MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190442

Generalizability of Readability Models for Medical Terms

Hanna Pylieva[a], Artem Chernodub[b], Natalia Grabar[c], Thierry Hamon[d,e]

[a] *Ukrainian Catholic University, Faculty of Applied Sciences, Kozelnytska st. 2a, Lviv, Ukraine*
[b] *Grammarly*
[c] *CNRS, Univ. Lille, UMR 8163 - STL - Savoirs Textes Langage, F-59000, Lille, France*
[d] *LIMSI, CNRS, Université Paris-Saclay, F-91405 Orsay, France*
[e] *Université Paris 13, Sorbonne Paris Cité, F-93430, Villetaneuse, France*

Abstract

Detection of difficult for understanding words is a crucial task for ensuring the proper understanding of medical texts such as diagnoses and drug instructions. We propose to combine supervised machine learning algorithms using various features with word embeddings which contain context information of words. Data in French are manually cross-annotated by seven annotators. On the basis of these data, we propose cross-validation scenarios in order to test the generalization ability of models to detect the difficulty of medical words. On data provided by seven annotators, we show that the models are generalizable from one annotator to another.

Keywords:

Natural Language Processing; Terminology; Health Information Systems

Introduction

Specialized areas, such as medical area, convey and use technical words, or terms, which are typically related to knowledge developed within these areas. In the medical area, this specific knowledge often corresponds to fundamental medical notions related to disorders, procedures, treatments, human anatomy, etc. For instance, technical terms like *blepharospasm* (abnormal contraction or twitch of the eyelid), *alexithymia* (inability to identify and describe emotions in the self), *appendicectomy* (surgical removal of the vermiform appendix from intestine), or *lombalgia* (low back pain) are frequently used in the medical area texts. Patients and their relatives usually have some difficulties in the understanding and using of such terms: they show indeed poor health literacy. Some existing studies stressed the difficulty in understanding medical notions and terms by non-expert users, and its impact on a successful healthcare process [1,2]. Yet, it is not uncommon that patients and their relatives must face very technical health documents and information. Examples of this kind are frequent and usually the non-expert users are at loss in such situations: understanding of information on drug intake [3,4], of clinical documents [5], of clinical brochures or informed consents [6], of information provided for patients by different websites [7,8], and communication between patients and medical staff [9,10]. These observations provide the motivation to our work: we address the needs of non-specialized users facing health information and propose to predict the readability of medical words.

In what follows, we first present some related work and introduce the material used as well as the proposed method. We then present and discuss the results. Finally, we conclude with some directions for future work.

Related Work

For studying the readability of medical documents, researchers usually exploit readability measures. Among these measures, it is possible to distinguish classical and computational readability measures [11]. Classical measures usually rely on number of letters and/or of syllables a word contains and on linear regression models [12,13], while computational readability measures may involve vector models and a great variability of features, among which the following have been used for processing the biomedical documents: combination of classical readability formulas with medical terminologies [14]; n-grams of characters [15], stylistic [16] or discursive [17] features, lexicon [18], morphological features [19], combinations of different features [5].

At a more fine-grained level, the readability of words has been addressed much less frequently. In the general language, some research actions are often performed as part of the NLP challenges, such as the SemEval NLP (*www.cs.york.ac.uk/semeval-2012*) challenge held in 2012. This challenge proposed the following task: for a short text and a target word, several possible substitutions satisfying the context have also been proposed. The objective was to rate and to order the substitutions according to their degree of simplicity [20]. The participants applied rule-based and/or machine learning systems. Combinations of various features, designed to detect the simplicity of words, have been used, such as: lexicon from spoken corpus and from Wikipedia, Google n-grams, WordNet [21]; word length, number of syllables, latent semantic analysis, mutual information and word frequency[22]; Wikipedia frequency, word length, n-grams of characters and of words, random indexing and syntactic complexity of documents [23]; n-grams and frequency from Wikipedia, Google n-grams [24]; WordNet and word frequency [25]. The best systems reached up to 0.60 Top-rank and 0.575 Recall. Another work has been done on scholar texts in French written for children with the purpose to differentiate between the texts from various scholar levels and to test various features suitable for that [26]. This system reached up to 0.62 classification accuracy.

In the medical area, we can mention three experiments: manual rating of medical words [27], automatic rating of medical words on the basis of their presence in different vocabularies [28], and exploitation of machine learning approach with various features

[29]. This last experiment achieved up to 0.85 F-measure on individual annotations.

Another issue is to know what are the most suitable data for the analysis of text readability. These data have indeed crucial impact on models created and on their usability. Several approaches have been proposed:

- exploitation of expert judgment, who have an idea on needs of population aimed in the study [30]. The main limitation is that experts may have difficulties to figure out what are the real needs of population;
- exploitation of text books created for population according to their readability levels, such as school books [26]. The main limitation is that such books are usually created by experts using theoretical basis and observations;
- exploitation of crowdsourcing involving large population [30]. The main limitation is that the population involved is uncontrolled and unknown;
- exploitation of eye-tracking methods for a more finegrained analysis of reading difficulties [31,32]. The main limitation is that only short text spans can be used;
- manual annotation by human annotators [33]. In this case, the annotators represent the population, they are part of the controlled population, they can perform more complicated tasks than in case of crowdsourcing, although they are usually less many than in crowdsourcing experiments.

Related to this issue is the question on generalizability of data and of models generated from these data. For instance, it has been observed that data from experts are difficult to generalize over the population [30].

We propose to study the generalizability of the automatic categorization models for a stronger distinction of readability of medical words and distinction of words which may present understanding difficulties to non-experts users. The medical data processed are in French. Seven human annotators participated in creation of the reference data.

Material

The source terms are obtained from Snomed Int [34], which is the most extensive terminology in French, such as available from the ASIP SANTE website (*esante.gouv.fr/services/referentiels/referentiels-d-interoperabilite/snomed-35vf*). Snomed contains 151,104 medical terms organized in eleven axes such as disorders, procedures, chemical products, living organisms, anatomy, social status. For our purpose, we use five axes: disorders, abnormalities, procedures, functions, and anatomy. The assumption is that studying the understanding of these terms is important because they are related to main medical notions and laymen must face then frequently. The 104,649 selected terms are lemmatized and tokenized into words resulting in 29,641 unique words.

The set of 29,641 unique words was annotated by seven French speakers, 25 to 65-year-old, without medical training and without specific medical problems. The annotators are expected to represent the average knowledge of medical words among the population as a whole. The annotators are presented with the list of terms and asked to assign each term to one of the three categories:

- *I can understand the word;*
- *I am not sure about the meaning of the word;*
- *I cannot understand the word.*

The assumption is that terms, which are not understandable by the annotators, are also difficult to understand by patients. The annotators were asked not to use dictionaries during the annotation process. Further to the annotation process, the most frequent category is I cannot understand the word, which gathers between 65 to 70% of terms.

Methods

We propose to tackle the problem through the supervised categorization: the purpose is to categorize terms according to whether they can be understood or not by lay people. The manual annotations provide the reference data. The categorization pipeline is the following: categorization features are computed, they are used for training the classifiers, and the results are evaluated using the cross-validation.

We exploit 11 types of automatically computed features:

- *Syntactic categories*. Syntactic categories and lemmas are computed by *TreeTagger* [35] and then checked by *Flemm* [36]. The syntactic categories are assigned to words within the context of their terms. If a given word receives more than one category, the most frequent one is kept as feature. Among the main categories we find for instance nouns, adjectives, proper names, verbs and abbreviations.
- *Presence of words in reference lexica*. We exploit two reference lexica of the French language: TLFi (*www.atilf.fr/*) and lexique.org (*www.lexique.org/*). TLFi is a dictionary of the French language covering XIX and XX centuries. It contains almost 100,000 entries. lexique.org is a lexicon created for psycholinguistic experiments. It contains over 135,000 entries, among which inflectional forms of verbs, adjectives and nouns. It contains almost 35,000 lemmas.
- *Frequency of words through a non specialized search engine*. For each word, we query the Google search engine in order to know its frequency attested on the web.
- *Frequency of words in the medical terminology*. We also compute the frequency of words in the medical terminology Snomed International.
- *Number and types of semantic categories associated to words*. We exploit the information on the semantic categories of Snomed International.
- *Length of words in number of their characters and syllables*. For each word, we compute the number of its characters and syllables.
- *Number of bases and affixes*. Each lemma is analyzed by the morphological analyzer *Dérif* [37], adapted to the treatment of medical words. It performs the decomposition of lemmas into bases and affixes known in its database and it provides also semantic explanation of the analyzed lexemes. We exploit the morphological decomposition information (number of affixes and bases).

- *Initial and final substrings of the words*. We compute the initial and final substrings of different length, from three to five characters.
- *Number and percentage of consonants, vowels and other characters*. We compute the number and the percentage of consonants, vowels and other characters (i.e., hyphen, apostrophe, comas).
- *Classical readability scores*. We apply two classical readability measures: Flesch [12] and its variant FleschKincaid [38]. Such measures are typically used for evaluating the difficulty level of a text. They exploit surface characteristics of words (number of characters and/or syllables) and normalize these values with specifically designed coefficients.
- *FastText word embeddings* [39] pre-trained on French Wikipedia corpus, which cover up to 56% of the words from our dataset. Tthe embeddings cluster together words that share common contexts and semantics, and can help in generalizing other features over contextually and semantically close words.

The ten first types of features, linguistic and non-linguistic, are called *standard features*, while the embeddings stand for themselves.

The supervised categorization is performed with decision tree (DT) classifier from the scikit-learn library (*scikit-learn.org*).

In the proposed experiments, we learn the model from all the annotations of a given annotator and then test the model on annotations provided by other annotators. In this way, we can measure the ability of the classifier to generalize on all known words, but for unknown annotators. This scenario is realistic to a real-world situation: the reference annotations can be obtained only from a couple of users, presumably representing the overall population, but not from all the possible users. Yet, it is necessary to predict the familiarity of medical words for all the potential users even if they did not participate in the annotations. Hence, the generalizability of models occupies the central position in these experiments.

Results and Discussion

The results obtained are presented in Table 1. The first two columns indicate the annotators. Data provided by each annotator are used for training the classifier (first column). The model generated is then tested on data from all the annotators including the reference annotator (second column). Three sets of such experiments are performed, depending on features exploited: standard features, word embeddings, and combination of all the features available. Each experiment is evaluated with several measures: *P* Precision, *R* Recall, *F* F-measure to evaluate the efficiency in prediction which medical words are understandable or not understandable for a given annotator: the darker background, the better the results.

We can do several observations on these results. Features used show an impact on the results. Thus, standard features usually show better results than embeddings. One explanation is that standard features include 24 individual features covering different aspects of linguistic and non-linguistic description of words, while word embeddings rely only on distribution of words and their similarity. Yet, combination of all the features (standard and embeddings) usually improves overall results, sometimes going to up to 2.9 improvement of F-measure. Our hypothesis is that there exists a robust nonlinear dependency between some subsets of standard features and subword-level components of word embeddings. Testing this hypothesis is the topic of our further research.

Recall values are always higher than Precision values. In each set of experiments, the best results are not obtained when the model of a given annotator is applied to own data. For instance, the *O1* model provides better results when tested on data from annotators *O2, O3* and *A8*. Similarly, the *A7* model shows better results when applied to data from annotators *O1*, *O2*, *O3* and *A8*. This is an important issue because it shows that the models acquired from one annotator can be successfully generalized over other annotators.

Besides, it seems that the annotators form two clusters according to the classification of difficult medical words: one cluster with four annotators (*O1, O2, O3, A8*) and one cluster with three annotators (*A1, A2, A7*). This issue may be related to the health literacy of annotators. This may indicate that the annotation models can be shared by people with similar skills and knowledge. Yet, to confirm this hypothesis, it is necessary to define the level of health literacy of annotators. This task is rather difficult because there is no existing tests created for computing the health literacy level for French-speaking healthy people. Another hypothesis is that some models may be better generalizable than other models. This hypothesis must also be verified with additional experiments.

Another important point is that, while the annotations go forward, the annotators usually show learning progress in decoding the morphological structure of terms and their understanding [40]. This progress is not taken into account in the current models.

Conclusions and Future Work

We proposed to address the detection of medical words which understanding may be difficult for non-specialized users of the medical area. We exploit for this machine learning algorithms, reference data from seven annotators, and several sets of NLP features: standard features (syntactic information, reference lexica, frequency, etc.), distributional features (word embeddings), and their combination. Our results provide several indications. Hence, the combination of all features is the most efficient. Concerning the generalization, we propose to learn model on a given annotator and then to apply it to data obtained from other annotators. This set of experiments indicates that models provide better results when tested on data from other annotators. We consider this to be a positive issue because it is important to be able to generalize annotations provided by a set of users on the whole population. Yet, these results may point out that the users should be apprehended through their health literacy, while currently there is no available tests for measuring it in French-language healthy people.

We have several directions for future work. For instance, we will train our own word embeddings specific to medical data in French, so that they suit better our data. We also plan to implement and test other deep learning/neural networks/NLP methods which use the morphological information of words, such as character-level recurrent neural networks and character embeddings together with 1D convolutions. In addition to the readability of medical words, we will also work on measuring health literacy of French-speaking people.

Acknowledgements

This work was partly funded by the French National Agency for Research (ANR) as part of the *CLEAR* project (*Communication, Literacy, Education, Accessibility, Readability*), ANR-17-CE19-0016-01. Also, we thank again the annotators who participated in preparing the reference data.

Table 1– Portability of models from one user to another

Train annotator	*Test annotator*	*Standard features*			*Embeddings*			*Standard features + embeddings*		
		P	R	F	P	R	F	P	R	F
O1	O1	77.2	82.5	79.7	67.0	72.5	69.3	79.0	82.4	80.2
O1	O2	78.6	81.7	80.1	70.3	74.0	71.2	82.0	84.2	82.8
O1	O3	81.2	85.0	83.0	70.7	75.4	72.6	84.9	87.6	85.9
O1	A1	71.0	74.7	71.2	62.1	63.8	58.8	74.1	75.4	72.2
O1	A2	70.6	78.4	74.0	61.9	68.5	63.3	75.0	80.1	76.2
O1	A7	72.6	77.5	74.2	63.0	66.6	61.9	76.2	78.9	75.8
O1	A8	82.3	84.9	83.5	73.1	76.8	74.5	85.7	87.8	86.6
O2	O1	77.0	82.2	79.1	67.3	72.8	69.6	80.2	83.9	81.1
O2	O2	78.9	82.0	80.0	69.9	73.5	71.3	79.5	81.9	80.3
O2	O3	81.1	85.4	83.0	71.1	75.3	73.0	83.5	86.8	84.7
O2	A1	71.1	72.1	68.2	61.7	64.5	60.2	74.0	75.1	71.5
O2	A2	70.8	77.3	72.7	61.8	68.9	64.2	76.0	79.8	75.5
O2	A7	72.7	75.6	71.8	62.6	67.0	62.8	75.9	78.3	74.9
O2	A8	83.0	86.2	84.4	73.7	77.1	75.3	85.4	88.2	86.7
O3	O1	77.4	82.8	79.7	67.1	72.7	69.4	81.3	84.9	82.4
O3	O2	79.0	82.2	80.2	70.4	74.1	71.6	82.1	84.2	82.8
O3	O3	81.2	85.5	83.2	70.4	74.9	72.3	83.0	85.9	84.2
O3	A1	71.8	73.3	69.5	61.7	64.1	59.6	75.1	75.4	72.1
O3	A2	71.2	78.0	73.5	61.8	68.7	63.9	76.8	80.2	76.3
O3	A7	73.2	76.5	72.9	62.4	66.6	62.2	77.2	78.8	75.8
O3	A8	82.6	85.8	84.1	73.7	77.2	75.2	86.0	88.0	86.9
A1	O1	77.2	82.5	79.8	66.5	67.9	66.6	76.9	79.5	77.6
A1	O2	78.6	81.6	80.1	69.2	69.0	68.5	78.8	79.6	78.9
A1	O3	81.2	84.9	82.9	70.7	69.6	69.2	81.8	82.0	81.0
A1	A1	70.9	74.7	71.3	59.4	64.6	61.8	72.4	75.1	72.9
A1	A2	70.5	78.3	74.0	60.6	66.4	63.2	73.7	78.6	75.0
A1	A7	72.6	77.5	74.2	61.3	66.1	63.6	75.1	79.2	76.5
A1	A8	82.2	84.8	83.5	72.3	70.4	70.4	81.5	81.0	80.5
A2	O1	77.3	82.6	79.8	67.2	72.6	69.6	81.0	82.8	81.8
A2	O2	78.6	81.6	80.1	70.4	74.0	71.9	82.0	82.0	82.0
A2	O3	81.2	84.9	83.0	71.0	75.2	73.0	84.9	85.4	85.1
A2	A1	70.9	74.6	71.2	61.5	64.6	60.4	76.5	76.5	74.7
A2	A2	70.6	78.4	74.0	61.2	68.4	63.7	74.7	77.8	75.6
A2	A7	72.6	77.5	74.2	62.4	67.0	63.0	77.6	78.9	77.3
A2	A8	82.2	84.8	83.4	73.8	77.0	75.3	85.6	85.3	85.4
A7	O1	77.1	82.5	79.7	67.6	73.2	69.9	79.4	81.9	80.3
A7	O2	78.5	81.6	80.0	70.6	74.2	71.8	80.6	81.4	80.9
A7	O3	81.0	84.9	82.9	71.3	75.7	73.3	83.1	83.8	83.0
A7	A1	71.0	74.4	70.9	62.1	64.8	60.3	75.8	78.0	75.7
A7	A2	70.5	78.2	73.8	62.0	69.1	64.3	75.3	79.6	76.5
A7	A7	72.6	77.4	74.0	62.2	67.0	63.1	74.5	77.5	75.3
A7	A8	81.9	84.7	83.3	73.7	77.2	75.3	82.8	82.7	82.4
A8	O1	77.0	82.4	79.6	67.2	72.7	69.6	80.8	84.4	81.7
A8	O2	78.4	81.5	79.8	70.4	74.0	71.7	82.0	84.7	83.0
A8	O3	80.9	84.9	82.8	71.0	75.2	72.9	84.7	87.6	85.6
A8	A1	71.0	74.2	70.7	61.4	64.3	60.0	73.7	75.0	71.5
A8	A2	70.4	78.1	73.7	61.7	68.8	64.1	75.0	80.1	75.9
A8	A7	72.6	77.2	73.7	62.2	66.6	62.5	75.7	78.2	74.9
A8	A8	81.9	84.9	83.4	73.6	77.0	75.1	84.2	86.5	85.2

References

[1] McGray A. Promoting health literacy. *J of Am Med Infor Ass* 2005;12:152–63.

[2] Eysenbach G. Poverty, human development, and the role of ehealth. *J Med Internet Res* 2007;9(4):34–4.

[3] Van der Stichele R. Promises for a measurement breakthrough. In: *Drug regimen compliance. Issues in clinical trials and patient management*, 1999:71-83.

[4] Patel V, Branch T, and Arocha J. Errors in interpreting quantities as procedures : The case of pharmaceutical labels. *Int Journ Med Inform* 2002;65(3):193–211.

[5] Zeng-Treiler Q, Kim H, Goryachev S, et al. Text characteristics of clinical reports and their implications for the readability of personal health records. In: MEDINFO, Brisbane, Australia. 2007:1117–21.

[6] Williams M, Parker R, Baker D, et al. Inadequate functional health literacy among patients at two public hospitals. *JAMA* 1995;274(21):1677–82.

[7] Oregon Practice Center . Barriers and drivers of health information technology use for the elderly, chronically ill, and underserved. Technical report, Agency for healthcare research and quality. Oregon Evidence-based Practice Center, 2008.

[8] Brigo F, Otte M, Igwe S, Tezzon F, and Nardone R. Clearly written, easily comprehended ? The readability of websites providing information on epilepsy. *Epilepsy & Behavior* 2015;44:35–9.

[9] Jucks R and Bromme R. Choice of words in doctor-patient communication: an analysis of health-related Internet sites. *Health Commun* 2007;21(3):267–77.

[10] Tran T, Chekroud H, Thiery P, and Julienne A. Internet et soins : un tiers invisible dans la relation médecin/patient ? *Ethica Clinica* 2009;53:34–43.

[11] François T and Fairon C. Les apports du TAL à la lisibilité du français langue étrangère. *TAL* 2013;54(1):171–202.

[12] Flesch R. A new readability yardstick. *Journ Appl Psychol* 1948;23:221–33.

[13] Gunning R. *The art of clear writing*. McGraw Hill, New York, NY, 1973. [14] Kokkinakis D and Toporowska Gronostaj M. Comparing lay and professional language in cardiovascular disorders corpora. In: Pham T., ed, WSEAS Transactions on BIOLOGY and BIOMEDICINE, 2006:429–37.

[15] Poprat M, Markó K, and Hahn U. A language classifier that automatically divides medical documents for experts and health care consumers. In: MIE. 2006:503–8.

[16] Grabar N, Krivine S, and Jaulent M. Classification of health webpages as expert and non expert with a reduced set of cross-language features. In: AMIA, 2007:284–8.

[17] Goeuriot L, Grabar N, and Daille B. Characterization of scientific and popular science discourse in French, Japanese and Russian. In: LREC, 2008.

[18] Miller T, Leroy G, Chatterjee S, Fan J, and Thoms B. A classifier to evaluate language specificity of medical documents. In: HICSS, 2007:134–40.

[19] Chmielik J and Grabar N. Détection de la spécialisation scientifique et technique des documents biomédicaux grâce aux informations morphologiques. TAL 2011;51(2):151-79.

[20] Specia L, Jauhar S, and Mihalcea R. Semeval-2012 task 1: English lexical simplification. In: *SEM 2012, 2012:347–55.

[21] Sinha R. Unt-simprank: Systems for lexical simplification ranking. In: *SEM 2012, 2012:493–6.

[22] Jauhar S and Specia L. UOW-SHEF: SimpLex – lexical simplicity ranking based on contextual and psycholinguistic features. In: *SEM 2012. 2012:477–81.

[23] Johannsen A, Martínez H, Klerke S, and Søgaard A. Emnlp@cph: Is frequency all there is to simplicity? In: *SEM 2012. 2012:408–12.

[24] Ligozat A, Grouin C, Garcia-Fernandez A, and Bernhard D. Annlor: A naïve notation-system for lexical outputs ranking. In: *SEM 2012, 2012:487–92.

[25] Amoia M and Romanelli M. SB: mmSystem - using decompositional semantics for lexical simplification. In: *SEM 2012. 2012:482–6.

[26] Gala N, François T, and Fairon C. Towards a French lexicon with difficulty measures: NLP helping to bridge the gap between traditional dictionaries and specialized lexicons. In: eLEX-2013, 2013.

[27] Zheng W, Milios E, and Watters C. Filtering for medical news items using a machine learning approach. In: AMIA, 2002:949– 53.

[28] Borst A, Gaudinat A, Boyer C, and Grabar N. Lexically based distinction of readability levels of health documents. In: MIE 2008, 2008. Poster.

[29] Grabar N, Hamon T, and Amiot D. Automatic diagnosis of understanding of medical words. In: EACL PITR Workshop, 2014:11–20.

[30] van Oosten P and Hoste V. Readability annotation: Replacing the expert by the crowd. In: Workshop on Innovative Use of NLP for Building Educational Applications, 2011:120–9.

[31] Yaneva V, Temnikova I, and Mitkov R. Accessible texts for autism: An eye-tracking study. In: ACM , ed, Int ACM SIGACCESS Conf on Comp & Accessibility, 2015:49–57.

[32] Grabar N, Farce E, and Sparrow L. Study of readability of health documents with eye-tracking approaches. In: Workshop on Automatic Text Adaption (ATA), 2018:1– 1.

[33] Grabar N and Hamon T. A large rated lexicon with French medical words. In: LREC. 2016:1–2.

[34] Côté RA, Rothwell DJ, Palotay JL, Beckett RS, and Brochu L. *The Systematised Nomenclature of Human and Veterinary Medicine: SNOMED International.* College of American Pathologists, Northfield, 1993.

[35] Schmid H. Probabilistic part-of-speech tagging using decision trees. In: Int Conf on New Methods in Language Processing, 1994:44–9.

[36] Namer F. FLEMM : un analyseur flexionnel du français à base de règles. Traitement automatique des langues (TAL) 2000;41(2):523–47.

[37] Namer F and Zweigenbaum P. Acquiring meaning for French medical terminology: contribution of morphosemantics. In: AMIA. 2004.

[38] Kincaid J, Fishburne R, Rogers R, and Chissom B. Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. Technical report, Naval Technical Training, U. S. Naval Air Station, 1975.

[39] Bojanowski P, Grave E, Joulin A, and Mikolov T. Enriching word vectors with subword information. *Transactions of the Assoc for Comp Linguistics* 2017;5(1):135–46.

[40] Grabar N and Hamon T. Understanding of unknown medical words. In: BIONLP workshop at RANLP, 2017:1–0.

Address for correspondence

Natalia Grabar
CNRS, Univ. Lille, UMR 8163 STL,
F-59000, Lille, France
e-mail: natalia.grabar@univ-lille.fr
url: http://natalia.grabar.free.fr/