UNCLASSIFIED

10TH INTERNATIONAL COMMAND AND CONTROL RESEARCH AND TECHNOLOGY SYMPOSIUM
THE FUTURE OF C2

CAPTURING AND MODELING DOMAIN KNOWLEDGE USING NATURAL LANGUAGE PROCESSING TECHNIQUES

Topic: Decisionmaking and Cognitive Analysis

Dr Alain Auger
DRDC Valcartier
2459 Pie-XI Blvd North
Val-Bélair, QC
G3J 1X5 CANADA
1 418 844-4000 x4821 / 1 418 844-4538
Alain.Auger@drdc-rddc.gc.ca

CAPTURING AND MODELING DOMAIN KNOWLEDGE USING NATURAL LANGUAGE PROCESSING TECHNIQUES

Alain Auger Ph. D.
Defence Research and Development Canada – Valcartier
2459, Pie-XI Blvd North, Val-Bélair
Quebec, CANADA G3J 1X5
Alain.Auger@drdc-rddc.gc.ca

ABSTRACT

Command and control (C2) and the decisionmaking domain are seriously threatened facing information overload and uncertainty issues. To make sense out of the flood of information, military have to create new ways of processing sensor and intelligence information, and of providing the results to commanders. Initiated in 2004 at Defence Research and Development Canada (DRDC), the SACOT[1] knowledge engineering research project is currently investigating, developing and validating innovative natural language processing (NLP) approaches as scientific means to capture knowledge objects contained in domain-specific electronic texts and turn them rapidly into broad domain ontologies to be used in third-party applications. Ontologies are key elements required to enable next generation of decision support and knowledge exploitation systems with new semantic capabilities. Major impediments to classic development of ontologies are that it is a time and budget consuming operation. It is also largely dependant of Subject Matter Experts' (SME) own limitations. Exhaustive elicitation of knowledge objects of a domain requires the application of NLP extraction techniques over textual data. This paper illustrates how recent advances in NLP techniques are implemented in the SACOT framework to automate elicitation of knowledge objects from unstructured texts and to support efficiently SMEs in ontology engineering tasks.

1. INTRODUCTION

Command and control (C2) and the decisionmaking domain are seriously threatened facing information overload and uncertainty issues. To make sense out of the flood of information, military have to create new ways of processing sensor and intelligence information, and of providing the results to commanders who must take timely operational decisions. Research in the field of Information and Knowledge Management (IKM) consists in investigating and advancing knowledge creation and discovery techniques through which information is collected and processed to support situation analysis and gain sufficient situational awareness to be able to project possible future courses of action or trends with confidence. In 2001, the Canadian Forces Future Army Capabilities report [DND, 2001] pointed out that "without some fundamental change, current army ISR[2] will be incapable of providing the degree of knowledge that will be required by future commanders." Therefore "all relevant data, information and knowledge must be available at all levels, but managed in a way that produces a current, rapid and coherent understanding of the battlespace, while at the same time allowing the various levels of command to process the relevant material for their specific purposes."

[1] SACOT: Semi-Automatic Construction of Ontologies from Texts
[2] Intelligence, Surveillance, Reconnaissance (ISR)

Ontologies are key elements required to enable decision support systems, knowledge exploitation and information retrieval systems with new semantic capabilities. Since Gruber [Gruber, 1993], the scientific community defines an ontology as a formal, explicit specification of a shared conceptualization. When domain knowledge is represented in a declarative formalism, such as in an ontology, the set of objects that can be represented is called the universe of discourse. This set of objects, and the formalized relationships among them, are reflected in the representational *vocabulary* [id.]. Domain ontologies provide vocabularies about the concepts within a domain and their relationships, about activities that take place in that domain and about theories and elementary principles governing that domain [Corcho *et al.*, 2003]. This paper illustrates how natural language processing techniques can support and automate domain ontologies engineering.

2. ONTOLOGIES FOR INTELLIGENT COMMAND AND CONTROL SYSTEMS

Research from the military community[3] clearly indicates that there are many needs and many potential uses for domain ontologies within many military areas similar to industry. With the development and maturity of the Semantic Web [Davies *et al.*, 2003], automated ontology engineering will provide the cornerstone technology, which shares a common understanding of a domain among humans, agents and machines. Ontologies for command and control systems will be instrumental in establishing a Common Operational Picture (COP) among units by making domain representations, situation analysis and assumptions more explicit. Agents assisting commanders with the command and control task will have the ability to "interpret" data and know its meaning and value based on domain ontologies. According to [Bowman *et al.*, 2001], in order for Artificial Intelligence (AI) to become truly useful in high-level military applications it is necessary to identify, document, and integrate into automated systems the human knowledge that senior military professionals use to solve high-level problems. This paper [*ibid.*] illustrates this statement by the development and use of a course of action ontology. If it is generally admitted that next generation of command and control systems shall integrate and use ontologies, existing technologies and methodologies to rapidly build such ontologies still remain very limited.

3. NATURAL LANGUAGE PROCESSING FOR ONTOLOGY ENGINEERING

Since knowledge objects of a given domain are expressed and conveyed in texts using domain-specific terminology, it is reasonable to think that mining and extracting this terminology will lead us to a certain domain representation model. Problem is how to reach high quality automated extraction of those knowledge objects in order to build reliable ontologies with them?

Initiated in 2004 at Defence Research and Development Canada (DRDC), the SACOT[4] knowledge engineering research project is currently investigating, developing and validating innovative natural language processing (NLP) approaches as scientific means to capture knowledge objects contained in open source electronic texts and turn them rapidly into broad domain ontologies to be used in third-party applications.

[3] See for instance [Bourry-Brisset, 2000; Chance & Hagenston, 2003; Gauvin *et al.*, 2004, Gouin *et al.*, 2003, Dorion & Bourry-Brisset, 2004, Bowman *et al.*, 2001]
[4] Semi-Automatic Construction of Ontologies from Texts (SACOT)

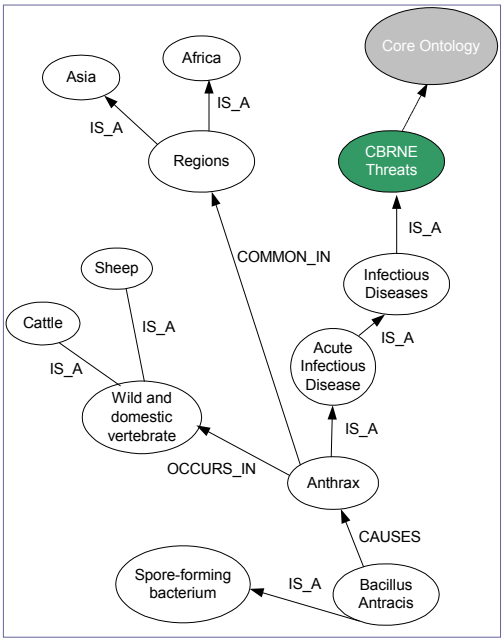

Fig. 1: Partial Draft Ontology of Infectious Diseases

As two of the core components of domain ontologies are **concepts** and **relations** among concepts, the SACOT project encompasses several NLP research areas. Identification and extraction of concepts contained in texts is supported by innovative terminology extraction techniques. Semantic relations existing among concepts are identified and extracted using other sets of natural language processing techniques. Using the two core components extracted from the electronic texts (concepts and semantic relations) and other reference material, draft ontologies are automatically compiled and generated.

Knowledge engineers can use this automated ontology-engineering environment as a knowledge framework in order to validate and enhance the draft ontologies. While validating the content of the draft ontologies, knowledge engineers will teach the system about which among all potential semantic relations identified in texts are most valuable and which are not relevant.

Essentially, domain ontologies are made of sets of concepts (classes) and the relationships or properties that can be expressed among those concepts. Figure 1 shows a partial draft ontology of infectious diseases generated from a local semantic network obtained from parsing a sample input text.

Since the building blocks of domain ontologies are concepts (e.g. ANTHRAX, BACILLUS ANTRACIS, SPORE-FORMING BACTERIUM, in Fig. 1) and relations among concepts (e.g. CAUSES, IS_A), three NLP techniques are being investigated in the SACOT framework to capture those elements: **terminology** extraction techniques, **named entities** extraction techniques and **semantic relations** extraction techniques. Those three extraction techniques will be presented in sections 5.2 to 5.4.

4. ONTOLOGY ENGINEERING METHODOLOGIES

Most of published ontology engineering methods[5] require interviews with Subject Matter Experts (SMEs) to elicit knowledge objects of a domain. In all the approaches relying heavily on SMEs, the extent of the domain represented in the ontology depends on the expertise and the degree of "expressiveness" of the available SMEs. This limitation might lead to unacceptable and poor performance of ontology-based information systems. Typically, domain terminology can contain from few hundreds (e.g. Professional Golfers' Association (PGA) Glossary of Golf) to several hundreds of thousands terms (e.g. up to 160,000 terms in a medical dictionary). It is unlikely that any SME interview will ever elicit the whole terminology of a domain. We need to turn to more exhaustive and objective data sources. The major impediments to classic development of ontologies are that it is a time and budget consuming operation and that it is largely dependant of SMEs' own knowledge limitations. Exhaustive elicitation of knowledge objects of a domain requires the application of NLP extraction techniques over textual data.

[5] [Corcho *et al.* 2003; Gómez-Pérez 1999; Gómez-Pérez *et al.* 2004; Sure 2003; Uschold and Grüninger 1996; Grüninger and Fox 1995]

5. SACOT ONTOLOGY ENGINEERING FRAMEWORK

5.1 The Overall Process

As mentioned, in traditional ontology engineering methodologies, SMEs are being interviewed at the beginning of the process to elicit knowledge objects. Methodology developed for the SACOT framework also includes early interviews with SMEs to identify domain-specific material (electronic texts, electronic dictionaries, if any, etc.). The SME is also playing the role of a knowledge engineer, being presented draft ontologies for validation. He also contributes to the maintenance of the ontology. Figure 2 below illustrates the overall ontology engineering process in SACOT.

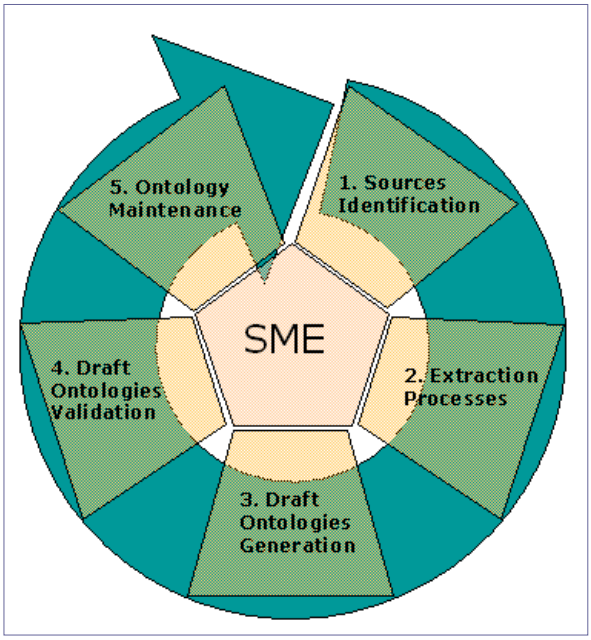

Fig. 2: SACOT's Ontology Engineering Process

1. **Sources Identification**. First step consists in gathering and formating all domain-specific information sources. SMEs are consulted to provide knowledge engineers with reference material that represents consensual knowledge sources among the SMEs community.
2. **Extraction Processes**. Domain-specific electronic texts are processed in three different extraction modules to identify knowledge objects. Next sections (5.2 – 5.4) describe those extraction processes.
3. **Draft Ontologies Generation**. Knowledge objects extracted during previous phase are then compiled into a draft ontology. Reference material such as core ontologies, lexical ontologies (e.g. WordNet), and domain-specific electronic dictionaries or thesaurus, if any, is used to guide the draft ontology generation process.
4. **Draft Ontologies Validation**. During this phase, SMEs are required to validate the content of the draft ontologies. An agent monitors the validation work. Rules are derived from the human-based validation work so that the system can learn from the validation process and prune future draft ontologies according to stored validation rules.
5. **Ontology Maintenance.** Finally, knowledge engineers use ontology management tools to manage versionning of the domain ontology which, in turn, is reused as reference material during the next extraction cycle.

5.2 Terminology Extraction

Terms are linguistic representations of concepts. Basically, terminology extraction is the process by which raw terminological units corresponding to specific morph-syntactic patterns are extracted from electronic texts[6]. Those extracted terminological units are considered as candidate terms and need further validation to determine whether they belong to a specific domain or are simply general vocabulary.

Nowadays, one of the most challenging problems in terminology extraction is the automation of the validation process by which raw candidate terms can be automatically assigned to specific domain terminology. Terminology extraction tools and techniques tend to generate huge amount of candidate

[6] Details on recent terminology extraction techniques can be found in [Jacquemin 2001] and [Bourigault *et al.* 2001].

terms requiring human validation. To be fully effective, validation of candidate terms needs to be automated. Otherwise, original *information overload* issues will simply be replaced by *candidate terms overload* ones.

Recent advances in computational terminology suggest the use of contrastive datasets and statistics as means to validate candidate terms [Drouin 2003, 2004]. Using this approach, candidate terms are extracted from two different domain-specific corpora. Resulting lists of candidates, together with their respective frequency ratio, are then compared. If the same candidate can be found in both lists with similar frequency ratios, the probability that it is not a domain-specific term is very high. When a candidate term can be found in both lists, statistical comparison of the frequencies observed in the two corpora is computed in order to elicit domain specific terminology.

Frequence	Term	Score
6619	terrorist	101,99
4209	terrorism	92,80
4587	nuclear	83,01
3018	biological	78,67
2520	weapon	68,01
1895	Iraq	61,35
2107	attack	57,79
1885	domestic	55,80
1200	department	47,57
1125	al	47,18
2266	military	46,97
1527	September	46,59
1048	Iraqi	46,23

Table. 1: Sample List of Terrorism Domain Candidate Terms

Table 1 shows a partial list of terms extracted using contrastive corpora. Scores quantify the observed deviation from a normal distribution. These deviations indicate that, considering the two corpora used to establish comparison, terms are statistically more related to the terrorism-related corpus than to the other corpus used. This is quite obvious with terms such as *nuclear*, *biological*, *weapon*.

SACOT's automatic terminology extraction and validation processes exploit contrastive datasets and implement approach proposed by [Drouin 2003, 2004].

5.3 Named Entities Extraction

Named Entities (NE) represent another important set of knowledge objects to be captured in texts. The following table introduce standard named entity categories that have been defined during the Message Understanding Conference (MUC-7) [Chinchor, 1997].

Entity	Description
ORGANIZATION	Named corporate, governmental, or other organizational entity
PERSON	Named person or family
LOCATION	Name of politically or geographically defined location (cities, provinces, countries, international regions, bodies of water, mountains, etc.)
DATE	Complete or partial date expression
TIME	Complete or partial expression of time of day
MONEY	Monetary expression
PERCENT	Percentage

Table 2: Standard Named Entities [Chinchor, 1997]

Named entities can also include street addresses, Uniform Resource Locator (URL), email addresses, symbols, and measures. Extending the concept of named entity itself, named entities categories can be considered as classes and corresponding retrieved information elements as instances or individual representations of those concepts. For instance, each different street address found in a text represents a different instance of the named entity category called STREET ADDRESS. From there, named entities themselves can be formalized using an ontology.

When it comes to domain-specific named entities extraction, standard categories proposed at the MUC-7 conference appear to be too generic. In the SACOT framework, the NE extraction module exploits the GATE[7] open source software. New named entities annotation schema have been defined and new grammar rules have been written and tested at DRDC Valcartier to handle morph-syntactic patterns specific to terrorism and to weapons of mass destruction (WMD) domains. New named entities classes such as TERRORISM_WEAPON have been defined as well. The two following figures show a list of ontology classes (Fig. 3) and how their corresponding named entities annotation schema is used to retrieve terrorism-related information in unstructured texts (Fig. 4).

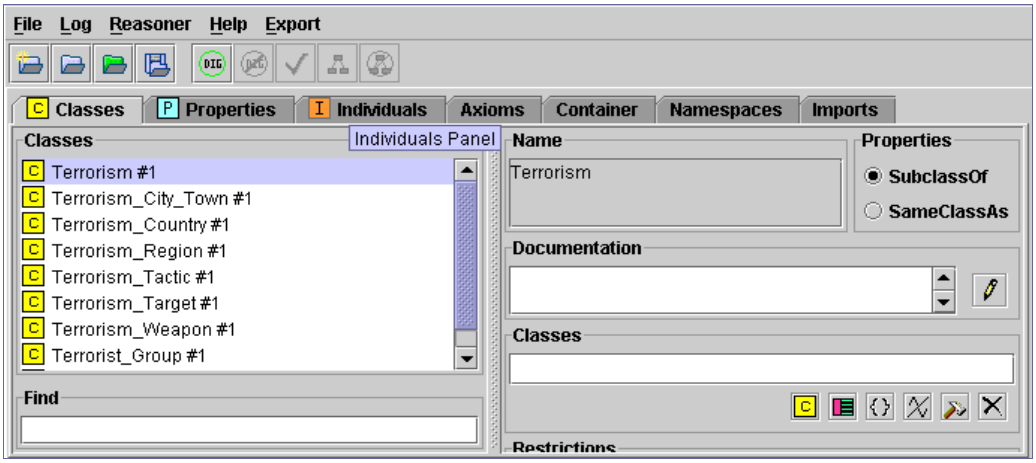

Fig. 3: Partial Terrorism-related Named Entities Ontology

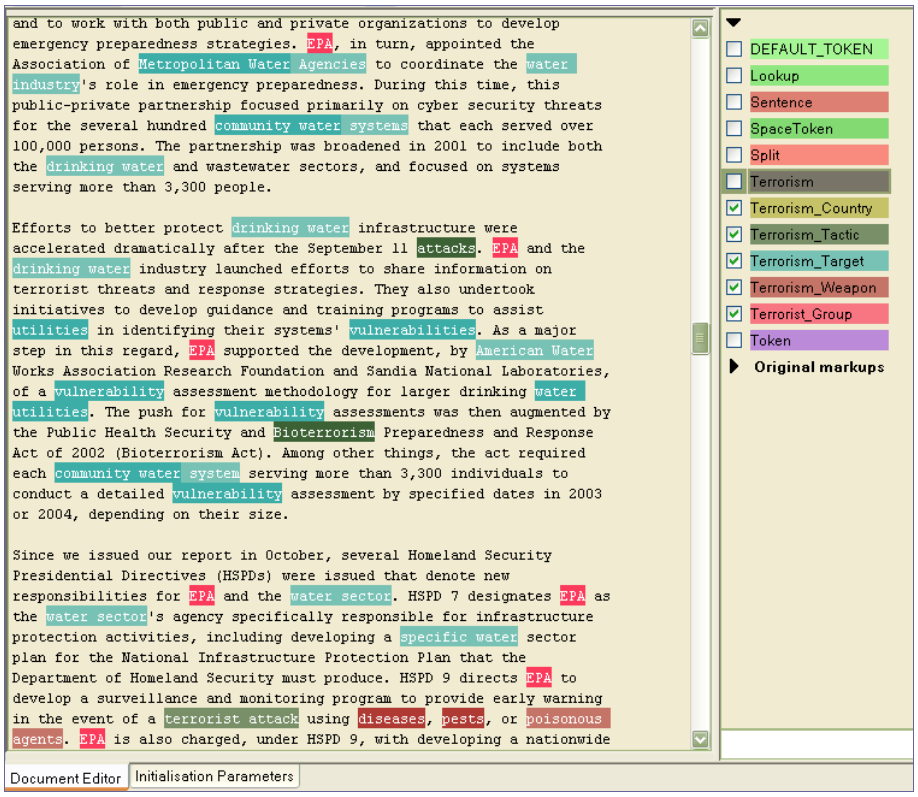

Fig. 4: SACOT's Terrorism-related Named Entities Automatic Identification Using GATE

[7] GATE: General Architecture for Text Engineering (http://gate.ac.uk)

5.3.1 From Instances to Named Entity Patterns

Building new named entity grammar rules requires analysis of lexical patterns. Since early 90's[8], collocation techniques applied to textual corpus have been widely used in the natural language processing community to identify recurrent co-occurring lexical items. Analysis of collocations can provide essential information about term variations such as in *car bombing*, *car-bombing*, and *carbombing*. Frequent collocations can lead to the discovery of different instances of the same class (e.g. *biological weapon*; *chemical weapon*; *nuclear weapon*, *radiological weapon*, etc.). These are all different instances of the class TERRORISM_WEAPON. Instead of enumerating all instances belonging to this class in the ontology, simple named entity grammar rules such as {`JJ + "weapon|weapons"`}[9] will easily capture them. This ilustrates how analysis of collocations can be used to create new grammar rules from specific morph-syntactic patterns that will capture instances of corresponding named entity categories. As illustrated in the two following figures (Fig. 5, 6), identification of recurring patterns is based on analysis of textual data[10].

```
United  States dropped atomic weapons on the Japanese cities of Hir
ey had a role in the attack.  Weapons of Mass Destruction (WMD) Ter
l-Owhali and Mohamed attended weapons   and explosive training at a
ssessed unlicensed automatic  weapons and silencers.  WASHINGTON, D
  Chemical bombs    Automatic weapon fire      0      0    0      0
y terrorist with an automatic weapon or one committed to a        s
nce1058  ELN  Bomb, automatic weapons  1  1  1990  Chile: U.S. emba
hern Pakistan411    Automatic weapons  1  0  2000  Pakistan: Jaish-
 Government troops  Automatic weapons  >100  ?  1991  Iraq: After t
 the complex firing automatic weapons and        throwing grenades
ned by Pan Am517    Automatic weapons gunfire  0  0  1977  Uganda:
um nitrate, over 70 automatic weapons, and 200 blasting caps1556
bombs, ammunition,  automatic weapons, grenades, and various explos
bombs, ammunition,  automatic weapons, grenades, and various explos
measures have been taken;  b. Weapons of Mass Destruction. Although
, chemical or bacteriological weapons,  assuming that they have any
st  groups, as well as bases, weapons, and protection to the Mujahe
rcraft.  Iraq provided bases, weapons, and protection to the MEK, a
IGENCE IS THE            BEST WEAPON AGAINST            INTERNATION
Good Intelligence is the Best Weapon Against International
and nine attempts to  use bio-weapons by Aum that should have been
cing an effective biological  weapon are not insurmountable, they a
tack, even with a biological  weapon. We can strengthen existing ca
tack, even with a biological  weapon. We can strengthen existing ca
e top drill for a biological  weapon."  "Much of the district are i
ned to reduce the biological  weapons threat. Security will be incr
o acquire and use biological  weapons on a mass scale face a major
like chemical and biological  weapons that once produced in the lab
with chemical and biological  weapons or materials, using low-tech
 existence of its biological  weapons program, Aum scientists seeme
 the prototypical biological  weapons agent - it is relatively easy
olved in chemical/biological  weapons (CBW) incidents: charismatic
created extensive biological  weapons programs including work on an
? With respect to biological  weapons, which pathogens deserve prio
 treaty governing biological  weapons. Other mechanisms exist, such
a Lederberg, ed., Biological  Weapons: Limiting the Threat, BCSIA S
rts, Brad (ed.), /Biological  Weapons: Weapons of the Future? /pp.
```

Fig. 5: Recurring Left Collocates for Word "Weapon(s)"

Left collocates for [weapon] = 274		
1	involving	507
2	biological	141
3	nuclear	136
4	chemical	92
5	their	89
6	of	79
7	CBRN	69
8	A	57
9	and	50
10	the	32
11	such	22
12	conventional	21
13	NBC	18
14	radiological	17
15	unconventional	17
16	these	16
17	for	13
18	frequency	13
19	to	13
20	CB	12
21	automatic	11
22	terrorist	10
23	Against	9

Fig. 6: Most Frequent Left Collocates for Word "Weapon(s)"

5.4 Semantic Relations Extraction

Once terms and named entities have been extracted and properly validated, next challenge consists in identifying the different semantic relations those elements share in texts. As stated in Bourigault *et al.* [2001], "it is generally admitted that texts contain several clues as to the meaning of terminological units. These clues can be automatically or semi-automatically detected and/or extracted to provide a

[8] See [Sinclair, 1991; Smadja, 1993]

[9] {`JJ + "weapon|weapons"`} means "a string made of any token having ADJ as part-of-speech category and followed by one of tokens WEAPON or WEAPONS"

[10] Corpus used to generate those figures contains 861916 tokens from open source terrorism-related documents.

better understanding of what terms mean." Expressed by surface linguistic forms, such clues represent explicit semantic relations markers and provide means to extract semantic networks from texts.

Early work from Hearst [1992] focused on automatic acquisition of hyponyms sharing taxonomic relationship. Hyperonymic and hyponymic relationships (IS_A) have been the most studied conceptual structures in the scientific literature. Nevertheless, the taxonomic relationship is only one among many other types of semantic relationships. In his work on retrieval strategies for defining contexts, Auger [1997] identified more than 150 semantic relation markers and proposed a taxonomy of semantic relation types (Fig. 7). More recently, Condamines and Rebeyrolle [2001] explored a number of conceptual relationships in order to build a terminological knowledge base from a corpus of electronic texts. Starting from previous studies from Morin [1999] and Séguéla [2001] on the hyperonymic relationship, Malaisé *et al.* [2005, 2004] extract defining contexts from texts to build differential ontologies. Barrière [2001] and Khoo *et al.* [2002] identified a wide variety of linguistic expressions for explicitly indicating cause and effect relationship in texts.

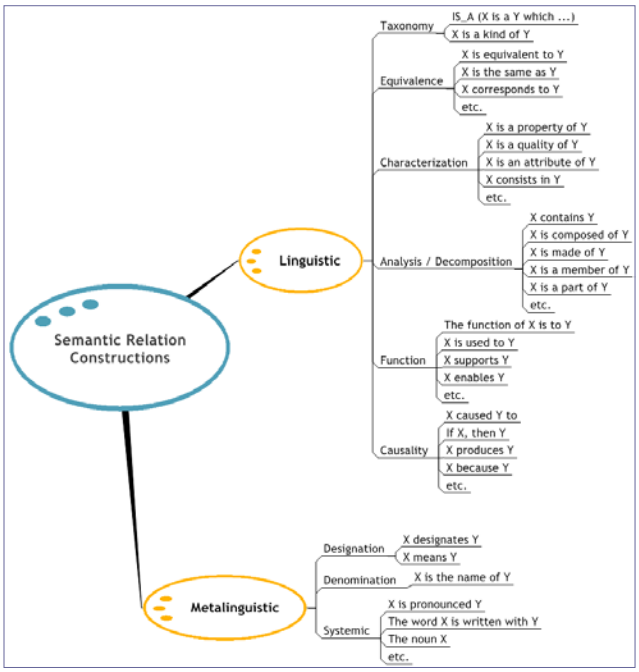

Fig. 7: Semantic Relation Types
(Adapted from Auger, 1997)

SACOT framework exploits several semantic relation markers to retrieve semantic relations among concepts in texts. Extracted concepts and relations are associated in triplet candidates $\{T_1, SemRel_1, T_2\}$ where $\{T_n\}$ is a term and $\{SemRel_n\}$ is a semantic relation. Figure 7 shows a partial view of the semantic relations taxonomy used in the SACOT knowledge engineering framework.

As an example, the following sentence:

`Anthrax` **`is a`**`n acute infectious disease` **`caused by`** `the spore-forming bacterium Bacillus anthracis.`

is rich in semantic relations markers. The semantic relation marker IS_A suggests that ANTHRAX is a kind of ACUTE INFECTIOUS DISEASE. This is typical taxonomic relationship. Therefore, according to this text portion, ANTHRAX can be said as being a member or instance of the class ACUTE INFECTIOUS DISEASE. Moreover, this ACUTE INFECTIOUS DISEASE itself shares a causality relationship (CAUSED_BY) with the instance BACTERIUM BACILLUS ANTHRACIS. Those semantic relations can be represented as in Figure 8.

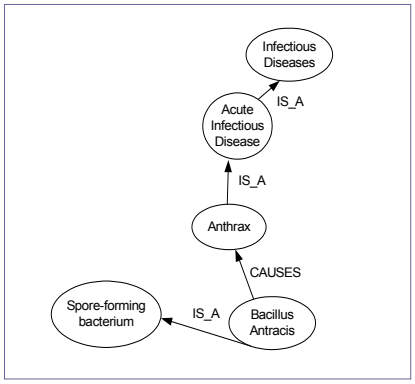

Fig. 8: Local Semantic Network

5.5 Compiling Draft Ontologies

In the SACOT framework, draft ontologies consist of local semantic networks integrating and structuring all knowledge objects captured during previous extraction processes. Those ontologies are considered as draft because they need to be validated by SMEs. Once validated, the knowledge objects of the draft ontology are merged to the domain-specific ontology being built. Since those new knowledge objects are now merged to the domain-specific ontology, SACOT framework will use them as reference material at next iteration of extraction processes.

In the next figure (Fig. 9), the three extraction modules of the SACOT knowledge engineering framework are applied to a sample input text to produce different validated lists. The extracted material is then linked to a local semantic network and, ultimately, validated by the SME as being part of the broader domain-specific ontology.

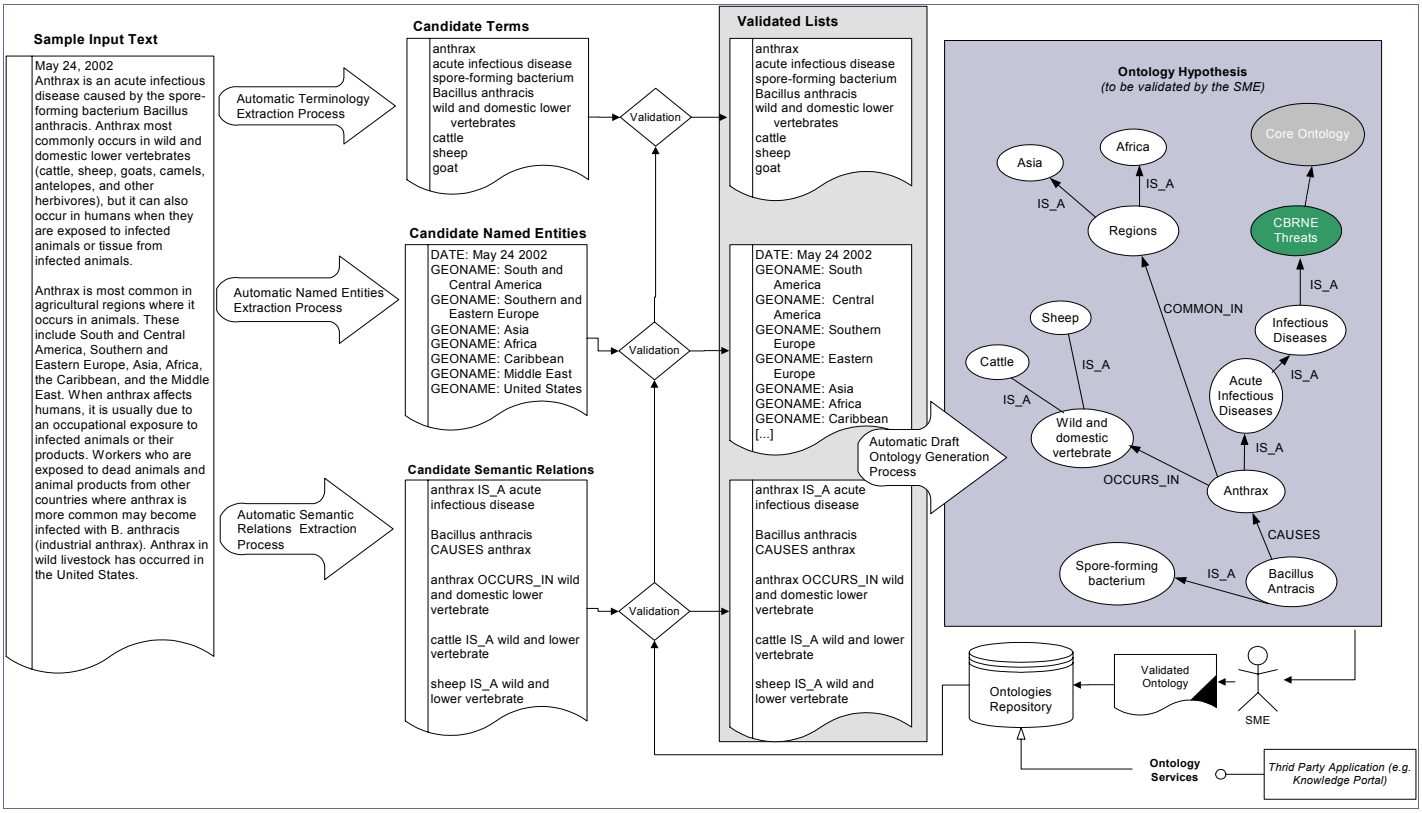

Fig. 9: Turning Electronic Texts into Domain Ontologies Using the SACOT Framework

6. CONCLUSION AND FUTURE WORK

The SACOT ontology-engineering framework significantly reduces time usually required to capture the knowledge objects of a domain in traditional, fully human-based, ontology building processes. It provides knowledge engineers with new means to leverage ever-increasing amount of domain-specific electronic texts and to rapidly build broad domain ontologies for new semantic-aware applications. Future work on the SACOT framework will investigate how learning algorithms could be efficiently used to monitor and learn from SMEs' validation work. Future work is also planned to use the SACOT framework in order to capture and structure knowledge objects from totally different domains. Finally, future work will also investigate post-processing of captured knowledge objects. More specifically, investigation will be conducted to develop and apply semantic link analysis over knowledge objects provided by the SACOT environment.

In the midterm, it is expected that outcomes of this new and integrated knowledge engineering framework will provide benefits for situational awareness portals, for ontology-based automatic document classification systems, for ontology-based data mining, for knowledge portals, for intelligent search engines and for any other application requiring semantic-level capabilities. Further integration efforts will be required to validate those expectations.

This paper have described how recent advances in natural language processing techniques are implemented in the SACOT framework to automate elicitation of knowledge objects from unstructured texts and to support efficiently Subject Matter Experts in ontology engineering tasks.

7. REFERENCES

[Auger, 1997] Auger Alain. *Stratégies de repérage des énoncés d'intérеt définitoire dans les bases de données textuelles*. Ph.D. Thesis. Neuchâtel, Switzerland: University of Neuchâtel, 1997. (Online)

[Aussenac-Gilles *et al.*, 2000] Aussenac-Gilles Nathalie, Biébow Brigitte Szulman Sylvie. "Revisiting Ontology Design: A Method Based on Corpus Analysis", *Knowledge Engineering and Knowledge Management. Methods, Models and Tools*. Rose Dieng and Olivier Corby (Eds.) 12th International Conference EKAW 2000. Juan-les-Pins, France. (2000).

[Barrière & Copeck, 2001] Barrière, Caroline and Copeck, Terry. "Building domain knowledge from specialized texts." *TIA'2001 : Terminologie et Intelligence Artificielle*, France, May 2001, p. 109-118

[Barrière, 2001] -----. "Investigating the Causal Relation in Informative Texts". *Terminology*, 7:2. 2001

[Barrière, 2002] Barrière, Caroline. *SeRT - A tool for Knowledge Extraction from Text*. Communication presented at CLiNE (Computational Linguistics in the North East), Montreal, May 2002

[Bourigault *et al.*, 2001] Bourigault Didier, Jacquemin Christian L'Homme Marie-Claude Eds. *Recent Advances in Computational Terminology*. Amsterdam - Philadelphia: John Benjamins, 2001.

[Bourry-Brisset, 2000] Bourry-Brisset Anne-Claire, 2000: Knowledge Modeling and Management for Command and Control Environments. *Command and Control Research and Technology Symposium (CCRTS)*, June 26-28, Montery, CA.

[Bowman *et al.*, 2001] Michael Bowman, Antonio M. Lopez, Jr., Gheorghe Tecuci. "Ontology Development for Military Applications". In Proceedings of the SouthEastern Regional ACM Conference, Atlanta, GA, March 16-17, 2001.

[Chance & Hagenston, 2003] Chance Samuel G., Hagenston Marty G., 2003: *Assessing the Potential Value of Semantic Web Technologies in Support of Military Operations*. Thesis. Naval Postgraduate School. Monterey California United States Navy. 289 p.

[Chincor, 1997] Chinchor. Nancy. Message Understanding Conference (MUC-7), Named Entity Task Definition Version 3.5. Available from: http://www.itl.nist.gov/iaui/894.02/related_projects/muc/proceedings/ne_task.html

[Condamines & Rebeyrolle, 2001] Condamines, Anne and Rebeyrolle, Josette. "Searching for and identifying conceptual relationships via a corpus-based approach to a Terminological Knowledge Base (CTKB): Method and Results", *Recent Advances in Computational Terminology*, Bourigault *et al*. (Eds), John Benjamins, Amsterdam, Philadelphia. (2001). p. 127-148

[Corcho *et al.*, 2003] Corcho Oscar, Fernández-López Mariano Gómez-Pérez Asunción. "Methodologies, Tools and Languages for Building Ontologies. Where Is Their Meeting Point?" *Data and Knowledge Engineering* 46 (2003): p. 41-64.

[Davies *et al.*, 2003] Davies John, Fensel Dieter, and Van Harmelen Frank (Eds). *Towards the Semantic Web. Ontology-Driven Knowledge Management*. John Wiley and Sons, LTD ed. England, 2003.

[DND, 2001]: Future Army Capabilities in *DLSC Report 01/01*.

[Dorion & Bourry-Brisset, 2004] Dorion, Éric, Bourry-Brisset, Anne-Claire. "Information Engineering in Support of Multilateral Joint Operational Interoperability", CCRTS, June 2004, San Diego, California

[Drouin, 2003] ---. "Term Extraction Using Non-Technical Corpora As a Point of Leverage." *Terminology* 9:1 (2003): 99-117.

[Drouin, 2004] Drouin Patrick. "Detection of Domain Specific Terminology Using Corpora Comparison." *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC), Lisbon, Portugal*. (2004).

[Gauvin *et al.*, 2004] Gauvin Marlène, Bourry-Brisset Anne-Claire and Auger Alain, 2004 : Context, Ontology and Portfolio: Key Concepts for a Situational Awareness Knowledge Portal in *Proceedings of HICSS-37. Hawaii's International Conference on System Sciences*.

[Gómez-Pérez *et al.*, 2004] Gómez-Pérez Asunción, Fernández-López Mariano, Corcho Oscar, *Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web*. Springer (2004).

[Gómez-Pérez, 1999] Gómez-Pérez Asunción. "Ontological Engineering. A State of the Art." *Expert Update* 2.3 (1999).

[Gouin *et al.*, 2003] Gouin Denis, Gauvin Marlène, and Woodliffe Elizabeth, 2003: COP 21 TD - Towards a Situational Awareness Knowledge Portal in *Proceedings of SPIE 2003 – Aerosense/ Defence Sensing, Simulation and Controls*, Orlando, 21-25 April 2003 Vol. 5101 - Battlespace Digitization and Network-Centric Systems III.

[Gruber, 1993] Gruber T. "A Translation Approach to Portable Ontologies." *Knowledge Acquisition* 5.2 (1993): 199-220.

[Grüninger & Fox, 1995] Grüninger M., Fox M. "Methodology for the Design and Evaluation of Ontologies." *Proceedings of IJCAI95's Workshop on Basic Ontological Issues in Knowledge Sharing* (1995).

[Hearst, 1992] Hearst, M.A. "Automatic Acquisition of Hyponyms from Large Text Corpora", *Proceedings of 14th International Conference on Computational Linguistics*, Nantes, France, (1992). p. 539-545

[Jacquemin, 2001] Jacquemin Christian. *Spotting and Discovering Terms Through Natural Language Processing*. Cambridge Massachussetts, London England: The MIT Press, (2001).

[Khoo *et al.*, 2002] Khoo Christopher, Chan Syin, Niu Yun, "The Many Facets of the Cause-Effect Relation", *The Semantics of Relationships. An Interdisciplinary Perspective*, Rebecca Green *et al.* (Eds.), Kluwer Academic Press. (2002). p. 51-70

[Maedche *et al.*, 2000] Maedche Alexander, Staab Steffen, "Mining Ontologies from Text", *Knowledge Engineering and Knowledge Management. Methods, Models and Tools*. Rose Dieng and Olivier Corby (Eds.) 12th International Conference EKAW 2000. Juan-les-Pins, France. (2000).

[Malaisé *et al.*, 2005] Malaisé Véronique, Zweigenbaum Pierre, Bachimont Bruno, « Mining Defining Contexts to Help Structuring Differential Ontologies », *Terminology*, 11:1, 2005 (to appear)

[Malsaisé *et al.*, 2004] Malaisé Véronique, Zweigenbaum Pierre, Bachimont Bruno, « Repérage et exploitation d'énoncés définitoires en corpus pour l'aide à la construction d'ontologies », 11ième édition de la Conférence sur le Traitement automatique des langues naturelles (TALN 2004), Fez, Maroc 19-21 avril 2004.

[Morin, 1999] Morin E., « Des patrons lexico-syntaxiques pour aider au dépouillement terminologique », *Traitement Automatique des Langues*, vol 40(1), 1999, p. 143-166.

[Séguéla, 2001] Séguéla P., *Construction de modèles de connaissances par analyse linguistique de relations lexicales dans les documents techniques*, Thèse de doctorat, Université Toulouse III, 2001.

[Sinclair, 1991] John Sinclair. *Corpus, Concordance, Collocation*, Oxford University Press. 1991

[Smadja, 1993] Frank Smadja. "Retrieving collocations from text". *Computational Linguistics*, 19(1):143-177, 1993.

[Sure, 2003] Sure York, Staab Steffen, and Studer Rudi. "On-To-Knowledge Methodology (OTKM)." *Handbook on Ontologies*. Staab S. and Studer R. (Eds). Series on Handbooks in Information Systems: Springer, 2003.

[Uschold & Grüninger, 1996] Uschold Mike, and Grüninger Michael. "Ontologies: Principles, Methods and Applications ." *Knowledge Engineering Review* 11.2 (1996): 93-155.