Decision trees as explicit domain term definitions.

Roberto Basili, Maria Teresa Pazienza, Fabio Massimo Zanzotto
University of Rome Tor Vergata,
Department of Computer Science, Systems and Production,
00133 Roma (Italy),
{basili,pazienza,zanzotto}@info.uniroma2.it

Abstract

Terminology Acquisition (TA) methods are viable solutions for the *knowledge bottleneck* problem that confines knowledge-intensive information access systems (such as Information Extraction systems) to restricted application scenarios. TA can be seen as a way to inspect large text collections for extracting concise domain knowledge. In this paper we argue that major insights over the notion of term can be obtained by investigating a more domain-based term definition. We propose a decision tree learning approach as an interesting model of the human TA activity. An incremental model is proposed to study the evolution of the term definition during the TA process over a particular implicit domain model. The experimental apparatus is based on robust text processing tools that support a large scale investigation. The good results suggest that the proposed automatic TA model can support the development of conceptual domain dictionaries as required by knowledge-based information systems.

1 Introduction

Terminology Acquisition (TA) methods are a viable solution for the *knowledge bottleneck* that confines knowledge-intensive information access systems (such as Information Extraction systems) to restricted application scenarios. TA is the study of methods to extract concise domain knowledge representation (i.e. terminological dictionaries or terminology knowledge bases, TKBs) by inspecting large text collections. These corpora embody domain knowledge in the most natural and effective ways. The major limitation for any TA process is the difficulty in capturing, in computational terms, the complex notion of the underlying cornerstone, i.e. the *term*.

Most automatic TA methods start from the definition of *what a term is* and use it against a domain corpus (Jacquemin, 1997). This latter represents source information for any decision about lexical items (i.e. legal terms of the domain) that do (or do not) meet the given definition. In this sense, the corpus expresses, implicitly, all the information needed for semantic characterization of the underlying domain: it is thus an *implicit domain model* (IDM)

In automatic TA, there is a general consensus in assuming a term as *a surface representation of a key domain concept* (Jacquemin, 1997). Since this definition is open to different "operational" interpretations, it has led to the design of different corpus-driven TA architectures. An *"operational" model* is obtained by specifying the *prototypes of admissible surface forms* and a notion of *relevance* of a candidate form able to capture the importance of the underlying concept for the target domain. The prototypes for the surface forms are usually specified via NP grammars in agreement with valid natural language interpretations. Generally the morpho-syntactic level is used where term prototypes may be specified for instance as `Adj Noun` or `Noun Noun` constraints able to select respectively surface forms as *joint venture* or *information access*. The notion of *relevance* for the domain relies generally on probabilistic properties. In (Daille, 1994), the simple frequency $f(s)$ of surface forms in the corpus is suggested as the most effective measure. Frequency f seems to reproduce the terminologist judgement better than other more complex statistical measures. However, as admittedly mentioned in (Daille, 1994), frequency alone is still far from being a perfect "*termhood*" function.

In this paper we propose to consider further information embedded in the underlying *im-*

plicit domain model (IDM). When terminological dictionaries are manually built, terminologists start from a general notion of term and apply it to the specific domain. As long as they look at the target collections their intuitive perception of the underlying domain improves. In fact, they tune their starting hypothesis along with their exposition to texts. In this process, the IDM usually consists of a domain collection together with an explicit pre-existing domain terminology, T_0. Two kinds of information, often neglected by other computational approaches, are here available: (1) *usage* of already accepted terms (terms in T_0) are embodied by the corpus and (2) *negative evidences*, derived through negative decisions, i.e. rejections. Frequent occurrences, but non-terminological, expressions increase the terminologists' perception of *what a non-term is*.

Typical uses of accepted (and refused) candidates refine incrementally an inner definition of terms. This, in a computational perspective, should be expressed via an *intentional term definition*. This is the purpose of the method described in this paper. Several observable properties can be derived from the collections (i.e. in the contexts of terms and non-terms). A predictive (intensional) model, able to correctly separate terms from non-terms, should be developed on the most relevant (i.e. distinctive) of such properties. In the following, two text fragments appear:

Example 1 .
a) The **vorticity equation** *governs the evolution of vorticity in a geophysical fluid. This is an* **equation** *used in large-scale geophysical fluid dynamics.*
b) The **generalized airfoil equation** *governs the pressure across an airfoil oscillating in a wind tunnel.*

Both expressions *vorticity equation* and *generalized airfoil equation* are here terminological with respect to a scientific domain. The syntagmatic structure of the sentences is similar. The expressions are both `subjects` of the verb *govern* and this is often true of technical definitions for physical laws. Such grammatical facts may be usefully adopted as selective criteria as they establish a domain specific notion of similarity. These decision rules should be embodied into the domain-specific *intensional term definition* (*itd*) that we aim to capture.

We then argue that major advances in terminology acquisition can be obtained by adopting the *intensional term definition*s as a concise operational notion. For this reason we settled a learning model within a cycle of TA acquisition. The resulting learning model is assumed to derive an *itd* as a decision tree representing the terminologist activity in an explicit and hierarchical way. The induction can be incrementally applied to the TA cycle and the psychological plausibility (as an heurism) of the resulting model can be studied.

In Section 2, the *itd* learning model is defined. The related feature space, introduced in Sec. 3, is based on the implicit domain model (i.e. the corpus plus a seeding terminological dictionary). It supports the application of machine learning algorithms such as (Quinlan, 1993). The natural language processing tools, responsible for mapping the textual material into the feature representations (Basili et al., 2000), are then described in Section 3.2. The results are analysed in Section 4. First, a discussion of the induced models is presented (Section 4.1). Then, performance in the TA task is measured over benchmarking data (Section 4.2).

2 Decision Tree Learning of *itd*s

The decision tree formalism is an interesting way for representing the heurisms used by the terminologists in assessing *"termhood"* of the incoming candidates as it represents the decision rules in a hierarchical fashion. As any categorisation method, a decision tree is a function that, given an object represented by a set of properties (i.e. attribute-value pairs), outputs a category chosen from a predetermined set. This latter is the classification decision over the input object. If Ω is the space where properties are represented and Θ the set of the target decisions, the decision tree DT is then a function:

$$DT : \Omega \rightarrow \Theta \qquad (1)$$

In this formalism, the decision strategy is represented by a tree where each internal node corresponds to a test on a given property, i.e. the test on the value of a given attribute. The categorisation is achieved when a leaf node is reached,

i.e. all the tests in the path are passed.

Given its nature, a decision tree imposes a hierarchy on the attributes. In fact, the more discriminating is an attribute with respect to the target competing concepts (decisions), the higher it should be modelled in the hierarchy since the decision can be taken more straightforwardly. Therefore, the inspection of an already built decision tree provides insights on which feature has been considered more important in the description of the target concepts. Applied to the problem of term definition, the decision tree should represent the internal hierarchy of choices that terminologists perform when observing the properties of a given term candidate. The classification decision they have to take is whether or not the candidate is a term, i.e. whether or not it is an instance of the *concept of term*.

Since in our model we assume that terminologists use as a source of discriminating hints the term contextual information, in a decision tree this information should be described. A sample decision tree based on such a kind of contextual information is depicted in Fig. 2. Here

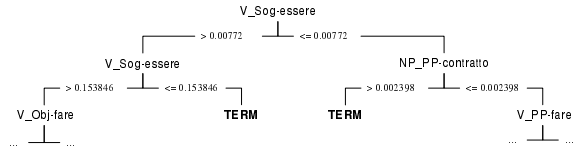

Figure 2: A sample decision tree.

four properties are considered. The property of the candidate of being: (1) `subject` of the verb *to be* (`V_Subj-essere`); (2) `object` of the verb *to make* (`V_Obj-fare`); (3) `prepositional modifier` of the verb *to make* (`V_PP-fare`); and (4) `prepositional modifier` of the noun *contract* (`NP_PP-contratto`). What is stated in the tree is that if the analysed candidate is "enough" correlated with the verb *to be* in a subject relation (i.e. the "correlation score" is between 0.00772 and 0.153846) it can be reasonable considered a term, otherwise the correlation with other features has to be evaluated. The noticeable information in the tree is that, in this particular term definition, the contextual relation with the verb *to be* has been considered as the more discriminating hint. In order to be a useful decision maker, the tree should represent the important properties of the notion of the term as well as the notion of non-term in the given environment, i.e. in the particular *implicit domain model*.

Standard and effective tools for the induction of decision trees are available (Quinlan, 1993). In particular, this latter method is able to infer regularities over feature space with continuous-valued attributes. This is necessary in the model we propose since we want to study the regular correlations of terms and non-terms with the other words in the domain contexts. It is worth noticing that the applicability of the tree learning method is possible due to the inclusion of the non-term concept in the model of TA postulated in this work.

The model of the overall process includes the following steps: (a) **Generation** of a *global feature vector* for knowledge item (i.e. a term or a non-term); (b) **Induction** of the target intensional definition as a *decision tree* that divides incoming candidates into terms and non-terms.

To better understand the terminologists' behaviour, the above process can be also modelled as an incremental approach. Newly accepted (or rejected) candidates allow a dynamic revision of the corresponding decision tree structure: a new learning process can be activated over the newly assessed instances.

3 Making use of *Implicit Domain Model*s in TA

The induction of concise domain-oriented term definition needs a suitable representation of the observations. This representation should be derivable from the implicit domain model. A suitable observation model should include all those selective properties characterizing the notion of term and non-term.

The aim here is to understand if and how regularities in the behaviour of terms in the corpus are used by terminologists as selective features for the final decision. Syntax will be used (in line with other works like (Grefenstette, 1993) or (Basili et al., 2001)) as linguistic level able to characterize the similarity among contexts.

In the next Sections the formal definitions of the feature vectors representing positive and negative instances are presented.

3.1 Sampling the *Implicit Domain Model*

When collecting evidences of a given term t across a domain corpus we need to determine whether or not different contexts are indicators of its syntactical behaviour. A first possibility is to collect only contexts where a valid surface form for t appears. However, in many cases terms are referred in an elliptic fashion. In the example 1.a), the second occurrence of the word *equation* is an elliptic occurrence of *vorticity equation*. As a consequence the context *This is an equation used in large-scale geophysical fluid dynamics.* describes the contextual behaviour of the *vorticity equation* term as well. Many simple terms (i.e. one-word terms) are elliptic references to complex terms (i.e. multiword terms). Generally, the term grammatical *head* (e.g. *equation* in *vorticity equation*) is used in elliptic references.

The syntactic, hereafter exogenous, behaviour of a term is driven by its semantics. The head $h(t)$ of a term t is usually its semantic carrier. This assumption is widely used in other term structuring approaches (cf. (Morin, 1999)). $h(t)$ is thus a good canonical candidate of t. Its occurrences in the corpus are representative of direct or elliptic occurrences of t. This is a computationally attractive approximation for estimating frequency. Moreover, as terms are expected to have unique interpretations in a coherent domain, terms t and t' such that $h(t) = h(t')$ will be considered equivalent with respect to their exogenous information. Accordingly, terms *vorticity equation* and *generalized airfoil equation* are equivalent with respect to the head *equation*.

The contribution of all contexts where a given head $h(t)$ appears forms an equivalence class, $C(t)$, in the corpus. A single (collective) representation, $v(t)$, for t can be thus derived from all $c \in C(t)$. This seemingly applies to "non-terms". In the next section, the definition for vectors $v(t)$, i.e. feature vectors populating the sample space, is given.

3.2 Syntactic feature spaces

The induction of a model for terms (or non terms) requires a suitable knowledge representation formalism in which the *global feature vectors* for each term equivalence class can be derived by their local contexts represented as *local feature vectors*. The global feature vectors should represent the exogenous behaviour of an entire term equivalence class. A model preserving the syntactic information together with the local lexicalisations is then proposed. In such a "syntactic lexicalised" model (Λ), the lexical item that governs the observed grammatical relation is stored in a local vector together with its grammatical type. For example, given the context *"The equation of mechanics governs the conservation of energy."* of *equation*, we can capture *equation* as the *subject* of the verb *to-govern*. In the syntactic lexicalised space Λ the different lexicalised information (`Syntactic_Type`, `governing_lemma`) will be considered as independent features. For example F_h^Λ=`(V-Subj,to-govern)` for t=*equation* or F_k^Λ=`(NP-PP,conservation)` for the t=*energy* can be derived from Ex. 1.

The above features can be obtained by shallow parsing of the corpus sentences. Notice that syntactic ambiguity in parsing may affect the above observations and frequency counts. Highly ambiguous (but frequent) phenomena (e.g. prepositional phrase attachments) may increase the values for irrelevant features. On the contrary, the pruning of all ambiguous relations may result in too poor evidences. In our approach we use the notion of plausibility of a grammatical relation within an eXtended Dependency Graph (XDG) representation scheme (see (Basili et al., 2000)). Ambiguous relations r in a dependency graph are given a score $pl(r)$ inversely proportional to the number of conflicting syntactic interpretations. The plausibility $pl(r)$ ranges in the $(0, 1]$ interval: $pl(r) = 1$ if r is unambiguous for the parser, and $pl(r) < 1$ otherwise.

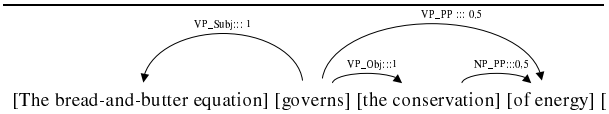

Figure 3: A sample XDG

Grammatical relations, local to the source sentence s, are thus a set $I(s)$ of triples (t, F, p) where p is the plausibility local to s of the relation between the term t and the feature F. The excerpt in Ex. 1.a) generates the XDG in figure

3, where
the relations (energy,`NP-PP-conservation`,0.5) and (energy,`VP-PP-govern`,0.5) are ambiguous. The i-th component (representing the feature F_i) of the local feature vector for t thus obtained as $\overline{v}_i^{\Lambda}(t,s) = \Sigma_{(t,F_i,p)\in I(s)} p$.

Once local vectors $\overline{v}^{\Lambda}(t,s)$ are available for sentence s, the global feature vectors in the two spaces are obtained as follows:

$$v^{\Lambda}(t) = \sum_{s \in C(t)} \overline{v}^{\Lambda}(t,s) \qquad (2)$$

where $C(t)$ include the corpus contexts (i.e. the equivalency class) of t.

The values a feature vector assigns to features F_i emphasize the strength of association between the t and F_i. Cumulative plausibility here replaces frequency counts to better model ambiguity in observations. Notice that, for the same F_i, the estimated frequency $\sum_{(t,F_i,p)\in C(t)} p$ produces the same ranking as mutual information $MI(t,F_i)$. Feature vectors $v^{\Lambda}(t)$ are finally normalized to obtain $\hat{v}^{\Lambda}(t)$. These normalized vectors $\hat{v}^{\Lambda}(t)$ are input to the decision tree learner. For sake of comparison, a frequency-based learner has been obtained (feature space Φ) by defining $\hat{v}^{\Phi}(t) = (rf(t))$ where $rf(t)$ is the relative frequency of t in the corpus. Such discrete space will simulate the behaviour of a quantitative model based on simple frequency.

The above spaces, i.e. the syntactic lexicalised and the frequency-based spaces, can be called here "pure". As better results can be obtained if different information is integrated (as also suggested in (Basili et al., 2001)): contextual information can be used in cooperation with the term frequency. An other space has been thus defined via juxtaposition of the underlying pure vectors, $v^{\Phi}(t)$, and $\hat{v}^{\Lambda}(t)$: the resulting space $\Phi \times \Lambda$ merges frequency and syntactic lexicalised information.

4 Experimental investigation

The aim of the investigation is twofold. Firstly, to establish that a domain-oriented term definition better models the terminologists' choices. Secondly, to analyse the upgrading of the model of the terminologists' term definition during the analysis. The two different lines of investigation have been carried out over a well-established implicit domain model. For what concerns the reported performances, a statistical validation has been obtained by n-fold cross validation. The source domain consists of a corpus of about 250,000 words on the Italian Civil laws, and of a corresponding thesaurus of 600 term equivalence classes built by a team of expert terminologists. The corpus has been processed by the CHAOS parser (Basili et al., 2000) producing about 3,000 different term equivalency classes. We assumed that the *only* valid term instances are those coded in the thesaurus. We have thus about 1/4 valid structures among the corpus-derived candidates.

4.1 DT as *itd*s

For the analysis of the intensional term definition, an incremental approach has been carried out (cf. Sec. 2). The seeding of the process (i.e. the pre-existing terminology of the initial implicit domain model) has been obtained collecting a 80% of the 600-term thesaurus as training info and the rest as test. Moreover, the terminologists incremental work has been simulated by training over increasing bags of non-terms. The learning process has been fed with an increasing number of non-terminology subsets (up to 20) and a decision tree has been derived for each subset. By adding the negative evidence (i.e. refused entries) as training examples we simulate the activity of the terminologists.

By inspecting the obtained trees we study the increasing awareness about the domain along with the term judgment. As expected, the trends described below are shared by the different trees derived via iterations in the n-fold cross validation.

In Fig. 4 and in Fig. 5, we report an excerpt of the decision trees derived, respectively, over the Λ and $\Phi \times \Lambda$ spaces. The reported 3 trees reflect different stages as they are built over increasing numbers of negative examples: `lex-1-3` to `lex-1-20` refer to 3/20 and 20/20 among the 2400 available negative examples. As the upper levels of the trees are shown, the figures show the most general rules. The trend (see Fig. 4) is that general features (e.g. being part of a predicative structures, i.e. `V_Obj-essere`) are initially retained as decision rules. However they lose importance as soon as more negative information is available. General prediction rules based on general verbs such as *essere (to be)*, *avere (to have)*, etc. are substituted

```
Iteration: <lex-1-3>

V_Obj-essere <= 0.166667 : TERM
V_Obj-essere > 0.166667 :
|   V_Obj-fare <= 0.05 : NON-TERM
|   V_Obj-fare > 0.05 : TERM


Iteration: <lex-1-4>

V_Obj-essere > 0.185185 : NON-TERM
V_Obj-essere <= 0.185185 :
|   V_Sog-avere > 0.00281691 : TERM
|   V_Sog-avere <= 0.00281691 :
|   |   V_Sog-dovere > 0.00262467 : TERM
|   |   V_Sog-dovere <= 0.00262467 :
|   |   |   NP_PP-effetto > 0.00673758 : TERM
|   |   |   NP_PP-effetto <= 0.00673758 : ----->


Iteration: <lex-1-20>

V_PP-intervenire > 0 : TERM
V_PP-intervenire <= 0 :
|   NP_PP-estinzione <= 0 :
|   |   NP_PP-nomina <= 0 :
|   |   |   V_PP-escludere <= 0.00136612 : ----->
|   |   |   V_PP-escludere > 0.00136612 :
|   |   |   |   NP_PP-cosa > 0.00250356 : NON-TERM
|   |   |   |   NP_PP-cosa <= 0.00250356 :
|   |   |   |   |   NP_PP-venditore > 0.0048077 : NON-TERM
|   |   |   |   |   NP_PP-venditore <= 0.0048077 :
|   |   |   |   |   |   V_PP-riconoscere <= 0.00333333 : TERM
|   |   |   |   |   |   V_PP-riconoscere > 0.00333333 : NON-TERM
|   |   NP_PP-nomina > 0 :
|   |   |   NP_PP-affittuario > 0 : NON-TERM
|   |   |   NP_PP-affittuario <= 0 :
|   |   |   |   NP_PP-scadenza <= 0.00735295 : TERM
|   |   |   |   NP_PP-scadenza > 0.00735295 : NON-TERM
|   NP_PP-estinzione > 0 :
|   |   NP_PP-persona <= 0.0277778 : TERM
|   |   NP_PP-persona > 0.0277778 : NON-TERM
```

Figure 4: Domain-oriented definition evolution in the Λ space

by more domain specific cues. Domain specific rules as the ones based on *intervenire (to intervene)*, *nomina (nomination)*, *estinzione (liquidation)*, etc. tend to appear higher in the hierarchy, i.e. they gain importance. Moreover, since the categorization capability of the trees augments (i.e. the error rate decreases from 40% to 14,25%), the induced (domain-specific) DT seems better modelling the terminologist judgement. Fig. 5 reports DTs based also on frequency. We can observe here a similar adaptation process. In fact, the general rules fully based on frequency (e.g. **freq-lex-3-2**) are replaced by more specific ones that do not depend only on frequency: on the contrary syntagmatic lexicalised decision rules emerge at the upper levels (e.g. **freq-lex-3-20**).

We observed the emergence of very specific rules (patterns) at the lower levels of the hierarchy as, for example, the following excerpt of tree (re-written in an IF...THEN...ELSE... fashion):

```
1. IF plausible(atto-NP_PP-X) THEN
1.1. IF plausible(apporre-V_Obj-X) THEN TERM
1.2. ELSE IF plausible(autorizzare-V_Obj-X)
          THEN TERM ELSE NON TERM
   ELSE ...
```

```
Iteration: <freq_lex-3-2>

Freq > 0.0348566 : TERM
Freq <= 0.0348566 :
|   Freq <= 0.0174283 : NON-TERM
|   Freq > 0.0174283 : TERM

Iteration: <freq_lex-3-4>

Freq <= 0.0348566 :
|   Freq <= 0.0174283 : NON-TERM
|   Freq > 0.0174283 :
|   |   V_PP-essere <= 0.0820313 : NON-TERM
|   |   V_PP-essere > 0.0820313 : TERM
Freq > 0.0348566 :
|   NP_PP-trasferimento > 0 : TERM
|   NP_PP-trasferimento <= 0 :
|   |   NP_PP-creditore > 0.00128699 : TERM
|   |   NP_PP-creditore <= 0.00128699 :
|   |   |   NP_PP-responsabilità > 0.00543479 : TERM
|   |   |   NP_PP-responsabilità <= 0.00543479 : ----->


Iteration: <freq_lex-3-20>

NP_PP-estinzione <= 0 :
|   NP_PP-deliberazione <= 0.00485437 :
|   |   V_PP-effettuare <= 0.000685865 :
|   |   |   NP_PP-trascrizione <= 0.000614251 : ----->
|   |   |   NP_PP-trascrizione > 0.000614251 :
|   |   |   |   Freq <= 0.278853 : NON-TERM
|   |   |   |   Freq > 0.278853 :   ----->
|   |   V_PP-effettuare > 0.000685865 :
|   |   |   V_PP-operare <= 0.00243307 : TERM
|   |   |   V_PP-operare > 0.00243307 : NON-TERM
|   NP_PP-deliberazione > 0.00485437 :
|   |   NP_PP-data <= 0.015641 : TERM
|   |   NP_PP-data > 0.015641 : NON-TERM
NP_PP-estinzione > 0 :
|   NP_PP-persona <= 0.0277778 : TERM
|   NP_PP-persona > 0.0277778 : NON-TERM
```

Figure 5: Domain-oriented definition evolution in the $\Phi \times \Lambda$ space

where **plausible(atto-Rel-X)** expresses the contraints that the candidate **X** must be observable (frequently) as a modifier of type **Rel** with the word **atto** (i.e. *legal act*). In this case, the rule applies to heads like *notaio* (*notary*) since structures like *atto di notaio* (*the act of notary*) and *autorizzare il notaio* (*to authorize a notary*) are frequent: they are thus accepted as terms (as for rules **1.** and **1.2**). On the contrary, an head like *ricevimento* (*the reception*) is refused. In fact, although *atto di ricevimento* (the act of reception) is frequent in the corpus, there are no frequent structures for contraints **1.1** and **1.2** (e.g. ** apporre un ricevimento* (*to pose a reception* and ** autorizzare un ricevimento* (*to authorize a reception*). Criteria like the above ones effectively capture the terminologist behaviour in a computationally attractive form.

4.2 Performance Evaluation

A general analysis of the average error rate ϵ (i.e. the percentage of misclassified items with respect to the terminological database avail-

Feature Space	Φ	Λ	$\Phi \times \Lambda$
Error Rate (%)	16,99	14,25	13,88

Table 1: Final error rate on the Φ, Λ and $\Phi \times \Lambda$.

able) has been also carried out. In each 5-fold cross-validation, the system considers an 80% of the corpus candidates as training items (divided evenly between positive terms in the thesaurus and negative items, i.e. nominals that are NOT in the thesaurus). The test is then run over the 20% remaining candidates and error rates are then reported as mean values. The syntactic lexicalised Λ space reaches superior performances with respect to the pure frequency (Φ). All the two learning processes make similar use of negative information. Moreover, the one depending more tightly on the domain evidence (Λ) outperforms a more domain independent notion of relevance (i.e. frequency). The exogenous grammatical information is very effective (i.e.+18% wrt Φ). This confirms the initial assumption: stable relations between particular lexicals in the domain (captured, in this case, with syntactic lexicalised feature model) produce better models for the inner perception of terms hold by the terminologists. Furthermore, the syntactic lexicalised model represents specific "shallow" semantic properties of terms as induced from the corpus. Combining different sources always outperforms "pure" systems: performances obtained in the $\Phi \times \Lambda$ are superior to the one obtained on the "pure" Λ.

5 Conclusion

In this paper, a terminology acquisition model based on the decision tree learning has been presented. The proposed approach makes use of contextual evidence observable for known terms as well as information about non terminological expressions. A lexico-syntactic representation of such information is used on a large scale within a robust text processing framework (Basili et al., 2000). Moreover, a decision tree machine learning algorithm (Quinlan, 1993) is applied for the empirical investigation. First, experiments aimed to simulate the development of an explicit domain-dependent model of termhood have been carried out. Results show that decision trees embed systematic information and emphasize correctly typical domain effects. Performance evaluation confirms the effectiveness of the overall approach either on a pure application of lexico-syntactic criteria as well as by combining it with more frequency oriented rules. An improvement of about 18% against the previously reported successfully methods has been obtained.

The method depicted above represents an original approach to automatic TA. Since it seems better to approximate the terminologist behaviour, it will play a relevant role in our future research on the induction of ontological knowledge from texts.

References

Roberto Basili, Maria Teresa Pazienza, and Fabio Massimo Zanzotto. 2000. Customizable modular lexicalized parsing. In *Proc. of the 6th International Workshop on Parsing Technology, IWPT2000*, Trento, Italy.

Roberto Basili, Maria Teresa Pazienza, and Fabio Massimo Zanzotto. 2001. Modelling syntactic context in automatic term extraction. In *Proc. of the 3th Conference on Recent Advances in Natural Language Processing, RANLP2001*, Tzigov Churck, Bulgaria.

Beatrice Daille. 1994. *Approche mixte pour l'extraction de terminologie: statistque lexicale et filtres linguistiques.* Ph.D. thesis, C2V, TALANA, Universitè Paris VII.

Gregory Grefenstette. 1993. Evaluation techniques for automatic semantic extraction: Comparing syntactic and window based approaches. In *Proceedings of the Workshop on Acquisition of Lexical Knowledge from Text*, Columus, OH, USA.

Christian Jacquemin. 1997. *Variation terminologique : Reconnaissance et acquisition automatiques de termes et de leurs variantes en corpus. Mémoire d'Habilitation Diriger des Recherches en informatique fondamentale.* Université de Nantes, Nantes, France.

Emmanuel Morin. 1999. *Extraction de liens sémantiques entre termes à partir de corpus de textes techniques.* Ph.D. thesis, Univesité de Nantes, Faculté des Sciences et de Techniques.

J.R. Quinlan. 1993. *C4.5: Programs for machine learning.* Morgan Kaufmann, San Francisco, CA.