

# Research on NLP for RE at CNR-ISTI: a Report

Stefania Gnesi  
ISTI-CNR  
Pisa, Italy  
stefania.gnesi@isti.cnr.it

Alessio Ferrari  
ISTI-CNR  
Pisa, Italy  
alessio.ferrari@isti.cnr.it

## Abstract

**[Team Overview]** The Formal Methods & Tools (FMT) group of CNR-ISTI focuses on the study and development of formal methods and tools to support software development processes. **[Past Research]** FMT started working on requirements formalisation through natural language processing (NLP) at the end of the nineties. This stream of research evolved into requirements analysis and defect detection by means of NLP, with a focus on ambiguity, and resulted in the development and application of the QuARS tool for requirements analysis. More recently, the group started working on the analysis of requirements elicitation interviews, in which ambiguity in spoken natural language and other communication defects are studied. **[Research Plan]** In the upcoming years, FMT will devote its effort to the diffusion of a dataset for requirements analysis, to the usage of NLP in product line engineering, and to research in NLP techniques applied to the analysis of interviews.

## 1 Team Overview

The Formal Methods & Tools (FMT) group (<http://fmt.isti.cnr.it/>) of the Institute of Information Science and Technologies, National Research Council (CNR-ISTI) is active in the fields of development and application of formal notations, methods and software support tools for the specification, design and verification of complex computer systems. These efforts are based mostly on the foundational concepts and techniques of process algebras, temporal logics, model checking approaches. Complementary to these more foundational research activities, natural language processing techniques (NLP) have been applied to the analysis of natural language (NL) requirements. The original activity on the application of NLP techniques to NL requirements started in the nineties with an attempt of translation of NL sentences describing safety and liveness properties into temporal logic formulae, to provide a help in their expression, which is not always an easy task due to the ambiguity that can be hidden in the NL description of the properties.

Starting from the experience done in the formalization of NL requirements, we moved to the applications of NLP techniques to the analysis of requirements documents, standards descriptions and public administration documents with the purpose of detecting linguistic defects that affect the quality of these documents, complementing this activity with the development of tools for performing the automatic analysis of NL defects. More recently, the experience acquired in NLP was used in product line engineering to devise techniques to identify commonalities and variabilities in NL requirements documents and we also started researching on linguistic problems in requirements elicitation interviews.

The remainder of the paper is structured as follows. In Sect. 2 we present the research done in our laboratory in the past years, and in Sect. 3 we outline the directions of research that we plan to explore.

## 2 Past Research on NLP for RE

### 2.1 Formalization of NL Requirements

The first approach in the application of NLP techniques in requirements engineering at FMT was the development of a prototype assistant, NL2ACTL [FAN94], for providing an automatic translation of NL sentences expressing properties of reactive systems, into formulae of an action-based temporal logic ACTL. NL2ACTL was realized using a general development environment, PDGE, for NLP and it was interfaced with a verification environment for the check of behavioural and logical properties of reactive systems. NL2ACTL was supported by a grammar which embodies the knowledge of particular conventions which allows the informative gaps of jargon expressions to be completed, while avoiding *ambiguities*. Moreover, it was built with the purpose of recognizing as many sentences as possible, and of defining a semantics for each sentence removing ambiguities in the properties description. NL2ACTL was implemented in Common Lisp running on a Macintosh SE 30 and it is no longer available.

### 2.2 Quality Analysis of NL Requirements: Techniques and Tools

NL requirements are widely used in software industry, at least as the first level of description of a system. Unfortunately they are often prone to errors and this is partially caused by interpretation problems due to the use of NL itself. An evaluation of NL requirements to address part of the interpretation problems due to linguistic problems was considered an interesting research problem. However, as any other evaluation process, the quality evaluation of NL software requirements needs the definition of a quality model.

Working in collaboration with Vincenzo Gervasi and Salvatore Ruggieri, we distinguished four quality types, namely *syntactic*, *structural*, *semantic*, and *pragmatic* [FAB98, FAB00, FAB01a, FAB01b]. The quality model was successively refined [BER06] to include those ambiguities, which were not already included in the initial quality model, described by Berry, Kamsties, and Krueger [BER03].

The quality model was the basis for implementing a tool, called QuARS<sup>1</sup> – Quality Analyzer for Requirement Specifications, for analyzing NL requirements in a systematic and automatic way [GNE05]. QuARS allows the requirements engineers to perform an initial parsing of the requirements for automatically detecting potential linguistic defects that can determine interpretation problems at the subsequent development stages of a software product. This tool is also able to partially support the consistency and completeness analysis by clustering the requirements according to specific topics [FAB04]. The quite simple clustering approach of QuARS was later improved by the application of a clustering algorithm to exploit lexical and syntactic relationships occurring between natural language requirements for grouping together similar requirements contained in a requirements document [FER12]. The QuARS approach, oriented to requirements, was later extended to address defects in public administration documents, leading to the development of online tool QuOD (Quality checker for Official Documents)<sup>2</sup>.

The approach provided by QuARS is mainly focused on lexical and syntactic quality aspects, while the pragmatic aspect, which depends on the reader of the requirements, is not taken into account. To address the challenge of pragmatic ambiguity, we proposed a novel algorithm that, based on the automatic construction of knowledge graphs that represent the domain-specific information of different subjects, is able to identify potential reader-dependent discrepancies in the interpretation of the same requirements [FER12], [FER14b]. More recently, with the use of Wikipedia crawling and word-embeddings, we devised an approach for measuring the degree of ambiguity of computer science terms (e.g., system, interface, code), when they are used in different domains. The approach, experimented on five different domains, paves the basis for domain-specific ambiguity detection.

The quality of NL requirements is not limited to linguistic defects, but is also affected by the structure of the specification, and by its completeness. In [FER13b], we proposed a novel clustering algorithm named Sliding Head-Tail Component (S-HTC), to improve the structure of requirements specifications, and we applied it to a requirements standard of the railway domain (583 requirements), showing the potential of the approach. In [FER14c], we defined a method to measure and improve the completeness of requirements documents, with respect to the input documents of the requirements definition phase, such as preliminary specifications, transcripts of meetings with the customers, *etc.*. The method is based on the automatic identification of concepts and relationships in the input documents, and in the assessment that such elements are transferred to the requirements document.

---

<sup>1</sup><http://quars.isti.cnr.it/>

<sup>2</sup><http://narwhal.it/quod/index.html>

### 2.3 Quality Analysis of NL requirements: Application Experiences

Use cases are powerful tools to capture functional requirements for software systems. They allow structuring requirements according to user goals, and provide a means to specify the interaction between a certain software system and its environment. As part of the ITEA project CAFFÉ [FAN02], we initiated with Nokia a collaboration on the use of methods based on a linguistic approach with the aim to collect metrics and perform a qualitative analysis on the natural-language-based use case modelling technique used by the company to specify functional requirements for the mobile phone software user interface [FAN02] [FAN03].

Empirical experiments to assess the impact in terms of effectiveness and efficacy of the automation in the requirements review process of a software company are needed to evaluate the usability and applicability of NLP based tools for the quality analysis of textual requirements [LAM07]. Three different experiments may be cited in this direction. One concerns the use of QuARS, in collaboration with Siemens CNX R&D Labs, on telecommunication requirement documents [BUC05]. In [BUC08] and [ROS17] two different experiences in the application of NLP techniques have been developed to automatically identify quality defects in natural language requirements in the Railway Domain. In [BUC08] an customization of QuARS, QuARS Express was used to evaluate the quality of a large set of requirements developed in a EU project. In [ROS17] we report the experience done within a collaboration between a world-leading railway signalling company, the University of Florence, and ISTI-CNR to investigate the feasibility of using NLP for defect identification in the requirements documents of the company. The experience shows that existing rule-based NLP approaches need to be incrementally tailored to the specific language of a company to achieve a sufficient degree of accuracy.

### 2.4 Analysing NL requirements to Identify of Commonalities and Variabilities in Product Lines

NL documents can be a precious source to identify variability information and this information can be later used to define feature models from which different systems can be instantiated and a company who wishes to enter an established market with a new, competitive product is required to analyse the product solutions of the competitors from available documentation as for example brochures. Identifying and comparing the *features* provided by the other vendors might greatly help during the market analysis. In this context, we have devised and applied a strategy, based on advanced information extraction NLP technologies, for mining common and variant features from available documents belonging to different vendors [FER13a, FER14a, FER15].

Still in the field of product line engineering, we used NLP techniques to identify variation points in single requirements documents, based on the following rationale. NL is intrinsically ambiguous, and this is seen often as a possible source of problems in the later interpretation of requirements. However, ambiguity or underspecification at requirements level can in some cases give an indication of possible variability, either in design choices, in implementation choices or configurability. Taking into account the results of our previous analyses conducted on different requirements documents with NLP analysis tools, a first classification of the forms of ambiguity that indicate variation points have been proposed starting from the analysis of documents describing real systems [FAN17]. One of the takeaway messages of [FER13a, FER14a, FER15, FAN17] was that the ambiguity defects found in NL requirement documents may be a means to extract variability issues. The underlying intuition is that often ambiguity in requirements is due to the (conscious or subconscious) need to postpone choices for later decisions in the implementation of the system.

### 2.5 Requirements Elicitation Interviews

Ambiguity and communication problems in requirements engineering do not affect solely the requirements documentation activity, but also requirements elicitation. Interviews are the most common and effective means to perform requirements elicitation, and in [FER15] we started studying the phenomenon of ambiguity in requirements elicitation interviews. Based on arranged interviews, we categorised the different types of ambiguities that were observed, and we proposed a novel classification of ambiguity, to take into account the pragmatic, subject-dependent facet of the phenomenon. The categorisation was later extended in [FER16b]. Stemming from this work, we identified which were the dominant cues of ambiguity [FER16a], showing that under-specified terms (e.g., system, area) were dominant with respect to, e.g., vague terms and pronouns, more common sources of ambiguity in written documents. We also noticed that half of the cases encountered in practice were not triggered by single terms, while entire sentences, and their communication context, should be taken into account to explain them. Therefore, in [FER17e], we proposed to use argumentation theory to explain these cases in a formal way.

One of the take away messages of [FER15] was that ambiguity in interviews could be seen as a resource to disclose tacit knowledge. Indeed, the occurrence of an ambiguity might reveal the presence of unexpressed, system-relevant knowledge that needs to be elicited. To leverage the usefulness of ambiguity, in [FER17b] we proposed a method to review interview

recordings, spotting out ambiguities that were not discovered during the interview. We also showed that different subjects identify highly different ambiguities in interviews. Within this streamline of research, we also went beyond ambiguity, and looked also at other defects in interviews, focusing on the mistakes that are committed by student analysts [FER17a].

Although the work on interviews is not strictly NLP-related, it paves the theoretical bases for future studies in speech processing (see Sect. 3), which may be made possible by the recent, and future, advances in this field.

### 3 Research Plan on NLP for RE

Given the recent advances in NLP technologies, FMT is in the first line to further apply and tailor available approaches to the requirements engineering context and our current and planned research activity mainly concerns (a) the delivery of a complete dataset of requirements documents to ease NLP experiments, (b) the assessment of techniques for commonality and variability identification, and (c) the study of NLP and speech processing techniques in the context of interviews.

- a) As highlighted in our recent vision paper [FER17c], one of the major challenges posed by novel NLP technologies, mostly based on machine learning, is the need for large datasets, which are required for training and testing machine learning algorithms. Therefore, we have defined a requirements dataset named PURE (Public REquirements dataset) [FER17f], and made it available to the community<sup>3</sup>. We plan to incrementally extend the dataset, also based on the contribution of other researchers, and we are formatting the different documents available to reach a common XML format.
- b) Another line of research that we are pursuing is related to the application of ambiguity detection techniques to the discovery of variation points in requirements documents. Specifically, we have recently applied the approach described in [FAN17] in three documents belonging to different domains [FAN18], and we are currently performing a more extensive, and empirically grounded, experimentation to validate the approach.
- c) Requirements elicitation interviews are another NLP-related topic of research to which we are devoting our efforts. Specifically, we performed an empirical evaluation of the interview review approach presented in [FER17b] in collaboration with Kennesaw State University and University of Sydney, and the results have been recently accepted for publication [SPO18]. A long-term research direction in interview analysis is the application of NLP and speech processing to interview transcripts and recordings. Specifically, we wish to automatically identify conversation topics in interview transcripts, by leveraging technologies for information extraction that we already applied in requirements [FER14c]. This will be useful to retrieve requirements sources later in the development process, when requirements are formalised into documents. It could be used also in those contexts in which requirements are never documented, and directly pass from their spoken form to software. Of course, this basic idea requires interviews to be transcribed, which is a time consuming process. In August 2017, Microsoft Research published a conversational speech recognition system that achieves an error-rate that is comparable to human transcribers [XIO17]. We argue that these advances can be leveraged by our future research in NLP for RE, and will enable a more in-depth understanding, and consequent control, of the process of transforming ideas into spoken sentences, written requirements, and, finally, software.

### References

- [FAN94] A. Fantechi, S. Gnesi, G. Ristori, M. Carenini, M. Vanocchi, P. Moreschini: Assisting Requirement Formalization by Means of Natural Language Translation. *Formal Methods in System Design* 4(3): 243-263, Springer, 1994.
- [FAB98] F. Fabbrini, M. Fusani, V. Gervasi, S. Gnesi, S. Ruggieri: On Linguistic Quality of Natural Language Requirements. 4th REFSQ, 57-62, Presses Universitaires de Namur, 1998.
- [FAB00] F. Fabbrini, M. Fusani, S. Gnesi, G. Lami: Software requirements verification by natural language analysis: a CNR initiative for italian SME's. In: *Ercim News*, vol. 40, 52 - 53, 2000.
- [FAB01a] F. Fabbrini, M. Fusani, S. Gnesi, G. Lami: The linguistic approach to the natural language requirements quality: benefit of the use of an automatic tool, 26th Annual NASA Software Engineering Workshop, 97-105, IEEE, 2001.
- [FAB01b] F. Fabbrini, M. Fusani, S. Gnesi, G. Lami, An automatic quality evaluation for natural language requirements, 7th REFSQ, 2001.
- [GNE05] S. Gnesi, G. Lami, G. Trentanni: An automatic tool for the analysis of natural language requirements. *Computer. Systems: Science & Engineering*. 20(1), CRL Publishing, 2005.
- [FAB04] F. Fabbrini, M. Fusani, S. Gnesi, G. Lami: Automatic clustering of non-functional requirements. *IASTED Conf. on Software Engineering and Applications 2004*, 672-677, IASTED/ACTA, 2004.

---

<sup>3</sup><http://fmt.isti.cnr.it/nlreqdataset/>

- [BER06] D.M Berry, A.Bucchiarone, S. Gnesi, G. Lami, G. Trentanni, A new quality model for natural language requirements specifications, 12th REFSQ, 2006.
- [FER12] A. Ferrari, S. Gnesi, G. Tolomei, A clustering-based approach for discovering flaws in requirements specifications. SAC,1043-1050, ACM, 2012.
- [FAN02] A. Fantechi, S. Gnesi, G. Lami, A. Maccari: Application of Linguistic Techniques for Use Case Analysis. 10th RE: 157-164, IEEE, 2002.
- [FAN03] A. Fantechi, S. Gnesi, G. Lami, A. Maccari: Applications of linguistic techniques for use case analysis. *Requir. Eng.* 8(3): 161-170, Springer, 2003.
- [LAM07] G. Lami, R. W. Ferguson: An empirical study on the impact of automation on the requirements analysis process. In: *Journal of Computer Science and Technology*, vol. 22 (3) pp. 338 - 347. Springer, 2007.
- [BUC05] A. Bucchiarone, S. Gnesi, P. Pierini: Quality Analysis of NL Requirements: An Industrial Case Study. 13th RE: 390-394, IEEE, 2005.
- [BUC08] A. Bucchiarone, S. Gnesi, G. Trentanni, A. Fantechi: Evaluation of Natural Language Requirements in the MODCONTROL Project, *ERCIM News* 2008(75), 2008.
- [ROS17] B. Rosadini, A.Ferrari, G. Gori, A. Fantechi, S. Gnesi, I. Trotta, S. Bacherini: Using NLP to Detect Requirements Defects: An Industrial Experience in the Railway Domain. 23rd REFSQ, LNCS 10153, 344-360, Springer 2017.
- [FER13a] A. Ferrari, G. O. Spagnolo, F. Dell'Orletta: Mining commonalities and variabilities from natural language documents. 17th SPLC:116-120, ACM, 2013.
- [FER14a] A. Ferrari, G. O. Spagnolo, G. Martelli, S. Menabeni: From commercial documents to system requirements: an approach for the engineering of novel CBTC solutions. *STTT* 16(6): 647-667, Springer, 2014.
- [FER15] A. Ferrari, G.O.Spagnolo, S. Gnesi, F. Dell'Orletta: CMT and FDE: tools to bridge the gap between natural language documents and feature diagrams. 19th SPLC: 402-410, ACM, 2015.
- [FAN17] A. Fantechi, S. Gnesi, L. Semini: Ambiguity defects as variation points in requirements. 11th VaMoS: 13-19, ACM, 2017.
- [FAN18] A. Fantechi, A. Ferrari, S. Gnesi, L. Semini: Hacking an Ambiguity Detection Tool to Extract Variation Points: an Experience Report, 12th VaMoS: 43-50, ACM, 2018.
- [FER15] A. Ferrari, P. Spoletini, S. Gnesi: Ambiguity as a resource to disclose tacit knowledge. 23rd RE: 26-35, IEEE, 2015.
- [FER16a] A. Ferrari, P. Spoletini, S. Gnesi: Ambiguity Cues in Requirements Elicitation Interviews. 24th RE: 56-65, IEEE, 2016.
- [FER17a] B. Donati, A. Ferrari, P. Spoletini, S. Gnesi: Common Mistakes of Student Analysts in Requirements Elicitation Interviews. 23rd REFSQ, LNCS 10153, 148-164, Springer,2017.
- [FER16b] A. Ferrari, P. Spoletini, S. Gnesi: Ambiguity and tacit knowledge in requirements elicitation interviews. *Requir. Eng.* 21(3), 333-355, Spinger, 2016.
- [FER17b] A. Ferrari, P. Spoletini, B. Donati, D. Zowghi, S. Gnesi: Interview Review: Detecting Latent Ambiguities to Improve the Requirements Elicitation Process. 25th RE, 400-405, IEEE, 2017.
- [FER12] A. Ferrari, S. Gnesi: Using collective intelligence to detect pragmatic ambiguities. 20th RE, 191-200, IEEE, 2012.
- [FER14b] A. Ferrari, G. Lipari, S. Gnesi, G. O. Spagnolo: Pragmatic ambiguity detection in natural language, *AIRE* 2014, 1-8, IEEE, 2014.
- [FER13b] A. Ferrari, S. Gnesi, G. Tolomei: Using Clustering to Improve the Structure of Natural Language Requirements Documents. 19th REFSQ, LNCS 7830, 34-49, Springer, 2013.
- [FER14c] A. Ferrari, F. Dell'Orletta, G. O. Spagnolo, S. Gnesi: Measuring and Improving the Completeness of Natural Language Requirements. 20th REFSQ: 23-38, LNCS 8396, Springer, 2014.
- [FER17c] A. Ferrari, F. Dell'Orletta, A. Esuli, V. Gervasi, S. Gnesi: Natural Language Requirements Processing: A 4D Vision. *IEEE Software* 34(6): 28-35, IEEE, 2017.
- [FER17d] P. Spoletini, A. Ferrari: Requirements Elicitation: A Look at the Future Through the Lenses of the Past. 25th RE 2017, 476-477, IEEE, 2017.
- [FER17e] Y. Elrakaiby, A. Ferrari, P. Spoletini, S. Gnesi, B. Nuseibeh: Using Argumentation to Explain Ambiguity in Requirements Elicitation Interviews. 25th RE: 51-60, IEEE, 2017.
- [BER03] D. M. Berry, E. Kamsties, M. M. Krieger: From Contract Drafting to Software Specification: Linguistic Sources of Ambiguity. University of Waterloo, 2017. <https://cs.uwaterloo.ca/~dberry/handbook/ambiguityHandbook.pdf>
- [FER17f] A. Ferrari, G. O. Spagnolo, S. Gnesi: PURE: A Dataset of Public Requirements Documents. 25th RE: 502-505, IEEE, 2017.
- [SPO18] P. Spoletini, A.Ferrari, M. Bano, D. Zowghi, S. Gnesi: Interview Review: an Empirical Study on Detecting Ambiguities in Requirements Elicitation Interviews. 24th REFSQ, to appear.
- [XIO17] W. Xiong, L. Wu, F. Alleva, J. Droppo, X. Huang, A. Stolcke: The Microsoft 2017 Conversational Speech Recognition System. Microsoft Technical Report MSR-TR-2017-39, 2017. <https://arxiv.org/abs/1708.06073>.