

Automatic Acquisition of Artifact Nouns in French

Xiaoqin Hu

Laboratory LDI
University of Paris 13
franhuxqin@hotmail.fr

Pierre-André Buvet

Laboratory LDI
University of Paris 13
pierreandre.buvet@gmail.com

Abstract

This article describes a method which allows acquiring artifact nouns in French automatically by extracting predicate-argument structures. Two strategies are presented: the supervised strategy and the semi-supervised strategy. In the supervised method, the semantic classes of artifact nouns are recognized by identifying the predicate-argument structures with the syntactic patterns of the given predicates. In the semi-supervised method, the extraction of predicate-argument structures is carried out from a semantic class of artifact nouns given in advance. The predicate candidates obtained from extracted predicate-argument structures are then intersected. Next, the syntactic patterns of predicates are automatically learned by probabilistic calculation. With the acquired predicates and the learned syntactic patterns, more artifact nouns are identified.

1 Introduction

The difficulties for automatic acquisition of terms might come from the linguistic techniques, the computational techniques or the limits of the natural language processing theory. Nowadays, many studies have been conducted for term extraction. This article presents a method for automatic acquisition of artifact nouns on the basis of syntactic-semantic analysis of predicates. Artifact nouns are the nouns of the artificial entities produced intentionally by human beings, with a view to a specific function. The automatic acquisition of artifact nouns is for completing the dictionary of semantic classes of laboratory LDI. There are two strategies for realizing this method: a supervised strategy in which the predicate-argument structures are ex-

tracted by the syntactic patterns of the given predicates and a semi-supervised strategy developed on the basis of the supervised strategy. The semi-supervised strategy consists of two steps. In the first step, it predicts which predicates are relevant by a probabilistic calculation. In the second step, it appeals to the supervised strategy. This article is organized as follows. Section 2 states the related work on term extraction in recent years. Section 3 presents the data model used in the proposed method. Section 4 explains in detail the proposed method including the semantic-syntactic analysis of appropriate predicates of artifact nouns. Section 5 presents the experiment results and the analysis of the results.

2 Related Work

In the as-built systems of term extraction, the part of linguistic model is often limited to morpho-syntactic descriptions and the part of statistical model, to a large extent, depends on the statistical knowledge. TERMINO (Lauriston, 1994) and LEXTER (Bourigault, 1996), two well-known semi-automatic systems of term extraction, are based on syntactic descriptions. The method of Hearst (1992) and the method of Snow et al. (2004) take advantage of morpho-syntactic patterns for automatically recognizing hyponyms and hypernyms. The statistical methods for term extraction can be based on Markov model (Jiang, 2012), co-occurrence, or vector support, etc. ANA (Enguehard, 1993) is a statistical method which is based on co-occurrence. Morlane-Hondère (2012) has presented a series of distributional methods realized with data mining techniques, such as mutual information, measures of association, log-likelihood or naive bays (Ibekwe-sanjuan, 2007). The method of Meilland and Bellot (2003) which extracts terms from annotated corpora, ACABIT (Daille, 1994) and the strategy of cooperation of many term extractors of Alecu et al. (2012) are

all the hybrid methods which combine linguistic model and statistical model. Furthermore, Seeker and Kuhn (2013) has proposed a method which identifies the morpho-syntactic patterns by statistical dependency Parsing, and Quiniou et al. (2012) has brought forward an approach aiming to identify the linguistic patterns via data mining techniques.

3 Data Model

3.1 Predicates, Arguments and Actualizers

The predicate is a linguistic unit defined as a language form of semantic relation between two entities. The entities linked by this relation are arguments. The actualizers are the linguistic elements which enable to register the predicates and arguments in grammatically correct statements. They can be grammatical units (such as prepositions, determiners...) or lexical units, such as modifying adjectives, adverbs, auxiliary verbs, support verbs, etc. The predicates semantically dominate the arguments.

3.2 Uses of Predicate

The predicates can be divided into verbal predicates, nominal predicates, adjective predicates, prepositional predicates and adverbial predicates in the conception of "uses of predicates" of Buvet (2009). The variations between the uses of predicate are morphosyntactic or interpretative. The interpretations of the uses result from a set of properties: type of state, type of action, processive aspect and stative aspect. A predicate can have one or more uses, for example, for the predicate *négozier* (*negotiate*), *négozier* (*negotiate*) is its verbal use, *négociation* (*negotiation*) is its nominal use and *négociable* (*negotiable*) is its adjective use.

3.3 Appropriate Predicates and Appropriate Relation

The appropriate predicates have a number of relatively limited semantic classes of arguments. This character of appropriate predicates allows predicting the semantic class to which their arguments belong. An appropriate predicate in a specific sense can define a semantic class of arguments. Nevertheless, the polysemy of most of the appropriate predicates necessitates delimiting the semantic class of arguments by gathering many appropriate predicates of one semantic class. For example, for the predicate *conduire* (*drive/take/lead*), it

can be used in the following senses: *conduire mon enfant à l'école* (*drive my child to school*), *conduire une voiture* (*drive a car*) or *conduire une entreprise* (*lead a company*). The polysemy of *conduire* (*drive/take/lead*) prevents from isolating the semantic class of transport. However, with another appropriate predicate *réparer* (*repair*), we can predict that the arguments which can appear after both *conduire* (*drive/lead*) and *réparer* (*repair*) belong to the semantic class of transport. A set of appropriate predicates that allows delimiting a semantic class of arguments is defined as the definitional appropriate predicates of this semantic class of arguments (Buvet, 2009; Mejri, 2009). The definitional appropriate predicates characterize the semantics of their class of arguments.

4 Presentation of Method

4.1 Corpora

The corpora used for the method are composed of texts coming from about ten French websites (e.g., <http://www.forum-auto.com/marques/index.htm>, <http://geekandfood.fr/blog/>, etc.). The websites are selected around various themes: automobile, household appliances and decoration, cooking, beauty, fashion, health, etc. The chosen texts include the comments, the discussions on forum and the articles on the blog. The volume of the corpora reaches 22,858 Ko. They comprise 3,754,334 words. The texts of different themes occupy about the same proportion in the corpora. The texts of the different genres (the comments, the discussions on the forum and the articles on the blog) also occupy about the same proportion respectively.

4.2 Work Tool: Unitex

In the proposed method, both the preprocessing and the extraction of predicate-argument structures are carried out with local grammars through Unitex. With the integrated linguistic resources (such as Dela, Delac, etc.), Unitex makes it possible to represent a local grammar in the form of finite state automaton.

4.3 Supervised Method

In the supervised method, the predicate-argument structures are recognized automatically from a set of predicates given in advance. A series of syntactic patterns are established on the basis of the syntactic-semantic analysis of the appropriate

predicates. The obtained arguments are then intersected for choosing the appropriate arguments of the given predicates.

4.3.1 Preprocessing

The constituents of multi-word expressions are often misrecognized by computer as the constituents of other syntactic structures, for example, in the sentence *Une fois acheté mon nouveau manteau, je suis rentré à la maison* (once my new coat bought, I returned to home), *fois acheté* (time bought) is often misrecognized as a noun phrase by computer with the syntactic pattern N+Adj. Nevertheless, *fois* (time) is the constituent of the multi-word expression *une fois* (once). To solve this problem, the following strategy is adopted: we appeal to the dictionary Delac in Unitex for labelling adjective multiword expressions, adverbial multi-word expressions, verbal multi-word expression and prepositional multi-word expressions ; these expressions are then replaced by the corresponding morphosyntactic label (like <ADV>, <ADJ>, <V> and <PREP>). Thus, the multi-word expression *une fois acheté* becomes <ADV> *acheté* after the preprocessing.

The given predicates are labelled by the tool Unitex considering the different uses of each predicate. The morphosyntactic disambiguation of predicates depending on context is conducted at the same time. Thus, the given predicates are labeled by identifying the corresponding verbal phrases with the syntactic patterns such as, avoir+été+ADV+Vpp, se+faire+V, aller+être+ADV+Vpp, se+être+ADV+Vpp, avoir+Vpp, Vpp+Det+N, Vpr+DET+N,. For some lexical units of which the parts of speech are often used as reference for the morphosyntactic disambiguation of other lexical units, if they have multiple parts of speech, their entries of the lesser-used parts of speech in Dela are eliminated. For example, to decide a lexical unit is a noun or not depends on whether the lexical unit follows an article or not while some articles in French have more than one morphosyntactic interpretations (e.g., *un* (*a*) can be an article or a noun). Thus, the entry of *un* (*a*) as noun considered less used is eliminated in the preprocessing.

4.3.2 Automatic Extraction of Predicate-argument Structures

In French, the nominal distribution of an appropriate predicate can be situated in the position

of subject, object complement (direct or indirect: the indirect object is introduced by the preposition in French), circumstantial complement of location or circumstantial complement of means. The syntactic position of nominal distribution often changes with the structural transformation of sentences (e.g., from an active sentence to a passive sentence). The analysis of syntactic-semantic distribution of appropriate predicates of artifact nouns is based on the elementary sentences of active form. The elementary sentences are the sentences containing only one conjugated verb. The complex sentences containing more than one conjugated verb can be obtained from a set of elementary sentences by the linguistic technique, i.e. transformation (Harris, 1976; Gross, 1986).

According to the syntactic position of nominal distribution of appropriate predicates, the appropriate predicates can be divided into four classes: the first class contains the appropriate predicates whose object complements (the object complement corresponds to the verb complement in English) are always artifact nouns; the second class contains the appropriate predicates whose object complements can be artifact nouns but whose circumstantial complements of means (it corresponds to the prepositional complement in English) are always artifact nouns; the third class includes the appropriate predicates whose object complements have less possibilities to be artifact nouns but whose circumstantial complements of location are always artifact nouns; the last class includes the appropriate predicates whose object complements are never artifact nouns but whose circumstantial complements (means or location) are always artifact nouns. Each class can be subdivided according to the syntactic features of the appropriate predicates. Table 1 lists all the classes that we made according to the syntactic-semantic distribution of appropriate predicates. For each class, some examples of predicates and corresponding syntactic patterns are given. In the formula expressions of syntactic-semantic distribution, V means verb, NAF indicates the artifact nouns and Nc refers to the nouns of other semantic classes.

Many other syntactic patterns are constructed considering language transformation from the basic syntactic patterns presented above. A series of graphs is established on the basis of the established syntactic patterns, and the predicate-argument structures are extracted through the tool

Classes	Appropriate Predicates	Syntactic-semantic Distribution
Class1		
Class_1	éteindre (turn off), inventer (invent), etc.	V+NAF
Class_1a	tirer (pull), retirer (remove), appuyer (support), etc.	V+dessus/dessous/sur...+NAF
Class_1b	jouer (play)	V+à/de+NAF
Class2		
Class_2	récurer (scrub), réparer (repare), tracter (tow), etc.	V+de/avec/par+NAF
Class_2a	découper (cut out), fouiller (dig), décasser (clean up), etc.	V+NAF/Nc+de/avec/par+NAF
Class_2b	équiper (equip), orner (decorate), etc.	V+NAF/Nc+de+NAF
Class3		
Class_3	ranger (arrange), installer (install), contenir (contain), etc.	V+NAF/Nc+sous/devant/sur/derrière...+NAF
Class_3a	transformer (transform)	V+NAF/Nc+en+NAF
Class_3b	connecter (connect)	V+NAF/Nc+à+NAF
Class4		
Class_4	verser(pour), enregistrer (record), etc.	V+Nc+dans+NAF
Class_4a	peigner (comb), maquiller(make up), farder (disguise), etc	V+Nc+avec/par/de+NAF
Class_4b	nourrir (nourish), alimenter (feed)	V+Nc+à+NAF
Class_4c	afficher (put up), placarder (placard), etc.	V+Nc+sur+NAF

Table 1: Analysis of syntactic-semantic distribution of appropriate predicates

Unitex. However, as the modifiers of a nominal phrase can be added without limits (especially when the modifiers are relative clauses), it is difficult to describe all types of constructions of sentence by the local grammar. In addition, an apposition often has a flexible position in one sentence. It can almost be inserted next to any noun phrase of a sentence. In the proposed method, the nominal phrases of more than five grams and the appositions are not taken into account.

4.3.3 Intersecting the Arguments

To intersect the arguments of different predicates is for finding the common arguments of the semantic class of predicates given in advance. As a semantic class of arguments is defined by a set of definitional appropriate predicates, the more an argument is shared by the given predicates, the more probably this argument belongs to the semantic class of the given predicates. The process of intersecting the arguments is shown in Figure 1. $Pred_i$ ($i=1, 2, 3$) refers to a predicate and Arg_j ($j=1, 2, 3$) means an argument. The grey parts are the intersection of arguments. In fact, in our method, not only the common arguments of the given predicates are selected, but also the arguments shared by most of the given predicates. The number of different predicates that co-occur with an argument is noted as the intersecting frequency of this argument. For example, in Figure 1, the

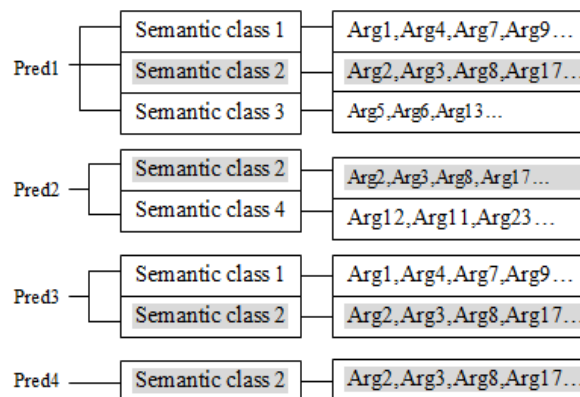


Figure 1: Intersecting the predicates

intersecting frequency of Arg2, Arg3, Arg8, and Arg17 is 4 because they are shared by all the four predicates and the intersecting frequency of Ar1, Arg4, Arg7, and Arg9 is 2 since they are shared by two predicates (Pred1 and Pred3).

4.4 Semi-supervised Method

In the semi-supervised method, the predicate-argument structures are identified with a semantic class of artifact nouns by local grammars. Then, all the predicates in the structures are extracted. Next, the predicates are intersected for determining which predicates belong to the semantic class of the given artifact nouns. The syntactic patterns associated with each predicate are also extracted. A probabilistic calculation is conducted in order

to choose the most appropriate syntactic pattern for each predicate. With the selected predicates and their syntactic patterns, the nominal distribution of appropriate predicates can be located and the artifact nouns of the semantic class are finally acquired after intersecting the artifact nouns. With the obtained artifact nouns, the processes can be iterated for getting more artifact nouns.

4.4.1 Extraction of Predicate-argument Structures from a Semantic Class of Arguments

In the semi-supervised method, the predicate-argument relations are extracted from a set of arguments. However, with a set of appropriate predicates given in advance, the syntactic-semantic distribution can be predicted, but from the arguments, it is not certain to predict which syntactic-semantic relation is associated with the given argument. Thus, the following solution is adopted: all the possible syntactic relations between the artifact nouns and their appropriate predicates are firstly predicated; a probabilistic calculation is then carried out to the predicted syntactic relations in order to choose the appropriate syntactic pattern for each predicate. If the semantic distribution of artifact nouns can be situated in the position of noun complement without preposition introducing it (which concerns the direct complement of objet) and in the position of noun complement introduced by preposition (which concerns the indirect complement of objet, the circumstantial complement of location and the circumstantial complement of means), there should be three necessary constituents for forming a syntactic pattern allowing predicting a predicate-argument relation: verb, noun complement without preposition and noun complement introduced by preposition. With these three constituents, four combinations for forming the desired syntactic patterns can be obtained: V+NAF, V+NAF+prep+NAF, V+Nc+prep+NAF and V+prep+NAF (V means verb, prep refers to preposition and NAF indicates the artifact nouns). Other syntactic patterns (such as V+ADV+NAF, NAF+be+V+prep+NAF, NAF+V+prep+NAF+NAF or V+ADV+ADV+prep+NAF+NAF) are derived from these four basic syntactic patterns through the transformation of natural language.

With the established syntactic patterns, a series of graphs is constructed and the predicate-

argument structures are labelled. In addition, the predicates, the arguments and the syntactic patterns associated with each predicate-argument structures are also labelled and extracted for the following processing.

4.4.2 Calculation of Syntactic Patterns

All the predicate-argument structures recognized by predicting the possible syntactic relations don't represent the real syntactic relation between a certain predicate and its arguments. For example, in *éteindre la lampe de poche* (turn off the flashlight), *éteindre* (turn off) can be identified by the syntactic pattern V+NAF+prep+NAF or V+NAF with the given artifact noun *lampe* (lamp) or *poche* (pocket); however, V+NAF+prep+NAF does not represent the syntactic-semantic distribution of the predicate *éteindre* (turn off). V+NAF+prep+NAF is misrecognized as the syntactic pattern of the predicate *éteindre* (turn off) because of the preposition *de* (of) which is a constituent of the compound noun *lampe de poche* (flashlight) rather than a preposition introducing a circumstantial complement. Thus, a probabilistic calculation of syntactic patterns is necessary for choosing the appropriate syntactic pattern for each predicate.

The syntactic pattern by which a predicate-argument structure is identified is recorded in the labels like $s=vactif_gnaf$, $s=vactif_gn_de_gnaf$,..., etc. The code *vactif* (*vpassif*) indicates the active (passive) form of the verb. The code *gn* means nominal phrase, and *gnaf* refers to a nominal phrase of artifact noun. The probability of having a direct object complement, $P(cod)$, is calculated by the formula:

$$P(cod) = \frac{c(gnaf) + c(gn)}{c(s)} \quad (1)$$

$c(gnaf)$ implies the frequency of occurrence of the syntactic patterns containing *gnaf* in the position of direct object complement. For example, $s=vactif_gnaf$, $s=gnaf_va$ and $gnaf_vpassif$ are all the syntactic patterns including *gnaf* in the position of direct object complement. $c(gn)$ indicates the frequency of occurrence of the syntactic patterns containing *gn* in the position of direct object complement. $c(s)$ indicates the frequency of occurrence of all the syntactic patterns associated with a predicate. The probability of having a direct object complement which is always artifact

noun, $P(codnaf)$, is calculated according to the formula:

$$P(codnaf) = \frac{c(gnaf)}{c(s)} \quad (2)$$

and the probability of having an object complement introduced by a preposition, $P(codi)$, is calculated as follows:

$$P(codi) = \frac{c(prepare)}{c(s)} \quad (3)$$

$c(prepare)$ refers to the frequency of occurrence of the syntactic patterns containing a preposition. For each predicate, if its $P(codnaf)$ is greater than $P(cod) - P(codnaf)$, the direct object complement of this predicate is considered to be always artifact nouns; if $P(cod)$ equals to zero, this predicate is not considered to have the direct object complement; if $P(prepare)$ is greater than 0.12, this predicate is considered as a predicate having an object complement introduced by preposition which is always artifact nouns. The threshold for $P(prepare)$ is decided after several tests and it allows obtaining a more accurate syntactic information for each predicate. According to these probabilities about the syntactic positions, the most appropriate syntactic pattern is chosen for each predicate from the four basic syntactic pattern candidates. Finally, the extracted predicates are classified into four groups according to their syntactic-semantic patterns: the group of V+NAF, the group of V+NAF+prep+NAF, the group of V+Nc+prep+NAF, and the group of V+prep+NAF.

4.4.3 Intersecting the Predicates

The aim of intersecting the predicates is to find out common predicates of the given artifact nouns. The more a predicate is shared by the given arguments, the more probably it belongs to the semantic class of the given arguments. For a predicate, the number of different artifact nouns which co-occur with this predicate is noted as the intersecting frequency of this predicate. The threshold for intersecting the predicates is set at 2 after several tests. This threshold allows giving a better result.

4.4.4 Elimination of Basic Predicates

In the result obtained after intersecting, many basic predicates occupy the top place of the list. The basic predicates have a large semantic spectrum. They are not appropriate predicates of arti-

fact nouns, but their nominal distribution cover the semantic class of artifact nouns. For the appropriate predicates which belong to the semantic class of given arguments, their frequencies of occurrence in the extracted predicate-argument structures (FC) and their frequencies of occurrence in the total corpus (FT) are more or less similar. On the contrary, for the basic predicates, there is a great disparity between their frequencies of occurrence in the extracted predicate-argument structures and their frequencies of occurrence in the total corpus, since the basic predicates have a larger and more general semantic spectrum. On the basis of this occurrence disparity, some of the basic predicates can be eliminated. The occurrence disparity (*Ecart*) is calculated as follows:

$$Ecart = \frac{FT - FC}{FT} \quad (4)$$

After several tests, we decided the threshold as 0.978 which gives a better result. If the *Ecart* surpasses the threshold, the corresponding predicate is considered as basic predicate.

4.4.5 Application of Supervised Method

With the filtered appropriate predicates and the learned syntactic-semantic patterns, a script is developed to automatically write the graphs for identifying the predicate-argument structures and labelling the arguments. Likewise, all the predicate-argument structures are extracted. The acquired arguments are then intersected. In this way, more artifact nouns are acquired from a small set of artifact nouns given in advance. The processes can be iterated for obtaining more artifact nouns.

5 Experiment and Evaluation

For the supervised method experiment, about one hundred appropriate predicates of artifact nouns are chosen, and a series of syntactic patterns are established on the basis of the syntactic-semantic distribution of the appropriate predicates. The semi-supervised method is tested with three semantic classes of arguments: container, cooker and road transport. For each semantic class, a list of arguments, including about twenty artifact nouns, is manually established. The evaluation is carried out by appealing to a dictionary of artifact nouns (including 13,400 entries) developed in the laboratory. The manual annotation is added because the dictionary is not complete. Firstly, the

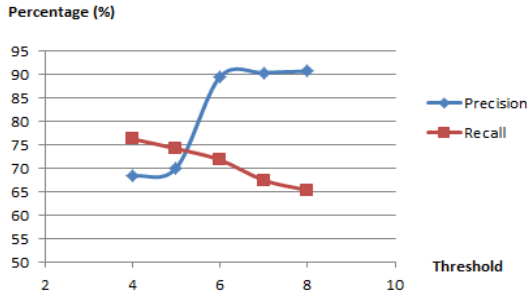


Figure 2: Experiment of threshold

artifact nouns in the corpus are labeled by the dictionary and the manual annotation. The result is considered as standard. Then, our method is applied for labelling the artifact nouns and another result is obtained. The result of our method is compared with the standard in order to calculate the precision, the recall and the F-measure.

For the supervised method, the threshold for intersecting the arguments is respectively set at 4, 5, 6, 7 and 8. Then, the precision, the recall and the F-measure are respectively calculated. The Evaluation results obtained with different thresholds are shown in Table 2. Figure 2 shows the comparison of the different evaluation results (F-measures) obtained with different thresholds. It is seen that the highest F-measure can be obtained when the threshold equals to 6.

Threshold	Precision	Recall	F-measure
4	68.31%	76.20%	72.03%
5	70.08%	74.16%	72.06%
6	89.40%	71.78%	79.63%
7	90.27%	67.45%	77.21%
8	90.59%	65.33%	75.91%

Table 2: Evaluation of supervised method

For the semi-supervised method, the experiment is firstly carried out with the artifact nouns of semantic classes "container". The processes of the semi-supervised method are iterated five times. The results obtained after each iteration are respectively evaluated. The threshold for intersecting the arguments is firstly set at 3. The result obtained by the semi-supervised method includes the grain terms. Table 3 shows the evaluation results obtained with different number of iterations, and Figure 3 shows the comparison of the evaluation results. It is found that the result obtained after three iterations has the highest F-measure. After four iterations, the precision falls down rapidly

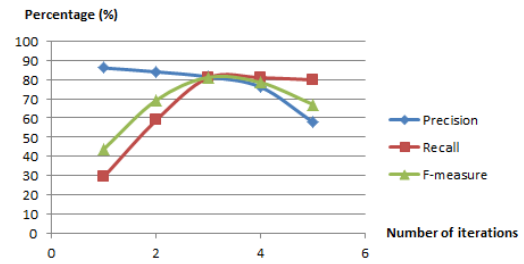


Figure 3: Experiment of iteration with semantic class "container"

and the recall reaches a relatively stable value. The noise is brought by the nouns of other semantic classes obtained in each iteration. Then, the threshold is set at 2, 4, 5 and 6 respectively. The same experiment presented above is repeated for each threshold. Figure 4 shows a comparison of the highest F-measures that can be obtained with different thresholds. For the other two semantic classes, the same experiment and evaluation are conducted. Finally, we choose 3, 2 and 3 as the threshold for intersecting the arguments of the semantic class "container", "cooker" and "road transport" respectively and select 3, 3 and 4 as the number of iterations for the semantic class "container", "cooker" and "road transport" respectively. Table 4 shows the evaluation results of each semantic class with the defined threshold and number of iterations. The different quantity of appropriate predicates of different semantic class in the corpus makes the performance of our method different.

Number of iterations	Precision	Recall	F-measure
1	86.12%	29.41%	43.85%
2	84.07%	58.82%	69.21%
3	81.34%	81.02%	81.20%
4	76.10%	81.02%	78.48%
5	57.79%	79.87%	67.06%

Table 3: Evaluation of iteration with semantic class "container"

Semantic classes	Precision	Recall	F-measure
Road transport	62.46%	58.53%	60.43%
Cooker	70.14%	76.87%	73.35%
Container	81.34%	81.02%	81.20%

Table 4: Evaluation of semi-supervised method

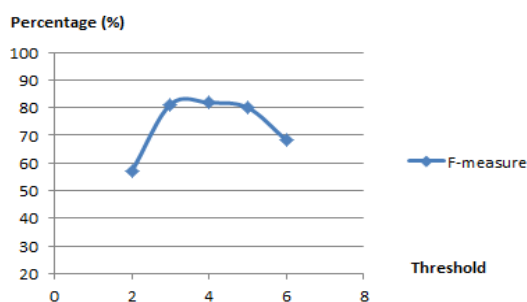


Figure 4: Experiment of threshold

6 Conclusion

The method in this article is based on the analysis of syntactic-semantic distribution of appropriate predicates of artifact nouns. The advantage of this method is that it allows locating not only the position of an artifact noun in each sentence but also the position of a nominal distribution which is composed of a semantic class of artifact nouns. A class of definitional appropriate predicate characterizes a semantic class of arguments and makes it possible to consider the polysemy. In addition, the identification of the nominal distributions of appropriate predicates also permits the identification of neologisms, misspelled artifact nouns or abbreviations. Although the performance of the proposed method is dependent on the accuracy and the completeness of the established local grammars, it allows obtaining lexicon resources with a relatively high precision and the obtained lexicon resources of semantic class can make a contribution to dialogue systems, natural language generation or other natural language processing applications.

References

- B. P. Alecu, Izabella Thomas, and Julie Renahy. 2012. La “multi-extraction” comme stratégie d’acquisition optimisée de ressources terminologiques et non terminologiques. In *Actes de la 19e conférence sur le Traitement Automatique des Langues Naturelles*, pages 511–518. Grenoble.
- D. Bourigault. 1996. Lexter: a natural language tool for terminology extraction. In *Proceedings of the 7th EURALEX International Congress*, pages 771–779. Göteborg.
- P.-A. Buvet. 2009. Des mots aux emplois : la représentation lexicographique des prédicats. *Le Français Moderne*, 77(1):83–96.
- B. Daille. 1994. Study and implementation of combined techniques for automatic extraction of terminology. In *The Balancing Act: Combining Symbolic and Statistical Approaches to Language, Proceedings of the Workshop of the 32nd Annual Meeting of the ACL*, pages 29–36. Las Cruces, New Mexico, USA.
- Chantal Enguehard. 1993. Acquisition de terminologie à partir de gros corpus. In *Actes Informatique & Langue Naturelle*, pages 373–384. Nantes.
- M. Gross. 1986. *Grammaire transformationnelle du français : Syntaxe du verbe ; Syntaxe du nom*. Cantilène.
- Z. S. Harris. 1976. *Notes du cours de syntaxe*. Le seuil.
- M. A. Hearst. 1992. Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the 14th conference on Computational linguistics*, volume 2, pages 539–545. Stroudsburg, PA, USA.
- Fidelia Ibekwe-sanjuan. 2007. *Fouille de textes*. Lavoisier.
- Jing Jiang. 2012. Information extraction from texte. In Charu C. Aggarwal and ChengXiang Zhai, editors, *Mining Text Data*, pages 18–22. Springer. US.
- A. Lauriston. 1994. Automatic recognition of complex terms: problems and the termino solution. *Terminology*, 1(1):147–170.
- Jean-Claude Meilland and Patrice Bellot. 2003. Extraction automatique de terminologie à partir de libellés textuels courts. In Geoffrey Williams, editor, *Linguistique de corpus*. Presses Universitaires de Rennes. France.
- S. Mejri. 2009. Le mot, problématique théorique. *Le Français Moderne*, 77(1):68–82.
- François Morlane-Hondère. 2012. *Une approche linguistique de l’évaluation des ressources extraites par analyse distributionnelle automatique*. Ph.D. thesis, Université Toulouse le Mirail. France.
- Solen Quiniou, Peggy Cellier, Thierry Charnois, and Dominique Legallois. 2012. What about sequential data mining techniques to identify linguistic patterns for stylistics? In *Computational Linguistics and Intelligent Text Processing*, pages 166–177. Heidelberg.
- Wolfgang Seeker and Jonas Kuhn. 2013. Morphological and syntactic case in statistical dependency parsing. *Computational Linguistics*, 39(1):23–55.
- Rion Snow, Daniel Jurafsky, and Y. Ng Andrew. 2004. Learning syntactic patterns for automatic hypernym discovery. In *Advances in Neural Information Processing Systems*, pages 1297–1305. British Columbia.