Concept Lattice Mining for Unsupervised Named Entity Annotation

Thomas Girault

Orange labs, 2 avenue Pierre Marzin, 22307 Lannion Cedex
IRISA, Campus universitaire de Beaulieu, 35042 Rennes Cedex, France
thomas.girault@orange-ftgroup.com

Abstract. We present an unsupervised method for named entity annotation, based on concept lattice mining. We perform a formal concept analysis from relations between named entities and their syntactic dependencies observed in a training corpus. The resulting lattice contains concepts which are considered as labels for named entities and context annotation. Our approach is validated through a cascade evaluation which shows that supervised named entity classification is improved by using the annotation produced by our unsupervised disambiguation system.

1 Introduction

Lexical ambiguity is a fundamental problem which is central in many tasks involving natural language processing (*e.g.* information retrieval, information extraction, ...). Our study focuses on a kind of lexical units (LU), named entities (NE), a generic denomination for proper names including persons, locations, organisations. As most LU considered outside a context, NE are ambiguous since their form can potentially refer to different meanings or objects. Our approach to disambiguation is based on *formal concept analysis* (FCA), a generic method for data analysis and knowledge representation which infers *formal concepts* from relational data. In this work, FCA is used to build a knowledge-base that is exploited for NE annotation.

The problem of ambiguity can be considered according to several Word Sense Disambiguation (WSD) approaches [1]. Knowledge-based approaches attempt to select the meaning of words using lexicons, dictionaries or thesauri (*e.g.* WordNet). Corpus-based approaches examine the occurrence of LU and their contexts using machine learning techniques. Supervised learning disambiguates LU according to pre-defined labels whereas unsupervised techniques discriminate the meanings of unlabelled LU thanks to similarity of their contexts.

Since corpus annotation is a tedious and costly task, this work is focused on unsupervised approaches. Among them, *formal concept analysis* (FCA) [2] has been selected : this symbolic unsupervised machine learning technique operates on relational data to infer *formal concepts* which can be structured into a *concept*

ISBN 978–80–244–2111–7, Palacký University, Olomouc, 2008.

lattice. FCA is applied on relations between NE and their syntactic dependencies extracted from English news wire articles. The sets of NE sharing the same syntactic dependencies constitute the formal concepts which are considered as units of meaning for the annotation of NE. The concept lattice obtained can be seen as a hierarchical knowledge-base modelling meaning overlapping on several levels of granularity. To our knowledge, these properties attached to concept lattices have not yet been exploited in an unsupervised WSD task. In this context, we propose a conceptual annotation method for NE disambiguation.

In this paper, we address the problem of exploiting a concept lattice for unsupervised NE annotation. We first introduce (Section 2) the problem of NE ambiguity by exposing few examples from our corpus in which relations between NE and their syntactic dependencies are extracted. These relations constitute a formal context from which FCA is performed (section 3). The resulting lattice contains formal concepts which are considered as labels for NE and dependency annotation (Section 4). Our approach is validated through a cascade evaluation (section 5) which shows that supervised NE classification is improved by using the annotation produced by our unsupervised disambiguation system.

2 Corpus-Based Methods for Word Sense Disambiguation

This section introduces corpus-based word sense disambiguation (WSD) with a small sample of a corpus where NE occurrences are semantically labelled. Supervised learning disambiguates LU according to labelled pre-defined meanings whereas unsupervised techniques discriminate the meanings of unlabelled LU thanks to similarity of their lexical contexts. Our unsupervised approach is built upon the study of syntactic relations between NE and other LU occurring in an utterance.

2.1 Tagset Granularity for Supervised NE Classification

Named Entity Recognition (NER) is a subtask of Information Extraction. Different NER systems were evaluated, among others, as a part of the Message Understanding Conferences [3] in 1995 and in the CoNLL 2003 shared task [4]. The most efficient NER systems are built upon supervised corpus-based learning for the detection and classification of NE. They rely on semantically annotated corpora which we can illustrate with the following examples (figure 2.1) :

1. India$_{loc}$ has acquired 120,000 tonnes of diesel in three cargoes, ...
2. Cricket - : India$_{loc}$ wins the toss and bat against Sri Lanka$_{loc}$.
3. Tennis - : Muster$_{per}$ upset, Philippoussis$_{per}$ wins, Stoltenberg$_{per}$ loses.
4. Schumacher$_{per}$ wins Belgian Grand Prix.
5. Clinton$_{per}$ wins democratic re-nomination.
6. Siam Commercial$_{org}$ wins agency bond auctions.

Fig. 1. Samples extracted from the English CoNLL-2003 annotated corpus.

The English CoNLL 2003 data is a collection of news wire articles from the Reuters Corpus in which the NE are manually labelled with respect to the coarse-grained semantic tagset {*person, location, organisation, miscellaneous*}.

The examples (1) and (2) illustrate a case of ambiguity : the NE "India" is labelled as location but a more fined granularity would distinguish the sport nation and the wholesale importer. In addition we could note that LU interacting with NE are ambiguous as well : the LU "wins" occurs with different meanings for the domains of politics, sport or business. Thus, we think that the original tagset should be enriched with a refined semantic description. However, a manual refinement would be a tedious and a costly task. In addition, we cannot define a general semantic tagset since it is domain dependent : for instance a biomedical semantic tagset should discriminate viruses and proteins and it would not be suitable to describe geographic entities such as rivers or mountains.

2.2 Unsupervised Corpus-Based Disambiguation

Instead of assigning predefined labels to LU, an alternative strategy is to discriminate their meanings by analysing their co-occurrences in the utterances of a corpus. This unsupervised approach is founded from the assumption that LU (NE in our case) which occur in similar contexts tends to have close meanings. Distributional methods [5] relying on Harris' hypothesis consider that the share of contexts having common syntactic patterns (*e.g.* subject-verb, modifier-noun) constitutes an indicator of semantic relatedness.

2.3 Named Entity Dependency Extraction

Before applying distributional hypothesis for NE disambiguation, the LU attached syntactically to NE need to be identified. We suppose that the NE frontiers have been already detected. Our method deals with two kinds of dependencies. External dependencies are mainly nouns, verbs and prepositions occurring before or after a NE. They are extracted with patterns defined manually relying on morphosyntactic tagging and phrase chunking available with the CoNLL-2003 corpus[1]. The patterns extracts expressions such as :

- noun + preposition + NE (*e.g.* [election of, Clinton], [results of, European Super League]);
- noun + NE (*e.g.* [champion, Pete Sampras]);
- NE + noun (*e.g.* [Russian, government]);
- NE + verb (*e.g.* [Clinton, signed], [India, wins]).

Internal dependencies correspond to non prepositional tokens occurring in the NE, such as first names or surnames. For example, the list of internal dependencies of *International Boxing Federation*, is {*international, boxing, federation*}.

This work on extraction provides a set of pairs (NE, syntactic dependency) where each element is potentially ambiguous.

[1] Morphosyntactic tagging and chunking have been generated automatically and are therefore noisy.

3 Formal Concept Analysis for Knowledge Base Acquisition

In this section, the approach for knowledge-base acquisition using FCA is exposed. We illustrate FCA with examples taken from our linguistic data. We then discuss the advantages of FCA for dealing with meaning overlapping and granularity of meanings.

3.1 Formal Context of Syntactic Relations

Classical distributional methods could deal with ambiguity of the whole set of LU. However, these methods consider them from a unique point of view whereas for our problem, the data seems more naturally represented according to two interconnected views as the figure (2) shows :

- a view on named entities which is associated to a set of objects $O = \{o_1, o_2, \cdots, o_m\}$.
- a view on their dependencies (*syntactic co-texts + internal components*) represented by a set of attributes $A = \{a_1, a_2, \cdots, a_n\}$

These views are connected by a relation $R \subseteq O \times A$, where $R(o, a)$ means that the object o has the attribute a (*i.e.* the NE o has the dependency a).

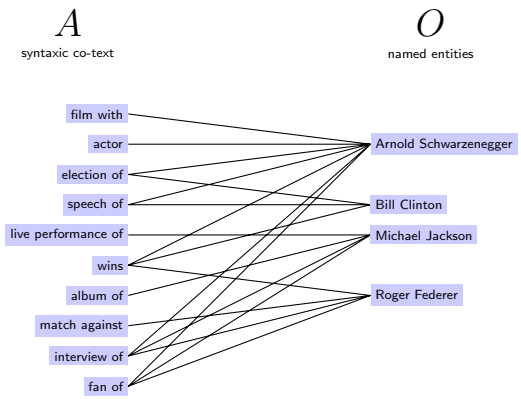

Fig. 2. Relations between NE and their dependencies.

In the FCA terminology, the triple $\mathbb{K} = (O, A, R)$ is called a formal context. It corresponds to a bigraph (from the figure (2)) of objects (NE) in relation with attributes (syntactic co-texts + internal components).

3.2 Formal Concept Analysis

For the understanding of the paper, we introduce standard definitions and notations of FCA [2]. For $E \subseteq O$ and $I \subseteq A$, we define two sets $E' \subseteq A$ and $I' \subseteq O$ extending them : $E' = \{a \in A | \forall o \in E : (o, a) \in R\}$ as the set of all attributes from I that are in relation with all objects from E and $I' = \{o \in O | \forall a \in I : (o, a) \in R\}$, the set of all objects from O that are in relation with all attributes from I. For instance, if $I = \{$speech of, election of$\}$ then $I' = \{$Bill Clinton, Arnold Schwarzenegger$\}$. For $E = \{$Michael Jackson$\}$, we have, $E' = \{$album, live performance of, interview of, fan of$\}$.

We can define a *formal concept* of the formal context $\mathbb{K}$ to be a pair (E, I) satisfying $E \subseteq O, I \subseteq A$, $E' = I$ and $I' = E$. E is called the *extent* and I is called the *intent* of concept. For instance, the pair ($\{$Bill Clinton, Arnold Schwarzenegger$\}$,$\{$wins, election of, speech of$\}$) is a formal concept. The concepts are partially ordered according to the relation $\leq$:

$(E_1, I_1) \leq (E_2, I_2) \Leftrightarrow E_1 \subseteq E_2 \Leftrightarrow I_2 \subseteq I_1$

For instance, we have $C_2 \leq C_0$ for the concepts $C_2 = (\{$Arnold Schwarzenegger,Roger Federer$\}, \{$wins,interview of, fan of$\})$ and $C_0 = (\{$Michael Jackson, Roger Federer, Arnold Schwarzenegger$\}, \{$interview of, fan of$\})$. The relation $\leq$ form a complete lattice $\mathcal{L}$, called the *concept lattice* of $\mathbb{K}$.

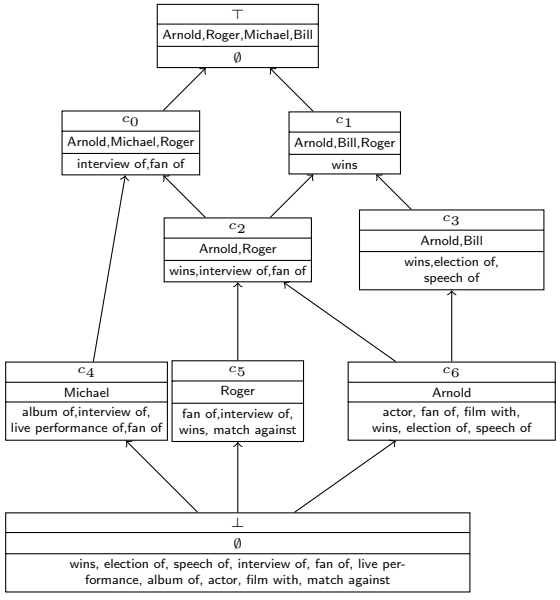

Fig. 3. Concept lattice for the formal context of figure (2). A concept box is contains a name, an extent and an intent.

3.3 The Concept Lattice : a Discriminative Knowledge Base

The general approach for building the concept lattice from linguistic data is similar to the work of Cimiano et al. [6]. The algorithm *AddIntent* [7] has been used for the construction of the lattice. It adopts an incremental procedure allowing dynamic lattice structuring according to new objects or attributes discovered from new utterances. Thus, a lattice could be seen as a knowledge-base already structured which could be adapted to a new corpus. This is an interesting property considering the weak evolutivity of classical lexical resources such as thesauri. According to this perspective, Priss [8] has been able to encapsulate the FrameNet thesauri within *relational concept analysis* framework.

As the figure (3) depicts, the concept lattice structure is organised according to several granularity layers. The upper part of the lattice is represented by general concepts grouping objects which share ambiguous attributes. The opposite part of the lattice has very specific concepts having ambiguous objects. The intermediate zone of the lattice provides concepts which seem more appropriate for LU disambiguation. Although the lattice model is generally considered as symbolic and discrete representation, the intent/extent overlapping reveals potential continuity of meanings. To our knowledge, these properties attached to concept lattices have not been exploited yet for an unsupervised WSD task.

4 Unsupervised Named Entity Annotation

In this section, we describe our FCA based methodology for annotation of relations between a NE and its context in an utterance. FCA is not only used to aggregate data, but also to perform a classification of NE. The unsupervised annotation is based on a selection of formal concepts according to n NE and its dependencies. We illustrate the method with an example and we finally introduce a dimensionality reduction method for the visualisation of formal concepts.

4.1 Concept Lattice Mining for Conceptual Annotation

Formal concepts are now considered as units of meaning potentially useful for LU annotation. As we noticed previously, the overlapping of intents and extents between formal concepts is linked to the intuition that some concepts are more similar than others since they share more objects or more attributes. Thus, the formal concepts could be associated to a metric space where the distance between two concepts measures a degree of semantic similarity.

In a new utterance, we suppose that a new NE $o \in O$ and its dependencies $Atts \subseteq A$ have been detected thanks to the extraction patterns (section 2.3). For a disambiguation task, we consider that the meaning of o relies on the meaning of its dependencies in $Atts$ occurring in the context : in other words, o can be annotated with a formal concept $x \in \mathcal{L}$ according to the concepts for the dependencies in $Atts$.

Our model for conceptual annotation of named entities is based on querying the concept lattice. In the lattice $\mathcal{L}$ the object o is associated to $Co = (\{o\}'', \{o\}')$

and similarly the concepts for the attributes of $Atts$ are the elements Ca_i from $C_{Atts} = \{(\{a_i\}', \{a_i\}'')|a_i \in Atts\}$. We are looking for a representative concept in the lattice which interpolates the concepts Co and Ca_i. We will call this concept x the *prototype* and we search it among the concepts containing o in their extent or at least one dependency a_i in their intent. More formally, $x \in \mathcal{L}(o, Atts)$ where $\mathcal{L}(o, Atts) = \{(E, I) \in \mathcal{L}|o \in E \vee Atts \cap I \neq \emptyset\}$. The prototype x is defined as the concept whose average dissimilarity to the concepts Co and Ca_i is minimal.

$$X = \underset{x \in \mathcal{L}(o,Atts)}{\operatorname{argmin}} \sum_{c \in C_{Atts} \cup \{Co\}} similarity(c, x) \tag{1}$$

In order to deal with similarities, we define two matrices $\mathcal{A}(o, Atts)$ and $\mathcal{O}(o, Atts)$ in which each row corresponds to a formal concept from $\mathcal{L}(o, Atts)$. The columns of $\mathcal{A}(o, Atts)$ are assigned to the intent of the concepts and similarly, the columns of $\mathcal{O}(o, Atts)$ are assigned to the extent of the concepts. Thus, the formal concepts are represented by a vector for extents and a vector for intents. Note that we can also consider $\mathcal{M}(o, Atts)$ which is the concatenation of the matrices $\mathcal{A}(o, Atts)$ and $\mathcal{O}(o, Atts)$.

Similarity measures can then be applied between the concept vectors of $\mathcal{A}(o, Atts)$, $\mathcal{O}(o, Atts)$ or $\mathcal{M}(o, Atts)$: measures such as Euclidean, cosine, correlation, Hamming or Jaccard can be chosen, depending of if we consider the vectors (and the formal context) as boolean or as weighted by the frequency counts of relations (cooccurrences) observed in the corpus. In the last case, the weights assigned to objects and attributes would be respectively $\sum_{a_i \in intent(C)} card(R(o, a_i))$ and $\sum_{o_i \in extent(C)} card(R(o_i, a))$.

4.2 Example from CoNLL Data

To illustrate the method, we propose to annotate the expression "English division" from which the pair $(o, Atts)$ = (English,{division}) is extracted. In a classical dictionary, the LU *division* is typically ambiguous because it can denotes, for instance, a group of military troops or a group of teams in an organised sport. The following list enumerates the concepts associated to $(o, Atts)$

1. (['SCOTTISH PREMIER DIVISION', 'SCOTTISH PREMIER', 'ENGLISH', 'FRENCH', 'SCOTTISH'], ['division', 'premier'])
2. (['DUTCH', 'ENGLISH', 'SCOTTISH'], ['division', 'results', 'league'])
3. (['ENGLISH', 'DUTCH'], ['division', 'draw', 'division leaders', 'league', 'results', 'result', 'news agency'])
4. (['ENGLISH', 'SCOTTISH'], ['league soccer', 'league', 'division', 'premier', 'results', 'league standings', 'league cup', 'summaries'])
5. (['ENGLISH', 'SCOTTISH PREMIER DIVISION', 'DUTCH', 'FRASER', 'MOROCCAN', 'SCOTTISH', 'SWISS', ...,'HUNGARIAN'], ['division'])
6. (['WELSH', 'ENGLISH'], ['division', 'referee', 'results'])
7. (['GERMAN', 'DUTCH', 'ENGLISH'], ['division', 'results', 'result'])
8. (['GERMAN', 'ENGLISH'], ['result', 'division', 'law', 'summaries', 'results'])
9. (['ENGLISH'], ['standings', 'premier', 'results', 'county', 'result', 'league cup', 'news agency', 'city', 'langage', 'soccer matches', 'play scores', ...])
10. (['WELSH', 'SCOTTISH', 'FRENCH', 'GERMAN', 'AUSTRIA', 'DUTCH', 'MOROCCAN', 'ENGLISH', 'SWISS', 'POLISH'], ['division', 'results'])
11. (['ENGLISH', 'FRENCH', 'SCOTTISH'], ['division', 'premier', 'results', 'summaries'])
12. (['AUSTRIA', 'DUTCH', 'ENGLISH'], ['division', 'draw', 'results'])
13. (['SWISS', 'ENGLISH'], ['division', 'results', 'league leaders'])
14. (['NATIONAL LEAGUE EASTERN DIVISION', 'DUTCH', 'ENGLISH', 'AMERICAN LEAGUE EAST DIVISION', 'NATIONAL LEAGUE CENTRAL DIVISION', 'SCOTTISH'], ['league', 'division'])
15. (['GERMAN', 'ENGLISH', 'FRENCH', 'SCOTTISH'], ['division', 'results', 'summaries'])

In the lattice $\mathcal{L}$(English,{division}), the object "English" is represented in the lattice by the concept C_9 and the attribute "division" is represented by the concept C_5. Most concepts appears to denote the *sport division* meaning and it remains to select an appropriate concept for the annotation of the query. The prototype calculation has been done on this example according to several similarity metrics. The Euclidean and hamming distances chosen among others for the similarity measures, have both selected the concept C_4 which seems a acceptable for the annotation.

4.3 Dimensionality Reduction for Visualisation of Formal Concepts

The technique presented here has not yet been linked to the disambiguation process. It illustrates our intuition that continuous semantic provided with distance fits with a high structured representation such as concept lattices. For a better understanding of this intuition, we propose to visualise formal concepts through a cartographic representation where distance between formal concepts translates the notion of semantic proximity.

We have describe previously a simple way to associate a set of formal concepts to matrices. Since the vectors associated to concepts potentially have a huge dimension, we propose to use dimensionality reduction methods on the matrix $\mathcal{M}(o, Atts)$. These methods are able to compress $\mathcal{M}(o, Atts)$ such as each vector/concept representation is reduced to two dimensions. Among these methods we have chosen *curvilinear component analysis* (CCA) [9] which can be seen as a non linear extension to *principal component analysis*. The first results of this method are depicted by the figure (4).

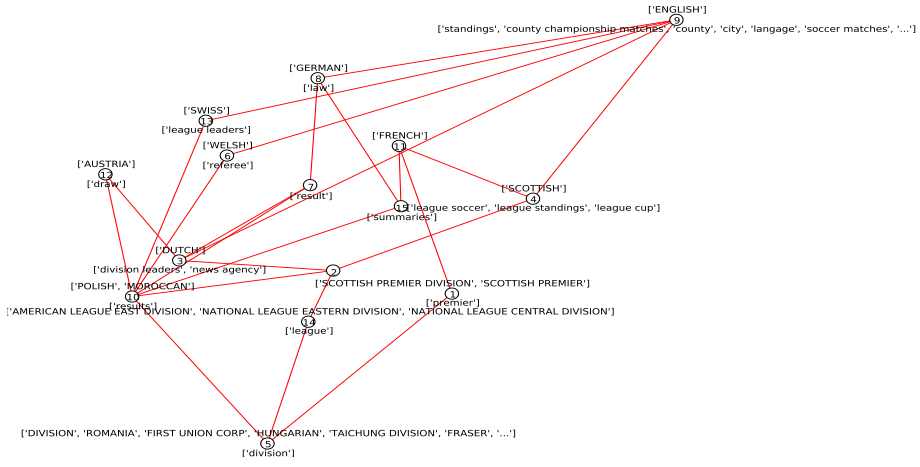

Fig. 4. Visualisation of formal concepts associated to the query *(English,* {*division*}*)* using CCA.

Reduced labelling has been used to improve the readability of the figure. In this scheme, the label for an object o is drawn above the *object concept* $\gamma(o) = (\{o\}'', \{o\}')$ while the label for an attribute a is drawn below the *attribute concept* $\mu(a) = (\{a\}', \{a\}'')$.

Our approach does not take advantage of the partial ordering between concepts that has been already computed. However, according to these figure, the general to specific ordering seems globally respected whereas it has not been taken into account for the rendering of the map : the most general and the most specific concepts occur to opposite sides of the map. The figure (4) also helps to understand where is the prototype C_4 among the other concepts resulting to the query.

5 Experiments and Evaluation

Previously, we have described an unsupervised method for conceptual annotation of NE. The evaluation of such unsupervised methods is subjective by nature since several concepts would be relevant to disambiguate a NE. In this section, we present a validation of our approach according to an existing task (supervised NE classification) that we are able to evaluate the performance. We then describe the cascade evaluation protocol [10] which considers the unsupervised conceptual annotation as a pre-processing step for a supervised NE classification task. We conclude the section with a study of the results obtained through this experiment.

5.1 CoNLL 2003 Data

The CoNLL-2003 named entity English data consists of three files : one training file (train), one development file (testa) and one test file (testb). Figure (5) gives an overview of the characteristics of the corpus.

	Articles	Sentences	Tokens	Locations	Misc	Organisations	Persons
Training corpus (train)	946	14987	203621	7140	3438	6321	6600
Development corpus (testa)	216	3466	51362	1837	922	1341	1842
Test (testb)	231	3684	46435	1668	702	1661	1617

Fig. 5. CoNLL 2003 corpus .

Our learning methods have been trained with the training and development data sets. The concept lattice obtained contains 14834 concepts for 8934 objects, 13983 attributes and 57170 relations in the formal context. The figure (6) depicts a conceptual annotation produced by our system on a CoNLL sample.

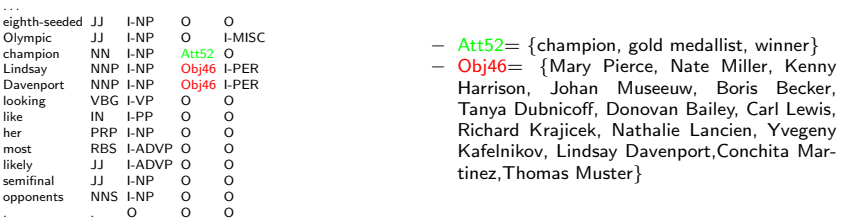

Fig. 6. Example of conceptual annotation in the CoNLL 2003 corpus.

The Euclidean measure has been used for the prototype determination of the intent matrix $\mathcal{A}$(Lindsay Davenport, {champion}) and for the extent matrix $\mathcal{O}$(Lindsay Davenport, {champion}). It selects two concepts C_{52} and C_{46} : the intent of C_{52} provides a disambiguation of "champion" and the extent of C_{46} gives an annotation for "Lindsay Davenport".

5.2 Cascade Evaluation

In the framework of cascade evaluation [10], unsupervised learning is considered as a pre-processing step for a supervised NE classification task that we are able to evaluate. This cascade process reveals whether the conceptual annotation provides interesting enrichments to improve the supervised task on the CoNLL 2003 corpus. The protocol consists in comparing errors produced by two classifiers A and B, when they perform on the test corpus (testb), after a training step on the same training data (train + testa).

The system A is a supervised classifier trained normally on the labelled training corpus. As Ehrmann and Jacquet proposed [11], the system B provides two annotations for NE. The first is given by our unsupervised annotation system exploiting the concept lattice learned on the unlabelled training corpus. This pre-processing step provides enrichments to the initial corpus description. The system B can then benefit from these additional enrichments during the supervised learning step in order to produce the second annotation layer.

5.3 Experimental Results with Transformation-Based Learning

We have adapted the *transformation-based learning* (TBL) algorithm [12] to design a supervised NER system. The algorithm initializes the NE labels with a language model classifier (unigram), trained on the training corpus. The goal is to correct this initial classification according to the original NE labels specified in the training corpus. The next steps follow an iterative process : it corrects the initial incorrect classification by inferring a sequence of transformation rules. They are successively applied over the corpus in order to improve progressively the NE classification.

The resulting rules are instantiated from a list of extraction patterns defined manually. These patterns are able to explore co-texts features in a window

of +/- 3 words : among the available features, we have considered the word, its morphosyntactic tag and the concept identifiers given by our unsupervised conceptual annotation method.

The figure (7) shows the results of the cascade evaluation. The left column indicates the performances reached by classifier A applied on the test corpus provided with morphosyntactic tagging. The right column corresponds to results obtained with the classifier B which has be used on the test corpus enriched with the conceptual annotation.

	A : TBL			B : conceptual annotation + TBL		
	Precision	Recall	$F_{\beta=1}$	Precision	Recall	$F_{\beta=1}$
Lieu	66.56%	66.19%	66.38	75.09%	65.65%	70.06
Organisation	52.22%	55.18%	53.66	61.55%	46.91%	53.24
Person	59.68%	68.62%	63.84	75.32%	57.82%	65.42
Misc.	83.58%	60.74%	70.35	85.21%	67.46%	75.30
Total	62.67%	63.61%	63.14	73.81%	59.27%	65.75

Fig. 7. Cascade evaluation results.

According to these results, the unsupervised annotation system increases the precision score to 11.14% and the $F_{\beta=1}$ (where $F_\beta = \frac{(1+\beta^2)\cdot(\text{precision}\cdot\text{recall})}{\beta^2\cdot\text{precision}+\text{recall}}$) measure to 2.61. However, a regression of 4.4% has been observed for recall.

6 Conclusion, Discussion and Future Work

We have presented an unsupervised method for named entity annotation, which is based on formal concept analysis. This method exploits a concept lattice structuring relations between named entities and their related lexical units, observed in text corpora. We have assumed that formal concepts are relevant units for the disambiguation of named entities. The selection of a concept for an annotation results of a query to the lattice. In addition, we have proposed a method based on dimensionality reduction for the visualisation of formal concepts. We have adapted the cascade evaluation protocol to validate the choice of concepts for annotation. It shows that a supervised named entity classifier improves its precision when it relies on the conceptual annotation produced by our unsupervised FCA-based system. Even if, our system does not reach the performances obtained by the best named entity recognizers, the first results are encouraging since some improvements are possible.

The syntactic extraction process could be improved by using a dependency parser : this could help to cover more syntactic patterns. It could also provide some additional information such as normalised forms (*e.g.* {is, was, were} → to be) or typed syntactic relations (*e.g.* subject-object, head-modifier).

The cascade evaluation framework, could compare our approach to other supervised and unsupervised classifiers : we would be particularly interested in the comparison with other FCA based classifiers [13]. At the present time, we are

working on a semi-supervised lattice based classifier in which formal concepts are tagged with the NE labels (persons, locations, organisations, miscellaneous) available in the training corpus. Thus, the lattice would then be usable directly as a supervised NE classifier which would be able to produce unsupervised conceptual annotation with additional supervised labelling.

References

1. Pedersen, T.: 6. In: Unsupervised corpus-based methods for WSD. Springer (2006) 133–166
2. Ganter, B., Wille, R.: Formal Concept Analysis, Mathematical Foundations. Springer-Verlag (1999)
3. Grishman, R., Sundheim, B.: Message understanding conference 6: A brief history. In: COLING. (1996) 466–471
4. Sang, E.F.T.K., Meulder, F.D.: Introduction to the conll-2003 shared task: Language-independent named entity recognition. CoRR **cs.CL/0306050** (2003)
5. Nazarenko, A., Zweigenbaum, P., Bouaud, J., Habert, B.: Corpus-based identification and refinement of semantic classes (1997)
6. Cimiano, P., Hotho, A., Staab, S.: Learning concept hierarchies from text corpora using formal concept analysis. Journal of Artificial Intelligence Research **24** (2005) 305–339
7. van der Merwe, D., Obiedkov, S.A., Kourie, D.G.: Addintent: A new incremental algorithm for constructing concept lattices. [15] 372–385
8. Priss, U., Old, L.J.: Modelling lexical databases with formal concept analysis. J. UCS **10**(8) (2004) 967–984
9. Hérault, J., Jausions-Picaud, C., Guérin-Dugué, A.: Curvilinear component analysis for high-dimensional data representation: I. theoretical aspects and practical use in the presence of noise. In: IWANN (2). (1999) 625–634
10. Candillier, L., Tellier, I., Torre, F., Bousquet, O.: Cascade evaluation of clustering algorithms. In Fürnkranz, J., Scheffer, T., Spiliopoulou, M., eds.: 17th European Conference on Machine Learning (ECML'2006). Volume LNAI 4212 of LNCS., Berlin, Germany, Springer Verlag (september 2006) 574–581
11. Ehrmann, M., Jacquet, G.: Vers une double annotation des entités nommées. Traitement automatique des langues **47**(3) (2006) 63–88
12. Brill, E.: Transformation-based error-driven learning and natural language processing: A case study in part-of-speech tagging. Computational Linguistics **21**(4) (1995) 543–565
13. Kuznetsov, S.O.: Machine learning and formal concept analysis. [15] 287–312
14. Girault, T.: Exploitation de treillis de Galois en désambiguïsation non supervisée d'entités nommées. In: 15ème conférence sur le Traitement Automatique des Langues Naturelles (TALN'08). (2008) 260–269
15. Eklund, P.W., ed.: Concept Lattices, Second International Conference on Formal Concept Analysis, ICFCA 2004, Sydney, Australia, February 23-26, 2004, Proceedings. In Eklund, P.W., ed.: ICFCA. Volume 2961 of Lecture Notes in Computer Science., Springer (2004)