

Helping term sense disambiguation with active learning

Pierre André Ménard

Centre de recherche informatique
de Montréal, Canada
pamenard@gmail.com

Caroline Barrière

Centre de recherche informatique
de Montréal, Canada
caroline.barriere@crim.ca

Jean Quirion

École de traduction
Université d'Ottawa, Canada
jquirion@uottawa.ca

Abstract

Our research highlights the problem of term polysemy within terminometrics studies. Terminometrics is the measure of term usage in specialized communication. Polysemy, especially within single-word terms as we will show, prevents using term corpus frequencies as appropriate statistics for terminometrics. Automatic *term sense disambiguation*, as a possible solution, requires human annotation to feed a supervised learning algorithm. Within our experiments, we show that although being polysemous, terms have a strong in-domain sense bias, making random sampling of annotation data less than optimal. We suggest the use of active learning and implement it within an annotation platform as a way of reducing annotation time.

1 Introduction

In our research, we investigate the measure of term usage in specialized communication, called Terminometrics (Section 2). The studied *terms* are provided by a *term bank*, and we are interested in the fact that many of these terms are polysemous, creating difficulties for our terminometrics study. We therefore investigate the presence of polysemy in term banks (Section 3). Polysemy found in term banks is problematic since it leads to term occurrences in corpora being possibly polysemous, preventing simple corpus frequencies to provide proper statistics. We confirm such polysemous occurrences within our specific terminometrics experiment in the nanotechnology domain (Section 4), as we analyse term sense human annotation results for a set of nanotechnology terms.

Results also show that terms, although polysemous, have a very strong bias toward their in-domain sense. In such biased case, a random sampling of annotation data is far from optimal, wasting much human effort. We therefore introduce active learning (Section 5) and implement it within an annotation platform (Section 6), to obtain a sense-annotated dataset in less time.

2 Terminometrics

Terminometrics is the measure of term usage in different types of communications (Quirion, 2006). Its purpose is to determine, for a particular concept, the relative corpus frequencies of its competing terms.

The protocol of terminometrics, as defined in Quirion (2003), consists in first deciding on a domain of interest and selecting its set of concepts (most often all) from a term bank. Then, for each particular concept, the individual number of occurrences of all its competing terms is counted within different corpora from the same domain gathered by terminologists to represent different communicative settings. Acknowledging the possible polysemy of competing terms, the protocol includes a human expert, to actually disambiguate a randomly selected subset of occurrences, and obtain better estimates of real frequencies.

A good example of this would be the concept of a *atomic cluster* within the *nanotechnology* domain. According to the term bank used, such notion can be expressed by the following 6 terms *atomic cluster*, *atom cluster*, *atomic aggregate*, *atom aggregate*, *cluster* and *aggregate*. In terminometrics, comparative studies of use of terms in specialized communications, government liter-

ature, specialized media, and general media are of interest, as they might reveal how some terms are used by the general public, while others are used by more official government documents.

Studying the occurrence in text of different synonyms of concepts would not be problematic if each one was monosemous. But unfortunately, that is not the case. For example, referring to Table 1, the term *cluster* is a competing term for multiple concepts, and simply counting its occurrences in text, without disambiguation, would not be indicative of its usage for any of them.

Obviously, human annotation is costly, and the possibility of performing automatic term sense disambiguation is quite appealing. In terminometrics, concepts are evaluated one at a time, reducing the disambiguation task to a binary decision. The annotation is not a selection among N senses, but rather a yes/no decision on whether the current instance represents the current concept or not. Furthermore, term disambiguation within terminometry cannot be dealt with similarly to more typical word-sense disambiguation or even term-sense disambiguation relying on knowledge contained in an external resource (Barrière, 2010) since the annotator, or the algorithm, is likely to only have access to the context of occurrences to perform term disambiguation.

3 Polysemy of specialized terms

Terms for the terminometrics studies are provided by term banks. Such repositories of terms are not often investigated for the study of polysemy. In Natural Language Processing, a typical task of word sense disambiguation requires a lexicographic resource, such as WordNet (Miller, 1995), to provide a repository of possible word senses in order to disambiguate words in texts (Pantel and Lin, 2002). No doubt that words are polysemous, even in specific domains (Chroma, 2011; Vogel, 2007), but less studies show and discuss the polysemy of terms.

Terms are single-word or multi-word expressions denoting particular concepts within particular domains. A term bank is organized by domains (e.g. biology, automotive, etc) and contains records corresponding to concepts. Each record contains at least one term, and often competing terms (synonyms) denoting that concept, possibly in more than one language. Examples of records

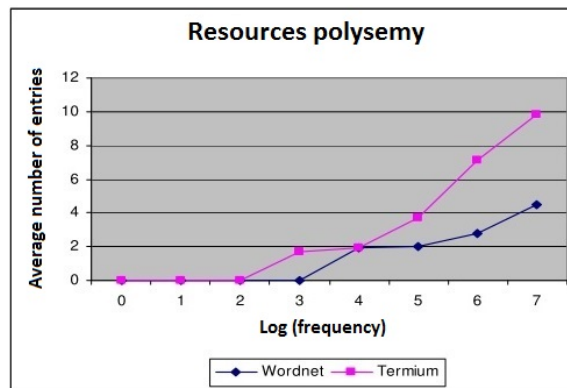


Figure 1: Degree of polysemy in Wordnet and Termium per term length

for the term *cluster*, as found in the Grand Dictionnaire Terminologique (GDT)¹ are shown in Table 1.

There might be a misconception that specialized language is less ambiguous, and would then not provide a proper challenge for word-sense disambiguation. A study by Barrière (2007), shows the contrary, as Wordnet and Termium² (the actual resource used in this experiment) were compared along different criteria. One criteria of comparison was coverage, and another one, more of our interest in this research, is the degree of polysemy in relation to word specificity. Word specificity was approximated by "hit counts", as found in a very large corpora (Waterloo Terabyte Corpus, used by Terra and Clarke (2003)), with words occurring from 1 to millions of times. Figure 1 shows their results. We see how for common words (hit counts in the $\log_{10}(freq) > 3$), the degree of polysemy in the term bank is even larger than in WordNet.

In our study, we wished to further characterize this degree of polysemy in terminological resources. We used a small set of 164 terms from the current experiment (presented in Section 4.1), and looked at the number of senses in two term banks: Termium and GDT. Figure 2 shows that specialized terms, especially short ones (1 to 3 words) can have many senses (records) and span many domains. This trend generally diminishes as the term length increases.

¹The GDT can only be accessed via a web interface at <http://www.granddictionnaire.com>.

²Termium term bank can be accessed online at <http://www.btb.termiumplus.gc.ca> or downloaded at <http://open.canada.ca/data/en/dataset/94fc74d6-9b9a-4c2e-9c6c-45a5092453aa>

Domain	Terms
nanotechnology	atomic aggregate, cluster , aggregate, atom aggregate, atom cluster, atomic cluster
	molecular aggregate, cluster , aggregate, molecule aggregate, molecule cluster
	nanoaggregate, cluster , aggregate, nanocluster, nanometer-size cluster, nanoscale aggregate, nanoscale cluster
seafood	crab section, section, crab cluster, cluster
software	cluster , document cluster
mining	vein system, vein set, cluster of veins, mining cluster, cluster
internet	service cluster, cluster of service, cluster
nanotechnology	scanning tunneling electron microscope, microscope , scanning tunneling microscope, STM
	atomic force microscope, microscope , AFM, SFM, scanning force microscope
	magnetic force microscope, microscope , MFM, SMM, scanning magnetic microscope
	scanning probe microscope, microscope , SPM, scanned-probe microscope

Table 1: Different records for "cluster" and "microscope".

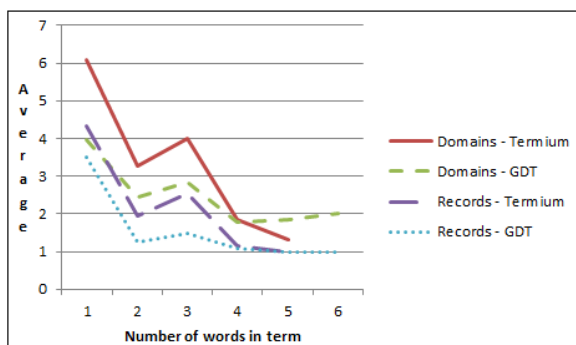


Figure 2: Degree of polysemy in Termium and GDT per term length

4 Experiment - Terminometrics in nanotechnology domain

Our current terminometrics study focuses on term usage in the nanotechnology domain within Canadian French. This domain, within the GDT term bank, contains 1,035 records (concepts)³, each with its competing terms. This set of terms is what we call our *nanotechnology term base* covering "the science of working with atoms and molecules to build devices that are extremely small" (Merriam-Webster dictionary).

To study the competing terms for the nanotechnology concepts, a corpus was built using documents from corporative, educational, news medias and government websites. These documents were retrieved first by selecting most of the organizations originating from the province of Québec, Canada, and whose core activities dealt with nanotechnology. This list was then vetted by an expert. Next, the websites of these organizations

³As the GDT expands everyday, this number might not represent its current status.

were downloaded. After such process, the corpus might still be noisy, but it does contain a majority of nanotechnology-related documents.

All terms in the nanotechnology term base are searched for in the corpus. For each of their occurrences, a window spanning 90 characters each side of the term is extracted. This text span becomes a contextualized instance to be annotated. Table 2 shows examples of these instances.

4.1 Human annotation process

For our current annotation experiment, a total of 164 terms taken from 29 records (among the 1,035 mentioned earlier) were selected along with the complete set of instances found in the nanotechnology corpus. Each term occurred between 75 to 2100 times in the corpus for a total of 17,227 instances for the whole term sample. This dataset was divided into two parts distributed between 2 PhD students in terminology. As shown in Table 2, annotators were presented text sample with a targeted term and were asked to indicate "yes" if the term was used in the correct nanotechnology sense and "no" otherwise. Prior to the annotation effort, the dataset was sorted by terms, as this was considered easier to annotate compared to an annotation by document order, which would ask the annotator to constantly switch between term definitions. They took a total of 82 hours (41 hours each) to annotate all the instances of the selected dataset. Each text sample was composed of the 90 characters prior to a term occurrence, the term occurrence as is, and another 90 characters following the term occurrence. The 90 characters window was adjusted to avoid word truncation.

The annotators were also asked to indicate the difficulty level of the provided answer: standard,

Annotation	Instance
Yes	... une technologie d'intégration par laquelle plusieurs nanostructures sont intégrées sur un même substrat . L'interface entre les dispositifs et d'autres systèmes (oxyde, verre) sera aussi étudiée. (... <i>an integration technology for which many nanostructures are integrated on a substrate. The interface between the components in other systems (oxyde, glass) will also be studied.</i>)
Yes	... dollars à Bromont dans une petite usine qui allait employer 200 personnes pour la production de substrats , que le dictionnaire définit comme un matériau sur lequel sont réalisés les éléments d'un ... (... <i>dollars at Bromont in a small factory which was going to employ 200 people for the production of substrates, which dictionary define as a material on which are realized elements of...</i>)
No	... et valoriser les boues de station d'épuration. L'investigation des possibilités d'acquérir ces substrats requiert l'inventaire des industries de la région, les quantités et les caractéristiques des ... (... <i>and valorize the epuration station's muds. Investigating the possibility of acquiring these substrates requires to inventoriate the region's industries, the quantity and features of...</i>)
Yes	... MNT Définition : Fabrication mécanique et contrôlée de structures moléculaires, par une approche ascendante qui consiste à les assembler, étape par étape, molécule par molécule, en se servant d'appareil ... (... <i>MNT Definition : Mechanical and controled fabrication of molecular structures by a bottom-up approach which consist of assembling, step (by step, molecule by molecule, by using tool...</i>)
No	... Quand il est possible de le faire, l'analyse de la demande d'énergie est fondée sur une approche ascendante agrégeant les demandes par usage, par secteur d'activités économiques, par région et par ... (<i>When it is possible to do it, the energy request analysis is founded on a bottom-up approach aggregating the requests by use, by economic activity sector, by regions and by...</i>)
No	... que beaucoup de problèmes rencontrés en pratique ne sont pas adressés par ces processus. L' approche ascendante de l'amélioration du processus consiste donc, selon ces mêmes auteurs, à implanter une équipe ... (... <i>that many issues encountered in practice are not adressed by these processes. The bottom-up approach of process improvement consist of, for these same authors, implanting a team...</i>)

Table 2: Instances for the terms *substrat* (substrate) and *approche ascendante* (bottom-up approach)

hard, hardest. Results showed that 626 instances (3.6%) needed a little more analysis while 222 instances (1.3%) were much harder to annotate with only the presented context. All the other instances were judged of standard difficulty meaning that the textual contexts of the term occurrences were sufficient for the disambiguation task. In anticipation of an automatic disambiguation algorithm which would only have access to the immediate context of the term, this confirmed that for most cases, it should be possible to disambiguate with a ± 90 characters window⁴.

4.2 Observations and results on polysemy

Analysis of the annotated instances reveals that 84.31% (14,524) of them occur in the correct nanotechnology sense of the term, and the remaining 15.69% (2,703 instances) are used with other meanings. To measure the overall polysemy in our dataset, we use the notion of entropy. Entropy is defined as a summation of all possible event probabilities multiplied by the log of their probability. In our current experiment, there are only two possible events, first the occurrence of a term in a correct sense, let us call that x , and second, the oc-

currence of a term in a different sense. If $P(x)$ is the probability of the correct sense, then $1 - P(x)$ is the probability of another sense. Then, we have the entropy, shown in Equation 1, as a sum over two possible events.

$$E(x) = P_x \log_2 P_x + (1 - P_x) \log_2 (1 - P_x) \quad (1)$$

The resulting function is at its maximum, a value of 1, with a probability of 50% and is equal to 0 with probabilities of either 0% or 100%. In our case, x is the rate of occurrence of an anticipated term sense in a corpus. A term with an entropy of 0 would mean it is not ambiguous, either all or none of the term's instances use the correct sense, and a term with an entropy of 1 would mean 50% of its instances are used in the correct sense, the remaining 50% of the instances using other meanings.

For example, the term *STM* (acronym of *scanning tunnelling microscope*) counts as a single-word term occurring a total of 341 times. Among those, 104 instances ($104/341=0.30499$) have the nanotechnology sense, which gives an entropy of 0.8873 as shown in Equation 2. This is a relatively high entropy level as it nears the 50% maximum. If the case would have been less ambiguous, for example 5 out of 341 instances, the entropy would have been 0.1103.

⁴This claim disregards the fact that humans certainly have much apriori knowledge which they use during the disambiguation task. Nevertheless, trigger of this apriori knowledge would still come from the limited context window.

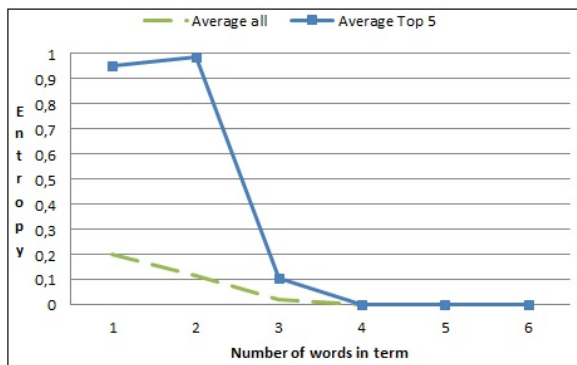


Figure 3: Average entropy per length on gold corpus

$$\begin{aligned}
 E(STM) &= 0.30499 \times \log_2(0.30499) + \\
 &\quad (1 - 0.30499) \times \log_2(1 - 0.30499) \quad (2) \\
 &= 0.8873
 \end{aligned}$$

The bottom dashed line (Figure 3) shows the average entropy over all terms having a particular word count. The top full line shows the average entropy for the 5 terms with the highest entropy (and thus the highest degree of ambiguity) of each length, emphasizing how a few terms account for much of the corpus polysemous instances. Examples of these very polysemous terms are *tunnelling*, *substrat*, or *top-down*.

These corpus results, showing an overall tendency for entropy to decrease with term length, are in line with our previous results presented in Figure 2 relating term length to the polysemy level within term banks. Nevertheless, these corpus results also show that the in-domain sense is much more likely than all other senses. This leads us to think that we should take advantage of the particularity of our task in selecting the annotation dataset, as we further describe in the next section.

5 Active learning for term sense annotation

The strong in-domain sense bias results shown in the previous section, indicate that random sampling, suggested by the terminometrics methodology, could lead to collecting a biased sample and provide a distorted analysis. Traditional machine learning algorithms trained on these unbalanced samples would suffer the same bias, as less information would be available to classify the minority class. This type of algorithm would likely produce

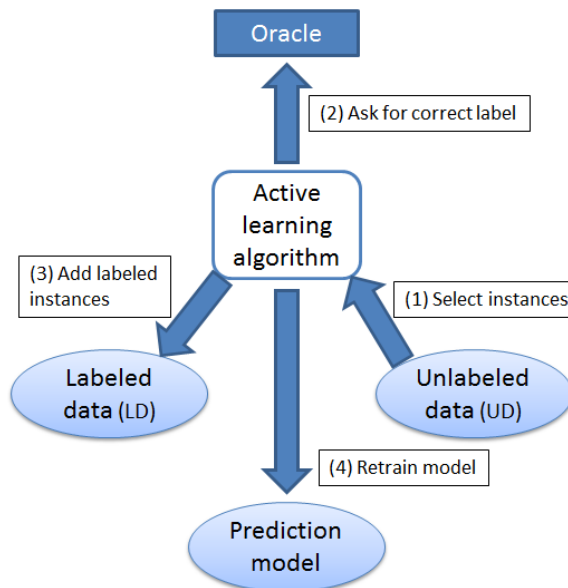


Figure 4: High level view of the active learning process.

a prediction model which would only target the majority class, overlooking instances potentially useful for terminometrics experts.

To sidestep this risk, we lean toward a learning approach called *active learning* which defines an iterative annotation process in order to reduce the risk of producing a biased prediction model. As shown in Figure 4, this four-step process implies the interaction with an oracle, typically a human annotator who needs to be familiar with the domain’s terminology and concepts being studied.

The active learning process starts with a set of unlabelled data (*UD*) containing, in the current context, individual occurrences of a term in a corpus, described by a group of features (e.g. a bag-of-words made of its co-occurring words in context). At this point, the labeled dataset (*LD*) is empty and there is no prediction model available. The active learning algorithm starts by selecting a group of instances, called the seed *S*, from *UD*. For each instance of *S*, the oracle is queried to specify a label, and the labeled example is then stored in *LD*. The oracle annotates the instance using one value of a predefined class label set, in this case {*yes*, *no*}, *yes* meaning the instance is used in the targeted sense, *no* if another other sense is used. When all instances in *S* are labeled, the active learning algorithm uses them to create a prediction model. It is important to note that there is no ideal size for the seed, but it should be suf-

ficient to enable the algorithm to train a relevant prediction model.

Once a prediction model is available, the process takes place in the same order, but with a variant. Instead of a seed, the algorithm superficially applies the prediction model to instances in *UD* (without labeling them or changing them to the labeled set) and pick an instance for which the model does not provide a sufficient level of confidence for its classification. It then submits this instance to the oracle who applies a label. Then, the newly labeled example is added to *LD*. The prediction model is then retrained and the process continues until the algorithm reaches an overall level of confidence for all instances in *UD*.

When this stopping criteria is reached, the active learning process is complete and the prediction model can be used to annotate the remaining instances in *UD*, if needed, or another similar dataset. Again, the level of confidence used as the stopping criteria must be empirically defined, as there is no ideal value. Of course, a higher confidence level might increase the annotation effort needed to produce the final prediction model, while a lower value might produce a less effective prediction model using fewer instances. Fine-tuning the confidence level helps to reduce the risk of training a biased prediction model on a predominant class in a dataset.

In our current implementation of active learning, we select a seed of 20 instances with random sampling which is then processed with RandomForest (Tin Kam Ho, 1995) as the prediction model. The oracle is then asked to annotate other blocks of 20 instances until the algorithm reaches its parametered confidence level. If this level is not reached after a total of 200 instances (including the seed), a final prediction model is trained and applied on *UD* in order to limit the effort to annotate each expression. The features for the classification process are extracted from the 90 characters window, which was judged as sufficient during the experiments (Section 4.1).

At this stage in our research, the current implementation provides a baseline on which we can later improve using different alternative models presented in the literature. Certainly, other research in word sense disambiguation has explored the empirical behaviour of active learning (e.g. (Chen et al., 2006)). Specific issues associ-

ated with active learning range from feature selection for particular disambiguation tasks (Palmer and Chen, 2005), model adaptation when changing domain between the training and application of the model (Chan and Ng, 2007), class imbalance problem (Zhu and Hovy, 2007) or deciding when the prediction algorithm stops asking for additional annotation (Zhu et al., 2008).

6 Terminometrics active-learning platform

We developed an annotation platform, shown in Figure 5, to facilitate terminometrics studies with an active learning component for term disambiguation. The platform implements the interactive active learning process described above to control and optimize the active learning between the prediction module and the human annotator. The platform will also enable future experiments within the field of terminometrics in which both the active learning algorithms and the human interaction can be further explored.

The user of this platform (typically the oracle in the active learning process) can create a corpus of documents, use this corpus to create an annotation project by defining a set of concepts, related terms and variations (plural, gender) and participate in the active learning process. At the end of the active learning process, the platform annotates the remaining instances in *UD* (see Figure 4) in order to estimate the distribution of occurrences of competing terms of a concept. This is used for the terminometrics analysis.

Aside from the *{yes, no}* classification, the interface offers two other choices; *undecided* and *reject*. The first choice allows the user to skip an instance and go to the next, while being able to later return to provide an answer. This could happen when the user wishes to see a larger context to perform the disambiguation. In fact, to help this process, the platform also provides an option to view an instance within its original document. The second choice, *reject*, removes the instance entirely from the unlabeled and labeled datasets. This is used typically when the user considers that the instance should not be used for the terminometrics final analysis.

In order to further reduce the annotation effort needed to perform a terminometrics study, other features, unrelated to active learning, were added

Prochaine annotation requise

Notion : nanotube de carbone **Terme :** nanotube de carbone **Langue :** FR **Position :** 1 / 4

Définition

nanotube de carbone

Contexte

Cependant, nous pouvons associer a` la mole`cule d'hydroge`ne un mo- | ment quadripolaire. | Le **nanotube de carbone** (Fig. 3) est une macromole`cule de carbone pouvant | 3 | Fig. 2 – Gauche : E`nergie potentielle

Décision : Oui Non Indécis Rejet

Texte complet

Précédent

Suivant

Liste des instances

Annotations du projet Nanotechnologie sur le corpus test

Mode d'annotation : En série Choisir un terme spécifique

#	Notion	Terme	Taille ▾	Nombre d'instances	Complet
189	laboratoire sur puce	puce microfluidique	19	1	Terminé
2	nanotube de carbone multifeuillet	nanotube multiparoi	19	3	Terminé
41	nanotube de carbone	nanotube de carbone	19	201	Non démarré
103	imagerie biologique	imagerie biologique	19	2	Non démarré
131	physique quantique	mécanique quantique	19	12	Non démarré
126	physique classique	physique classique	18	7	Non démarré

Détails

Exporter stats

Stats des notions

Gérer notions/termes

Figure 5: Annotation interface for terminometrics.

to the platform. The first is a language-based document filter which can be applied during the corpus creation to try to remove documents which are not suited for the targeted analysis. Each document is analysed with a language detection algorithm to extract a confidence level associated with its deduced language. It then enables the user to keep only the documents which are above a specific threshold and exclude the remaining from the corpus to be annotated. Of course, documents with no text, such as files containing only images, are also removed.

Another effort reduction feature is the duplicate context detection which takes place at the creation

of an annotation project. The source issue is that a sentence or a whole paragraph (or sometimes complete documents) can be found in several locations within a corpus created from web sites. While each occurrence of a term (or its variations) is stored and kept for an accurate assessment of its rate of occurrence in a corpus, only unique contexts (the term occurrence and a ± 90 characters window) are used for the active learning process. For example, if the first context of "substrat" shown in Table 2 was found with the same prior and post context in five documents in a corpus, the oracle would be asked at most once to annotate this instance (if it is selected for annotation by the

algorithm), but it would count as five occurrences in the terminometrics analysis.

The platform also facilitates the management of terminometrics studies by providing many features: an integrated storage and search capability on domain-specific corpora, a user interface specifically designed to facilitate annotation by providing in-context display of a term to validate, an access to a term list with the possibility for addition and removal of terms, and so on. This is an improvement over the traditional manual handling of documents and term lists, instances generation and annotation, traditionally done with folders and spreadsheets. While the upper limits of the platform have not been tested explicitly, the current experiment was done with a term list of 1,036 entries on a corpus of over 220,000 documents. As far as the sizes of the corpora and vocabulary are concerned, the platform is mainly limited by the speed and capacity of the computer that runs it.

7 Conclusion and future work

In this article, we introduced *term sense disambiguation*, a close cousin to word sense disambiguation, but much less studied within the NLP community. We showed how terms, especially single-word terms, are polysemous, both in term banks and in specialized corpus.

We presented the idea of using active learning within our terminometrics application, in which the in-domain sense bias is quite strong. So far, we have implemented a simple active learning algorithm, and will move toward more complex ones in the near future. The annotation platform, ready for experimentation, will allow terminologists to further complete, in less time, the annotation process of the nanotechnology domain and other domains. This will provide test data, on which we can measure the different gains in terms of time and accuracy of our current and future active learning approaches.

Furthermore, we plan to push further our exploration of term disambiguation. In fact, although lexicographic and terminological resources are organized differently, the distinction between terms and words is not always that "clear-cut". Many single-word terms exist also as common words. Some specialized terms also migrate from specific domains to the general language (Meyer, 2000) when a specialized domain becomes more part of

the day-to-day life of people (e.g. computer domain). We believe there is much room to further study term polysemy in term banks, in specialized corpus and also in more general corpus where both specialized and common senses might be present.

One of the envisioned experiments is to annotate semi-automatically a whole corpus to be able to compare the current approach to a supervised learning method. This will enable us to evaluate the contribution of active learning on the raw performance of disambiguation and time reduction of the annotation task. A new dataset related to a domain different than nanotechnology will also be defined for this experiment to avoid evaluating the approach on the dataset used for development.

Acknowledgments

We thank the annotators Julián Zapata and Barış Bilgen.

References

- Caroline Barrière. 2007. La désambiguïsation du sens en traitement automatique des langues (TAL): l'apport de ressources terminologiques et lexicographiques. In Marie-Claude L'Homme and Sylvie Vandaele, editors, *Lexicographie et Terminologie: compatibilité des modèles et méthodes*, pages 113–140. Presses de l'Université d'Ottawa.
- Caroline Barrière. 2010. Recherche contextuelle d'équivalents en banque de terminologie. In *Traitement Automatique des Langues Naturelles 2010*.
- Yee Sang Chan and Ht Ng. 2007. Domain adaptation with active learning for word sense disambiguation. *Acl*.
- Jinying Chen, Andrew Schein, Lyle Ungar, and Martha Palmer. 2006. An empirical study of the behavior of active learning for word sense disambiguation. *Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*, pages 120–127.
- Marta Chroma. 2011. Synonymy and Polysemy in Legal Terminology and Their Applications to Bilingual and Bijural Translation. *Research in Language*, 9:31–50.
- Ingrid Meyer. 2000. Computer Words in Our Everyday Lives : How are they interesting for terminography and lexicography ? In *Euralex'2000, International Congress on Lexicography*, pages 39–58, Stuttgart, Germany.
- George A. Miller. 1995. WordNet: a lexical database for English. *Communications of the ACM*, 38(11):39–41.

- M Palmer and Jy Chen. 2005. Towards robust high performance word sense disambiguation of English verbs using rich linguistic features. *Natural Language Processing - Ijcnlp 2005, Proceedings*, 3651:933–944.
- Patrick Pantel and Dekang Lin. 2002. Discovering word senses from text. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '02*, pages 613–619, New York, NY, USA. ACM.
- Jean Quirion. 2003. Methodology for the design of a standard research protocol for measuring terminology usage. *Terminology*, 9(c):29–49.
- Jean Quirion. 2006. Terminometrics - an Evaluation Tool of/for Term Standardization. In *TSTT'2006 - International Conference on Terminology, Standardization and Technology Transfer*, pages 19–24, Beijing, China.
- Egidio Terra and C.L.A. Clarke. 2003. Frequency Estimates for Statistical Word Similarity Measures. In *Proceedings of the NAACL 2003*, page 165.
- Tin Kam Ho. 1995. Random decision forests. *Proceedings of 3rd International Conference on Document Analysis and Recognition*, 1:278–282.
- Radek Vogel. 2007. Synonymy and polysemy in accounting terminology: fighting to avoid inaccuracy. In *Proceedings of the English for Specific Purposes Terminology and Translation Workshop, Košice 13-14 September 2007*. Univerzita P.J. Šafárika.
- Jingbo Zhu and EH Hovy. 2007. Active Learning for Word Sense Disambiguation with Methods for Addressing the Class Imbalance Problem. *EMNLP-CoNLL*.
- Jingbo Zhu, Huizhen Wang, and Eduard Hovy. 2008. Learning a Stopping Criterion for Active Learning for Word Sense Disambiguation and Text Classification. *International Joint Conference on Natural Language Processing*, pages 366–372.

