Chapter 5
From Textual Scenarios to Message Sequence Charts
Leonid Kof
Technische Universität München, Germany

ABSTRACT
Requirements engineering, the first phase of any software development project, is the Achilles’ heel of the whole development process, as requirements documents are often inconsistent and incomplete. In industrial requirements documents natural language is the main presentation means. In such documents, the system behavior is specified in the form of use cases and their scenarios, written as a sequence of sentences in natural language. For the authors of requirements documents some facts are so obvious that they forget to mention them. This surely causes problems for the requirements analyst. By the very nature of omissions, they are difficult to detect by document reviews: Facts that are too obvious to be written down at the time of document writing, mostly remain obvious at the time of review. In such a way, omissions stay undetected. This book chapter presents an approach that analyzes textual scenarios with the means of computational linguistics, identifies where actors or whole actions are missing from the text, completes the missing information, and creates a message sequence chart (MSC) including the information missing from the textual scenario. Finally, this MSC is presented to the requirements analyst for validation. The book chapter presents also a case study where scenarios from a requirement document based on industrial specifications were translated to MSCs. The case study shows feasibility of the approach.

DOI: 10.4018/978-1-60566-758-4.ch005
Some kind of requirements document is usually written at the beginning of every software project. The majority of these documents are written in natural language, as the survey by Mich, Franch, and Novi Inverardi (2004) shows. This results in the fact that the requirements documents are imprecise, incomplete, and inconsistent. The authors of requirements documents are not always aware of these document defects. From the linguistic point of view, document authors introduce three defect types, without perceiving them as defects (Rupp (2002), p. 169:

- **Deletion**: “...is the process of selective focusing of our attention on some dimensions of our experiences while excluding other dimensions. Deletion reduces the world to the extent that we can handle.”
- **Generalization**: “…is the process of detachment of the elements of the personal model from the original experience and the transfer of the original exemplary experience to the whole category of objects.”
- **Distortion**: “…is the process of reorganization of our sensory experience.”

It is one of the goals of requirements analysis, to find and to correct the defects of requirements documents. This book chapter focuses on the “deletion”-defects in scenarios. Deletion manifests itself in scenarios in the form of missing action subjects or objects or even in whole missing actions. One of the reasons for the deletion may be the fact that some information is too obvious for the author of the requirements document, so that she finds it unnecessary to write down this information.

Missing information can affect any part of the requirements document. In the case of scenarios it is especially crucial because in this case some portions of the desired system behavior are just omitted. It is the goal of the approach presented in this book chapter, to identify missing parts of the scenarios written in natural language and to produce message sequence charts (MSCs) containing the reconstructed information. These MSCs must be validated and can then be used in further project run.

- **Terminology**: For the remainder of the book chapter we use the following terminology: A scenario is a sequence of natural language sentences. Every sentence consists of several segments (subordinate phrases). An MSC consists of a set of actors, a sequence of messages sent and received by these actors, and a sequence of conditions interleaved with the message sequence. Figure 1 illustrates the introduced terminology.

- **Outline**: The remainder of the book chapter is organized as follows: Section “Case Study: the Instrument Cluster” introduces the case study used to evaluate the presented approach, and Section “Computational Linguistics Prerequisites” introduces the applied linguistic tools. Section “From Textual Scenarios to Message Sequence Charts” is the core of the whole book.