

Mining metalinguistic activity in corpora to create lexical resources using Information Extraction techniques: the MOP system

Carlos Rodríguez Penagos

Language Engineering Group, Engineering Institute
UNAM, Ciudad Universitaria A.P. 70-472
Coyoacán 04510 Mexico City, México
CRodriguezP@iingen.unam.mx

Abstract

This paper describes and evaluates MOP, an IE system for automatic extraction of metalinguistic information from technical and scientific documents. We claim that such a system can create special databases to bootstrap compilation and facilitate update of the huge and dynamically changing glossaries, knowledge bases and ontologies that are vital to modern-day research.

1 Introduction

Availability of large-scale corpora has made it possible to mine specific knowledge from free or semi-structured text, resulting in what many consider by now a reasonably mature NLP technology. Extensive research in Information Extraction (IE) techniques, especially with the series of Message Understanding Conferences of the nineties, has focused on tasks such as creating and updating databases of corporate joint ventures or terrorist and guerrilla attacks, while the ACQUILEX project used similar methods for creating lexical databases using the highly structured environment of machine-readable dictionary entries and other resources. Gathering knowledge from unstructured text often requires manually crafting knowledge-engineering rules both complex and deeply dependent of the domain at hand, although some successful experiences using learning algorithms have been reported (Fisher et al., 1995; Chieu et al., 2003).

Although mining specific semantic relations and subcategorization information from free-text has been successfully carried out in the past (Hearst, 1999; Manning, 1993), automatically extracting lexical resources (including terminological definitions) from text in special domains has

been a field less explored, but recent experiences (Klavans et al., 2001; Rodríguez, 2001; Cartier, 1998) show that compiling the extensive resources that modern scientific and technical disciplines need in order to manage the explosive growth of their knowledge, is both feasible and practical. A good example of this NLP-based processing need is the MedLine abstract database maintained by the National Library of Medicine¹ (NLM), which incorporates around 40,000 Health Sciences papers each month. Researchers depend on these electronic resources to keep abreast of their rapidly changing field. In order to maintain and update vital indexing references such as the Unified Medical Language System (UMLS) resources, the MeSH and SPECIALIST vocabularies, the NLM staff needs to review 400,000 highly-technical papers each year. Clearly, neology detection, terminological information update and other tasks can benefit from applications that automatically search text for information, e.g., when a new term is introduced or an existing one is modified due to data or theory-driven concerns, or, in general, when new information about sublanguage usage is being put forward. But the usefulness of robust NLP applications for special-domain text goes beyond glossary updates. The kind of categorization information implicit in many definitions can help improve anaphora resolution, semantic typing or acronym identification in these corpora, as well as enhance “semantic re-rendering” of special-domain ontologies and thesaurii (Pustejovsky et al., 2002).

In this paper we describe and evaluate the MOP² IE system, implemented to automatically create Metalinguistic Information Databases (MIDs) from large collections of special-domain

¹ <http://www.nlm.nih.gov/>

² Metalinguistic Operation Processor

research papers. Section 2 will lay out the theory, methodology and the empirical research grounding the application, while Section 3 will describe the first phase of the MOP tasks: accurate location of good candidate metalinguistic sentences for further processing. We experimented both with manually coded rules and with learning algorithms for this task. Section 4 focuses on the problem of identifying and organizing into a useful database structure the different linguistic constituents of the candidate predications, a phase similar to what are known in the IE literature as Named-Entity recognition, Element and Scenario template fill-up tasks. Finally, Section 5 discusses results and problems of our experiments, as well as future lines of research.

2 Metalanguage and term evolution in scientific disciplines

2.1 Explicit Metalinguistic Operations

Preliminary empirical work to explore how researchers modify the terminological framework of their highly complex conceptual systems, included manual review of a corpus of 19 sociology articles (138,183 words) published in various British, American and Canadian academic journals with strict peer-review policies. We look at how term manipulation was done as well as how metalinguistic activity was signaled in text, both by lexical and paralinguistic means. Some of the indicators found included verbs and verbal phrases like *called*, *known as*, *defined as*, *termed*, *coined*, *dubbed*, and descriptors such as *term* and *word*. Other non-lexical markers included quotation marks, apposition and text formatting.

A collection of potential metalinguistic patterns identified in the exploratory Sociology corpus was expanded (using other verbal tenses and forms) to 116 queries sent to the scientific and learned domains of the British National Corpus. The resulting 10,937 sentences were manually classified as metalinguistic or otherwise, with 5,407 (49.6% of total) found to be truly metalinguistic sentences. The presence of three components described below (*autonym*, *informative segment and markers/operators*) was the criteria for classification. Reliability of human subjects for this task has not been reported in the literature, and was not evaluated in our experiments.

Careful analysis of this extensive corpus presented some interesting facts about what we have termed “Explicit Metalinguistic Operations” (or EMOs) in specialized discourse:

A) EMOs usually do not follow the *genus-differentia* scheme of aristotelian definitions, nor conform to the rigid and artificial structure of dictionary entries. More often than not, specific information about language use and term definition is provided by sentences such as: (1) This means that they ingest oxygen from the air via fine hollow tubes, known as tracheae, in which the term *trachea* is linked to the description *fine hollow tubes* in the context of a globally non-metalinguistic sentence. Partial and heterogeneous information, rather than a complete definition, are much more common.

B) Introduction of metalinguistic information in discourse is highly regular, regardless of the specific domain. This can be credited to the fact that the writer needs to mark these sentences for special processing by the reader, as they dissect across two different semiotic levels: a metalanguage and its *object* language, to use the terminology of logic where these concepts originate.³ Its constitutive markedness means that most of the times these sentences will have at least two indicators present, for example a verb and a descriptor, or quotation marks, or even have preceding sentences that announce them in some way. These formal and cognitive properties of EMOs facilitate the task of locating them accurately in text.

C) EMOs can be further analyzed into 3 distinct components, each with its own properties and linguistic realizations:

- i) An **autonym** (*see note 3*): One or more self-referential lexical items that are the logical or grammatical subject of a predication that needs not be a complete grammatical sentence.

³ At a very basic semiotic level natural language has to be split (at least methodologically) into two distinct systems that share the same rules and elements: a metalanguage, which is a language that is used to talk about another one, and an object language, which in turn can refer to and describe objects in the mind or in the physical world. The two are isomorphic and this accounts for reflexivity, the property of referring to itself, as when linguistic items are mentioned instead of being used normally in an utterance. Rey-Debove (1978) and Carnap (1934) call this condition *autonymy*.

- ii) An **informative segment**: a contribution of relevant information about the meaning, status, coding or interpretation of a linguistic unit. Informative segments constitute what we state about the autonymical element.
- iii) **Markers/Operators**: Elements used to mark or made prominent whole discourse operation, on account of its non-referential, metalinguistic nature. They are usually lexical, typographic or pragmatic elements that articulate autonyms and informative segments into a predication.

Thus, in a sentence such as (2), the [autonym] is marked in square brackets, the {informational segment} in curly brackets and the <marker-operators> in angular brackets:

- (2) {The bit sequences representing quanta of knowledge} <will be called “>[Kenes]<”>, {a neologism intentionally similar to 'genes'}.

2.2 Defaults, knowledge and knowledge of language

The 5,400 metalinguistic sentences from our BNC-based test corpus (henceforth, the EMO corpus) reflect an important aspect of scientific sublanguages, and of the scientific enterprise in general. Whenever scientists and scholars advance the state of the art of a discipline, the language they use has to evolve and change, and this build-up is carried out under metalinguistic control. Previous knowledge is transformed into new scientific common ground and ontological commitments are introduced and defended when semantic reference is established. That is why when we want to structure and acquire new knowledge we have to go through a resource-costly cognitive process that integrates, within coherent conceptual structures, a considerable amount of new and very complex lexical items and terms.

It has to be pointed out that non-specialized language is not abundant⁴ in these kinds of metalinguistic exchanges because (unless in the context of language acquisition) we usually rely on a lexical competence that, although subsequently modified and enhanced, reaches the plateau of a generalized lexicon relatively early in our adult life. Technical terms can be thought of as semantic anomalies, in the sense that they are *ad hoc*

⁴ Our study shows that they represent between 1 and 6% of all sentences across different domains.

constructs strongly bounded to a model, a domain or a context, and are not, by definition, part of the far larger linguistic competence from a first native language. The information provided by EMOs is not usually inferable from previous one available to the speaker's community or expert group, and does not depend on general language competence by itself, but nevertheless is judged important and relevant enough to warrant the additional processing effort involved.

Conventional resources like lexicons and dictionaries compile established meaning definitions. They can be seen as repositories of the default, core lexical information of words or terms used by a community (that is, the information available to an average, idealized speaker). A Metalinguistic Information Database (MID), on the other hand, compiles the real-time data provided by metalinguistic analysis of leading-edge research papers, and can be conceptualized as an anti-dictionary: a listing of exceptions, special contexts and specific usage, of instances where meaning, value or pragmatic conditions have been spotlighted by discourse for cognitive reasons. The non-default and highly relevant information from MIDs could provide the material for new interpretation rules in reasoning applications, when inferences won't succeed because the states of the lexico-conceptual system have changed. When interpreting text, regular lexical information is applied by default under normal conditions, but more specific pragmatic or discursive information can override it if necessary, or if context demands so (Lascari-des & Copestake, 1995). A neologism or a word in an unexpected technical sense could stump a NLP system that assumes it will be able to use default information from a machine-readable dictionary.

3 Locating metalinguistic information in text: two approaches

When implementing an IE application to mine metalinguistic information from text, the first issue to tackle is how to obtain a reliable set of candidate sentences from free text for input into the next phases of extraction. From our initial corpus analysis we selected 44 patterns that showed the best reliability for being EMO indicators. We start our processing⁵ by tokenizing text, which then is

⁵ Our implementation is Python-based, using the

run through a cascade of finite-state devices based on identification patterns that extract a candidate set for filtering. Our filtering strategies in effect distinguish between useful results such as (3) from non-metalinguistic instances like (4):

(3) Since the shame that was elicited by the coding procedure was seldom explicitly mentioned by the patient or the therapist, Lewis **called** it unacknowledged shame.

(4) It was Lewis (1971;1976) who **called** attention to emotional elements in what until then had been construed as a perceptual phenomenon .

For this task, we experimented with two strategies: First, we used corpus-based collocations to discard non-metalinguistic instances, for example the presence of *attention* in sentence (4) next to the marker *called*. Since immediate co-text seems important for this classification task, we also implemented learning algorithms that were trained on a subset from our EMO corpus, using as vectors either POS tags or word forms, at 1, 2, and 3 positions adjacent before and after our markers. These approaches are representative of wider paradigmatic approaches to NLP: symbolic and statistical techniques, each with their own advantages and limitations. Our evaluations of the MOP system are based on test runs over 3 document sets: a) our original exploratory corpus of sociology research papers [5581 sentences, 243 EMOs]; b) an online histology textbook [5146 sentences, 69 EMOs] ; and c) a small sample from the MedLine abstract database [1403 sentences, 10 EMOs].

Using collocational information, our first approach fared very well, presenting good precision numbers, but not so encouraging recall. The sociology corpus, for example, gave 0.94 precision (P) and 0.68 recall (R), while the histology one presented 0.9 P and 0.5 R. These low recall numbers reflect the fact that we only selected a subset of the most reliable and common metalinguistic patterns, and our list is not exhaustive. Example (5) shows one kind of metalinguistic sentence (with a copulative structure) attested in corpora,

but that the system does not attempt to extract or process:

(5) “Intercursive” power , on the other hand , is power in Weber's sense of constraint by an actor or group of actors over others.

In order to better compare our two strategies, we decided to also zoom in on a more limited subset of verb forms for extraction (namely, *calls*, *called*, *call*), which presented ratios of metalinguistic relevance in our MOP corpus, ranging from 100% positives (for the pattern *so called* + *quotation marks*) to 77% (*called*, by itself) to 31% (*call*). Restricted to these verbs, our metrics show precision and recall rates of around 0.97, and an overall F-measure of 0.97.⁶ Of 5581 sentences (96 of which were metalinguistic sentences signaled by our cluster of verbs), 83 were extracted, with 13 (or 15.6% of candidates) filtered-out by collocations.

For our learning experiments (an approach we have called contextual feature language models), we selected two well-known algorithms that showed promise for this classification task.⁷ The naive Bayes (NB) algorithm estimates the conditional probability of a set of features given a label, using the product of the probabilities of the individual features given that label. The Maximum Entropy model establishes a probability distribution that favors entropy, or uniformity, subject to the constraints encoded in the feature-label correlation. When training our ME classifiers, Generalized (GISMax) and Improved Iterative Scaling (IIS-Max) algorithms are used to estimate the optimal maximum entropy of a feature set, given a corpus.

1,371 training sentences were converted into labeled vectors, for example using 3 positions and POS tags: ('VB WP NNP', 'calls', 'DT NN NN') /'YES'@[102]. The different number of positions considered to the left and right of the markers in our training corpus, as well as the nature of the features selected (there are many more word-types than POS tags) ensured that our 3-part vector introduced a wide range of features against our 2 possible YES-NO labels for processing by our algorithms. Although our test runs using only collocations showed initially that structural regulari-

NLTK toolkit (nltk.sf.net) developed by E. Loper and S. Byrd at the University of Pennsylvania, although we have replaced stochastic POS taggers with an implementation of the Brill algorithm by Hugo Liu at MIT. Our output files follow XML standards to ensure transparency, portability and accessibility

⁶ With a β factor of 1.0, and within the sociology document set

⁷ see Ratnaparkhi (1997) and Berger et al. (1996) for a formal description of these algorithms

ties would perform well, both with our restricted lemma cluster and with our wider set of verbs and markers, our intuitions about improvement with more features (more positions to the right of left of the markers) or a more controlled and grammatically restricted environment (a finite set of surrounding POS tags), turned out to be overly optimistic. Nevertheless, stochastic approaches that used short range features did perform very well, in line with the hand-coded approach.

The results of the different algorithms, restricted to the lexeme *call*, are presented in Table 1, while Figures 1 and 2 present best results in the learning experiments for the complete set of patterns used in the collocation approach, over two of our evaluation corpora.

Type	Positions	Tags/ Words	Features	Accuracy	Precision	Recall
GISMax	1	W	1254	0.97	0.96	0.98
IISMax	1	T	136	0.95	0.96	0.94
IISMax	1	W	1252	0.92	0.97	0.9
GISMax	1	T	138	0.91	0.9	0.96
GISMax	2	T	796	0.88	0.93	0.92
IISMax	2	T	794	0.86	0.95	0.89
IISMax	3	W	4290	0.87	0.85	0.98
GISMax	3	W	4292	0.87	0.85	0.98
IISMax	2	W	3186	0.86	0.87	0.95
GISMax	2	W	3188	0.86	0.87	0.95
NB	1	T	136	0.88	0.97	0.84
NB	2	T	794	0.87	0.96	0.84
NB	3	W	4290	0.73	0.86	0.77

Table 1. Best metrics for “call” lexeme sorted by F-measure and classifier accuracy

Figure 1. Best metrics for Sociology corpus

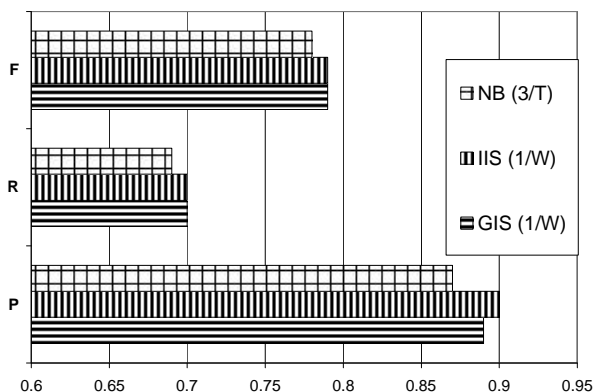
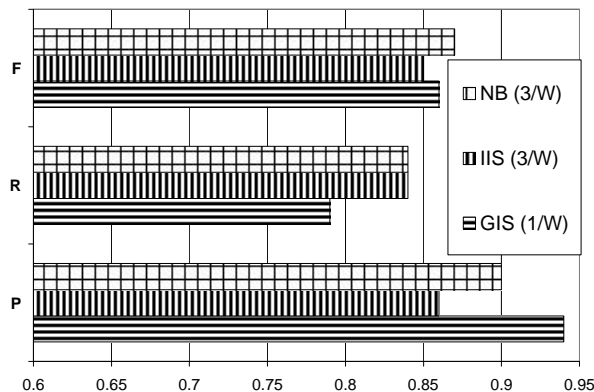


Figure 2. Best metrics for Histology corpus



Figures 1 & 2. Best results for filtering algorithms.⁸

Both Knowledge-Engineering and supervised learning approaches can be adequate for extraction of metalinguistic sentences, although learning algorithms can be helpful when procedural rules have not been compiled; they also allow easier transport of systems to new thematic domains. We plan further research into stochastic approaches to fine tune them for the task.

One issue that merits special attention is why some of the algorithms and features work well with one corpus, but not so well with another. This fact is in line with observations in Nigam et al. (1999) that naive Bayes and Maximum Entropy do not show fundamental baseline superiorities, but are dependent on other factors. A hybrid approach that combines hand-crafted collocations with classifiers customized to each pattern’s behavior and morpho-syntactic contexts in corpora might offer better results in future experiments.

4 Processing EMOs to compile metalinguistic information databases

Once we have extracted candidate EMOs, the MOP system conforms to a general processing architecture shown in Figure 3. POS tagging is followed by shallow parsing that attempts limited PP-attachment. The resulting chunks are then tagged semantically as Autonyms, Agents, Markers, Anaphoric elements or simply as Noun Chunks,

⁸ Legend: P: Precision; R: Recall; F: F-Measure. NB: naive Bayes; IIS: Maximum Entropy trained with Improved Iterative Scaling; GIS: Maximum Entropy trained with Generalized Iterative Scaling. (Positions/Feature type)

using heuristics based on syntactic, pragmatic and argument structure observation of the extraction patterns.

Next, a predicate processing phase selects the most likely surface realization of informational segments, autonyms and makers-operators, and proceeds to fill the templates in our databases. This was done by following different processing routes customized for each pattern using corpus analysis as well as FrameNet data from *Name conferral* and *Name bearing* frames to establish relevant arguments and linguistic realizations.

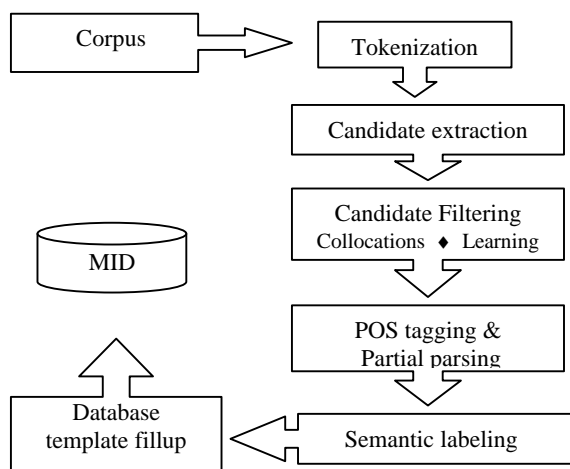


Figure 3. MOP Architecture

As mentioned earlier, informational segments present many realizations that distance them from the clarity, completeness and conciseness of lexicographic entries. In fact, they may show up as full-fledged clauses (6), as inter- or intra-sentential anaphoric elements (7 and 8, the first one a relative clause), supply a categorization descriptor (9), or even (10) restrict themselves semantically to what we could call a sententially-unrealized “existential variable” (with logical form $\exists x$) indicating only that certain discourse entity is being introduced.

- (6) In 1965 the term soliton was coined *to describe waves with this remarkable behaviour*.
- (7) This leap brings cultural citizenship in line with *what* has been called the politics of citizenship .
- (8) *They* are called “endothermic compounds.”
- (9) One of the most enduring aspects of all social

theories are *those conceptual entities* known as structures or groups.

- (10) A $\exists x$ so called cell-type-specific TF can be used by closely related cells, e.g., in erythrocytes and megakaryocytes.

We have not included an anaphora-resolution module in our present system, so that instances 7, 8 and 10 will only display in the output as unresolved surface element or as existential variable place-holders,⁹ but these issues will be explored in future versions of the system. Nevertheless, much more common occurrences as in (11) and (12) are enough to create MIDs quite useful for lexicographers and for NLP lexical resources.

- (11) The Jovian *magnetic field exerts an influence out to near a surface*, called the “magnetopause”.

- (12) Here we report the discovery of *a soluble decoy receptor*, termed decoy receptor 3 (DcR3)...

The correct database entry for example 12 is presented in Table 4.

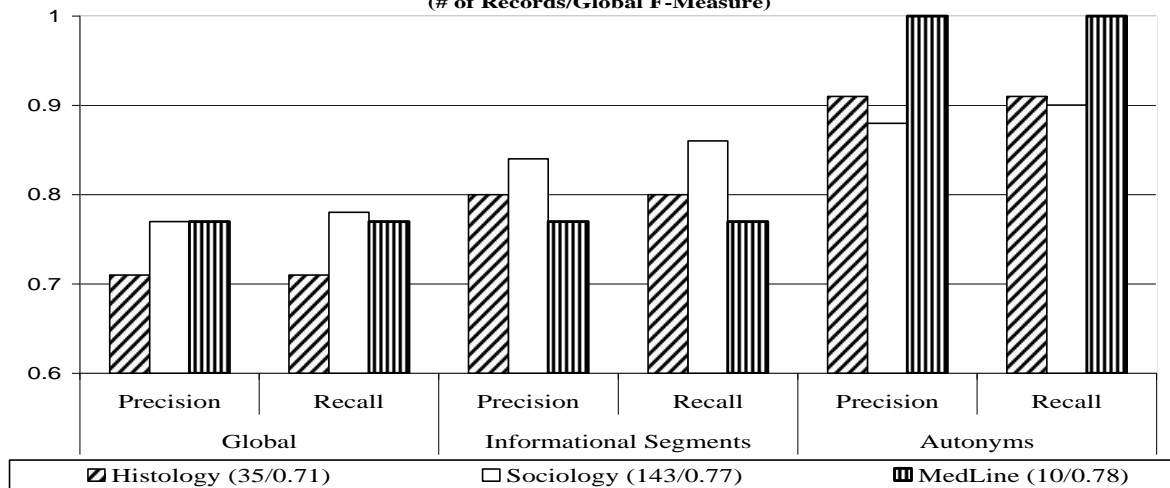
Reference:	MedLine sample # 6
Autonym:	decoy receptor 3 (DcR3)
Information	a soluble decoy receptor
Markers/ Operators:	termed

Table 4. Sample entry of MID

The final processing stage presents metrics shown in Figure 4, using a β factor of 1.0 to estimate F-measures. To better reflect overall performance in all template slots, we introduced a threshold of similarity of 65% for comparison between a golden standard slot entry and the one provided by the application. Thus, if the autonym or the informational segment is at least 2/3 of the correct response, it is counted as a positive, in many cases leveling the field for the expected errors in the prepositional phrase- or acronym-attachment algorithms, but accounting for a (basically) correct selection of superficial sentence segments.

⁹ For sentence (8) the system would retrieve a previous sentence: (“A few have *positive enthalpies of formation*”). to define “endothermic compounds”.

Figure 4. Metrics for 3 corpora
(# of Records/Global F-Measure)



5 Results, comparisons and discussion

The DEFINDER system (Klavans et al, 2001) at Columbia University is, to my knowledge, the only one fully comparable with MOP, both in scope and goals, but some basic differences between them exist. First, DEFINDER examines user-oriented documents that are bound to contain fully-developed definitions for the layman, as the general goal of the PERSIVAL project is to present medical information to patients in a less technical language than the one of reference literature. MOP focuses on leading-edge research papers that present the less predictable informational templates of highly technical language. Secondly, by the very nature of DEFINDER's goals their qualitative evaluation criteria include readability, usefulness and completeness as judged by lay subjects, criteria which we have not adopted here. Neither have we determined coverage against existing online dictionaries, as they have done. Taking into account the above-mentioned differences between the two systems' methods and goals, MOP compares well with the 0.8 Precision and 0.75 Recall of DEFINDER. While the resulting MOP "definitions" generally do not present high readability or completeness, these informational segments are not meant to be read by laymen, but used by domain lexicographers reviewing existing glossaries for neological change, or, for example, in machine-readable form by applications that attempt automatic categorization for semantic re-rendering of an expert ontology, since definitional contexts provide sortal information as a natural part of the

process of precisely situating a term or concept against the meaning network of interrelated lexical items. The Metalinguistic Information Databases in their present form are not, in full justice, lexical knowledge bases comparable with the highly-structured and sophisticated resources that use inheritance and typed features, like LKB (Copestake et al., 1993). MIDs are semi-structured resources (midway between raw corpora and structured lexical bases) that can be further processed to convert them into usable data sources, along the lines suggested by Vossen and Copestake (1993) for the syntactic kernels of lexicographic definitions, or by Pustejovsky et al. (2002) using corpus analytics to increase the semantic type coverage of the NLM UMLS ontology. Another interesting possibility is to use a dynamically-updated MID to trace the conceptual and terminological evolution of a discipline.

We believe that low recall rates in our tests are in part due to the fact that we are dealing with the wider realm of metalinguistic information, as opposed to structured definitional sentences that have been distilled by an expert for consumer-oriented documents. We have opted in favor of exploiting less standardized, non-default metalinguistic information that is being put forward in text because it can't be assumed to be part of the collective expert-domain competence (Section 2.1). In doing so, we have exposed our system to the less predictable and highly charged lexical environment of leading-edge research literature, the cauldron where knowledge and terminological systems are forged in real time, and where scienti-

fic meaning and interpretation are constantly debated, modified and agreed. We have not performed major customization of the system (like enriching the tagging lexicon with medical terms), in order to preserve the ability to use the system across different domains. Domain customization may improve metrics, but at a cost for portability.

The implementation we have described here undoubtedly shows room for improvement in some areas, including: adding other patterns for better overall recall rates, deeper parsing for more accurate semantic typing of sentence arguments, etc. Also, the issue of which learning algorithms can better perform the initial filtering of EMO candidates is still very much an open question. Applications that can turn MIDs into truly useful lexical resources by further processing them need to be written. We plan to continue development of our proof-of-concept system to explore those areas. DEFINDER and MOP both show great potential as robust lexical acquisition systems capable of handling the vast electronic resources available today to researchers and laymen alike, helping to make them more accessible and useful. In doing so, they are also fulfilling the promise of NLP techniques as mature and practical technologies.

References

- ACQUILEX projects*, final report available at: <http://www.cl.cam.ac.uk/Research/NL/acquilex/>
- Berger, A., S. Della Pietra et al., 1996. *A Maximum Entropy Approach to Natural Language Processing*. Computational Linguistics, vol. 22, no. 1.
- Carnap, R. 1934. *The Logical Syntax of Language*. Routledge and Kegan, Londres 1964.
- Cartier, E. 1998. *Analyse Automatique des textes: l'exemple des informations définitives*. RIFRA 1998. Sfax, Tunisia.
- Chieu, Hai Leong, Ng, Hwee Tou, & Lee, Yoong Keok. 2003. *Closing the Gap: Learning-Based Information Extraction Rivaling Knowledge-Engineering Methods*. 41st ACL. Sapporo, Japan.
- Copestake, A., Sanfilippo, A., Briscoe, T. and de Pavia, V. 1993. *The ACQUILEX LKB: An introduction*. In: Inheritance, Defaults and the Lexicon. Cambridge University Press.
- Fisher, D., S. Soderland, J. McCarthy, F. Feng, and W. Lehnert. 1995. *Description of the UMass system as used for MUC-6*. In Proceedings of MUC-6
- Hearst, M. 1998. *Automated discovery of wordnet relations*. In Christiane Fellbaum, editor, WordNet: An Electronic Lexical Database. MIT Press, Cambridge, MA
- Klavans, J. and S. Muresan. 2001. *Evaluation of the DEFINDER System for Fully Automatic Glossary Construction*, proceedings of the American Medical Informatics Association Symposium 2001
- Lascarides, A. and Copestake A. 1995. *The Pragmatics of Word Meaning*, Proceedings of the AAAI Spring Symposium Series: Representation and Acquisition of Lexical Knowledge: Polysemy, Ambiguity and Generativity, Stanford CA.
- Manning, Ch. 1993. *Automatic acquisition of a large subcategorization dictionary from corpora*, In Proceedings of the 31st ACL, Columbus, OH.
- Nigam, K., Lafferty, J., and McCallum, A. 1999. *Using Maximum Entropy for Text Classification*, IJCAI-99 Workshop on Machine Learning for Information Filtering, pp. 61-67
- Pustejovsky J., A. Rumshisky and J. Castaño. 2002. *Rerendering Semantic Ontologies: Automatic Extensions to UMLS through Corpus Analytics*. LREC 2002 Workshop on Ontologies and Lexical Knowledge Bases. Las Palmas, Canary Islands, Spain.
- Ratnaparkhi A. 1997. *A Simple Introduction to Maximum Entropy Models for Natural Language Processing*, TR 97-08, Institute for Research in Cognitive Science, University of Pennsylvania
- Rey-Debove, J. 1978. *Le Métalangage*. Le Robert, Paris.
- Rodríguez, C. 2001. *Parsing Metalinguistic Knowledge from Texts*, Selected papers from CICLING-2000 Collection in Computer Science (CCC); National Polytechnic Institute (IPN), Mexico.
- Vossen, P. and Copestake, A. 1993. *Untangling Definition Structure into Knowledge Representation*. In: Inheritance, Defaults and the Lexicon.