Semantic Filtering of Textual Requirements Descriptions

Jorge J. García Flores[1]

LaLICC

Université Paris-Sorbonne

96, bd Raspail, 75006 Paris, France

jorge.gflores@paris4.sorbonne.fr

Abstract. This paper explores the use of semantic filtering techniques for the analysis of large textual requirements descriptions. Our approach makes use of the contextual exploration method in order to identify specific linguistic markers which could bring to light the structure of semantic knowledge within natural language requirements, in particular those statements considered as relevant from requirements engineering perspective: concepts relationships, aspecto-temporal organisation, cause and control statements. We investigate to what extent filtering with these criteria can be the base of requirements analysis and validation processing, and what kind of software tools are necessary to support contextual exploration systems for this purpose.

1. INTRODUCTION

Many software development projects fail because of a deficient requirements strategy. In the context of software systems development, requirements engineering (RE) studies the processes, techniques and methods that would allow a system to meet the purpose for which it was intended [Nus00]. RE activities include requirements elicitation, analysis, validation and documentation. The realisation of these activities, in an iterative way, all along the system development life cycle may lead to a full understanding of the achievements expected from the system in its environment [Zav97].

The first step in the RE process is often a textual description where users and clients describe their needs. Textual requirements descriptions (TRD) play an important role in RE not only because they state natural language (NL) requirements, but also because they collect necessary knowledge from the application domain and they provide the basis for an effective communication between stakeholders. Use cases descriptions, scenarios, user stories, transcriptions of conversations for requirements elicitation [Gog94] and even rough sketches [Jac95] are examples of textual requirements descriptions.

The purpose of this paper is to explore the use of a linguistic technique, the Contextual Exploration (CE) Method [Des97] to support RE activities, in particular requirements analysis and validation. This method organizes linguistic knowledge under the form of hypothesis that could confirm or refute the attribution of a certain semantic value to a sentence, according to the presence of linguistic clues. A semantic viewpoint is a set of sentences that, because of their semantic values, convey a particular semantic notion. This paper proposes that the following viewpoints may be useful to support RE analysis and validation: 1) relations between concepts, 2) aspecto-temporal organisation, 3) control and 4) causality.

The use of linguistic knowledge for natural language requirements processing is not new. In the past, linguistic instruments have been used to extract conceptual schema from NL requirements ([Rol92], [Fli03]), to analyse and validate requirements through a conceptual representation ([Bur97], [Ben99], [Amb97]) or to improve the linguistic quality of TRD documents ([Osb96], [Fab01]). However, few of these approaches take into account the size of the input text. Large TRD documents raise interesting questions for NL requirements processing techniques. What should be the processing's scope? Is it

[1] PhD. Research funded by the Mexican Government (CONACYT)

useful to process the whole mass of documents, or is it better to limit it to some extent? If there are formal representations involved, how do they deal with such amount of input text? What are the effects of large TRD on the NL processing strategy?

One could suppose that, faced with a large set of TRD, conceptual schema generators would produce a substantial profusion of diagrams without amplifying cognition, at least if no further processing is undertaken. Experiences on NL requirements validation [Ger02] suggest that, when using formal methods, partial processing of TRD (in the referred case study, three pages from a 250-pages document) is more effective than translating the entire document into a formal representation and having to maintain both representations, the formal and the natural language one, in parallel all along the development process [Kem90]. Works on requirements extraction from ethnographical observation documents [Ray00] deal with their copious TRD by means of statistical NL processing tools. The CASE tool for NL requirements developed by [Hop97] proposes to manually selecting meaningful paragraphs before processing, in order to reduce the size of the input text.

This paper postulates that filtering relevant text fragments according to semantic criteria enhances large TRD processing. The underlying idea behind this approach is that natural language structures requirements within a text by semantic and discursive means. The use of the Contextual Exploration Method aims at bringing to light this semantic structure in order to 1) improve readability of the whole TRD by offering partial views of it, as well as 2) base further NL requirements processing (requirements analysis, conceptual schema generation, partial validation, etc.) on this new abridged text unit made of relevant TRD fragments.

The paper is organised as follows: section 2 introduces the NL processing strategy, the Contextual Exploration Method, which organizes linguistic knowledge and allows text filtering according to semantic viewpoints. Section 3 is devoted to the discussion of how semantic filtering would facilitate TRD processing. It outlines the TRD semantic filtering cycle and proposes an architecture for a TRD semantic filtering system. Section 4 presents the conclusions and sketches future work.

2. CONTEXTUAL EXPLORATION METHOD

In the frame of the Cognitive and Applicative Grammar (GAC in French) linguistic theory [Des90], the contextual exploration (CE) method [Des97] was originally designed to solve lexical and grammatical polysemy problems. The method rests on the principle that domain independent linguistic elements structure text meaning. Its purpose is to access the semantic content of texts in order to extract relevant information according to a certain task.

2.1 Contextual Exploration Rules

According to contextual exploration, all signs occurring in a text (the *textual context)* must be taken into account to determine the semantic value of a sentence. This example illustrates how indeterminacy is solved[2]:

(1) ***Malgré*** *toutes les précautions, le lendemain il* ***était pris*** (In spite of all precautions, he was captured the day after)

(2) ***Sans*** *toutes les précautions*, *le lendemain il* ***était pris*** (Without all precautions, he was captured the day after)

From the aspecto-temporal point of view, the tense of “was captured” (*était pris*) is the linguistic marker of a semantic value that can be NEW_STATE (he was captured) or UNREALIZED (he was not captured). However, the tense itself is not enough to decide which one of both values must be assigned, so the context has to be analysed in order to get more clues. In this case *malgré* (in spite) and *sans* (without) are the clues that determine the semantic values of these sentences.

Linguistic markers correspond to plausible hypothesis that must be confirmed by the presence of certain clues. The heuristic mechanism is based in rules. A rule **R** is defined as follows:

[2] This exemple is taken from [Des97]

$$\mathbf{R_k = [K, \{I_p, C_p\}, D_k]}$$

where K is a class of linguistic marker, $\{I_p, C_p\}$ a finite set of couples of the **p**-th class of clue (**I** from *indice*) to search for in the context C_p and take the decision D_k. [Min02]. Rules are declared in a language intended to separate their linguistic definition from their computer-based implementation.

As an example, the following rule is meant to assign the semantic value INGREDIENCE to a sentence like *Le disque dur* ***est*** *en général un* ***élément de*** *l'unité centrale* (The hard drive is generally an element of the central unit):[3]

> **Head**: Ringrconst01
> **Task** : Relations Statiques
> **Trigger:** &ILIngNomConst
> **Body**
> E1 := CreateSpace(Previous (I))
> E2 := CreateSpace (Afterwards (I))
> L1 := &tiret
> L2 := &2-poincts
> L3 := &etre
> L4 := &LIngPrep2
> **Conditions**
> There_is **x** belonging_to E1 that_class_of **x** belongs_to (L1 OR L2 OR L3)
> There_is **y** belonging_to E2 that_class_of **y** belongs_to (L4)
> **Actions**
> SemanticLabel(ParentSentence(I), "*est_une_partie_de*")
> SearchArgument2(ParentSentence(I), E1, E2 , x1→x2)
> **End**

The class of markers triggering this rule is *&ILIngNomConst.* In the linguistic repository, this class contains 54 nouns, like *membre* (member), *pièce* (piece), *partie* (part), *organe* (organ), etc. The occurrence of any of these markers triggers the rule, which looks in the context (afterwards and previous) for a flexional form of the verb to-be (clue-class *&etre*) , the prepositions *de* (of), *par* (by) or *avec* (with) from the clue-class *&LIngPrep2*, and punctuation sign ":" (*&2-poincts*). Therefore, a sentence like *Disque dur* ***: partie de*** *l'unité centrale* would get the INGREDIENCE label as well.

Linguistic clues may not only be lexical entries. Punctuation signs, text structural elements, or even discourse acts, like a definition or a conclusion, are allowed in a clue class. The context's scope is not the same for all clues. Desclés et al. [Des97] characterize four kinds of context:

- Context C1 is limited to the clause the sentence S where the semantic value is to be assigned.
- Context C2, necessary to solve anaphora, is limited to sentences belonging to the same paragraph, located just before the sentence S containing the anaphoric clue.
- Context C3 includes several sentences before and after the sentence S. It is limited, either by the beginning of a textual section or by textual organisation clue.
- The context C4 must be explored to identify textual segments formatted with special cues as bullet lists, enumerated.

Rules are organized in tasks. The rule from the above example is part of the task *Relations Statiques* (Static Relations), collected by Le Priol [LeP00]. This task specifies a set of 238 rules, 6149 markers and 1777 clues in order to assign a set of 14 semantic labels, corresponding to the semantic values defined as static relations in the Cognitive and Applicative Grammar.

So far, the CE method has been applied to domain knowledge extraction [Jou97], thematic announcements extraction [Car96], automatic summarisation [Ber96], filtering of causality relations [Jac98], relations between concepts [LeP00] and aspecto-temporal relations [Cha03]. CE linguistic

[3] This rule was written by F. Le Priol in LJava language [LeP00].

repository has also been used for Information Extraction under different semantic viewpoints in [Lau02].

2.2 Contextual Exploration Systems

The construction of a linguistic repository for CE entails two kinds of activities: linguistic knowledge acquisition and text extraction. The former is intended to allow linguistic markers collection and CE rules definition by a linguist user; the latter to allow text extraction, and therefore CE rules execution, by a non-linguist user. Any implementation of the CE method must support both activities [Min02]. The architecture of a CE system framework is shown in Fig. 1.

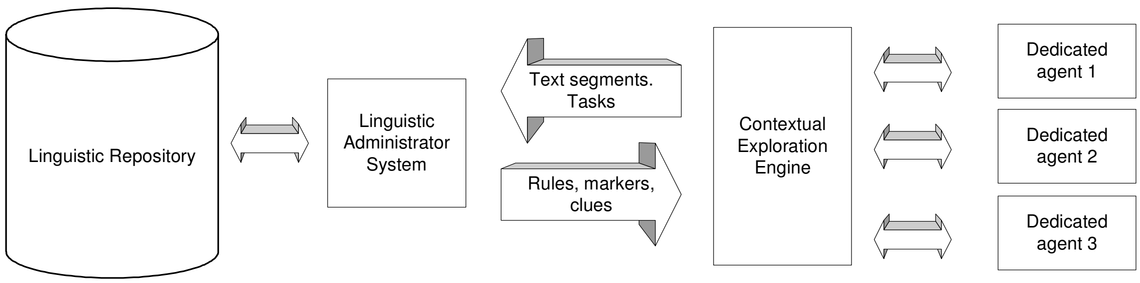

Figure 1: Architecture of a CE system [Min02]

The linguistic repository stores only domain independent linguistic knowledge. The CE engine executes rules associated to a certain task on a text collection. The results are processed and presented to the user by dedicated agents. Domain knowledge extraction, automatic summarization and filtering of causal relations are examples of a dedicated agents.

NL processing in CE systems differs from the typical *grammar* + *lexicon* approach in that no syntactic or lexical analysis is performed. However, morphological variations are coded as part of some classes, as illustrated in Table 1, where the *&etre* and *&faire* classes represent morphological variations of the verbs "to be" and "to do", while the *&inclusion_contenu* class contains morphological variations of "to contain" in past perfect (*conntenu/e/s*).

MARKER	CLASS	TASK
&faire+fonctionner	&qualitative_causality	Causality
&faire+marcher	&qualitative_causality	Causality
&faire+progresser	&qualitative_causality	Causality
&faire+surgir	&qualitative_causality	Causality
&etre+contenu+en	&inclusion_content	Static Relations
&etre+contenue+en	&inclusion_content	Static Relations
&etre+contenues+en	&inclusion_content	Static Relations
&etre+contenus+en	&inclusion_content	Static Relations

Table 1: Extract from the database of linguistic markers

CE cycle has two phases:

1) A configuration phase, where a linguistic study is made in order to identify the net of possible semantic values and organize them in an *abductive* way [Des96], i.e. as plausible hypothesis that must be confirmed by the presence of certain clues or, in other words, as CE rules.
2) A production phase, where any user can execute CE task and rules on real texts.

So far, two CE systems have been developed: ContextO [Cris04] and Semantext [Ben00]. Both of them support the CE cycle. Yet, Semantext differs from ContextO in that, in sake of performance, it analyses text as a flat chain of characters, while ContextO processes it in a hierarchical way (title, section, subsection, paragraph, etc).

2.3 Contextual Exploration and NL requirements processing

The CE method present specific aspects that distinguish it from other semantic tagging techniques used in the context of NL requirements processing:

- CE is oriented to large document sets (it has been used for automatic summarization by [Ber96]), so it can handle large TRD while taking account of the linguistic context, unlike [Ray97] and other statistical approaches.
- CE markers and clues are not restricted to linguistic elements, but also typographical and structural ones. For instance, the ":" punctuation sign followed by a bulleted list can indicate the plausibility of a "part-of" relation (see section 3). Despite the importance of structural elements (titles, subtitles, etc.) in industrial TRD documents, little research has been done on their exploitation by NL requirements processing tools.
- CE linguistic repository separates domain independent linguistic elements (verbs, prepositions, articles, etc.) from domain dependent ones (lexical entries). This characteristic makes possible the reutilization of linguistic resources in different application domains, like in [Osb96] and [Fab01], but with an important difference: organizing criteria for linguistic hierarchies are neither syntactic nor lexical, but semantic (causality, control, aspecto-temporal organization and conceptual relationships).
- CE semantic tags are not tight to any software design "methodology". The difference between CE an other "methodology neutral" approaches ([Amb97], [Ben99], [Fli03]) is that CE semantic tags are based on semantic relationships and primitives from the GAC linguistic theory [Des90], so they produce an intermediate text unit (the filtered text) enriched with semantic values. Section 3 explores how this semantic values can be exploited in the RE process.

3. SEMANTIC FILTERING OF TEXTUAL REQUIREMENTS DESCRIPTIONS

3.1 Semantic Viewpoints

Semantic filtering is defined as the interactive view of semantically relevant text fragments extracted from a set of documents [Min02]. It depends on which text fragments can be considered as semantically relevant. Research on automatic summaries evaluation [Min97] indicates that relevance criteria are subjective and depend on the reader's point of view. The purpose of semantic filtering is to throw light on the semantic and discursive structures that organize knowledge taking account of the reader's point of view.

The idea of applying semantic filtering to large TRD is intended, on the one hand, to improve readability of large TRD by drawing together a small number of statements that share a certain semantic viewpoint or address the same topic [Hull02]. On the other hand, this approach aims at extracting relevant statements from the TRD to shape a new abridged text unit that, because of its size and relevance, would facilitate further processing, like requirements extraction, automatic generation of conceptual schema or partial validation.

However, as it has been mentioned above, relevance criteria are subjective and depend for the most part on the reader's point of view. For instance, a reader interested in data modelling would certainly pay particular attention to "part-of" relations between domain entities. A relevance criterion for this kind of reader could be to filter only those statements where there are linguistic markers of a "part-of" relation, like the following[4] (from a TRD of an insurance system):

(1) *A joint policy* ***includes*** *the joiner's age, the joiner's gender and his smoker condition.*

(2) *The agent displays billing mode, effective date, due date, bill number and total premium amount, which* ***are part of*** *the policy's billing detail.*

(3) ***For each*** *mandatory rider, the agent should specify the* ***following*** *values****:***

[4] This work is being developed for the French language but, in sake of a better understanding English examples are used.

-Face amount,
-Due period
-Increased Periods

According to CE rules, this sentences would have the INGREDIENCE semantic label, even if linguistic markers (in bold) from (1) and (2) looks stronger than those of (3), where the part-of relation is not explicitly indicated, but its plausibility may be confirmed by non-lexical clues, like the typographical sign ":" or the presence of a list.

Therefore, from the requirements engineering point of view, relevance must be defined in terms of those activities supported by the TRD. In this paper the following viewpoints are considered as semantically relevant:

I. Relations between concepts

II. Aspecto-temporal organization

III. Control

IV. Causality

Each one of this viewpoints conveys a semantic concepts that, according to the GAC linguistic model, may structure and organize meaning [Des90], and each one represents an important aspect in requirements analysis as well. Much effort have been devoted to build conceptual schema from TRD's relations between concepts for requirements analysis ([Rol92], [Fil68], [Amb97]}. The value of extracting event's and processes (II) temporal organisation (the "dynamic" aspect) for NL requirements validation has been remarked by [Bur97]. Control issues (III), i.e. specifying if actions are "environment controlled" or "machine controlled", are of primary importance in requirements engineering [Zav97], and a precise understanding of the causal organisation (IV) of actions is necessary to specify the rules that a system must obey [Kov98].

The following example shows different views of one paragraph according to these semantic viewpoints (linguistic markers and clues are underlined)[5]:

(4) *When the start button is pressed, if there is an original in the feed slot, the photocopier makes N copies of it, and places them in the output tray. N is the number currently registering in the count display. If the start button is pressed while photocopying is in progress, it has no effect. The number N in the count display updates in response to button pressed according to the state table*

equality

(4-I) *When the start button is pressed, if there is an original in the feed slot, the photocopier makes N copies of it, and places them in the output tray. N is the number currently registering in the count display. If the start button is pressed while photocopying is in progress, it has no effect. The number N in the count display updates in response to button-presses according to the state table*

Relations between concepts viewpoint

event

(4-II) *When the start button is pressed, if there is an original in the feed slot, the photocopier makes N copies of it, and places them in the output tray. N is the number currently registering in the count display. If the start button is pressed while photocopying is in progress, it has no effect. The number N in the count display updates in response to button-presses according to the state table.*

Aspecto-temporal organisation viewpoint

environment controlled

machine controlled

(4-III) *When the start button is pressed, if there is an original in the feed slot, the photocopier makes N copies of it, and places them in the output tray. N is the number currently registering in the*

[5] This example is taken from a requirements document in [Kov98].

count display. If the start button is pressed while photocopying is in progress, it has no effect. The number N in the count display updates in response to button pressed according to the state table

Control viewpoint

(4-IV) *When the start button is pressed, if there is an original in the feed slot, the photocopier makes N copies of it, and places them in the output tray. N is the number currently registering in the count display. If the start button is pressed while photocopying is in progress, it has no effect. The number N in the count display updates in response to button pressed according to the state table*

Causality viewpoint

In the above example, sentences in grey are not relevant from a particular semantic viewpoint. The boxes indicate semantic values assigned by CE rules according to linguistic markers and clues (underlined). Every filtered sentence has at least one semantic value. The following example shows the XML output filtered text with its semantic values according to the CE aspecto-temporal relations task [Cha03], which is used in this paper's approach for the aspecto-temporal organisation viewpoint:


```
<REF_ENONC>
    <P1 value="event"> When the start button is pressed, </P1>
    <REF_POSS>
        <P2 value="state"> if there is an original in the feed slot,</P2>
        <event_in_process>
            <P3 value="event"> the photocopier makes N copies of it,</P3>
        </event_in_process>
        <P4 value="event">
            <event_post_event>and places them in the output tray.</event_post_event>
        </P4>
    </REF_POSS>
    <REF_POSS>
        <P5 value="event"> If the start button is pressed </P5>
        <event_in_process>
            <P6 value="process"> while photocopying is in progress, </P6>
        </event_in_process>
        <P7 value="state"> it has no effect. </P7>></REF_POSS>
</REF_ENONC>
```


3.2 Post-filtering processing

Every viewpoint would produce, after TRD filtering, a two-folded output: the filtered text and its associated semantic values. The following are examples of the post-filtering process:

- To continue filtering the output text under application domain criteria. For instance, to select all statements from the "Relation between concepts" viewpoint which contains *billing detail* and *premium amount,* like in example number (2).
- To apply analytic rules based on "Relation between concepts" semantic values. For instance, to verify that, in INGREDIENCE statements, the relationship between an element and its parts does never gets reversed.
- To look for conflicts between viewpoints. For instance, suppose that the following statements are filtered by control and cause viewpoints:

  (5-III) *When the product's lifecycle is over, the system should* ***trigger*** *a premium-collection event.*

  (5-IV) *The system can* ***prevent*** *a premium-collection event but only an agent can* ***cause*** *it.*

  In statement (5-III) there are clues of a machine-controlled situation over the premium-collection event, while (5-IV) states that this event can only be caused by the agent.

- To look for semantic patterns. For instance, a high proportion of cause and control statements may be sign of a Jackson's "Control problem" frame [Jac95].

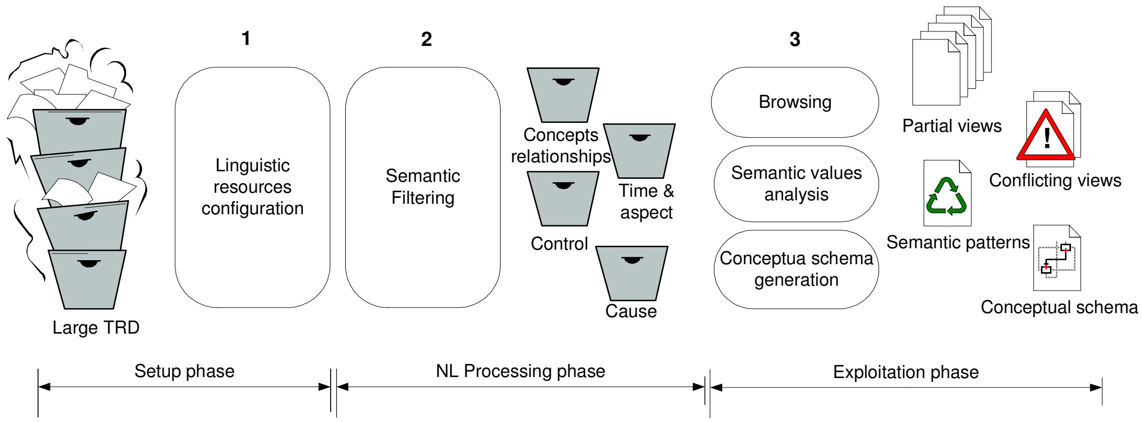

Figure 2: TRD filtering cycle

As shown in Fig. 2, the configuration of linguistic resources is needed before filtering. This setup phase is intended assimilate the application's domain glossary into the linguistic resources and to do fine-tuning on CE rules in order to adapt them to the style and the language of TRD. Then, TRD documents are then filtered according to the above-mentioned semantic viewpoints. An interactive presentation of the partial TRD views is required in order to browse the TRD and make a first evaluation on the filtering quality. Conflicting views and semantic values analysis need a set of semantic-value-based rules, other that CE rules, that allow inter-viewpoint analysis based on semantic values (for instance, in the above example, the comparison between cause and control viewpoints). Existing conceptual schema generation tools, like Circe [Amb97] or ColorX [Bur97], could be used at this point to process the filtered text regardless of the semantic values.

The required tools to support requirements engineering process by TRD semantic filtering are summarized in Table 2.

GOAL	ACTION	TOOL
TRD readability	Analysing TRD according the different viewpoints and support further filtering under lexical criteria.	TRD viewpoint browser
Requirements analysis	Continue filtering under application domain lexical criteria.	TRD viewpoint browser
Requirements analysis	Mapping the filtered text to conceptual schema	Conceptual schema generation tool
Requirements analysis	Semantic patterns analysis	Semantic value (SV) analyser + SV-based rules
Requirements validation	Validating according to one viewpoint's semantic values	Semantic value (SV) analyser + SV-based rules
Requirements validation	Conflicting viewpoint analysis	Semantic value (SV) analyser + SV-based rules

Table 2: TRD semantic filtering tools

3.3 An architecture proposal

Following the above-mentioned TRD filtering cycle (Fig. 2), the proposed architecture supports the linguistic configuration phase with the rules configuration tool, that would allow to configure CE rules, to add application domain's glossary in the linguistic repository and to specify non CE rules, that is, rules based on semantic values that would allow operations between viewpoints (for instance, how control is handled on all the elements of an INGREDIENCE relation). The NL processing phase would be supported by a CE system like the above-mentioned Semantext or ContextO (see section 2), which would receive large TRD as an input and would filter it according to CE rules calculating semantic values. TRD Viewpoint browser is intended to allow a user to browse between the source TRD and partial viewpoint-based views, and to allow further filtering according to application domain criteria. The Semantic filtering broker supports exploitation tools by handling semantic value and SV-based rules requests.

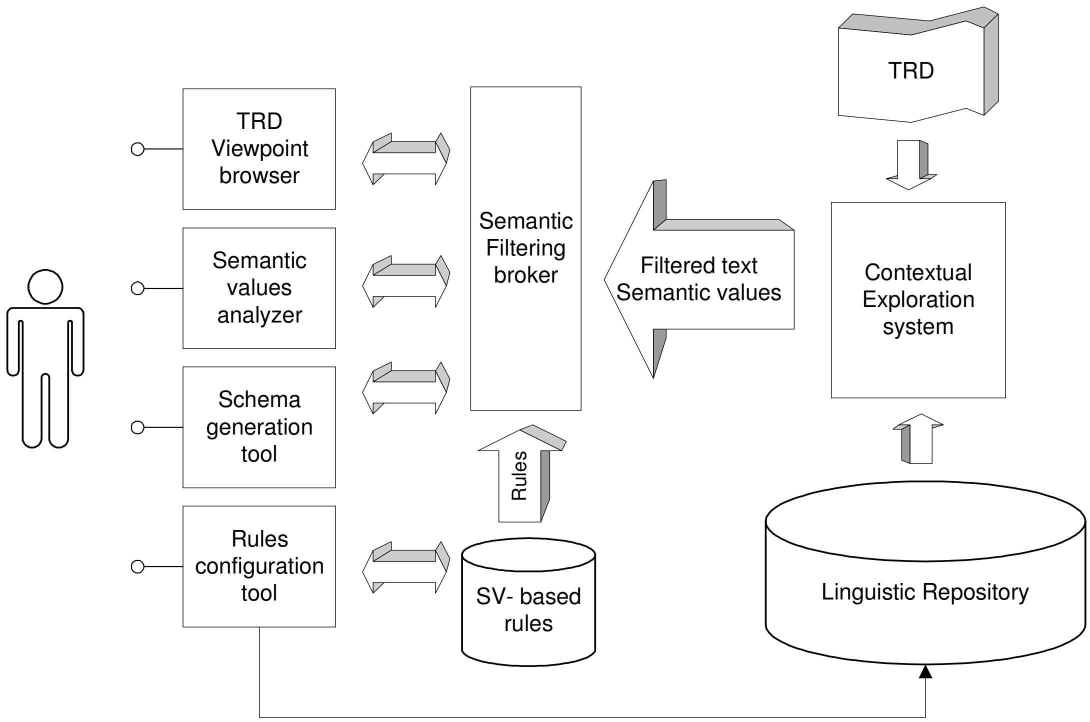

Figure 3: Architecture for TRD semantic filtering

4. CONCLUSION AND FURTHER WORK

This paper has proposed a semantic-based approach for NL requirements processing, which filters semantically relevant sentences from large TRD making use of linguistic-based rules. It has exposed the contextual exploration method, which organizes rules, linguistic markers and clues, assigning semantic values according to four major viewpoints, considered as relevant from a requirements engineering perspective: relations between concepts, aspecto-temporal organization, control and causality. Furthermore, this paper has outlined how semantic viewpoints could improve requirements analysis and validation, as well as the overall readability of the TRD documents by means of semantic filtering tools. Finally, it has proposed an architecture for a TRD semantic filtering platform.

Currently, work is being done to evaluate the precision of CE rules, markers and clues (most of them issued from linguistic studies on scientific corpus) on industrial requirements documents, as well as on the development of the semantic filtering broker. Further work will focus on the definition of a declarative language for semantic-value based rules in a way that could allow interoperability between viewpoints, as well as in the study of the CONTROL primitive of the GAC linguistic theory, which so far hasn't been used for contextual exploration.

Once the platform is ready to test, it will be necessary to evaluate, on the one hand, to which extent semantic filtering improves large TRD readability and, on the other hand, if the proposed semantic viewpoints facilitate requirements analysis and validation. Based on evaluation methods of the CREWS/L'écritoire project [Ben99] and automatic summaries evaluation experiences [Min97], this papers author concludes that an empirical evaluation would be necessary, where system analyst could use the proposed semantic filtering tools on real TRD.

ACKNOWLEDGMENTS

Thanks to Joël Bourgeoys, Marie-Josée Goulet, Jean-Luc Minel, and Camille Salinesi, for their feedback on earlier versions of this paper, and to Marie Chagnoux for her help with the aspecto-temporal aspects.

REFERENCES

[Amb97] Ambriola,V. and Gervasi,V., Processing Natural Language Requirements. In *12th International Conference on Automated Software Engineering*, pp. 36-45, IEEE Press, U.S., 1997.

[Ben00] Ben Hazez, S., Minel, J-L., Designing Tasks of Identification of Complex Linguistics Patterns used for Text Semantic Filtering . In *Proceedings of Content-based multimedia information access RIAO 2000 International Conference*, France, 2000.

[Ben99] Ben Achour, Extraction des besoins par analyse de scénarios textuels. PhD thesis, Université Paris 6,1999.

[Ber96] Berri, J., Contribution à la méthode d'exploration contextuelle. Applications au résumé automatique et aux représentations temporelles. PhD thesis, Université Paris-Sorbonne, 1996.

[Bur97] Burg, J.F.M., Linguistic Instruments in Requirements Engineering, IOS Press, Amsterdam, 1997.

[Car96] Cartier, E., SERAPHIN, marqueurs sémiotiques et linguistiques, règles d'exploration contextuelle et phase d'ajustement, *Rapport Interne 96/1 du CAMS-LaLICC,* Université Paris-Sorbonne, 1996.

[Cha03] Chagnoux, M., Ben Hazez, S., Desclés, J.-P., Identification automatique des valeurs temporelles dans les textes. In *10e conférence sur le traitement automatique des langues (TALN'2003)*, France, 2003.

[Cris04] Crispino, G., Une plate-forme informatique de l'Exploration Contextuelle : modélisation, architecture et réalisation (ContextO). PhD thesis, Université Paris-Sorbonne, 2004.

[Des90] Desclés, J-P., Langages applicatifs, langues naturelles et cognition, Hermes, Paris, 1990.

[Des96] Desclés, J-P., Abduction and non-observability in language and cognitive sciences, *Colloque sur les hypothèses cachées dans les sciences, Académie Internationale de Philosophie des Sciences, Parma,1996*.

[Des97] Desclés, J-P., Cartier, E., Jackiewicz, A., Minel, J-L., Textual Processing and Contextual Exploration Method. In *Proceedings of the First International and Interdisciplinary Conference on Modeling and Using Context (Context 97)*, pp. 189-197, Rio de Janeiro, 1997.

[Fab01] Fabbrini, F., Fusani, M., Gnesi, S., Lami, G. An Automatic Quality Evaluation for Natural Language Requirements. In *Seventh International Workshop on Requirements Engineering: Foundation for Software Quality (REFSQ'2001)*

[Fli03] Fliedl, G., Kop, C., Mayr, H.C., From Scenarios to KCPM Dynamic Schemas: Aspects of Automatic Mapping. In Düsterhöft, A. and Thalheim, B., editors, *Proceedings of the 8th International Conference on Applications of Natural Language and Information Systems (NLDB'2003),* pp. 91-105, GI Edition, Germany, 2003.

[Ger02] Gervasi, V. and Nuseibeh, B., Lightweight Validation of Natural Language Requirements. *Software Practice and Experience*, 32, pp. 113-133, 2002.

[Gog94] Goguen, J., Requirements engineering as the reconciliation of social and technical issues. In Goguen, J. and Jirotka, M., editors, *Requirements Engineering: Social and technical issues.* Academic Press, London, 1994.

[Hop97] Hoppenbrouwers, J., et al., NL Structures and Conceptual Modelling: Grammalizing for KISS. In: R.P. van de Riet, editor, *Journal of Data and Knowledge Engineering*, volume 23, pp. 79-92, 1997.

[Hull02] Hull, M. E. C., Jackson, K.; Dick, A. J. J., Requirements Engineering, Springer, Berlin, 2002.

[Jac95] Jackson, M., Software Specifications and Requirements: a lexicon of practice, principles and prejudices, Addison-Wesley, New York, 1995.

[Jac98] Jackiewicz, A., L'expression de la causalité dans les textes. Contribution au filtrage sémantique par une méthode informatique d'exploration contextuelle. PhD thesis, Université Paris-Sorbonne, 1998.

[Jou97] Jouis, C., Widad, M., Biskri, I., Le Priol, F., Vers la spécification et l'extension des relations terminologiques : typologie et insertion dans un système de significations des relations. In *5e journée Scientifique du Réseau Lexicologique Terminologie Traduction, LTT'97*, France, 1997.

[Kem90] Kemmerer R.,A. Integrating formal methods into the development process. *IEEE Software*, 7(5):37–50, 1990.

[Kov98] Kovitz, B., Practical Software Requirements, Manning, U.K., 1998.

[Lau02] Laublet P., Nait-Baha L., Jackiewicz A., Djioua B., Collecte d'information textuelles sur le Web selon différents points de vue, In Céline Paganelli, editor, *Interaction homme-machine et recherche d'information (chapitre 8)*, Hermes , Paris, 2002.

[LeP00] Le Priol, F., Extraction et capitalisation automatiques de connaissances a partir de documents textuels. PhD thesis, Université Paris-Sorbonne, 2000.

[Min97] Minel, J-L., Nugier, S., Piat, G., How to Appreciate the Quality of Automatic Text Summarization? Examples of FAN and MLUCE Protocols and their ; Results on SERAPHIN, In *Workshop Intelligent Scalable Text Summarization EACL 97*, pp. 25-30, 1997.

[Min02] Minel, J-L., Filtrage sémantique. Du résumé automatique à la fouille de textes, Hermes, Paris, 2002.

[Nus00] Nuseibeh, B. and Easterbrook, S., Requirements Engineering: A Roadmap. In Finkelstein, A., editor, *The Future of Software Engineering,* ICSE, London, 2000.

[Osb96] Osborne,M. and MacNish, C.K., Processing natural language software requirement specifications. In *2nd IEEE International Conference on Requirements Engineering (ICRE'96)*, IEEE Press, U.S., 1996.

[Ray00] Rayson, P., Garside, R., and Sawyer, P. Assisting requirements engineering with semantic document analysis. In *Proceedings of Content-based multimedia information access RIAO 2000 International Conference*, pp. 1363 – 1371, France, 2000.

[Rol92] Rolland, C. and Proix, C., Natural Language Approach to Conceptual Modelling. In Loucopoulos, P. and Zicari, R., editors, Conceptual Modelling, Database and CASE: an integral view of information systems. Wiley, 1992.

[Zav97] Zave, P. and Jackson, M., Four dark corners in requirements engineering, *ACM Computing Surveys,* 6(1), pp.1-30, January 1997.