



# Business processes modeling: effect of notation method in the ambiguity level

## Modelagem de processos de negócio: efeito do método de notação no nível de ambiguidade

Antônio Eduardo Carvalho e Silva<sup>1</sup>   
Rogerio Atem de Carvalho<sup>2</sup>   
Hudson Silva Ferreira<sup>2</sup>

**How to cite:** Silva, A. E. C., Carvalho, R. A., & Ferreira, H. S. (2019). Business processes modeling: effect of notation method in the ambiguity level. *Gestão & Produção*, 26(1), e1624. <https://doi.org/10.1590/0104-530X1624-19>

**Abstract:** The aim of this controlled and randomized experimental study was to analyze the functional requirement specification. The possible impacts of process modeling notation method in the ambiguity level of natural language specification were evaluated. The notations were used as a tool to express the user's needs about buying process in e-commerce applications - typical Enterprise Information System use case. From a low-fidelity prototype, that represents the purchasing process in a virtual store, 43 undergraduate computer science students were grouped according to the notations performed: natural (Brazilian Portuguese) language control group (CG); finite state machine experimental group (FSMEG); and combined notation experimental group (CNEG) that used *manual annotation of semantic roles based on finite state machine*. It was used the concept that deals with ambiguity as inconsistent information that leads to multiple interpretations of functional requirement specification document. By a Test-based Reading technique, associated with the use of appropriate quality metric, a one-way analysis of variance (completely randomized design) was conducted to know if notation method, as a primary factor, affects the ambiguity level. The inter-group comparison suggests that: combined notation is the best method to minimize ambiguity levels in requirement specification; and that requirement specification expressed in finite state machine generates the higher ambiguity level among assessed notations. However, these trends are not statistically significant. Generally, it was observed that the independent variable, notation method, does not affect the ambiguity level in the process described in the functional requirement specification.

**Keywords:** Management of production systems; Business processes; Requirements specification; Ambiguity level.

**Resumo:** A proposta deste estudo experimental controlado e aleatorizado foi analisar a especificação de requisitos funcionais. Avaliaram-se os possíveis impactos do método de notação de modelagem de processo no nível de ambiguidade presente na especificação em linguagem natural. As notações foram utilizadas como instrumento para manifestar as necessidades do usuário quanto ao processo de compra em aplicações de comércio eletrônico - caso de uso típico de sistemas de informação empresarial. A partir de um protótipo de baixa fidelidade, que representa o processo de compra numa loja virtual, 43 estudantes de graduação em ciência da computação foram agrupados de acordo com as notações desempenhadas: grupo controle linguagem natural (GC); grupo experimental máquina de estados finitos (GEMEF); e o grupo experimental notação combinada (GENC), que utilizou anotação manual de papéis semânticos baseados em máquina de estados finitos. Foi utilizado o conceito que trata da ambiguidade como informações inconsistentes que levam a múltiplas interpretações do documento de especificação dos requisitos funcionais. Pela técnica de leitura baseada em teste, associada à utilização da métrica de qualidade apropriada, foi conduzida uma análise de variância de fator único com delineamento completamente casualizado para saber se o método de notação, como fator primário, afeta o nível de ambiguidade. A comparação intergrupo sugere que: a notação combinada é o melhor método para reduzir o nível de ambiguidade da especificação de requisitos; e que a especificação de requisitos expressa em máquina de estados finitos gera o maior nível de ambiguidade dentre as notações avaliadas. No entanto, estas tendências não são estatisticamente significativas. De forma geral, observou-se que a variável independente, método de notação, não afeta o nível de ambiguidade do processo descrito na especificação de requisitos funcionais.

**Palavras-chave:** Gestão de sistemas de produção; Processos de negócio; Especificação de requisitos; Nível de ambiguidade.

<sup>1</sup> Laboratório de Engenharia de Produção – LEPROD, Centro de Ciência e Tecnologia – CCT, Universidade Estadual do Norte Fluminense Darcy Ribeiro – UENF, Av. Alberto Lamego, 2000, Parque Califórnia, CEP 28013-602, Campos dos Goytacazes, RJ, Brasil, e-mail: eduardo.oak80@gmail.com

<sup>2</sup> Núcleo de Sistemas de Informação – NSI, Instituto Federal de Educação, Ciência e Tecnologia Fluminense – IFF, Rua Dr. Siqueira, 273, Parque Dom Bosco, CEP 28030-130, Campos dos Goytacazes, RJ, Brasil, e-mail: ratem@iff.edu.br; silvaferreira.hsf@gmail.com

Received Sept. 14, 2017 - Accepted May 21, 2018

Financial support: Secretaria de Educação Profissional e Tecnológica do Ministério da Educação (SETEC/MEC).

## 1 Introduction

How can user's needs be translated to a programming language? In agile methodology, the user stories used to specify requirements serve to guide the system development throughout its life cycle. However, considering the principles of classical methodology, the requirements specification happens in early and well-defined stage of systems development process (Leffingwell, 2011; Cohn, 2004). Regardless of adopted methodology for system development, usually the requirements are specified with an ambiguous language – the natural language (NL).

It is possible to analyze the NL ambiguity from two different perspectives. On the one hand, the ambiguity can be considered a benefit that motivates dialectics, ensuring the philosophical thinking accuracy. On the other hand, through the scientific perspective adopted in this study, the ambiguity phenomenon is an adverse factor that affects the requirements specification (Piantadosi et al., 2012; Raffin, 2009; Walia & Carver, 2009; Sutcliffe et al., 1999; Card, 1998). To get rid of ambiguity, in this context, represents an important step to fill a gap between the business model and the processes automation. The benefit of semantic annotation for this purpose is notorious for making the process executable by machine. An example is reported within the semantic business process modeling (SBPM) where there is an integration between model and metamodel to form an ontology that defines the semantic roles (Yan, 2007; Lautenbacher et al., 2008).

The problem in the intrinsic NL's ambiguity can make the organization knowledge flow difficult between the parties engaged in the process. This communication gap is especially harmful in an enterprise information system (EIS) development, because there is a close relation between the business logic model (BLM) and the code that provides the system. Therefore, to the user figure as a stakeholder, bears, for association, the roles of business user and system user. This complex user role needs means that enable its integration to the development process (Coughlan & Macredie, 2002; Coughlan et al., 2003; Van Grondelle et al., 2010; Jørgensen et al., 2007; Carvalho et al., 2010, 2011, 2012).

From all this considerations that assume the ambiguity as a problem to be solved, derives the following question which motivate this research: *How to reduce the ambiguity level in the functional requirements specification predicted in the development process of enterprise information systems?*

The general objective of this research is to analyze the effect of process modeling notation method in the ambiguity level of specification of EIS's functional requirements. The specific objectives were: (i) to identify the data distribution profile related to the ambiguity level of specification of EIS's functional

requirements; (ii) to propose a combined notation method for processes modeling; (iii) to apply the combined notation method on the business process modeling; (iv) to compare the effect of the process modeling notation method in the ambiguity level of requirements specification.

## 2 Notation methods for business process modeling

It is known that formal notation shows a potential reduction of ambiguity inherent to the NL and the restriction imposed by the deterministic finite automats (DFA's), as the finite state machine (FSM), reduces the description potential of NL (Lopes, 2008; Menezes, 2008a; Hopcroft et al., 2003). However, it is necessary to create mechanisms to extract the value of the expressed requirements in NL. For this value to be appropriately recognized in the information system universe it is necessary to facilitate the knowledge flow between the consumers and other parts interested in the process, stimulating the dialogue to improve the organizational processes continuously (Coughlan & Macredie, 2002; Coughlan et al., 2003; Van Grondelle et al., 2010; Jørgensen et al., 2007; Piantadosi et al., 2012; Leffingwell, 2011; Carvalho et al., 2010, 2011, 2012).

Within the EIS's, where the consumer role is empowered by the business and system user figures, its active participation in the productive process is essential in the pursuit of the business mission (Kotler et al., 2012).

The behavior driven development (BDD), forseen in the agile methodology, enables the directed testing implementation from the user's necessity expressed by the triad Given-When-Then (GWT). In this context, the direct connection "client's necessity-information system (IS)" would spare the need of the BLM formalization with a modeling notation. Actually, the BLM would be implicit in the GWT triad predicted on BDD. This fact is due to the possible analogy between the triad and formal FSM notation, where *Given* would be a state, *When* an action and *Then* the resulting state. The GWT triad would, therefore, characterize a kind of combined notation (CN) which restricts the NL with the FSM disguised by the GWT structure. However, the NL used to manifest the client's needs, can affect, with its inherent ambiguity, knowledge flow quality both on the translation that occurs on the axis "BLM-IS" and on the direct connection "client's needs-IS". Therefore, business language driven development (BLDD), conceptual model that comes from the BDD, had as an additional proposition the use of graph representation. This representation contributes to promote the direct connection "client's need-IS", seeking to preserve the knowledge flow quality and

a closer relation description-code (Carvalho et al., 2010, 2011, 2012; Turnquist, 2011).

## 2.1 Natural language

It should be ensured the means that support the user's communication with the other actors of the business/system development process. However, the problem of NL's ambiguity is a limiting factor for this communication. The challenge in promoting the user's involvement in the knowledge modeling is represented by the lack of user's training in the formal knowledge representation techniques (Van Grondelle et al., 2010; Carvalho et al., 2010, 2011, 2012).

The controlled NL is a method that helps reducing the ambiguity level of NL, collaborating to the user's inclusion in the knowledge life cycle (Ross, 2009; Van Grondelle et al., 2010). Besides this method, some automated approaches that use the natural language processing (NLP) also contribute for this reduction (Jurafsky & Martin, 2000; Manning & Schütze, 2000; Bird et al., 2009).

## 2.2 Finite state machine

The FSM is used as a notation method for business process modeling (Van der Aalst, 2011). It's a DFA composed by a quintuple  $M = (Q, \Sigma, \delta, q_0, F)$ . Where:  $Q$  is the set of states;  $\Sigma$  is the set of actions or entries;  $\delta$  is the set of transitions;  $q_0$  is the initial state;  $F$  is the final state (Jurafsky & Martin, 2000).

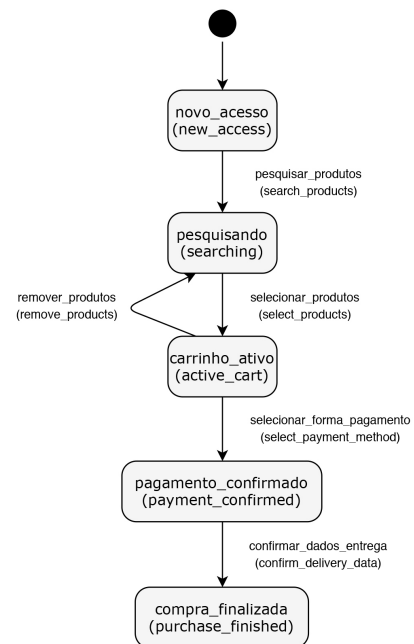
Besides its formal representation, the FSM can be symbolized by a graph or by a state transition table. In the first case, the automaton is represented by a graph  $G(V, A)$ , where  $V$  is a non-empty and finite set of vertices (nodes) and  $A$  is an ordered pairs of  $V$ , called edges (arcs). The kind of graph that represents a FSM is a directed graph, because the edges have an associated direction. An arrow in the graph representation indicates this direction. In the second case, the automaton can be represented by a state transition table, composed by the states, arranged in the lines, and through the entries (actions) arranged in columns (Jurafsky & Martin, 2000; Hopcroft et al., 2003; Menezes, 2008b).

## 2.3 Combined notation

Some languages are, potentially, more ambiguous than others. In an extreme, the formal process modeling notations, like the FSM and the Petri net, have a quite defined semantics and, therefore, tend to be less ambiguous. In the other extreme, the NL presents itself with a lot of inherent ambiguity. It is believed that the NL's substitution by a formal notation could contribute for a considerable reduction in the ambiguity level in system requirement specification.

Although, this substitution could affect, in a harmful way, the comprehension. Therefore, it is suggested that the best approach would be the combination of NL with formal notations. So, the advantages of both languages – natural and formal – could be preserved. With the application of this combined notation method, that characterizes itself by NL's restriction with the use of a formal process modeling notation, there is likely to be a balance between the NL's comprehension and the ambiguity reduction potential through the formal notation (Davis et al., 1993).

What is proposed to CN is the description of the process in NL with manual annotation of semantic roles. The set of semantic roles proposed was constituted by the labels predicted in the FSM formal notation (initial\_state, state, action) and by the delimiter "end", used to indicate the end of the previous label coverage. The labels were assigned to the syntagmas that feature, indeed, the process (figure 1). The point was to verify if the executable semantics of this formal notation – metamodel – make the non-ambiguous information recover possible (Oren et al., 2006; Van der Aalst, 2011; Lautenbacher et al., 2008). This way, instances of the state and action classes can be identified in text snippets. The labels were attributed in the LibreOffice Writer (Version: 4.2.4.2) text editor with the command Insert > Comment or through the shortcut Ctrl + Alt + C (LibreOffice, 2013).



**Figure 1.** Low-fidelity prototype of the buying process in a virtual store: graph representation of a Finite State Machine – experimental study. Note: translation in parentheses.

### 3 Perspective-based reading

It's a reading technique used in system inspection, which aim is to analyze a software description from stakeholders perspective to identify fails. The perspective-based reading (PBR) is systematic, focused, customizable and capable of improving from the application and training experience. It can be used to inspect various document types, including requirement specifications, code and project models. It was created to reduce human influence in inspection results. Experimental researches suggest that the PBR technique is significantly more effective in relation to traditional reading techniques as checklist-based reading and the ad-hoc reading (Lahtinen, 2012).

The main idea of the PBR technique is to inspect a document from different reviewers perspectives. In PBR the perspectives come from the interested parts, that is, the most relevant people during the project life cycle. The reason to adopt distinctive perspectives is the probability that a document will be best in quality when the interested parts do not detect any defect on it. Typical perspectives for the inspection of a requirement specification document could include the designer, tester and/or user perspective. Because each reader is responsible for only one document point of view, any potential mistake is rigorously analysed. When only one perspective is chosen, the reviewers tend to use their specific knowledge in a more natural way, and therefore, more effective. On the other hand, when the reviewers evaluate multiple perspectives simultaneously, they are assigned to the most appropriate perspective (Lahtinen, 2012). For each one of these perspectives there is a specific form with a series of related questions that guide the reading. This way, when using different perspectives there would be a better comprehension of the defects, once the aspects related to each perspective could be identified in a more effective way (Basili et al., 1996; Sørungård, 1997; Shull et al., 2000; Sommerville, 2007a).

The defects identified by the PBR are the following: missing information; ambiguous information; incorrect fact; extraneous information; and others (Sørungård, 1997). It's a customizable technique because it allows the reviewers – developers, testers and users – to adequate to the desired purpose (Shull et al., 2000). Nevertheless, in this paper scope, the only used perspective was the test one. The called Test-based Reading (TBR) is focused in aspects related to the systems testability, from the information that is presented in the requirement specification (Basili et al., 1996). A PBR scenario is an instruction document with a few pages. For each perspective, one or more scenarios are written so the reviewers could fulfill specific and reproducible actions, besides responding to a set of questions to identify defects (Lahtinen,

2012). A scenario is made of three parts: introduction, instructions and questions.

## 4 Methodology

### 4.1 Target population

This experimental research was conducted with a sample of 43 subjects obtained from a target population constituted by undergraduate students in Technology in Systems Analysis and Development/Information Systems from Instituto Federal de Educação, Ciência e Tecnologia Fluminense (IFF) in the years 2013/2014. The mentioned courses are related to the “Computer Science” area, as classified by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

### 4.2 Dependent variable

The dependent variable chosen was the ambiguity level in the specification of the functional system requirements. The choice of this variable, which in this study was designated simply by ambiguity level, is due to many factors. Among them, it is possible to highlight the following: (a) the non-ambiguous specification is one of the criteria of quality in the requirement description predicted in the software/systems development process (Davis et al., 1993). So, it is recommended that the requirement specification to be non-ambiguous (IEEE, 1998); (b) the availability of a specific reading technique of the requirement specification in NL with the purpose of identifying defects, designated by PBR (Basili, 2013; Sørungård, 1997; Lanubile et al., 1998); (c) the requirement specification ambiguity is broadly reported in the scientific literature as a problem in software/system development process (Walia & Carver, 2009; Sutcliffe et al., 1999; Card, 1998); (d) the ambiguity is described as a kind of defect subject to identification from the use of the PBR technique (Sørungård, 1997); (e) after the ambiguous requirement identification, it is possible to measure the ambiguity level from a quality metric applicable to the software requirement specification (Bokhari & Siddiqui, 2011).

As the recommendation is that the software requirement specification is non-ambiguous, the quality metric reported in the scientific literature, which demonstrates adequacy to this purpose is:

$$Q_a = n_{ui} / n_r \quad (1)$$

where  $n_{ui}$  is the requirement number to which all the reviewers (testers) presented identical interpretations; and  $n_r$  is the total requirements number.

The requirement unit was defined by the transition of FSM because it contains a testable state-action-state triad. This way, it was possible to obtain the percentage



of the requirements interpreted in only one way by all testers. The reference values of Equation 1 are presented in an interval scale  $Q_a$  which varies from 0 (zero) to 1 (one), where the values closer to 0 (zero) represent non-ambiguous requirements; and the values closer to 1 (one) represent ambiguous requirements.

### 4.3 Independent variable

The independent variable used in this study was process modeling notation method. Different methods were applied during the specifications description of the system functional requirements. This variable was designated simply by the term *notation method* and presented 3 levels: NL, FSM and CN. Because it is a non-artificial language, the NL was used as the first level of the independent variable. The other notations, with artificial characteristics, were attributed to the FSM and CN's levels.

FSM is the most basic formal notation of process modeling (Jurafsky & Martin, 2000). The third level of notation method, designated by CN, was composed by a notation combination. This combination characterizes itself by the label attribution of FSM to the NL's description. This manual annotation of semantic roles was destined to the NL's syntagms that were considered relevant by the following aspects: (1) they characterized the process requirements; and (2) they are subject of characterization by the quintuples labels M that composes the FSM notation.

The referred variable choice was motivated by the following factors: (a) it is believed that the NL substitution for a formal notation like the FSM reduces considerably the ambiguity level in the software requirement specification. As this effect occurs over the comprehensibility, it is recommended a combination between the NL with more formal languages like the FSM. Therefore, the advantages of natural and formal languages are preserved (Davis et al., 1993); (b) it is suggested that, in NL, there is an inherent ambiguity and that the use of the FSM language presents itself as an effect the reduction of the intrinsic ambiguity level, by the fact that it demonstrates a well-defined semantics (Davis et al., 1993; IEEE, 1998); (c) during the phase of requirement specification there are two common mistakes: the ones resulting of not knowing the requirements and those resulting by not knowing how to specify them adequately (Davis et al., 1993; Sommerville, 2007b). So, the proposal of CN as a notation method, is to demonstrate the feasibility of specifying the requirements adequately concerning to the ambiguity level.

### 4.4 Pilot-study

It was performed a populational pilot-study with single-blind characteristic in order to adjust the experimental design. The first motivation to this approach was to gather experience with the TBR

technique application (topic 4.8) and give support to guide the choice of the statistical inference technique more appropriate to the experimental study. In this preliminary study, it was analysed the normality assumption in relation to the ambiguity level in the textual process description in brazilian portuguese NL. This description happened during the functional requirements specification of a hypothetic EIS. From a low-fidelity prototype with a high level of abstraction, the subjects were motivated to describe the buying process in a virtual store (Silva et al., 2013). The implications of the result in this preliminary study were very important for the present experimental study.

The tool used to the statistical analyses was RStudio: an Integrated Development Environment (IDE), which is a multiplatform free software that improves the productivity, with an appropriate user interface for the use of R language (Bryer & Speerschneider, 2012; R Core Team, 2013).

On that occasion, when specifying the requirements the subject did not need to restrict the functionalities description predicted in the graph. This worked only as a starting point so the subjects could express the functionalities of what they understand as a buying system in a virtual store. After the data gathering from each subject in the sample, the sentences were analysed with the TBR technique, using the quality metric  $Q_a$ , specified before, to assess requirements ambiguity level.

Because it is a populational study, there was no group division in the first moment. Besides, as all single-blind study, the subject had no information about the phenomenon observed during the experiment – i.e. ambiguity level – only the researchers had such information (Silva et al., 2013).

### 4.5 Experimental study procedures

A randomized, controlled, single-blind trial was conducted. The independent variable, notation method, was designated as the Analysis of Variance (ANOVA) single factor. This way, the sample obtained from the target population was divided in 3 (three) subject groups, each one representing a level of the single factor: NL Control Group (CG); FSM Experimental Group (FSMEG); and CN Experimental Group (CNEG). To the CG was asked the brazilian portuguese NL description of a virtual store buying process, the same way it was done in the pilot-study. To the FSMEG was asked the description of the same process, but, this time using FSM notation to specify the functional requirement. To the CNEG the description was also made in NL, but the subjects in this group assigned FSM labels (initial\_state, state, action) to the NL's description. The overview of the experimental procedure adopted in the system functional requirement evaluation is represented in Table 1.

As only one independent variable was chosen for the study realization, it was applied the One-way ANOVA. The dependent variable was the ambiguity level in the requirement specification.

The low-fidelity prototype used in the experimental study (Figure 1) can be characterized by a more reduced abstraction level in relation to the prototype used in the pilot-study (Silva et al., 2013).

It is possible to represent the transitions  $\delta$  by the own formal notation described in Table 2 or by the state transition table represented in Table 3. The transitions expressed in Table 2 were the 5 requirements evaluated by the testers. These requirements express the prototype adequately. Table 3 illustrates the task to be done by the subjects of the experiment that were assigned to

the FSMEG. When adopting it, a desirable example of the FSM description of the buying process in a virtual store can be observed, as represented by the low-fidelity prototype in Figure 1.

Figure 2 illustrates the application of the CN as an evaluation pattern for the CNEG experimental group. Adopting this pattern makes possible to observe a desirable example of NL description in a virtual store buying process, as represented by the low-fidelity prototype exposed on Figure 1. Besides NL it is also possible to observe the predicted labels in the formal FSM notation being attributed to the syntagms that characterize, indeed, the business process. The label “end” is only an adjustment to indicate the end of the preceded label comprehensiveness.

**Table 1.** Experimental procedure adopted in the evaluation of the system functional requirement specification.

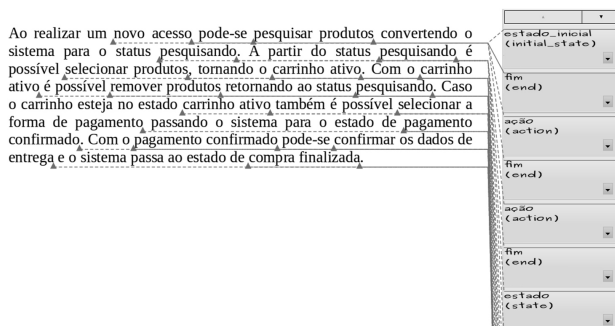
	Experimental Procedure		
	Group	Notation Method	Evaluated Notation
CG	Control	single	NL
FSMEG	Experimental	single	FSM
CNEG	Experimental	combined	NL + FSM

**Table 2.** Low-fidelity prototype requirements represented by the FSM notation transitions – experimental study.

N°	Transitions
1	$\delta(\text{novo\_acesso, pesquisar\_produtos}) = \text{pesquisando}$
2	$\delta(\text{pesquisando, selecionar\_produtos}) = \text{carrinho\_ativo}$
3	$\delta(\text{carrinho\_ativo, remover\_produtos}) = \text{pesquisando}$
4	$\delta(\text{carrinho\_ativo, selecionar\_forma\_pagamento}) = \text{pagamento\_confirmado}$
5	$\delta(\text{pagamento\_confirmado, confirmar\_dados\_entrega}) = \text{compra\_finalizada}$

**Table 3.** State Transition table: FSM notation representation that characterize the buying process in a virtual store – experimental study.

$\delta$	Actions ( $\Sigma$ )					
	pesquisar_ produtos	selecionar_ produtos	remover_ produtos	selecionar_ forma_ pagamento	confirmar_ dados_ entrega	
novo_acesso	pesquisando	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$
pesquisando	$\emptyset$	carrinho_ativo	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$
carrinho_ativo	$\emptyset$	$\emptyset$	pesquisando	pagamento_confirmado	$\emptyset$	$\emptyset$
pagamento_confirmado	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	compra_finalizada	$\emptyset$
compra_finalizada	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$



**Figure 2.** Virtual store buying process description in combined notation, used as an evaluation pattern of the CNEG experimental group. Note: translation in parentheses.

#### 4.6 Experimental design

The ANOVA application depends on the accomplishment of three assumptions: normality, independency and homoscedasticity (Larson & Farber, 2010). The Completely Randomized Design (CRD) was adopted to execute the predicted routine in the One-way ANOVA. The choice of this kind of design is due to the comparative goal proposed in this study. It is probable that other factors also affect the ambiguity level, but the idea is knowing if the primary factor *notation method*, *a priori*-defined, significantly affects the ambiguity level, despite the eventual existence of other nuisance factors (NIST, 2013).

The use of CRD contributes greatly to the independency assumption fulfillment. This fact is due to the randomness of assigning primary factor levels to experimental units. Therefore, the design is completely randomized or casual (Bailey, 2008b; NIST, 2013). This way, all the levels of the primary factor have the same chance to be applied in any experimental unit – experiment subjects (Bailey, 2008a, b).

Finally, so there isn't any initial restriction concerning to the correct application of One-way ANOVA, the following procedure was adopted: after the data collection it was verified the homoscedasticity assumption fulfillment. Before comparing the averages, it is necessary to check the supposition that the homogeneity of variance is reasonable. With this idea, the Levene's test is applicable (NIST, 2013; Bisquerra et al., 2007).

The *post-hoc* Tukey's test was used to identify if there are statistically significant inter-group differences and between which groups these differences present themselves. Among the most common *post-hoc* tests, the Tukey's test was chosen because it preserves the pre-defined significance level of 5%, despite the increasing of the number of averages to be analysed. That is, the significance level *a priori*-defined is kept, despite the addition of the number of groups to be analysed.

#### 4.7 Evaluator attributions

In the data collection of the pilot-study, the prototype description was requested only in NL. The same way happened with the CG prototype description of the experimental study. However, to the experimental groups CNEG and FSMEG it was given a brief orientation about the designated task. Besides, during all the procedure, the experimental groups had a reference material available with the description of another process, using the group-specific notation method for which it was designated. No group had any information about the task asked to the other

group and there was no limitation concerning to the tasks time of execution.

#### 4.8 The reader-tester attributions

From the application of the TBR and from the requirements specified by the subjects, the reader-tester adopted the following procedure: (1) in case of defects, these were classified, following the specific scenario of the reader-tester; (2) the low-fidelity prototype and its requirements are taken as reference (Table 2); (3) the defects identified and classified during the reading were reported in an appropriate form.

As mentioned before, in the topic 4.2 that is related to the dependent variable, the requirement unit was defined by the FSM transition, by the fact that it contains a testable triad state-action-state. If, by any reason, the requirement specified by the subject doesn't cover the content expressed in the triad, the defect is accounted. In the CNEG evaluation, only the labeled syntagms were evaluated by the testers.

The defects identifiable by the reader-tester were: missing information; ambiguous information; incorrect fact; extraneous information; and others. A form was provided to the reader-tester describing the procedure to be used to detect these defects in the requirements specification (Lahtinen, 2012).

### 5 Research hypotheses

#### 5.1 Basic hypothesis

From the problem explicitly mentioned in topic 1, it was possible to design the study's general pair of hypotheses in a broad way.  $H_{0(i)}$ : *The notation method, used in process modeling, doesn't affect the ambiguity level of the process described in the functional requirement specification.*  $H_{1(i)}$ : *The notation method, used in process modeling, affects the ambiguity level of the process described in the functional requirement specification.*

#### 5.2 Secondary hypotheses

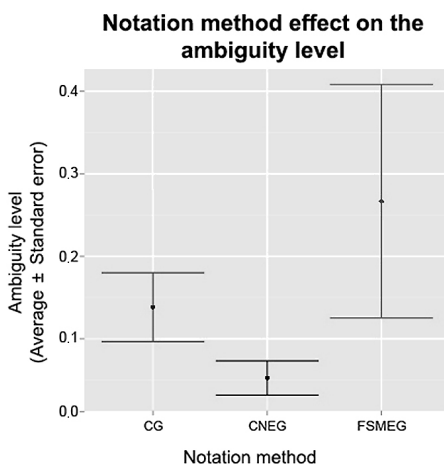
To the fulfillment of the specific objective predicted in topic 1 (item i), it was conducted a pilot-study that took the following hypotheses as its starting point: (i)  $H_{0(p)}$ : *The sample isn't from a normally distributed target population concerning to the ambiguity level in the functional requirements specification in NL.*  $H_{1(p)}$ : *The sample is from a normally distributed target population concerning to the ambiguity level in the functional requirements specification in NL.* After the proposition of a combined notation method to the process modeling, it was conducted an experimental study for the fulfillment of the specific objectives iii and iv, predicted in topic 1. This experimental study took the following hypotheses as its starting

point: (ii)  $H_{0(ii)}$ : There are no inter-group difference in the ambiguity level of the process described in the functional requirements specifications.  $H_{1(ii)}$ : There are inter-group difference in the ambiguity level of the process described in the functional requirements specifications.

## 6 Results

### 6.1 Parametric analysis of variance

In the One-way ANOVA and in CRD, the basic hypotheses  $H_{0(i)}$  was not rejected considering the significance level of 5%, because  $p > \alpha$  ( $0.07 > 0.05$ ) and  $F < F$  critical ( $2.89 < 3.24$ ). This result suggests that the notation method, used in process modeling, does not affect the ambiguity level of the process described in the functional requirement specification ( $p > 0.05$ ). On Figure 3 the average level of ambiguity and the standard error are presented: CG ( $0.14 \pm 0.04$ ), CNEG ( $0.05 \pm 0.02$ ) and FSMEG ( $0.27 \pm 0.14$ ). The comparison of averages suggests that: the CN is the best method to reduce the ambiguity level on requirements specification; the requirement specification expressed in the FSM generates the highest ambiguity level among the evaluated notations. However, these trends aren't statistically significant ( $p > 0.05$ ). Table 4 summarizes the statistical data obtained.



**Figure 3.** Effects of the process modeling notation method in the ambiguity level of the functional requirement specification of enterprise information systems.

**Table 4.** Summary of the statistical data observed.

Group	n	Average	Median	Ceiling	Standard-deviation ( $\sigma$ )	Variance ( $\sigma^2$ )
CG	13	0.14	0.20	0.40	0.15	0.02
CNEG	19	0.05	0.00	0.20	0.09	0.01
FSMEG	9	0.27	0.00	1.00	0.42	0.18

### 6.2 Non-parametric analysis of variance

The Kruskal-Wallis Test pointed results similar to the One-way ANOVA. The basic hypothesis  $H_{0(i)}$  was not rejected considering the significance level of 5%, because  $p > \alpha$  ( $0.20 > 0.05$ ). So, this result also suggests that the notation method, used in process modeling, does not affect the ambiguity level of the process described in the functional requirement specification ( $p > 0.05$ ).

### 6.3 Homogeneity of variance

By the standard Levene's Test, which uses the average as a parameter, the homoscedasticity hypothesis was rejected because  $p < \alpha$  ( $0.00006 < 0.05$ ). On the other hand, in the Levene's Test calculated by the median (Brown-Forsythe Test), the homoscedasticity hypothesis was not rejected because  $p > \alpha$  ( $0.056 > 0.05$ ).

### 6.4 Normality

In the Shapiro-Wilk Test the null hypothesis states that the sample is from a normally distributed population. On the other hand, the null hypothesis  $H_{0(p)}$  (topic 5.2, item i) in the present study is contrary, because it affirms that the sample is not from a normally distributed population.

The result of the adherence test for each experimental group is expressed in Table 5. The Shapiro-Wilk null hypothesis was rejected in the significance level of 5%, because in all the experimental groups the calculated W is less than the critical tabulated W. This fact implied in not rejecting the null hypothesis  $H_{0(p)}$  of this study, suggesting that none of the groups is from a normally distributed population.

### 6.5 Residual analysis

The quantile-quantile analysis revealed the non-adherence of the ambiguity level profile in groups CNEG and FSMEG with the Gauss curve. The non-adherence was also observed in CG. The normality scape can be attributed to the fact that the data that compose the ambiguity level metric are of discrete type.

### 6.6 Post-hoc analysis

The *post-hoc* Tukey's Test suggests that there is no statistically significant inter-group difference concerning to the ambiguity level of the process



**Table 5.** Result of the Shapiro-Wilk adherence test in each experimental group.

Group	n	W	W*	P	p*
CG	13	0.79	W < 0.866	0.0052260	P < 0.05
CNEG	19	0.55	W < 0.901	0.0000016	P < 0.05
FSMEG	9	0.66	W < 0.829	0.0004289	P < 0.05

p\* = comparison of the calculated p with the significance level ( $\alpha = 0.05$ ); W\* = comparison of the W with critical W; n = number of individuals in the sample; W = statistical test; P = calculated probability.

**Table 6.** Post-hoc Tukey’s Test: paired comparison of the average ambiguity level.

Inter-group difference	p	p*	Type II Error ( $\beta$ )	Test Power
CNEG-CG	0.533	P > 0.05	0.59	0.41
FSMEG-CG	0.383	P > 0.05	0.58	0.42
FSMEG-CNEG	0.056	P > 0.05	0.65	0.35

p\* = comparison of the calculated p with the significance level ( $\alpha = 0.05$ ).

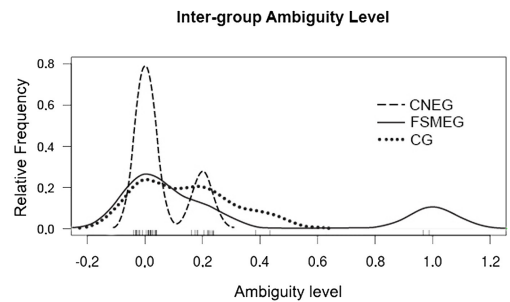
described in the functional requirement specification ( $p > 0.05$ ). Therefore, the paired comparison result of the average level of ambiguity suggests that the null hypothesis  $H_{0(ii)}$  (topic 5.2, item ii) was not rejected.

When adopting the significance level of 5%, there was no statistically significant difference between the analysed pairs (Table 6). As the null hypothesis was not rejected, it highlighted the importance of evaluating the Type II Error ( $\beta$ ) – probability of not rejecting a false null hypothesis. This way, the statistical Test Power was estimated by  $1 - \beta$ , which is, the probability of finding an inter-group difference that, in fact, proceeds.

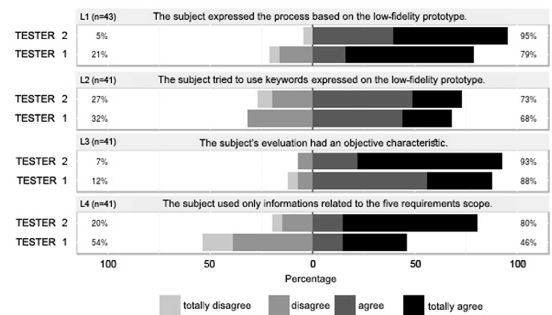
The distribution profile of the inter-group ambiguity level is shown on Figure 4 by the relative frequency polygon. Taking as a reference the x-axis, one can notice that there was an overlap of the empirical residues concentration. Besides, it is also possible to notice the positive asymmetry of the distribution. In general terms the asymmetry degree observed was 2.61.

### 6.7 Likert scale

Four critical issues were observed by the testers during the sample analysis. These issues concerning the application of the TBR technique were evaluated in a qualitative way by the use of the Likert scale shown in Figure 5. In general, it is possible to notice the optimism of tester 2 in relation to tester 1 regarding to the critical issues. The first issue served as a criterion for discarding subjects from the sample when both testers agreed with the fact that the subject had not described the buying process based on the low-fidelity prototype. This way, 4.65% of the 43 subjects were excluded from the following analyses. Therefore, the evaluation of the other Likert items were done with the remaining 41 subjects. In relation to the second item shown in Figure 5, it is possible to observe that even between the valid subjects, there



**Figure 4.** Inter-group ambiguity level distribution.



**Figure 5.** Likert scale with critical issues experienced in the use of the TBR technique during the data analysis routine of the experimental study.

was a consensus between testers that 17.07% of the sample did not use keywords. This fact affected the testability of the 5 requirements predicted on Table 2. Another consensus between testers was that 85.37% of the evaluations had an objective character. At this point, the subjectivity level was of 4.88%. Finally, in 19.51% of the cases it was possible to notice the presence of information outside the scope of five evaluated requirements.

## 7 Discussion

In the pilot-study 83.87% of sample subjects were approached by the same appraiser, minimizing possible discrepancies in the approach. Despite the subjectivity level of the evaluation has been relatively low, there was no consensus between the testers in 29.17% of the cases for this item. The Spearman's rank correlation was of 0.45 in this item, suggesting a low relation between the testers opinion when evaluating the subjectivity. The difficulty in evaluating it may have been affected by the fact that 58.33% of the subjects presented information beyond the scope of the five requirements being evaluated (Silva et al., 2013).

When analysing the fulfillment of the ANOVA assumptions reported on topic 4.6, based on the pilot and experimental studies results, it is possible to consider the following aspects: (1) *Normality*: the Shapiro-Wilk adherence test and the quantile-quantile residual analysis demonstrate that the values distribution of the ambiguity level variable did not present a normal profile in none of the analysed groups ( $p < 0.05$ ). The non-normality of the functional requirements specification in NL, observed in the pilot-study, confirmed itself in the experimental study control group and was also observed in the experimental groups CNEG and FSMEG; (2) *Independence*: the data sampling was randomized to ensure the independence of the observations. The randomization predicted in the CRD process ensured that an observation occurrence did not affect the likelihood of another observation occurrence; (3) *Homoscedasticity*: the Levene's test calculated by the median (Brown-Forsythe test) demonstrated that the homoscedasticity hypothesis was not rejected ( $p > 0.05$ ). This test was adopted because it presents a good robustness in non-normal distribution data. The standard Levene's test, which uses the average as a parameter, rejected the homoscedasticity hypothesis. However, it was not considered because it is more appropriate to distributions that present a normal profile – which did not occurred (NIST, 2013). This way, there was no inter-groups difference in relation to the variance. Therefore, the variances are homogeneous, which reduces the risk that the variances attributed to individual differences would be bigger than those related to the factor application itself. This homogeneity is of great importance to favor the conclusion validity, but it compromises external validity because it represents a threat to the random heterogeneity of the subjects.

Based on the mentioned aspects, it is possible to infer that only the normality assumption was not accomplished. As the assumptions to apply the ANOVA (parametric analysis) were not completely accomplished, the non-parametric analysis would be the most appropriate. Nonetheless, as the Shapiro-Wilk test is very sensitive to normality deviations, the rejection of the normality hypothesis is very probable.

This way, it was decided to carry out the parametric (One-way ANOVA and CRD) and non-parametric (Kruskal-Wallis test) analyses. The results obtained by both of them were similar.

In both studies (pilot and experimental) there were constraints concerning to the resource availability for the experimental conducting, simultaneously, with all the subjects, which motivated the application of the experiment in different occasions. However, in the case of experimental study, all subjects of the sample were approached by the same evaluator, minimizing possible discrepancies in the approach. Despite the subjectivity level of the evaluation has been low, there was no consensus between the testers in 9.76% of the cases in this item. The Spearman's rank correlation was of 0.40 in this matter, suggesting a low relation between the testers' opinion when evaluating the subjectivity. The difficulty in evaluating it may have been affected by the fact that 19.51% of the subjects presented information outside the scope of the five requirements being evaluated.

In the pilot-study the main factor that may have negatively affected all the evaluated Likert items was the complaisance in the instruction to the subjects about how the task should have been performed. On that occasion, since a low-fidelity prototype with a high abstraction level was used, the subjects were allowed to describe buying process tasks that could be implicit in the model (Silva et al., 2013). However, in the experimental study the subjects were instructed to describe strictly what was represented in the low-fidelity prototype. Besides, the last used prototype was characterized by a lower abstraction level in order to represent the buying process more realistically. Therefore, the data suggest combined and positive effect in reducing complaisance when instructing and in the prototype abstraction level in all the evaluated Likert items.

The TBR technique showed a favorable applicability while specific tool for the proposed analysis. However, the ambiguity level was underestimated by the usage of the  $Q_a$  metric associated to the TBR technique. This situation was evidenced by the positive asymmetry of the distribution, which observed value was 2.61 (Figure 4). Since the numerator of  $Q_a$  ( $n_{ui}$ ) is only counted when there is agreement between testers regarding the requirement ambiguity, the opinion of the most optimistic tester tends to prevail. Beyond that, the probability of assigning ambiguity to a requirement is very low on mathematical terms. There are 5 possible labels on TBR, demonstrated by the following sample space:  $TBR = \{\text{missing information, ambiguous information, incorrect fact, extraneous information, others}\}$ . Between the testers there may be  $T = \{\text{agreement, disagreement, no-consensus}\}$ , adding 3 other possible labels. The ambiguity was only counted in  $n_{ui}$  if the TBR label was "ambiguous

information” and there was “agreement” between both testers. Being  $TBR(\text{ambiguous information}) = 20\%$  and  $T(\text{agreement}) \cong 33.33\%$ , the probability of counting the ambiguity in  $n_{ui}$  is defined by  $P(\text{ambiguity}) = TBR(\text{ambiguous information}) \cap T(\text{agreement}) \cong 6.67\%$ . An alternative to minimize this problem would be reducing the number of possible labels on TBR. Something as  $TBR = \{\text{ambiguous information, other defects, no defects}\}$ . This way, the probability accounting for ambiguity in  $n_{ui}$  would reach approximately 11.11%, forcing the asymmetry reduction in the data distribution profile to favor the performance of comparative analyses from a more favorable perspective.

Some situations appear to happen circumstantially (Figure 3). It is the case of low relative ambiguity of requirements specification in NL, in face of the pre-established concept that NL presents a high ambiguity. This counterpoint is also observed in the high relative ambiguity of the formal notation, in face of the pre-established concept that the formal notation presents low ambiguity. In a purely qualitative analysis and independent of statistical significance, the results of this study point out that the inherent ambiguity of NL and the absence of ambiguity inherent to the formal notation are highly dependent on the usage. It is likely that these potential and intrinsic properties of the notation methods tend to manifest themselves according to the user’s degree of proficiency and/or maturity.

Therefore, the notation method training effect on the ambiguity level should be evaluated. The NL may have presented a relatively low level of ambiguity because it is a more usual notation for subjects. The formal notation, despite presenting a good potential for ambiguity reduction, needs to be practiced to evidence its effects. This way, the evaluation of training factor effect on the ambiguity level is essential to estimate more robustly the “combined notation effect” as an alternative notation method. The measurement of the ambiguity level variable is a difficult task because it is very sensitive to the intervening factors. It is necessary that the experimental scope be very well designed and defined so that the reliability and the scientific reproducibility of the evaluations are preserved. The zeal for these aspects makes the environment more assertive for the phenomenon observation.

Because subjects are undergraduate students in computer science, and the functional requirements specification could also be performed by programmers, testers, analysts and engineers from various fields, one cannot generalize the results of this study. However, the availability of human resources greatly restricts the selection of a sample that represents the desirable target population reliably.

## 8 Final considerations

From the problems raised in the pilot-study, the results obtained indicate that the TBR technique demonstrated a satisfactory applicability to analyze the ambiguity level of the EIS’s functional requirements specification. In addition, the results suggest that the sample is not from a normally distributed target population in terms of ambiguity level in NL. In the experimental study, the effect of the process modeling notation method on the ambiguity level of the requirements specification were observed. Usually, in systems development, the requirements are expressed in NL and it is desirable that their specification presents the lowest possible ambiguity level. So, this experimental research proposed a combination of NL, widely used for this purpose, with another notation method to induce the ambiguity level reduction - the Finite State Machine.

The results indicate that there was no statistically significant evidence to affirm that the notation method, used in process modeling, affects the ambiguity level. Besides, there was no statistically significant inter-group difference in the ambiguity level. Thus, it can not be said that the combination of NL and FSM reduces the ambiguity level of EIS’s functional requirements specification.

In conducting further studies in this knowledge field, it is important to consider: the effect of training in the notation method to induce adjustments in the ambiguity level; and reducing the number of possible labels in the TBR in order to increase the mathematical probability of consensus (agreement/disagreement) between testers with respect to defects. This reduction may contribute to minimize the asymmetry in the distribution profile of the ambiguity level, making it possible to perform comparative analyses from a more favorable perspective. Such action should contribute to avoid that the ambiguity level is underestimated and highly conditioned to the most optimistic tester’s opinion. In addition, it is relevant to keep reduced the complaisance when instructing and the low-fidelity prototype abstraction level to improve the analysis quality.

The semantic annotation of the requirements specification in NL with SBPM labels could be considered in future studies. When there is a close relationship between the labels of the formal notation (metamodel) and the domain ontology, perhaps there is a possibility of retrieving information more effectively and less ambiguously, facilitating the organizational knowledge flow. Therefore, the comparative and experimental analysis of the ambiguity level from the semantic annotation based on different formal notations of business process modeling would be plausible. On the other hand, the subjects’ proficiency degree needs to be pondered because there are formal notations with many labels and this can make it difficult to choose the most appropriate label for the syntagma during the annotation, especially if the individuals are not trained, or there is no defined ontology.

## References

- Bailey, R. A. (2008a). Forward look. In R. A. Bailey, *Design of comparative experiments* (Cap. 1, pp. 1-16). Cambridge: Cambridge University Press.
- Bailey, R. A. (2008b). Unstructured experiments. In R. A. Bailey, *Design of comparative experiments* (Cap. 2, pp. 19-41). Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511611483.003>.
- Basili, V. R. (2013). *Lab package for the empirical investigation of perspective-based reading*. College Park: University of Maryland. Recuperado em 25 de março de 2013, de <http://www.cs.umd.edu/projects/SoftEng/ESEGE/>
- Basili, V. R., Green, S., Laitenberger, O., Lanubile, F., Shull, F., Sørumgård, S., & Zelkowitz, M. V. (1996). The empirical investigation of perspective-based reading. *Empirical Software Engineering: an International Journal*, 1(2), 133-164. <http://dx.doi.org/10.1007/BF00368702>.
- Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with python*. Sebastopol: O'Reilly Media.
- Bisqueria, R., Sarriera, J. C., & Matínez, F. (2007). Provas de homoscedasticidade. In R. Bisqueria, J. C. Sarriera & F. Martínez (Eds.), *Introdução a estatística: enfoque informático com o pacote estatístico SPSS* (Cap. 5, pp. 83-87). Porto Alegre: Bookman.
- Bokhari, M. U., & Siddiqui, S. T. (2011). Metrics for requirements engineering and automated requirements tools. In *Proceedings of the 5th National Conference on Computing for Nation Development (INDIACOM)*. New Delhi: Bharati Vidyapeeth's Institute of Computer Applications and Management.
- Bryer, J., & Speerschneider, K. (2012). *Likert: functions to analyze and visualize likert type items. R package version 1.1*. Recuperado em 7 de julho de 2013, de <http://jason.bryer.org/likert>
- Card, D. N. (1998). Learning from our mistakes with defect causal analysis. *IEEE Software*, 15(1), 56-63. <http://dx.doi.org/10.1109/52.646883>.
- Carvalho, R. A., Carvalho, F. L. C., & Manhães, R. S. (2011). Business language driven development: joining business process models to automated tests. In *Proceedings of the VIFIP TC8 International Conference on Enterprise Information Systems*. Aalborg: LNBIP.
- Carvalho, R. A., Manhães, R. S., & Silva, F. L. C. (2010). *Introducing business language driven development*. Ithaca: Cornell University. Relatório técnico. Recuperado em 25 de maio de 2013, de <http://arxiv.org/abs/1011.2238>
- Carvalho, R. A., Silva, F. L. C., & Manhães, R. S. (2012). Advances in enterprise information systems II. In C. Möller & S. S. Chaudhry (Eds.), *Business language driven development: joining business process models to automated tests* (pp. 167-177). Leiden: CRC Press.
- Cohn, M. (2004). *User stories applied for agile software development: an overview* (Cap. 1, pp. 3-15). Boston: Addison-Wesley.
- Coughlan, J., & Macredie, R. D. (2002). Effective communication in requirements elicitation: a comparison of methodologies. *Journal of Requirements Engineering*, 7(2), 47-60. <http://dx.doi.org/10.1007/s007660200004>.
- Coughlan, J., Lycett, M., & Macredie, R. D. (2003). Communication issues in requirements elicitation: a content analysis of stakeholder experiences. *Information and Software Technology*, 45(8), 525-537. [http://dx.doi.org/10.1016/S0950-5849\(03\)00032-6](http://dx.doi.org/10.1016/S0950-5849(03)00032-6).
- Davis, A., Overmyer, S., Jordan, K., Caruso, J., Dandashi, F., Dinh, A., Kincaid, G., Ledebner, G., Reynolds, P., Sitaram, A., Ta, A., & Theofanos, M. (1993). Identifying and measuring quality in a software requirements specification. In *Proceedings of the 1st International Software Metrics Symposium*. New York: IEEE. <http://dx.doi.org/10.1109/METRIC.1993.263792>.
- Hopcroft, J. E., Ullman, J. D., & Motwani, R. (2003). Autômatos finitos. In J. Ullman & J. Hopcroft (Eds.), *Introdução à teoria de autômatos, linguagens e computação* (Cap. 2, pp. 39-88). Rio de Janeiro: Elsevier.
- Institute of Electrical and Electronics Engineers – IEEE. (1998). *IEEE STD 830: IEEE recommended practice for software requirements specifications* (pp. 1-40). Piscataway: IEEE.
- Jørgensen, H. D., Karlsen, D., & Lillehagen, F. (2007). Product based interoperability: approaches and requirements. In *Proceedings of the Conference Coordination of Collaborative Engineering: State of the Art and Future Challenges 5th International Workshop on Challenges in Collaborative Engineering (CCE'07)* (pp. 33-44). Cracow: CCE.
- Jurafsky, D., & Martin, J. H. (2000). *Speech and language processing: an introduction to natural language processing, computational linguistics and speech recognition*. New Jersey: Prentice Hall.
- Kotler, P., Kartajaya, H., & Setiawan, I. (2012). O marketing da missão junto aos consumidores. In P. Kotler, H. Kartajaya & I. Setiawan, *Marketing 3.0: as forças que estão definindo o novo marketing centrado no ser humano* (Cap. 3, pp. 57-78). Rio de Janeiro: Elsevier.
- Lahtinen, J. (2012). Application of the perspective-based reading technique in the nuclear I&C context. In VTT Technical Research Centre of Finland (Ed.), *CORSICA Work Report 2011* (VVT Technology, 9). Finland: VTT Technical Research Centre of Finland.
- Lanubile, F., Shull, F., & Basili, V. R. (1998). Experimenting with error abstraction in requirements documents. In *Proceedings of 5th International Software Metrics Symposium (METRICS'98)* (pp. 114-121). Bethesda: IEEE Computer Society. <http://dx.doi.org/10.1109/METRIC.1998.731236>.



- Larson, R., & Farber, B. (2010). Testes qui-quadrado e a distribuição F. In R. Larson & B. Farber, *Estatística aplicada* (Cap. 10, pp. 439-483). São Paulo: Pearson Prentice Hall.
- Lautenbacher, F., Bauer, B., & Seitz, C. (2008). Semantic business process modeling: benefits and capability. In *Proceedings of the AAAI 2008 Stanford Spring Symposium: AI Meets Business Rules and Process Management (AIBR)*. California: Stanford University.
- Leffingwell, D. (2011). A brief history of software requirements methods. In D. Leffingwell, *Agile software requirements: lean requirements practices for teams, programs, and the enterprise* (Cap. 1, pp. 3-28). Boston: Addison-Wesley.
- LibreOffice. (2013). Recuperado em 2 de abril de 2013, de <http://www.libreoffice.org/>
- Lopes, E. (2008). Modalidades de gramática. In E. Lopes, *Fundamentos da linguística contemporânea* (Cap. 5, pp. 183-228). São Paulo: Pensamento-Cultrix.
- Manning, C. D., & Schütze, H. (2000). *Foundations of statistical natural language processing*. Cambridge: The MIT Press.
- Menezes, P. B. (2008a). Hierarquia de classes de linguagens e conclusões. In P. B. Menezes, *Linguagens formais e autômatos* (Cap. 9, pp. 195-201). Porto Alegre: Bookman.
- Menezes, P. B. (2008b). Linguagens regulares. In P. B. Menezes, *Linguagens formais e autômatos* (Cap. 3, pp. 41-79). Porto Alegre: Bookman.
- National Institute of Standards and Technology – NIST. SEMATECH. (2013). *e-handbook of statistical methods*. Gaithersburg: NIST. Recuperado em 17 de março de 2013, de <http://www.itl.nist.gov/div898/handbook/>
- Oren, E., Möller, K. H., Scerri, S., Handschuh, S., & Sintek, M. (2006). *What are semantic annotations?* Ireland: DERI Galway. Technical report.
- Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, 122(3), 280-291. <http://dx.doi.org/10.1016/j.cognition.2011.10.004>. PMID:22192697.
- R Core Team. (2013). *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Recuperado em 7 de julho de 2013, de <http://www.R-project.org>
- Raffin, F. (2009). Rigor e ambiguidade. In *Pequena introdução à filosofia* (Cap. 15, pp. 181-191). Rio de Janeiro: Editora FGV.
- Ross, R. G. (2009). RuleSpeak® sentence forms: specifying natural-language business rules in english. *Business Rules Journal*, 10(4). Recuperado em 28 de maio de 2013, de <http://www.BRCommunity.com/a2009/b472.html>
- Shull, F., Rus, I., & Basili, V. (2000). How perspective-based reading can improve requirement inspections. *Computer*, 33(7), 73-79. <http://dx.doi.org/10.1109/2.869376>.
- Silva, A. E. C., Carvalho, R. A., & Ferreira, H. S. (2013). Nível de ambiguidade da especificação dos requisitos funcionais de sistemas de informação empresarial: um estudo-piloto. In *Anais do XX Simpósio de Engenharia de Produção (SIMPEP)*. Bauru: Unesp. Recuperado em 4 de junho de 2014, de [http://www.simpep.feb.unesp.br/abrir\\_arquivo\\_pdf.php?tipo=artigo&evento=8&art=933&cad=17616&opcao=com\\_id](http://www.simpep.feb.unesp.br/abrir_arquivo_pdf.php?tipo=artigo&evento=8&art=933&cad=17616&opcao=com_id)
- Sommerville, I. (2007a). Verificação e validação. In I. Sommerville, *Engenharia de software* (Cap. 22, pp. 341-453). São Paulo: Addison-Wesley.
- Sommerville, I. (2007b). Requisitos de software. In I. Sommerville, *Engenharia de software* (Cap. 6, pp. 79-94). São Paulo: Addison-Wesley.
- Sørumgård, S. (1997). *Verification of process conformance in empirical studies of software development* (Tese de doutorado). Trondheim: Department of Computer and Information Science, The Norwegian University of Science and Technology.
- Sutcliffe, G. A., Economou, A., & Markis, P. (1999). Tracing requirements errors to problems in the requirements engineering process. *Requirements Engineering*, 4(3), 134-151. <http://dx.doi.org/10.1007/s007660050024>.
- Turnquist, G. L. (2011). Testing customer stories with behavior driven development. In G. L. Turnquist, *Python testing cookbook* (Cap. 4, pp. 117-163). Birmingham: Packt Publishing.
- Van der Aalst, W. M. P. (2011). Process modeling and analysis. In W. M. P. Aalst, *Process mining: discovery, conformance and enhancement of business processes* (Cap. 2, pp. 29-57). Heidelberg: Springer-Verlag. [http://dx.doi.org/10.1007/978-3-642-19345-3\\_2](http://dx.doi.org/10.1007/978-3-642-19345-3_2).
- Van Grondelle, J., Heller, R., Van Haandel, E., & Verburg, T. (2010). Involving business users in formal modeling using natural language pattern sentences. In *Proceedings of the 17th International Conference on Knowledge Engineering and Management by the Masses (EKAW'10)* (pp. 31-43). Berlin: Springer-Verlag. [http://dx.doi.org/10.1007/978-3-642-16438-5\\_3](http://dx.doi.org/10.1007/978-3-642-16438-5_3).
- Walia, G. S., & Carver, J. C. (2009). A systematic literature review to identify and classify software requirement errors. *Information and Software Technology*, 51(7), 1087-1109. <http://dx.doi.org/10.1016/j.infsof.2009.01.004>.
- Yan, Z. (2007). Semantic business process modeling. In *Proceedings of the Knowledge Web PhD Symposium*. Innsbruck: KWEPSY.