √√ | - | - | V | √ | - | - | √ | V | √ | - | | √ | - | - | - | 1 |
| **suivre traitement** | | 2 | | **13** | N | √ | √√ | - | √ | x | - | - | - | - | V | √ | - | | √ | - | - | - | 1 |
| **Tot Nb of constr.** | **38** | | | | 15 | 6 | 7 | 3 | 7 | 4 | 2 | 0 | 2 | 4 | 12 | 11 | 0 | 2 | 12 | - | - | - | 15 |
| **Nb of constructions described / nb of constructions found (in each dictionary) / total nb of constructions analysed in the study** | | | | | **16/17/38** | | | | | **4/7/38** | | | | | **13/14/38** | | | | | **0/15/38** | | | |

√ = the option applies to the construction
**x** = the option does not apply to the construction
- = there is nothing provided

√√ = there is a section dedicated to medical usages of the entry and it contains the searched construction vs. √ the medical section exists but does not contain the searched construction

**ac** = active voice, **p** = passive voice, **V** = verb, **N** = noun, **O** = others, **P** = past participle, **pp** = prepositional phrase, **freq** = frequency of the verb in the corpus, **occ** = number of occurrences of the construction

| | |
|---|---|
| 🟦 | Frequencies of the most frequent constructions in the forum subcorpus |
| 🟥 | Frequencies of the most frequent constructions in the expert subcorpus |