# A Large Rated Lexicon with French Medical Words

**Natalia Grabar**[1], **Thierry Hamon**[2]
[1]CNRS, UMR 8163, F-59000 Lille, France;
Univ. Lille, UMR 8163 - STL - Savoirs Textes Langage, F-59000 Lille, France
[2]LIMSI, CNRS, Université Paris-Saclay, Bât 508, rue John von Neumann, Campus Universitaire, F-91405 Orsay
Université Paris 13 - Sorbonne Paris Cité, 99 avenue J.B. Clément, F-93430 Villetaneuse, France
natalia.grabar@univ-lille3.fr, hamon@limsi.fr

**Abstract**

Patients are often exposed to medical terms, such as *anosognosia*, *myelodysplastic*, or *hepatojejunostomy*, that can be semantically complex and hardly understandable by non-experts in medicine. Hence, it is important to assess which words are potentially non-understandable and require further explanations. The purpose of our work is to build specific lexicon in which the words are rated according to whether they are understandable or non-understandable. We propose to work with medical words in French such as provided by an international medical terminology. The terms are segmented in single words and then each word is manually processed by three annotators. The objective is to assign each word into one of the three categories: *I can understand*, *I am not sure*, *I cannot understand*. The annotators do not have medical training nor they present specific medical problems. They are supposed to represent an average patient. The inter-annotator agreement is then computed. The content of the categories is analyzed. Possible applications in which this lexicon can be helpful are proposed and discussed.
The rated lexicon is freely available for the research purposes. It is accessible online at
*http://natalia.grabar.perso.sfr.fr/rated-lexicon.html*.

**Keywords:** Medical area, Medical terminology, Readability, Annotation, Inter-annotator agreement, French

## 1. Introduction

The medical field has a very specific terminology (*anosognosia*, *myelodysplastic*, *hepatojejunostomy* or *anastomosis*) commonly used by medical professionals. For its correct understanding, an extensive knowledge is required, as indeed it appears difficult to understand information contained in drug package inserts (Patel et al., 2002), in websites (Rudd et al., 1999; Berland et al., 2001; McCray, 2005; Oregon Practice Center, 2008), and more generally in patients and medical doctors communication (McCray, 2005; Jucks and Bromme, 2007; Tran et al., 2009).
In this work, we present a lexicon with French medical words, which have been rated by human annotators on a scale with three positions: *I can understand*, *I am not sure*, *I cannot understand*. Study of the understanding of words used in medical field is the first step towards the simplification of medical texts. Indeed, before the simplification can be performed, it is necessary to know which textual units may show understanding difficulty and should be simplified.

## 2. Related Work

The understanding of words is a complex notion closely linked to different research fields, such as linguistics (Section 2.1.), psycholinguistics (Section 2.2.) and Natural Language Processing (NLP) (Section 2.3.).

### 2.1. Linguistics

In linguistics, the question is closely related to the lexical complexity and composition. It has been indeed observed that several factors may be involved in semantic complexity:

1. Knowledge of the components of complex words, which permits to correctly segment the words and then to understand them (*e.g. appendicitis*, *otitis*, *reticulitis*);

2. Variety of morphological patterns and relations among components, (Booij, 2010) which permits to apply patterns such as *[N1N2]* in *erythrocyte* and *ovocyte*, and then to induce the relation between the components and the semantics of the whole lexeme;

3. Polysemous components (Iacobini, 1997; Amiot and Dal, 2005), which may make the understanding more complicated;

4. Difference in the order of components (Iacobini, 2003), which gives the clues for the interpretation of compounds (*e.g. snow tyre* is a kind of *tyre*, *erythrocyte* is a kind of *cyte/cell*).

### 2.2. Psycholinguistics

In psycholinguistics, specific research questions are studied and some of them can be related to those studied in linguistics:

- Knowledge of components within complex words and their decomposition, which main purpose is to study how the complex words are processed and recorded in brain and then to make hypotheses on how these words are decoded and produced by speakers. Thus, it has been demonstrated that several factors can facilitate reading and production of complex words: hyphenation (Bertram et al., 2011), space character (Frisson et al., 2008), presence of other morphologically related words (Lüttmann et al., 2011), use of primes (Bozic et al., 2007; Beyersmann et al., 2012), of pictures (Dohmes et al., 2004; Koester and Schiller, 2011) and of supportive contexts (Cain et al., 2009);

- Order of components and variety of morphological patterns, which focus on the position of the components. It has been observed that these factors show a stable influence for the recognition of complex words (Libben et al., 2003; Holle et al., 2010; Feldman and Soltano, 1999). The notion of the morphological headedness has been isolated (Jarema et al., 1999; Libben et al., 2003): the related work indicates that the morphological head plays an important role in decomposition of complex words, in detection of decomposition patterns, and more generally in lexical activity.
- Impact of the word length and of the types of affixes (Meinzer et al., 2009) and the frequency of bases (Feldman et al., 2004).

### 2.3. Natural Language Processing

In the NLP area, a variety of work is dedicated to the word understanding and readability studies. The purpose is to decide whether given documents are accessible for a given reader. The readability measures are widely used for evaluating complexity of documents. Among the existing measures, it is possible to distinguish classical and computational measures:

- Classical measures usually exploit information on number of characters and/or syllables within words, and on linear regression models (Flesch, 1948; Gunning, 1973);
- Computational measures can use vector models and a great variety of descriptors, among which the following have been utilized: combination of classical readability measures with medical terminologies (Kokkinakis and Toporowska Gronostaj, 2006); n-grams of characters (Poprat et al., 2006); stylistic (Grabar et al., 2007) and discursive (Goeuriot et al., 2007) descriptors; lexicon (Miller et al., 2007); morphological information (Chmielik and Grabar, 2011); combination of different kinds of descriptors (Wang, 2006; Zeng-Treiler et al., 2007; Leroy et al., 2008; François and Fairon, 2013).

## 3. Description of the Source Terminology

The source terms are obtained from the medical terminology Snomed International (Côté, 1996) in French, available from the ASIP SANTE website[1]. The purpose of this terminology is to provide an extensive description of the medical field. Snomed contains 151,104 medical terms structured into eleven semantic axes such as disorders and abnormalities, procedures, chemical products, living organisms, anatomy, social status, etc. We keep here five axes related to the main medical notions (disorders, abnormalities, procedures, functions, anatomy), which are the categories the most often used in medical texts. The objective is not to consider axes such as chemical products (*hydrogen sulfide*) and living organisms (*Sapromyces, Acholeplasma laidlawii*) that group very specific terms, hardly known by laymen, and that can be easily categorized as non-understandable by patients.

[1] *http://esante.gouv.fr/services/referentiels/referentiels-d-interoperabilite/snomed-35vf*

## 4. Pre-processing of the Source Lexicon

The 104,649 selected terms are tokenized in words (or tokens), POS-tagged and lemmatized by TreeTagger (Schmid, 1994) and then corrected by Flemm (Namer, 2000). The syntactic categories are assigned to words within the context of their terms. In order to make syntactically acceptable structures, the terms are transformed in sentence. For instance, *infarctus du myocarde* becomes *C'est un infarctus du myocarde* ((*This is a myocardial infarction*)). When hyphenated, the prefixes are not separated from their bases (*e.g. anti-virus*), the compounds are not tokenized either (*e.g. canaliculo-rhinostomie*). Similarly, the names of chemicals remain non-tokenized *acétylgalactosaminyl-O-glycosyl-glycoprotéine*.
In this way, we obtain 29,641 unique words. For instance, *infarctus du myocarde* (*myocardial infarction*) gives three words in French (*infarctus, du, myocarde*).
¿From the morphological point of view, this dataset contains three kinds of words:

- compound words which contain several bases: *abdominoplastie (abdominoplasty), dermabrasion (dermabrasion)*;
- constructed (or derived) words which contain one base and at least one affix: *cardiaque (cardiac), acineux (acinic), lipoïde (lipoid)*;
- simple words which contain one base, no affixes and possibly inflections (when the lemmatization fails): *acné (acne), fragment (fragment)*.

In addition to French words, this dataset contains also borrowings from other languages (Latin, Greek, English), and abbreviations.
The stopwords are also removed during the pre-processing step.

## 5. Annotation Process

The 29,641 words from the original set are annotated by three French speakers, 25-40 year-old, without medical training, without specific medical problems, but with linguistic background. To our opinion, these annotators represent the average knowledge of medical words among the population as a whole. One of the authors participated in the annotation process.
The annotators are presented with a list of terms and asked to assign each word to one of the three categories:

1. *I can understand the word*;
2. *I am not sure about the meaning of the word*;
3. *I cannot understand the word.*

The assumption is that the words, which are not understandable by the annotators, are also difficult to understand by patients.
The annotators were asked not to use dictionaries during this annotation. Given the large number of words to process, the annotation process needed one to two months.

| *Categories* | *A1* | *A2* | *A3* | *Una.* | *Maj.* |
|---|---|---|---|---|---|
| | *Nb. (%)* | *Nb. (%)* | *Nb. (%)* | *Nb. (%)* | *Nb. (%)* |
| *1. I can understand* | 8,099 (28,0%) | 8,625 (29,0%) | 7,529 (25,0%) | 5,960 (26,0%) | 7,655 (27,0%) |
| *2. I am not sure* | 1,895 (6,0%) | 1,062 (4,0%) | 1,431 (5,0%) | 61 (0,3%) | 597 (2,0%) |
| *3. I cannot understand* | 19,647 (66,0%) | 19,954 (67,0%) | 20,681 (70,0%) | 16,904 (73,7%) | 20,511 (71,0%) |
| *Total annotated words* | 29,641 | 29,641 | 29,641 | 22,925 | 28,763 |

Table 1: Number (and percentage) of words assigned to reference categories by human annotators (A1, A2 and A3).

## 6. Description and Discussion of the Rated Lexicon with Medical Words

Description and discussion of the rated lexicon is done following several points: presentation of the annotation output (Section 6.1.), discussion of the annotation process (Section 6.2.), analysis and discussion of annotations (Section 6.3.), presentation of possible applications and expected impact of the lexicon (Section 6.4.).

### 6.1. Annotation Output

The results of the manual annotations are presented in Table 1. On the basis of the annotations, we can create five sets of data: one for each annotator (*A1, A2* and *A3*), the *Unanimity* set in which all the annotators agree, and the *Majority* set in which at least two annotators agree.

We can see that the datasets corresponding to the three annotators provide very similar distribution of words among the three categories. The less frequent category is *I am not sure* (between 4 and 6% of the whole set of words), which means that the decision on the understanding of words from specialized areas is quite easy to perform for non-expert annotators. The most frequent category is *I cannot understand*: it gathers 66 to 70% of words. This means that highly specialized areas, such as medicine, contain a large number of specialized expert words which may require explanations in order to be correctly understood and used by patients.

### 6.2. Discussion of the Annotation Process

The annotation process implies only that the annotators give their opinion on their understanding of words and that they assign each word to one of three categories. Given the large size of the set of words (almost 30,000 words), it is impossible to verify whether this understanding is real or not. For collecting a more correct judgment on the understanding of words, it would be necessary to require that annotators also provide an explanation to the processed words, which should be then checked out (Zeng et al., 2005) by experts. As we indicated, such an approach is hardly possible when processing such a large set of words. The categories are defined is such a way that they may include both (un)known words and words felt as such by the annotators. Several situations can occur, such as:

- when a word is known and understood, then this word is to be assigned to the *I can understand* category;
- when a word has already been read or heard by the annotator, but if its meaning is not known, if the word itself cannot be parsed, interpreted, and if the global meaning cannot be deduced by the annotator, then this word is to be assigned to the *I cannot understand* category;
- when a word is unknown but if its components (bases, affixes) can be parsed, interpreted, and the global meaning can be then deduced by the annotator, then this word can be assigned to the *I can understand* category;
- when a word is unknown and if its components (bases, affixes) cannot be parsed, interpreted, and the global meaning cannot be deduced by the annotator, then this word can be assigned to the *I cannot understand* category.

### 6.3. Analysis and Discussion of Annotations

The inter-annotator agreement is computed with the Cohen's Kappa (Cohen, 1960), applied to pairs of annotators, which values are then leveraged to obtain the unique average value; and Fleiss' Kappa (Fleiss and Cohen, 1973), suitable for the processing of data provided by more than two annotators. The standard interpretation of the scores are for instance (Landis and Koch, 1977): substantial agreement between 0.61 and 0.80, almost perfect agreement between 0.81 and 1.00. On our dataset, the inter-annotator agreement shows substantial agreement: Fleiss' Kappa 0.735 and Cohen's Kappa 0.736.

Regarding the *Majority* and *Unanimity* sets, they indicate that 878 words do not have the majority agreement, while up to 6,716 words do not have the unanimity agreement.

An analysis of the content of these categories indicates that *I cannot understand* category contains several types of words, some of which have already been noticed in the related work (Section 2.):

- neoclassical compounds (*e.g. coproporphyrie, abiotrophie, dermacentorose, dysurie, abomasopexie, angiomyoliposarcome, fistulo-végétant*). As indicated in Section 2., compounds are also addressed in studies in linguistics and psycholinguistics because such words show complex morphological structure and are, for this reason, difficult to understand. Indeed, several factors induce this complexity (*e.g.* knowledge of the components, morphological patterns and relations, polysemy of components);
- abbreviations (*e.g. ADPase, Pro-leu, Fyx, Gln-glu, Hga*). Abbreviations correspond to specific representations of usually complex terms. Unless already

known by a reader, the abbreviations provide very little information on their content and meaning. We assume that this is the main reason why abbreviations are systematically categorized as non-understandable;

- Latin borrowings (*e.g. Chrysocoma, Adiaspiromycoses, aborta, urtica, aberrans, abdominalis, dolens*). Since Latin borrowings are not commonly used in the modern French language, the current usage provides no or little examples with their occurrences and, for this reason, Latin borrowings may remain opaque for the speakers. Even if some of them show similar surface form with the corresponding French words, such as {*abdominalis*, *abdomen*}, the Latin grammatical and casual system may prevent their parsing and understanding;

- proper names (*e.g. Christiansborg, Malacarne, Glasgow, Anton-Babinski*). The difficulty with the proper names is that they are often used within specific contexts (*e.g.* disorder or laboratory test names). Besides, they are closely connected to the research work performed by the researchers who gave their name to a given disease or examination, to the place in which this research or discovery have been performed, etc. Unless already known, it remains difficult to understand such words and terms properly;

- anatomical terms (*e.g. coracohuméral, abdominalis, diaphragmatique, acral, endaurale*). The main difficulty with the anatomical terms and words is that they convey very specific and precise meaning, are seldom used by non-experts in medicine and keep close links and resemblance with the corresponding Latin or Greek words;

- chemicals (*e.g. N-acétylgalactosaminyltransférase, abequosyltransférase, P-crésol, aminocaproate, diméthylsulfoxide, UDP-N-acétylmuramoyl-L-alanyl-D-glutamyl-L-lysine*). As for the chemical names, these are artificially coined words and terms (Klinger et al., 2008). Usually, their objective is to describe or represent the structure of the corresponding chemical products and molecule. As noticed above for other categories of words, the names of the components of chemicals may also be borrowed from Latin. Besides, the names of chemicals obey to specific morphological rules (World Health Organization, 2006), which remain opaque to non-experts.

### 6.4. Possible Applications and Impact of the Lexicon

This graded lexicon is freely available for the research purposes. It is accessible online[2].
The proposed lexicon can be used in different ways:

- One of the questions asked when working with specialized areas is *How many specialized words the medical area contains?* Our work and the lexicon built can answer such questions precisely, because the *I cannot understand* category may be associated with specialized words, while the *I can understand* category may be associated with non-specialized words. Hence, the rate of specialized and technical words within the medical area reaches up to 70% (around 20,000 words), while up to 30% of words (around 8,000 words) may be considered as general-language words. Of course, the number of technical words and terms will increase if more medication and chemical names are considered;

- This lexicon provides a good material for training and fitting supervised or non-supervised machine-learning algorithms (Gala et al., 2013; Grabar et al., 2014);

- This lexicon, build with material from specialized medical area, can be compared with similar general-language lexicon (Gala et al., 2013) and with similar lexica from other specialized areas if available. Similarly, cross-language comparison may be performed if similar rated lexica exist in other languages. This may help in making contrastive studies across languages, discourses and domains;

- In relation with the readability task, the lexicon can be used for making the diagnosis of difficulty and understanding of a given piece of text (Wang, 2006; Zeng-Treiler et al., 2007; Borst et al., 2008; Leroy et al., 2008);

- Going beyond the diagnosis of difficulty and understanding of a given piece of text, the lexicon can be used for the detection of zones in documents which contain complex and non-understandable terms and words. An example of this utilization is shown in Figure 1, in which potentially difficult words are marked in red. The recognition of the zones can be done by direct projection of the rated lexicon or through the application of a machine-learning algorithm trained with the rated lexicon and then applied to raw text (Grabar et al., 2014);

- In relation with the simplification task, we assume the words from the *I cannot understand* category must be provided with explanations and definitions in order to make them understandable by non-expert users of the medical and health texts. In this way, this lexicon may be used prior to the simplification tasks in order to guide the building of vocabulary with terms and words which meaning should be explained;

- In the patient-medical doctor communication, this lexicon may focus attention of the medical staff on notions which should be explained to patients in order to make the communication more successful and easy.

## 7. Conclusion

In this work, we propose to build a lexicon with French medical words, which are assigned to three categories (*I can understand*, *I am not sure*, *I cannot understand*). The

[2] `http://natalia.grabar.perso.sfr.fr/rated-lexicon.html`

Histoire de la maladie
Le patient a été hospitalisé le 18 / 7 / 11 à PELLEGRIN pour un AVC ischémique dans le territoire profond de l' artère cérébrale postérieure droite , thrombolysé à H + 3 .

Le patient présente , comme déficit , une hypoesthésie gauche et une parésie gauche ( force motrice à 1 / 5 au membre supérieur gauche et 2 / 5 au membre inférieur gauche ) , un NIHSS à 8 , une désorientation tempora-spatiale et une vigilance fluctuante . Dans les suites , est survenu un OAP post thrombolyse , probablement iatrogène ( scanner injecté et NaCl afin de visualiser la zone de thrombolyse ) .

Le patient est donc transféré en réanimation : l' OAP est résolutif sous VNI et oxygénothérapie .

La majoration de l' insuffisance rénale nécessite 2 cures de dialyse . Mr K . est ensuite transféré en post-réanimation devant l' évolution favorable et revient en service de neurologie à Pellegrin pour suite de la prise en charge .

Le 11 / 8 / 2011 , le patient présente une douleur thoracique associée à une désaturation à 83 % , il est donc transféré en Unité de soins intensifs cardiologiques . Une embolie pulmonaire basale droite est mise en évidence par une scintigraphie pulmonaire . Une anticoagulation curative par CALCIPARINE est mise en place .

Figure 1: Detection of non-understandable words within clinical discharge summaries.

lexicon contains almost 30,000 words. Three annotators participate in the annotation process. The annotators have linguistic background, but do not have medical training, nor do they have specific medical problems. To our opinion, these annotators may represent the average knowledge of medical words among the population as a whole.
The content of the obtained lexicon is then presented and discussed. Besides, we also outline some possible usages and applications for which the lexicon can be helpful. The lexicon is freely available for the research purposes. It is accessible online at `http://natalia.grabar.perso.sfr.fr/rated-lexicon.html`.

## 8. Bibliographical References


Amiot, D. and Dal, G. (2005). Integrating combining forms into a lexeme-based morphology. In *Mediterranean Morphology Meeting (MMM5)*, pages 323–336.

Berland, G., Elliott, M., Morales, L., Algazy, J., Kravitz, R., Broder, M., Kanouse, D., Munoz, J., Puyol, J., and et al, M. L. (2001). Health information on the internet. accessibility, quality, and readability in english ans spanish. *JAMA*, 285(20):2612–2621.

Bertram, R., Kuperman, V., Baayen, H. R., and Hyönä, J. (2011). The hyphen as a segmentation cue in triconstituent compound processing: It's getting better all the time. *Scandinavian Journal of Psychology*, 52(6):530–544.

Beyersmann, E., Coltheart, M., and Castles, A. (2012). Parallel processing of whole words and morphemes in visual word recognition. *The Quarterly Journal of Experimental Psychology*, 65(9):1798–1819.

Booij, G. (2010). *Construction Morphology*. Oxford University Press, Oxford.

Borst, A., Gaudinat, A., Boyer, C., and Grabar, N. (2008). Lexically based distinction of readability levels of health documents. In *MIE 2008*. Poster.

Bozic, M., Marslen-Wilson, W. D., Stamatakis, E. A., Davis, M. H., and Tyler, L. K. (2007). Differentiating morphology, form, and meaning: Neural correlates of morphological complexity. *Journal of Cognitive Neuroscience*, 19(9):1464–1475.

Cain, K., Towse, A. S., and Knight, R. S. (2009). The development of idiom comprehension: An investigation of semantic and contextual processing skills. *Journal of Experimental Child Psychology*, 102(3):280–298.

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

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37–46.

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

Dohmes, P., Zwitserlood, P., and Bölte, J. (2004). The impact of semantic transparency of morphologically complex words on picture naming. *Brain and Language*, 90(1-3):203–212.

Feldman, L. B. and Soltano, E. G. (1999). Morphological priming: The role of prime duration, semantic transparency, and affix position. *Brain and Language*, 68(1-2):33–39.

Feldman, L. B., Soltano, E. G., Pastizzo, M. J., and Francis, S. E. (2004). What do graded effects of semantic transparency reveal about morphological processing? *Brain and Language*, 90(1-3):17–30.

Fleiss, J. and Cohen, J. (1973). The equivalence of weighted Kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, 33:613–619.

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 23:221–233.

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

Frisson, S., Niswander-Klement, E., and Pollatsek, A. (2008). The role of semantic transparency in the processing of english compound words. *Br J Psychol*, 99(1):87–107.

Gala, N., François, T., and Fairon, C. (2013). Towards a french lexicon with difficulty measures: NLP helping to bridge the gap between traditional dictionaries and specialized lexicons. In *eLEX-2013*.

Goeuriot, L., Grabar, N., and Daille, B. (2007). Caractérisation des discours scientifique et vulgarisé en français, japonais et russe. In *TALN*, pages 93–102.

Grabar, N., Krivine, S., and Jaulent, M. (2007). Classification of health webpages as expert and non expert with a reduced set of cross-language features. In *AMIA*, pages 284–288.

Grabar, N., Hamon, T., and Amiot, D. (2014). Automatic diagnosis of understanding of medical words. In *Workshop on Predicting and Improving Text Readability for*

*Target Reader Populations*, pages 11–20, Gothenburg, Sweden.

Gunning, R. (1973). *The art of clear writing*. McGraw Hill, New York, NY.

Holle, H., Gunter, T. C., and Koester, D. (2010). The time course of lexical access in morphologically complex words. *Neuroreport*, 21(5):319–323.

Iacobini, C. (1997). Distinguishing derivational prefixes from initial combining forms. In *First mediterranean conference of morphology*, Mytilene, Island of Lesbos, Greece, septembre.

Iacobini, C., (2003). *Composizione con elementi neoclassici*, pages 69–96.

Jarema, G., Busson, C., Nikolova, R., Tsapkini, K., and Libben, G. (1999). Processing compounds: A cross-linguistic study. *Brain and Language*, 68(1-2):362–369.

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

Klinger, R., Kolárik, C., Fluck, J., Hofmann-Apitius, M., and Friedrich, C. (2008). Detection of iupac and iupac-like chemical names. In *ISMB 2008*, pages 268–276.

Koester, D. and Schiller, N. O. (2011). The functional neuroanatomy of morphology in language production. *NeuroImage*, 55(2):732–741.

Kokkinakis, D. and Toporowska Gronostaj, M. (2006). Comparing lay and professional language in cardiovascular disorders corpora. In Australia Pham T., James Cook University, editor, *WSEAS Transactions on BIOLOGY and BIOMEDICINE*, pages 429–437.

Landis, J. and Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33:159–174.

Leroy, G., Helmreich, S., Cowie, J., Miller, T., and Zheng, W. (2008). Evaluating online health information: Beyond readability formulas. In *AMIA 2008*, pages 394–8.

Libben, G., Gibson, M., Yoon, Y. B., and Sandra, D. (2003). Compound fracture: The role of semantic transparency and morphological headedness. *Brain and Language*, 84(1):50–64.

Lüttmann, H., Zwitserlood, P., and Bölte, J. (2011). Sharing morphemes without sharing meaning: Production and comprehension of german verbs in the context of morphological relatives. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 65(3):173–191.

McCray, A. (2005). Promoting health literacy. *J of Am Med Infor Ass*, 12:152–163.

Meinzer, M., Lahiri, A., Flaisch, T., Hannemann, R., and Eulitz, C. (2009). Opaque for the reader but transparent for the brain: Neural signatures of morphological complexity. *Neuropsychologia*, 47(8-9):1964–1971.

Miller, T., Leroy, G., Chatterjee, S., Fan, J., and Thoms, B. (2007). A classifier to evaluate language specificity of medical documents. In *HICSS*, pages 134–140.

Namer, F. (2000). FLEMM : un analyseur flexionnel du français à base de règles. *Traitement automatique des langues (TAL)*, 41(2):523–547.

Oregon Practice Center. (2008). 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.

Patel, V., Branch, T., and Arocha, J. (2002). Errors in interpreting quantities as procedures : The case of pharmaceutical labels. *International journal of medical informatics*, 65(3):193–211.

Poprat, M., Markó, K., and Hahn, U. (2006). A language classifier that automatically divides medical documents for experts and health care consumers. In *MIE 2006 - Proceedings of the XX International Congress of the European Federation for Medical Informatics*, pages 503–508, Maastricht.

Rudd, R., Moeykens, B., and Colton, T., (1999). *Annual Review of Adult Learning and Literacy*, page ch 5.

Schmid, H. (1994). Probabilistic part-of-speech tagging using decision trees. In *International Conference on New Methods in Language Processing*, pages 44–49.

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

Wang, Y. (2006). Automatic recognition of text difficulty from consumers health information. In IEEE, editor, *Computer-Based Medical Systems*, pages 131–136.

World Health Organization. (2006). *The use of stems in the selection of International Nonproprietary Names (INN) for pharmaceutical substances*. World Health Organization, Geneva, Switzerland.

Zeng, Q. T., Kim, E., Crowell, J., and Tse, T. (2005). A text corpora-based estimation of the familiarity of health terminology. In *ISBMDA 2006*, pages 184–92.

Zeng-Treiler, Q., Kim, H., Goryachev, S., Keselman, A., Slaugther, L., and Smith, C. (2007). Text characteristics of clinical reports and their implications for the readability of personal health records. In *MEDINFO*, pages 1117–1121, Brisbane, Australia.