

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE INFORMÁTICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

RENATO CÉSAR BORGES FERREIRA

**Uma Abordagem Semiautomática para  
Identificação de Elementos de Processo de  
Negócio em Texto em Linguagem Natural**

Dissertação apresentada como requisito parcial  
para a obtenção do grau de Mestre em Ciência da  
Computação

Orientador: Prof. Dr<sup>a</sup>. Lucinéia Heloisa Thom

Porto Alegre  
2017

## CIP — CATALOGAÇÃO NA PUBLICAÇÃO

Ferreira, Renato César Borges

Uma Abordagem Semiautomática para Identificação de Elementos de Processo de Negócio em Texto em Linguagem Natural / Renato César Borges Ferreira. – 2017.

103 f.: il.

Orientador: Lucinéia Heloisa Thom.

Dissertação (mestrado) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação. Porto Alegre, BR–RS, 2017.

1. Regras de Mapeamento. 2. Notação e Modelo de Processo de Negócio. 3. Gerenciamento de Processos de Negócio. 4. Elementos de Processo. 5. Modelo de Processo. 6. Processamento de Linguagem Natural. 7. Modelagem de Processos. I. Thom, Lucinéia Heloisa. II. Uma Abordagem Semiautomática para Identificação de Elementos de Processo de Negócio em Texto em Linguagem Natural.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. Rui Vicente Oppermann

Vice-Reitor: Prof. Jane Fraga Tutikian

Pró-Reitor de Pós-Graduação: Prof. Celso Giannetti Loureiro Chaves

Diretor do Instituto de Informática: Prof. Carla Maria Dal Sasso Freitas

Coordenador do PPGC: Prof. João Luiz Dihl Comba

Bibliotecária-chefe do Instituto de Informática: Beatriz Regina Bastos Haro

*“O sofrimento é passageiro,  
desistir é pra sempre.”*

— LANCE ARMSTRONG

## AGRADECIMENTOS

Agradeço inicialmente a DEUS e a JESUS CRISTO, no qual me fortaleço todos os dias.

Gostaria de agradecer a minha família, Evaldo, Suelise e Evelise, por todo apoio e força dada antes e durante o meu período em Porto Alegre. Com certeza, sem eles, não estaria aqui. Vocês são o meu espelho. Agradeço também aos meus avós, Anita e Alair, pelas orações realizadas antes e durante o mestrado.

Gostaria de agradecer a minha namorada, Janine, pelas orações, compreensões e apoio durante a realização do mestrado.

Um agradecimento especial a dona Graça pelas orações, apoio e de me receber tão bem aqui em Porto Alegre.

Gostaria de agradecer também a minha orientadora, prof<sup>a</sup> Lucinéia Heloisa Thom, pela confiança, auxílio, conselhos, paciência e pela disposição em realizar este trabalho juntamente comigo.

Aos docentes da UFRGS, pela disposição em auxiliar este trabalho. É um privilégio discutir este trabalho com vocês.

Aos colegas de laboratório 213: Carlos, Vinícius, Lorayne, Diego, Simone e Gabriel, pelas conversas e experiências vividas.

Finalmente, a todos, que diretamente ou indiretamente contribuíram para a realização do sonho de tornar mestre em computação. MUITO OBRIGADO!

## RESUMO

Para permitir um efetivo gerenciamento de processos de negócio, o primeiro passo é o desenvolvimento de modelos de processo adequados aos objetivos das organizações. Tais modelos são utilizados para descreverem papéis e responsabilidades dos colaboradores nas organizações. Além disso, a modelagem de processos é de grande importância para documentar, entender e automatizar processos. As organizações, geralmente provêm documentos não estruturados e de difícil entendimento por parte dos analistas. Neste panorama, a modelagem de processos se torna demorada e de alto custo, podendo gerar modelos de processo que estão em desacordo com a realidade prevista pelas organizações. A extração de modelos ou fragmentos de processo a partir de descrições textuais pode contribuir para minimizar o esforço necessário à modelagem de processos. Neste contexto, esta dissertação propõe uma abordagem para identificar elementos de processo de negócio em texto em linguagem natural de forma semiautomática. Baseado no estudo de processamento de linguagem natural, foi definido um conjunto de regras de mapeamento para identificar elementos de processo em descrição textual. Além disso, para avaliar as regras de mapeamento e viabilizar a abordagem proposta, foi desenvolvido um protótipo capaz de identificar elementos de processo em texto de forma semiautomática. Para medir o desempenho do protótipo proposto, foram utilizadas métricas de recuperação de informação, tais como precisão, revocação e medida-F. Além disso, foram aplicados dois questionários com o objetivo de verificar a aceitação perante os usuários. As avaliações apresentam resultados promissores. A análise de 70 textos, apresentou, em média, 73,61% de precisão, 70,15% de revocação e 71,82% de medida-F. Além disso, os resultados do primeiro e segundo questionários apresentaram, em média, 91,66% de aceitação dos participantes. A principal contribuição deste trabalho é propor regras de mapeamento para identificar elementos de processo em texto em linguagem natural para auxiliar e minimizar o tempo necessário à modelagem de processos realizada pelos analistas de processo.

**Palavras-chave:** Regras de Mapeamento. Notação e Modelo de Processo de Negócio. Gerenciamento de Processos de Negócio. Elementos de Processo. Modelo de Processo. Processamento de Linguagem Natural. Modelagem de Processos.

# **A Semi-Automatic Approach to Identify Business Process Elements in Natural Language Text**

## **ABSTRACT**

To enable effective business process management, the first step is the design of appropriate process models to the organization's objectives. These models are used to describe roles and responsibilities of the employees in an organizations. In addition, business process modeling is very important to report, understand and automate processes. However, the documentation existent in organizations about such processes is mostly unstructured and difficult to be understood by analysts. In this context, process modeling becomes highly time consuming and expensive, generating process models that do not comply with the reality of the organizations. The extracting of process models from textual descriptions may contribute to minimize the effort required in process modeling. In this context, this dissertation proposes a semi-automatic approach to identify process elements in natural language text. Based on the study of natural language processing, it was defined a set of mapping rules to identify process elements in text. In addition, in order to evaluate the mapping rules and to demonstrate the feasibility of the proposed approach, a prototype was developed able to identify process elements in text in a semiautomatic way. To measure the performance of the proposed prototype metrics were used to retrieve information such as precision, recall, and F-measure. In addition, two surveys were developed with the purpose of verifying the acceptance of the users. The evaluations present promising results. The analyses of 70 texts presented, on average, 73.61% precision, 70.15% recall and 71.82% F-measure. In addition, the results of the first and second surveys presented on average 91.66% acceptance of the participants. The main contribution of this work is to provide mapping rules for identify process elements in natural language text to support and minimize the time required for process modeling performed by process analysts.

**Keywords:** Mapping Rules, Business Process Model and Notation, Business Process Management, Process Element, Process Model, Natural Language Processing, Process Modeling.

## LISTA DE FIGURAS

Figura 1.1	Exemplo de um processo de negócio.....	12
Figura 2.1	Ciclo de vida de BPM.....	20
Figura 2.2	Categoria básica de elementos da BPMN.....	23
Figura 2.3	Exemplo de modelo de processo modelado em BPMN. ....	24
Figura 2.4	Hierarquia de sentenças, cláusulas, frases e palavras .....	26
Figura 2.5	Árvore sintática da sentença do exemplo .....	27
Figura 2.6	Relação de dependência da ferramenta <i>Spacy</i> .....	29
Figura 3.1	Uma Abordagem para Identificação de Elementos de Processo de Negócio em Texto em Linguagem Natural.....	36
Figura 3.2	Saída de dados separados por sentenças.....	37
Figura 3.3	Visão geral estrutural da etapa de análise sintática de texto para gerar uma sentença marcada .....	38
Figura 3.4	Vetor de objetos desenvolvido pela classe <i>DOC</i> do analisador sintático. ....	39
Figura 3.5	Representação estrutural da etapa de análise lógica de texto. ....	49
Figura 3.6	Separação de <i>palavras-chave</i> das regras de desvio exclusivo (XOR) .....	51
Figura 3.7	Separação de <i>palavras-chave</i> das regras de desvio paralelo (AND) .....	52
Figura 4.1	Primeira etapa do questionário. ....	63
Figura 4.2	Segunda etapa do questionário. ....	64
Figura 4.3	Terceira etapa do questionário. ....	65
Figura 4.4	Continuação da terceira etapa do questionário. ....	67
Figura 4.5	Resultados da segunda etapa do primeiro questionário através dos 22 participantes selecionados.....	69
Figura 4.6	Resultados da terceira etapa do primeiro questionário através dos 22 participantes selecionados.....	70
Figura 4.7	Resultados da terceira etapa do primeiro questionário através dos 22 participantes selecionados.....	70
Figura 4.8	Resultados obtidos a partir da primeira parte do primeiro questionário.....	71
Figura 4.9	Resultados obtidos a partir da segunda parte do primeiro questionário. ....	72
Figura 4.10	Resultados obtidos da terceira parte do primeiro questionário.....	73
Figura 4.11	Segunda e terceira etapa do questionário.....	74
Figura 4.12	Quarta etapa do questionário. ....	75
Figura 4.13	Resultados obtidos a partir da primeira parte do segundo questionário. ....	79
Figura 4.14	Resultados obtidos a partir da segunda e terceira parte do segundo questionário.....	80
Figura 4.15	Resultados obtidos a partir da quarta parte do segundo questionário.....	81
Figura 4.16	Primeira e Segunda Sentenças Identificadas no Texto 1.....	83
Figura 4.17	Terceira,Quarta e Quinta Sentenças Identificadas no Texto 1 .....	84
Figura 4.18	Sexta e Sétima Sentença Identificada .....	84
Figura 4.19	Oitava e Nona Sentenças Identificadas no Texto 1 .....	85
Figura 4.20	Medidas dependem do resultado da consulta (circular/laranja e cinza) .....	86
Figura 4.21	Exemplos de Matrizes de Confusão .....	87
Figura 4.22	Resultados gerais das métricas de avaliação classificadas por classes .....	92

## LISTA DE TABELAS

Tabela 2.1	Categorias léxicas de um analisador sintático. ....	28
Tabela 2.2	Estado da Arte para Identificar Elementos de Processo em Texto em Linguagem Natural .....	32
Tabela 2.3	Comparação dos trabalhos relacionados com a abordagem apresentada nesta dissertação.....	34
Tabela 3.1	Regras para identificação de atividades. ....	41
Tabela 3.2	Regras para identificação de eventos. ....	42
Tabela 3.3	Regras para identificação de desvio exclusivo (XOR). ....	43
Tabela 3.4	Regras para identificação de desvio paralelo (AND). ....	44
Tabela 3.5	Regras para identificação de raias. ....	44
Tabela 4.1	Resultados de desempenho do protótipo para a classe de atividades. ....	90
Tabela 4.2	Resultados de desempenho do protótipo para a classe de eventos. ....	90
Tabela 4.3	Resultados de desempenho do protótipo para a classe de desvio exclusivo (XOR). ....	90
Tabela 4.4	Resultados de desempenho do protótipo para a classe de desvio paralelo (AND). ....	91

## LISTA DE ALGORITMOS

1	Algoritmo de Execução da Análise Sintática de Texto.....	39
2	Algoritmo de Verificação de Palavras-Chave do Desvio Exclusivo (XOR) .....	47
3	Algoritmo de Verificação de Palavras-Chave do Desvio Paralelo (AND).....	48
4	Algoritmo de Verificação de Palavras Sinônimas do Desvio Exclusivo (XOR).....	50
5	Algoritmo de Identificação de desvio exclusivo (Regra 3).....	51
6	Algoritmo de Verificação de Palavras Sinônimas de Desvio Paralelo (AND).....	52
7	Algoritmo de Definição de Verbo Principal da Sentença.....	53
8	Algoritmo de Identificação de desvio paralelo (Regra 1) .....	54
9	Algoritmo de Verificação do Tempo Verbal.....	55
10	Algoritmo de Identificação do Sujeito .....	55
11	Algoritmo de Identificação do Objeto.....	56
12	Algoritmo de Identificação de Atividades (Regra 1) .....	57
13	Algoritmo de Identificação de Eventos (Regra 2).....	57

## LISTA DE ABREVIATURAS E SIGLAS

7PMG	<i>Seven Process Modeling Guidelines</i>
BPM	<i>Business Process Management</i> – Gerenciamento de Processos de Negócio
BPMN	<i>Business Process Management Notation</i> – Notação e Modelo de Processos de Negócio
BPMS	<i>Business Process Management System</i> – Sistema de Modelagem de Processos de Negócio
ERP	<i>Enterprise Resource Planning</i>
EPC	<i>Event Process Chain</i>
GATE	<i>General Architecture for Text Engineering</i>
IA	Inteligência Artificial
NLTK	<i>Natural Language Toolkit</i>
OMG	<i>Object Management Group</i>
PLN	Processamento de Linguagem Natural
POS	<i>Part of speech</i>
RASP	<i>Robust Accurate Statistical Parsing</i>
RI	<i>Recuperação de Informação</i>
SI	Sistemas de Informação
TI	Tecnologia da Informação
UML	<i>Unified Modeling Language</i> – Linguagem de Modelagem Unificada
WfMC	<i>Workflow Management Coalition</i>
WfMS	<i>Workflow Management Systems</i> – Sistema de Gerenciamento de <i>Workflow</i>
YAWL	<i>Yet Another Workflow Language</i>

## SUMÁRIO

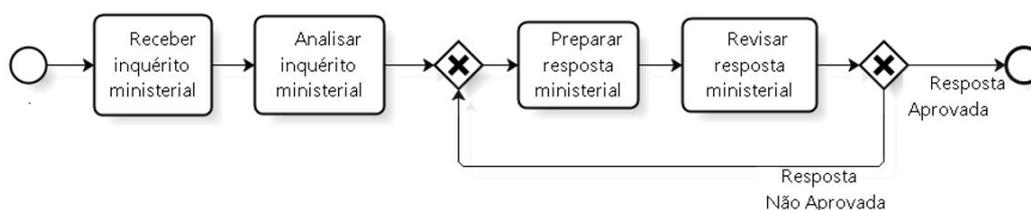
<b>1 INTRODUÇÃO</b> .....	<b>12</b>
<b>1.1 Motivação e Justificativa</b> .....	<b>14</b>
<b>1.2 Objetivos e Hipóteses</b> .....	<b>15</b>
<b>1.3 Contribuições da Pesquisa</b> .....	<b>16</b>
<b>1.4 Organização do Texto</b> .....	<b>17</b>
<b>2 FUNDAMENTOS DE GERENCIAMENTO DE PROCESSOS DE NEGÓ- CIO E PROCESSAMENTO DE LINGUAGEM NATURAL</b> .....	<b>18</b>
<b>2.1 Ciclo de Vida de BPM</b> .....	<b>19</b>
<b>2.2 Notação e Modelo de Processo de Negócio (BPMN)</b> .....	<b>22</b>
<b>2.3 Processamento de Linguagem Natural (PLN)</b> .....	<b>25</b>
2.3.1 Análise Sintática .....	26
2.3.2 Analisador Sintático - <i>Spacy</i> .....	29
<b>2.4 Trabalhos Relacionados</b> .....	<b>30</b>
<b>2.5 Considerações Finais</b> .....	<b>33</b>
<b>3 UMA ABORDAGEM PARA IDENTIFICAÇÃO DE ELEMENTOS DE PRO- CESSO DE NEGÓCIO EM TEXTO EM LINGUAGEM NATURAL</b> .....	<b>35</b>
<b>3.1 Entrada de Dados</b> .....	<b>35</b>
<b>3.2 Análise Sintática de Texto</b> .....	<b>38</b>
<b>3.3 Análise Lógica de Texto</b> .....	<b>40</b>
3.3.1 Definição de Regras de Mapeamento .....	40
3.3.2 Aplicação das Regras de Mapeamento para Identificação de Elementos de Processo de Negócio em Texto em Linguagem Natural .....	46
<b>3.4 Saída de Texto</b> .....	<b>58</b>
<b>3.5 Considerações Finais</b> .....	<b>60</b>
<b>4 AVALIAÇÃO DA ABORDAGEM PROPOSTA E RESULTADOS OBTIDOS</b> .....	<b>61</b>
<b>4.1 Avaliação Através de Questionário</b> .....	<b>61</b>
4.1.1 Primeiro Questionário Aplicado .....	62
4.1.2 Resultados Obtidos Através da Aplicação do Primeiro Questionário .....	68
4.1.3 Segundo Questionário Aplicado .....	73
4.1.4 Resultados Obtidos Através da Aplicação do Segundo Questionário .....	77
<b>4.2 Resultados Obtidos a partir do Desenvolvimento do Protótipo</b> .....	<b>82</b>
4.2.1 Avaliação de Desempenho .....	85
4.2.1.1 Métricas Seleccionadas.....	85
4.2.1.2 Conjunto de Dados Seleccionados.....	88
4.2.1.3 Resultados de Desempenho .....	89
<b>4.3 Discussão dos Resultados</b> .....	<b>91</b>
<b>5 CONCLUSÃO</b> .....	<b>93</b>
<b>5.1 Publicações</b> .....	<b>96</b>
<b>REFERÊNCIAS</b> .....	<b>97</b>

## 1 INTRODUÇÃO

Organizações públicas ou privadas, para manterem-se competitivas no mercado, buscam uma melhor interação com clientes e parceiros de negócio, além de ofertar um melhor produto ou serviço, atingir uma melhor padronização e eficiência na execução de seus processos de negócio (nesta dissertação também denominado como processo(s)). A automação de processos realizados em uma organização fornece maior controle sobre custos, tempo, erros e redundância na execução (THOM, 2012; THOM; REICHERT; IOCHPE, 2009).

Um processo de negócio é uma *"coleção de eventos, sequencia de atividades, tomadas de decisões, que envolvem um número de atores e objetos, que coletivamente trazem valor ao cliente"* (DUMAS et al., 2013). A Figura 1.1 traz um processo, cujo objetivo é aprovar ou reprovado uma resposta ministerial. Quando um inquérito ministerial é recebido, primeiramente, este é registrado no sistema. Após, o pedido é analisado, de modo que uma resposta ministerial possa ser preparada. Após ser preparada, a resposta é revisada pelo registrador. Se não for aprovada, esta deve novamente ser revista pelo chefe de gabinete. O processo termina assim que a resposta for aprovada (DUMAS et al., 2013).

Figura 1.1: Exemplo de um processo de negócio



Fonte: Adaptado de (DUMAS et al., 2013)

O Gerenciamento de Processos de Negócios (BPM – *Business Process Management*) tem sua base centrada na gestão dos processos de negócio de uma organização, possibilitando a sua representação e suas atividades (WESKE, 2007; WESKE, 2012). Segundo Leopold (2013), através de BPM, organizações podem se adaptar com flexibilidade em um ambiente de negócios em constante mudanças. Além disso, BPM traz diversas melhorias para a organização, tal como a padronização de processos, melhoria, qualidade e rapidez na execução das atividades (THOM, 2012; THOM; REICHERT; IOCHPE, 2009).

O ciclo de vida de BPM inclui as etapas de modelagem, configuração, execução e validação de um processo de negócio (WEBER; SADIQ; REICHERT, 2009; DUMAS et

al., 2012). Cada etapa do ciclo de vida é de extrema importância para automatização de processos dentro da organização.

A modelagem de processos pode ser considerada uma das etapas mais importantes e complexas do ciclo de vida de BPM (DUMAS et al., 2013). Tal etapa tem como finalidade facilitar a concepção do processo e associar o conhecimento entre os envolvidos em sua execução, assim como identificar possíveis problemas organizacionais. Nesta etapa, os processos de negócio executados na organização são modelados com o uso de uma notação gráfica, como por exemplo a Notação e Modelo de Processo de Negócio (BPMN – *Business Process Modeling and Notation*) (OMG, 2013). Conforme Dumas et al. (2013) a modelagem de processos é pré-requisito para a análise, redesenho e automação de processos de negócio em um projeto de BPM. Uma modelagem de processos incorreta compromete as próximas etapas do ciclo de vida de BPM, ou seja, uma correta automatização de processos origina-se de uma precisa modelagem de processos.

Na fase de modelagem de processos, existem diversos métodos de identificação de processos, tais como: entrevistas com usuários, *workshops* sobre o tema e documentos existentes na organização para iniciar a modelagem dos processos (DUMAS et al., 2013). Uma das vantagens da análise de documentos é que o analista de processos pode usá-los para se familiarizar com o processo em seu ambiente, e também formular hipóteses. Os documentos podem ter diversas fontes, incluindo relatórios, formulários, cartas, notas de *call centers*, pesquisas, investigação, políticas empresariais, manuais, sistemas de gerenciamento de conhecimento, mensagens de e-mails, dados de eventos (*logs*) de sistemas de informação, páginas web, documentos de textos e entrevistas (DUMAS et al., 2013). Porém, tais métodos podem apresentar limitações devido a problemas de comunicação entre analistas e usuários, falta de documentação e padronização dos processos e ausência de informações de usuários (DUMAS et al., 2013). Os profissionais de gestão de conteúdo consideram que 85% das informações nas organizações é armazenada na forma de documentos textuais (BLUMBERG; ATRE., 2003). Além disso, a quantidade de informações textuais estão crescendo de maneira mais rápida do que os dados estruturados tradicionais (WHITE, 2003).

Neste contexto, para desenvolver modelos de processo, os analistas de processo (nesta dissertação também denominado analista(s)) obtém informações abstratas sobre como esses processos são implementados. Para modelar o modelo de processo inicial (modelo *as-is*), o analista geralmente coleta várias informações sobre o processo através dos métodos mencionados com o propósito de obter conhecimento preliminar do processo

(DUMAS et al., 2013). A definição do modelo de processo inicial em um projeto de BPM, consome cerca de 60% do tempo total do projeto (HERBST, 1999).

### 1.1 Motivação e Justificativa

Diversos trabalhos (FRIEDRICH; MENDLING; PUHLMANN, 2011; CHUENG; KOLIADIS; GHOSE, 2007; SANTORO; GONÇALVES; BAIÃO, 2009), demonstraram que a extração de modelos de processo a partir de texto em linguagem natural pode minimizar o esforço dos analistas de processo para capturar, principalmente através de entrevistas com usuários tais modelos. Tais entrevistas raramente levam à compreensão de todo o processo, pois muitas vezes descrevem apenas o conhecimento de partes isoladas do processo.

A maior parte dos trabalhos avaliados consideram que os documentos textuais (textos linguagem natural) são descritos de forma que apenas os analistas de processo são capazes de extrair modelos de processo a partir destes (LEOPOLD, 2013; MEITZ; LEOPOLD; MENDLING, 2013; LEOPOLD; MENDLING; POLYVYANYYY, 2014). Isto significa que os textos possuem palavras-chave (por exemplo, *control flow*, *parallel gateway*, *exclusive gateway*, etc) e frases (sentenças) que denotam elementos de processo de negócio ou possuem partes do texto com uma linguagem extremamente técnica. Portanto, os analistas de processo iniciantes apresentam dificuldades para extrair modelos ou fragmentos de processo a partir de texto em linguagem natural (MENDLING; STREMBECK; RECKER, 2012; MENDLING; REIJERS; CARDOSO, 2007; SCHREPFER et al., 2009; REIJERS; MENDLING, 2008).

Desta forma, diante da grande quantidade de documentos textuais existentes nas organizações (BLUMBERG; ATRE., 2003), os analistas têm dificuldades em extrair informações relevantes para modelagem de processos, pois estes documentos estão dispersos nas organizações, o que dificulta a extração considerando esses documentos. Além disso, a maior parte dos documentos disponíveis sobre as operações de uma organização, não está organizada na forma de processos, isto é, existem documentos não estruturados nas organizações, dificultando a extração de informações relevantes a processos de negócio (DUMAS et al., 2013). Com isso, os textos em linguagem natural não estão preparados para serem usados diretamente pelas ferramentas de extração de modelos de processo. Esse fator contribui para a complexidade de identificação de elementos de processo de negócio em texto em linguagem natural.

Assim, para possibilitar que os analistas modelem processos a partir de texto em linguagem natural, uma técnica de transformação automática é necessária para obter potenciais informações já existentes nas organizações. Tal técnica é denominada de Processamento de Linguagem Natural (PLN), a qual permite automatizar a análise morfológica, sintática e semântica de textos em linguagem natural (INDURKHYA; DAMERAU, 2010).

A fim de identificar elementos de processo em texto, foram definidas regras de mapeamento. *Uma regra de mapeamento é a relação entre elementos de processo da BPMN, classes gramaticais e relações de dependência da gramática do idioma.* As regras foram baseadas no conjunto básico de elementos da BPMN (objetos de fluxo e divisões). Essas regras apoiam-se em classes gramaticais da língua inglesa (por exemplo, verbo, substantivo, adjetivo etc) e dependências de palavras (por exemplo, sujeito, objeto direto, indireto etc). Desta forma, aplicando técnicas e ferramentas de PLN, juntamente com as regras de mapeamento de texto em linguagem natural propostas nesta dissertação, é possível identificar semiautomaticamente elementos de processo de negócio em texto em linguagem natural. Tal identificação pode minimizar o esforço e tempo necessários pelo analista para extrair modelos de processo a partir de textos e agilizar assim a realização do projeto BPM. Portanto, o desafio de pesquisa abordado nesta dissertação visa desenvolver uma abordagem para identificar de forma semiautomática elementos de processo de negócio em texto em linguagem natural baseados em regras de mapeamento, descrever um protótipo de implementação e avaliá-lo.

## 1.2 Objetivos e Hipóteses

O objetivo geral desta dissertação é: *propor uma abordagem semiautomática para identificar elementos de processo de negócio em texto em linguagem natural.* As hipóteses de tal proposta são:

- É possível desenvolver uma abordagem para identificar elementos de processo de negócio, considerando os elementos notacionais da BPMN, em texto em linguagem natural.
- Tal abordagem auxiliará e minimizará o esforço necessário dos analistas de processo na etapa de modelagem do ciclo de vida de BPM.

A partir das hipóteses definidas, são descritos os seguintes objetivos específicos:

- Propor regras de mapeamento para identificar elementos de processo de negócio em

texto em linguagem natural.

- Implementar um protótipo fundamentado em técnicas e ferramentas de PLN tendo como base as regras de mapeamento desenvolvidas nesta dissertação, para identificar elementos de processo de negócio de forma semiautomática.
- Avaliar, através de pesquisa de opinião, as considerações de possíveis usuários quanto ao uso das regras de mapeamento definidas nesta dissertação e aplicabilidade delas em texto.
- Medir o desempenho do protótipo implementado por meio de métricas de recuperação de informação, tais como a precisão, revocação (*recall*) e medida-F (*F-measure*).

Busca-se, com a realização deste trabalho, contribuir com a comunidade acadêmica, oferecendo uma abordagem para identificação de elementos de processo em texto. Além disso, apresentar um protótipo que identifique de forma semiautomática tais elementos baseado nas regras de mapeamento propostas nesta dissertação. Portanto, tal abordagem auxiliará os analistas de processo na identificação de elementos e economizará o tempo da intensiva tarefa de modelagem de processos. Além disso, o trabalho proposto nesta dissertação representa o primeiro passo visando o desenvolvimento de uma abordagem para gerar texto orientado a processo para criação de modelos de processo de negócio a partir do mesmo (FERREIRA; THOM, 2016).

### 1.3 Contribuições da Pesquisa

As principais contribuições desta dissertação, as quais são detalhadas nas conclusões são:

- **Definição de Regras de Mapeamento:** entre texto em linguagem natural e elementos de processos de negócio. Tais regras possibilitam identificar elementos de processo em texto em linguagem natural. Essas regras foram baseadas em classes gramaticais da língua inglesa e dependências de palavras, e em seguida, aplicadas semi-automaticamente em texto em linguagem natural. As regras foram definidas com base no conjunto básico de elementos da BPMN. Tais elementos foram: Atividades, Eventos, *Desvio Exclusivo (XOR)*, *Desvio Paralelo (AND)* e *Raias*.
- **Protótipo:** Implementação de um protótipo para identificar elementos de processo de negócio de forma semiautomática em texto em linguagem natural. Tal imple-

mentação foi baseada nas regras de mapeamento e construída por duas ferramentas de PLN, denominadas como *Spacy*<sup>1</sup> e *PyDictionary*<sup>2</sup>.

- **Pressuposições:** Em um trabalho anterior, Ferreira and Thom (2016), foi conduzida uma abordagem introdutória para gerar texto orientado a processo a partir de texto em linguagem natural. Com base nesta, foi observado que o texto em linguagem natural deve ser pré-processado antes da extração de modelos de processo. Portanto, a abordagem apresentada nesta dissertação é um pré-requisito para gerar texto orientado a processo.

## 1.4 Organização do Texto

Esta dissertação está dividida em 5 capítulos, considerando esta introdução. Os demais capítulos estão dispostos da seguinte forma:

- o capítulo 2 expõe informações gerais sobre o uso, técnicas e objetivos de BPM, além de PLN, e apresenta seus conceitos básicos que foram utilizados nesta dissertação. Para BPM foi destacada a importância da aplicação desta disciplina em organizações e a relevância do uso de uma notação gráfica para modelagem de processos. Na área de PLN, foram apresentadas 3 análises de processamento de linguagem natural, tais como: morfológica, sintática e semântica. Além disso, são apresentados os trabalhos relacionados a esta dissertação, ilustrando as principais contribuições.
- o capítulo 3 apresenta a abordagem proposta e as regras de mapeamento expondo a metodologia de criação, definição, e conceitos inerentes ao seu entendimento. Ainda neste capítulo, são demonstrados os algoritmos de identificação baseados nas regras de mapeamento.
- o capítulo 4 mostra a avaliação da abordagem proposta nesta dissertação, através de pesquisas com usuários, métricas de desempenho e e apresenta uma análise crítica das avaliações realizadas;
- o capítulo 5 apresenta as conclusões deste trabalho, com destaque as principais contribuições, limitações e possibilidades de extensões do mesmo.

---

<sup>1</sup><https://spacy.io/>; último acesso em 28-12-2016

<sup>2</sup><https://pypi.python.org/pypi/PyDictionary/1.3.4>; último acesso em 28-12-2016

## 2 FUNDAMENTOS DE GERENCIAMENTO DE PROCESSOS DE NEGÓCIO E PROCESSAMENTO DE LINGUAGEM NATURAL

Este capítulo apresenta os conceitos fundamentais necessários para o entendimento desta dissertação. É apresentado o ciclo de vida de BPM e a notação BPMN. São discutidos conceitos de PLN, suas aplicações e relações com modelos de processos. São discutidos também os trabalhos relacionados à abordagem desenvolvida.

Segundo Dumas et al. (2013), BPM objetiva organizar a realização do trabalho nas organizações visando assegurar resultados consistentes e aproveitar oportunidades de melhoria. BPM se preocupa com a gestão de eventos, atividades, decisões e invocações de processos de negócio, buscando agregar valor às organizações e aos clientes.

Tecnologia da informação (TI) em geral e sistemas de informação (SI) (por exemplo, *Enterprise Resource Planning (ERP)* e *Workflow Management Systems (WfMSs)*) em particular, merecem um papel importante no gerenciamento de processos de negócios, pois cada vez mais atividades que uma organização realiza são apoiadas por sistemas de informação. As atividades de processos de negócio podem ser realizadas por colaboradores da organização manualmente ou com o apoio de SI. Há também atividades do processo que podem ser implementadas automaticamente por SI, sem qualquer envolvimento humano (por exemplo, softwares de geração de boletos de pagamento) (WESKE, 2007).

Uma organização pode atingir seus objetivos de negócio de forma eficiente e eficaz apenas se as pessoas e outros recursos empresariais, como SI, apresentarem cooperação. Portanto, os processos de negócio são um conceito importante para facilitar esta colaboração eficaz e produtiva para ambas as partes.

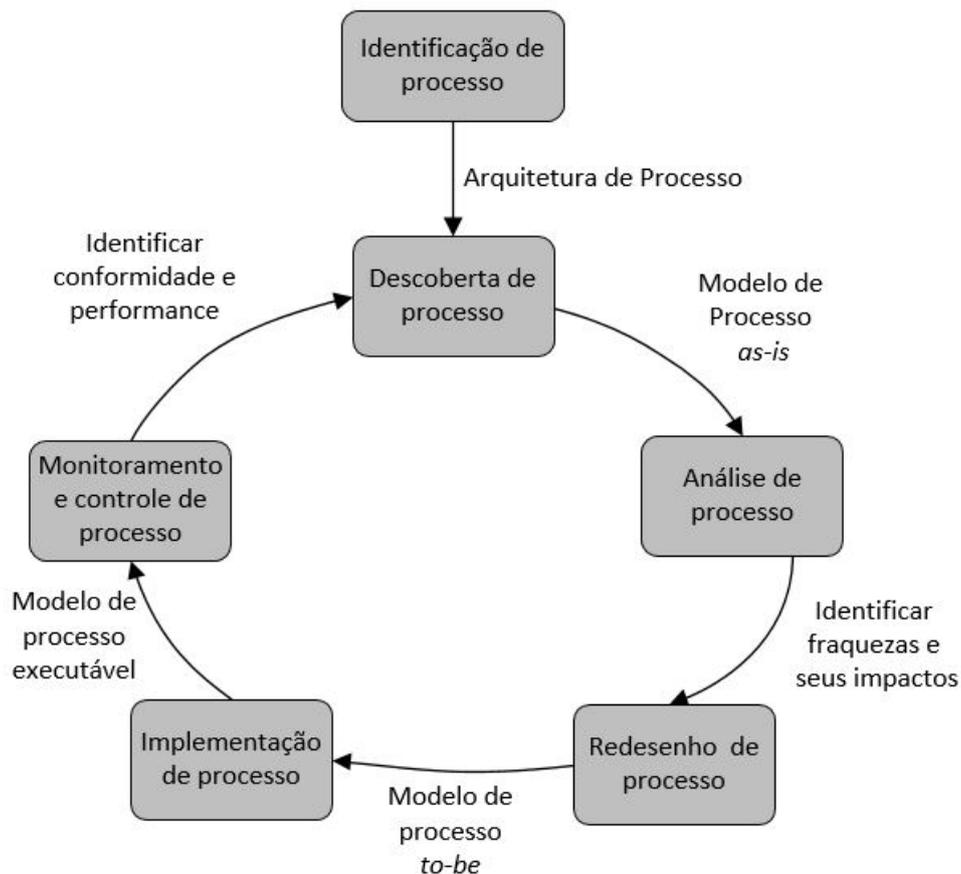
Para atingir maior gerenciamento e controle desses processos de negócio é necessário de conjunto de técnicas para a análise e melhoria contínua dos processos executados nas organizações. Portanto, BPM pode ser definido como um conjunto de métodos, técnicas e ferramentas para descoberta, análise, redesenho, execução e monitoramento de processos de negócio (DUMAS et al., 2013; WESKE, 2007). Através de BPM as organizações atingem uma transparência em todas as etapas do processo, maior controle administrativo, redução de custos, automação de processos e geração de evidências (por exemplo, documentos de textos, *logs* de SI etc) (ROSA et al., 2011). Conforme Fantinato et al. (2012), BPM pode ser visto como uma vantagem competitiva para as organizações.

## 2.1 Ciclo de Vida de BPM

Com base nas definições de BPM e processo de negócio, as atividades desta disciplina, são tipicamente organizadas no contexto de um ciclo de vida. Existem diversos autores que apresentam uma outra abordagem de ciclo de vida de BPM, tais como: (AALST, 2013; MUEHLEN; ROSEMANN, 2004; BROCKE; ROSEMANN, 2010; MENDLING, 2008). Nesta dissertação foi seguido o ciclo de vida proposto por Dumas et al. (2013), em que demonstra claramente como as diferentes etapas do ciclo de vida estão relacionadas aos modelos de processo. A Figura 2.1 ilustra o ciclo de vida proposto, incluindo as saídas que são produzidas por cada atividade do ciclo. No total, o ciclo de vida constitui de seis fases. Tais fases são:

- **Identificação de processo:** Com base em um objetivo ou problema de negócio, os processos relevantes são identificados, delimitados e relacionados entre si. O resultado desta identificação é uma arquitetura de processos que fornece uma visão geral dos processos e suas relações.
- **Descoberta de processo** (também denominada como modelagem *as-is*): Nesta fase, os processos selecionados são documentados em seus estados atuais. O resultado é um conjunto de modelos de processo (*as-is*).
- **Análise de processo:** Questões associadas ao processo são identificadas, documentadas e sempre que possível quantificadas, utilizando medidas de performance. Como saída têm-se um conjunto de questões estruturadas, que são organizadas conforme o seu impacto, e também em função do esforço requerido para realizar determinada ação. Essas questões são utilizadas para guiar a melhoria dos processos de negócio.
- **Redesenho de processo:** (também denominada como melhoria de processo): Objetiva identificar mudanças no processo que possam auxiliar na solução de problemas levantados na fase de análise e permite a organização conhecer os objetivos de performance relacionados ao processo.

Figura 2.1: Ciclo de vida de BPM



Fonte: Adaptado de (DUMAS et al., 2013).

Portanto, com base nas limitações do processo atual, nova versão do processo é desenvolvida. Os resultados estão documentados em um conjunto de modelos de processo (*to-be*). Nesta fase é importante observar que a análise de processo e redesenho de processo estão fortemente relacionadas. Técnicas de análise de processos podem ser utilizadas para identificar limitações de processos *to-be*. Assim, podem haver várias iterações de análise de processo e redesenho processo.

- **Implementação de processo:** As modificações necessárias para tornar o processo *as-is* em um processo *to-be* são preparadas e executadas. Esta fase contempla, principalmente a automação de processos, no qual abrange o uso de tecnologia da informação como suporte ao processo *to-be*, como por exemplo, um BPMS.
- **Monitoramento e controle de processo:** A fim de avaliar o desempenho do processo redesenhado, ele precisa ser monitorado e controlado. Por conseguinte, os dados da execução do processo são coletados e analisados com as medidas de desempenho pré-definidas. Com o objetivo de minimizar possíveis problemas, é importante monitorar e controlar continuamente o processo. Uma vez que um pro-

cesso não está executando corretamente com relação às medidas pré-definidas, uma nova iteração do ciclo de vida é realizada.

Considerando os diferentes resultados das etapas do ciclo de vida, torna-se claro que a modelagem de processos é uma atividade essencial no contexto de BPM. A modelagem é usada para documentar os processos *as-is*, para realizar análises subsequentes dos modelos de processo *to-be*.

Os modelos de processo fornecem uma representação abstraída de várias instâncias de processos de negócios que são as interações dentro de uma organização conduzida para criar valor (WESKE, 2007). Inclui atividades conduzidas por humanos e/ou softwares que são capazes de mostrar interdependências.

Para visualizar graficamente modelos de processo, podem ser utilizadas diferentes notações, tais como EPC (DAVIS; BRABANDER, 2007), YAWL (AALST; HOFSTEDE, 2005), BPMN (OMG, 2013) (ver seção 2.2), Diagrama de Atividades (WESKE, 2012; RUMBAUGH; JACOBSON; BOOCH, 2004) ou Redes de Petri (PETERSON, 1981).

O processo de modelagem é resultado de competências complementares necessárias para a criação de um modelo. Tais competências necessitam de: familiaridade com o processo, e conhecimento profundo da linguagem de modelagem aplicada (BPMN, EPC etc). Na prática, o problema é que essas competências são frequentemente distribuídas entre vários *stakeholders* (RITTGEN, 2007; DUMAS et al., 2013). Em particular, existem geralmente dois papéis que estão envolvidos no processo de modelagem: O *especialista de domínio* e o *analista de processos*. O especialista de domínio possui conhecimento detalhado das operações do processo de negócio em questão. Isto deve, por exemplo, resultar do fato de que o especialista de domínio é um participante do processo (WESKE, 2012). O analista de processos apresenta as competências necessárias para a atividade de modelagem de processos. Isso inclui, entre outros, o conhecimento da linguagem de modelagem e a capacidade de organizar adequadamente a informação no contexto de um modelo de processo (LEOPOLD, 2013). Como resultado desta distribuição de competências, o modelo é criado em conjunto por um especialista de domínio e um analista de processos. Por outro lado, estudos práticos demonstram que nem sempre isso acontece. Em muitos projetos de BPM, existem ainda analistas envolvidos que não estão suficientemente treinados (ROSEMANN, 2006).

## 2.2 Notação e Modelo de Processo de Negócio (BPMN)

Para a abordagem desenvolvida nesta dissertação, foi decidido utilizar a BPMN versão 2.0.2 como notação e modelo de processo de negócio. A BPMN<sup>1</sup> é um padrão oficial apoiado pela *Object Management Group* (OMG, 2013). A BPMN fornece uma grande variedade de elementos notacionais e boa aceitação por usuários leigos. Além disso, tal notação é baseada em diagrama de fluxos para definição de processos, além de permitir um mecanismo para a geração de um processo executável (BPEL) a partir de um diagrama de processo. O padrão OMG, é suportado por diversas ferramentas de modelagens de processos, tais como: *Bizagi Process Modeller*<sup>2</sup>, *Signavio*<sup>3</sup>, *TIBCO Business Studio*<sup>4</sup>, *IBM Websphere Business Modeler*<sup>5</sup>, *ARIS*<sup>6</sup>, *Oracle BPA*<sup>7</sup> entre outros. A BPMN considera um conjunto básico de elementos notacionais, conforme Figura 2.2:

- **Objetos de fluxo:** utilizados para definir o comportamento de um processo de negócio. Há três objetos de fluxo:
  - **Evento:** define algo que "acontece" no decorrer de um processo de negócio. Tal elemento afeta o fluxo e, normalmente, têm uma causa ou um impacto, além de requerer ou permitir alguma reação (DUMAS et al., 2013). Tal elemento abrange diversos aspectos de processo (por exemplo, o início de uma atividade, o fim de uma atividade, a mudança de estado em um documento) (OMG, 2011). Os eventos são representados por círculos, conforme os elementos *a* e *k* da Figura 2.3 e são divididos em três tipos:
    - *Evento de início:* indica o início de um processo (elemento *a* da Figura 2.3);
    - *Evento de fim:* indica o fim de um processo (elemento *k* da Figura 2.3);
    - *Evento intermediário:* indica algo que acontece em algum ponto do processo entre seu início e seu fim.
  - **Atividade:** é uma unidade de trabalho a ser executada no contexto de um processo de negócio, que pode ser atômica ou um subprocesso. Uma atividade

<sup>1</sup> A versão atual da BPMN está traduzida para pelo menos 10 idiomas. A versão em português esta disponível em: <https://goo.gl/kQnv9>; (THOM; IOCHPE, 2010)

<sup>2</sup> <http://www.bizagi.com/>; último acesso em: 04-01-2017

<sup>3</sup> <http://www.signavio.com/>; último acesso em: 04-01-2017

<sup>4</sup> <http://www.tibco.com/>; último acesso em: 04-01-2017

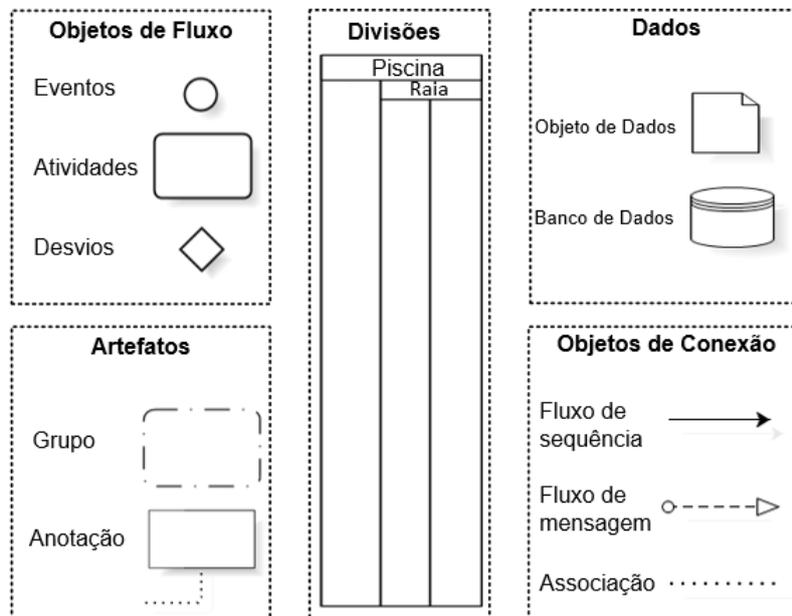
<sup>5</sup> <http://www.ibm.com/developerworks/downloads/ws/wbimod/>; último acesso em: 04-01-2017

<sup>6</sup> <http://www.ariscommunity.com/aris-express>; último acesso em: 04-01-2017

<sup>7</sup> <https://goo.gl/LQcHQK>; último acesso em: 04-01-2017

atômica é uma tarefa<sup>8</sup>, a qual representa uma ação que pode ser executada por uma pessoa ou sistema (DUMAS et al., 2013).

Figura 2.2: Categoria básica de elementos da BPMN



Fonte: Adaptado por (WESKE, 2012) e (OMG, 2013)

No exemplo de processo da Figura 2.3, os elementos *b*, *c*, *d*, *e*, *h*, *g*, *j* são considerados atividades, pois representam essas características. Um subprocesso<sup>9</sup> é conjunto de tarefas que possuam um propósito específico dentro de um processo. Além disso, podem ser abstraídas em uma outra unidade de processo e representadas no processo maior por um único objeto do tipo atividade (OMG, 2013). No contexto da Figura 2.3 não há um subprocesso definido. Porém, considerando hipoteticamente que durante a execução da atividade *e*, seja necessário executar um conjunto de atividades, então seria possível afirmar que *e* se trata de um subprocesso.

- Desvio: é usado para definir possíveis caminhos a serem tomados durante a execução de um processo. Tais caminhos podem ser de forma sequencial, paralela, ou por meio de uma seleção exclusiva. Um único desvio (*Gateway*) obtém múltiplas entradas e saídas. Analistas de processos e ferramentas de modelagem analisam as melhores práticas para modelar e aplicar os desvios da BPMN em modelos de processos (MENDLING; REIJERS; AALST, 2010;

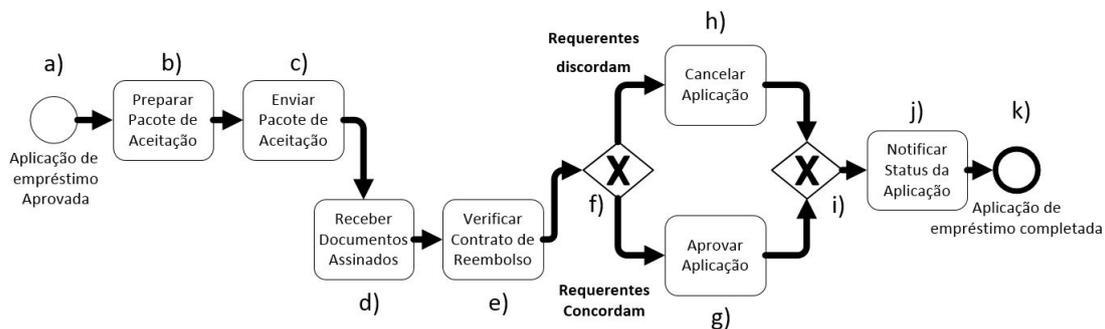
<sup>8</sup>As tarefas podem ser classificadas como: *Human Task*, *Service Task*, *Send Task/Receive Task*, *Manual Task*, *Script Task* e *Business Rule Task*

<sup>9</sup>Subprocessos são representados visualmente como retângulos com bordas arredondadas, semelhantes as tarefas atômicas, porém apresentam um símbolo [+] na base inferior.

MENDLING; NEUMANN; AALST, 2007; KINDLER, 2006). Na Figura 2.3, os elementos *f* e *i* representam desvios do processo.

- **Dados:** são os dados consumidos e/ou gerados pelo processo. É subdividido em: Objeto de dados e Banco de dados.
- **Artefatos:** representa informações adicionais sobre o processo. Tais elementos podem ser usados para representar as entradas ou saídas de uma atividade.

Figura 2.3: Exemplo de modelo de processo modelado em BPMN.



Fonte: Adaptado de (DUMAS et al., 2013)

- **Objetos de conexão:** representam a forma como os objetos de fluxo se conectam. São divididos em três tipos:
  - Fluxo de sequência: descreve a ordem do fluxo, geralmente é utilizado dentro de piscinas (*Pools*);
  - Fluxo de mensagem: demonstra o fluxo mensagens entre o emissor e o receptor. Normalmente utilizado entre piscinas;
  - Associação: Utilizada para associar textos, dados e outros artefatos aos objetos do fluxo.
- **Divisões:** Utilizados para agrupar os demais elementos notacionais. Fazem parte desse grupo: Piscina e Raia (*Lane*).

Vários trabalhos discutem a importância da utilização de rótulos (*labels*) em elementos no processo de modelagem (MENDLING; REIJERS; AALST, 2010; LAWRIE et al., 2006; SHARP; MCDERMOTT, 2001). Em um modelo de processo BPMN, aplicar rótulo aos elementos permite maior entendimento do processo (DUMAS et al., 2013; MENDLING; REIJERS; AALST, 2010; MENDLING, 2013; MENDLING; REIJERS; RECKER, 2010). Dumas et al. (2013) e Mendling, Reijers and Aalst (2010) descrevem que os rótulos para atividades devem começar com um *verbo* na forma imperativa, seguido de um *substantivo*, normalmente referenciado pelo objeto da sentença, por exem-

plo, "*Aprovar Ordem de Compra*". Para eventos deve começar com um *substantivo* (normalmente o objeto da sentença), seguido de um verbo na forma de passado do particípio da língua inglesa, por exemplo, "*Invoice emitted*". Nesta dissertação, tais *labels* são extremamente importantes para a identificação de elementos de processo em documentos textuais.

### 2.3 Processamento de Linguagem Natural (PLN)

Uma linguagem desenvolvida naturalmente sem planejamento e modelagem é considerada uma linguagem natural, por exemplo, a língua inglesa. Uma linguagem natural poderia ser considerada como a contrapartida de um código de computação (JOHN, 1993).

Técnicas e métodos na área de Linguística Computacional e Processamento de Linguagem Natural, tentam analisar e extrair informações relevantes a partir de textos em linguagem natural ou fala (discurso). Portanto, se refere, por exemplo, ao reconhecimento e síntese da fala em línguas naturais como a língua inglesa (JURAFSKY; MARTIN, 2000).

Jurafsky and Martin (2009) definem que o objetivo do processamento da linguagem natural é assegurar que "*computadores executem tarefas relevantes envolvendo a linguagem humana, tarefas como permitir a comunicação homem-máquina, melhorar a comunicação humano-humano ou simplesmente fazer processamento relevante de texto ou fala*". Portanto, o campo de pesquisa de PLN, busca-se da interação de máquinas e humanos, além de modelar o processamento em motivações, crenças, modos de raciocínio e outros estados mentais necessários ao processo de comunicação através da linguagem falada e/ou escrita.

Dada a importância do assunto, a complexidade das linguagens naturais - natureza ambígua - e a grande quantidade de problemas que podem abranger, muitas técnicas foram desenvolvidas para lidar com essa complexidade ao longo dos anos, como modelos estatísticos e semântica baseada em regras (ROTH, 1998). Diferentes métodos são comumente aplicados a tipos específicos de aplicações, como a sumarização de texto usando esquemas de ponderação, tradução automática (*machine translation*) aplicando a tradução automática estatística (por exemplo, *Google Translate*) e BPM usando a sintaxe e análise sintática das palavras.

Um sistema de processamento de linguagem natural descreve três níveis de conhe-

cimento (INDURKHYA; DAMERAU, 2010):

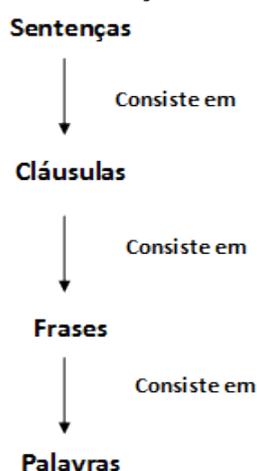
- Morfológico: estuda a construção das palavras, com seus radicais e afixos, que correspondem a partes estáticas e variantes das palavras, como as inflexões verbais;
- Sintático: trata das relações entre as palavras em uma frase (sentença), decidindo a que grupo de classificação pertence a palavra, de acordo com uma gramática (INDURKHYA; DAMERAU, 2010);
- Semântico: baseia-se nos resultados dos outros dois níveis (morfológico e sintático). O objetivo é definir o significado das palavras e sentenças com base no conhecimento de sua estrutura, relações e papel (INDURKHYA; DAMERAU, 2010);

Nesta dissertação, é destacado o conhecimento sintático das palavras, sendo este descrito detalhadamente na próxima seção.

### 2.3.1 Análise Sintática

A análise sintática estabelece como as palavras podem ser combinadas em sentenças, cláusulas e frases (KATE; GORDON, 2009). Conforme a Figura 2.4, uma sentença pode consistir em uma ou mais cláusulas, uma cláusula consiste em uma ou mais frases, e uma frase contém uma ou mais palavras.

Figura 2.4: Hierarquia de sentenças, cláusulas, frases e palavras



Fonte: Adaptado de (KATE; GORDON, 2009)

Para Ilustrar esses conceitos, considere o seguinte exemplo de frase:

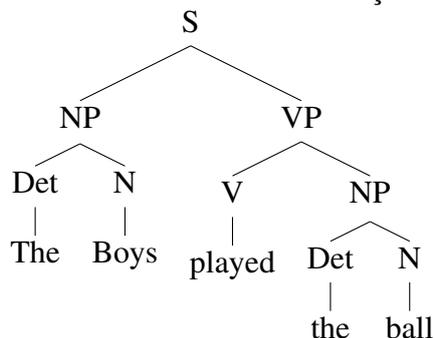
- *The boys played the ball.*

Ao nível de palavras, podemos identificar diferentes categorias léxicas, ou as chamadas *parts of speech*. As principais *parts of speech* incluem substantivos (N), verbos (V), adjetivos (Adj), advérbios (Adv), preposições (Prep) e determinantes (Det). *Parts of speech tags* fornecem informações significativas sobre uma palavra em seu contexto. Além disso, fornece informações sobre a inflexão<sup>10</sup> de uma palavra (KOK; BROUWER, 2011). A lista com todas as categorias léxicas explicadas detalhadamente pode ser vista em *Guidelines for the Clear Style Constituent* (CHOI; PALMER, 2012). Todavia, nesta dissertação é apresentada uma visão geral sobre as categorias mais utilizadas, que são descritas na Tabela 2.1. Assim, a sentença de exemplo pode ser descrita como segue:

- The – Det
- Boys – N
- Played – V
- the – Det
- ball – N

Além do *parts of speech*, também é possível atribuir funções gramaticais para as palavras. Em geral, as sentenças da língua inglesa contêm um sujeito e um predicado. Na sentença do exemplo, o sujeito é dado pelo *Boys* e pelo predicado *Played*. Normalmente, as frases são enriquecidas pela adição de objetos e advérbios. Consequentemente, a sentença de exemplo contém ainda o objeto *Ball* para especificar que o sujeito (*Boys*) da sentença realmente jogaram a bola. A Figura 2.5 representa a árvore sintática do exemplo.

Figura 2.5: Árvore sintática da sentença do exemplo



Fonte: Os autores

Na área de processamento de texto, um objetivo corresponde a automatização do *parts of speech*, o reconhecimento da estrutura sintática, isto é, como as palavras formam uma frase, e relações gramaticais entre palavras dentro de uma sentença. Exemplos

<sup>10</sup>Inflexão é a modificação de uma palavra para expressar diferentes categorias gramaticais, tais como modo, tempo, voz, aspecto, pessoa, número, gênero e caso (BRINTON, 2000).

Tabela 2.1: Categorias léxicas de um analisador sintático.

- **Agente (agent)** – Descreve o complemento de um verbo passivo. Encontra-se geralmente em uma frase prepositiva, introduzida pela palavra "by".
- **Complemento causal (ccomp)** – Um complemento clausal de um verbo ou adjetivo reflete uma cláusula dependente com um sujeito que funciona como um objeto do verbo, ou adjetivo.
- **Conjunção (conj)** – Descreve uma relação de conjunção entre dois elementos, como "and" e "or".
- **Objeto Direto (dobj)** – Um objeto direto que segue diretamente um verbo na sentença.
- **(Mark)** – Aponta para a palavra que identifica um complemento clausal adverbial, tais como "if", "only", "because" ou "after".
- **Negação - (neg)** – O modificador de negação indica que uma palavra é negada, por exemplo, usando "not".
- **Substantivo (nn)** – Um substantivo de modificador composto demonstra o substantivo que modifica a raiz de uma sentença nominal.
- **Sujeito nominal (nsubj)** – representa o sujeito ativo de uma sentença.
- **Sujeito nominal passivo (nsubjpass)** – corresponde o sujeito sintático em uma cláusula passiva.
- **Preposição (prep)** – preposição de um verbo, adjetivo ou substantivo ou qualquer frase prepositiva que atue para modificar o significado do verbo, adjetivo ou substantivo Marnee and Manning (2009).
- **Modificador de cláusula relativa (rcmod)** – Indica a presença de uma cláusula relativa que modifica uma palavra.
- **Complemento causal (xcomp)** – Um complemento clausal representa uma cláusula sem sujeito interno que modifica a raiz de uma sentença adjetiva (ADJP) ou sentença adverbial (ADVP) ou frase verbal (VP) ou sentença declarativa invertida (SINV).

---

Fonte: Adaptada de Choi and Palmer (2012)

de tais analisadores são o *UC Berkley parser*<sup>11</sup>, *Stanford Parser*<sup>12</sup> e o *Spacy*<sup>13</sup>. Outros analisadores (*POS-taggers*) são, por exemplo, *Brill tagger* (BRILL, 1992) ou disponível gratuitamente como *NLTK*<sup>14</sup>, *OpenNLP*<sup>15</sup>, *GATE*<sup>16</sup>, ou o *RASP system* (BRISCOE; CARROLL; WATSON, 2006), em que além de ser um analisador, também contém outras utilidades como *tokenization* e utilitários de transformação.

---

<sup>11</sup><http://nlp.cs.berkeley.edu/projects/historical.shtml#Parsing>; último acesso em 28-12-2016

<sup>12</sup><http://nlp.stanford.edu/software/lex-parser.shtml>; último acesso em 28-12-2016

<sup>13</sup><https://spacy.io/>; último acesso em 28-12-2016

<sup>14</sup>[www.nltk.org/](http://www.nltk.org/); último acesso em 28-12-2016

<sup>15</sup><https://opennlp.apache.org/>; último acesso em 28-12-2016

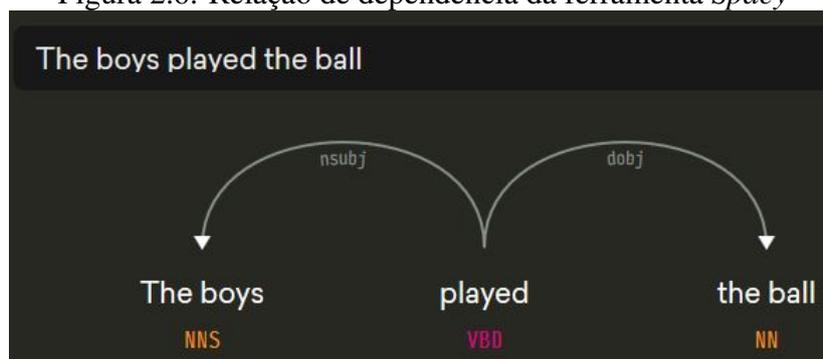
<sup>16</sup><https://gate.ac.uk/>; último acesso em 28-12-2016

### 2.3.2 Analisador Sintático - *Spacy*

O objetivo de um analisador sintático é determinar a estrutura (sintática) do texto de entrada na ferramenta. De acordo com Allen (1995), para analisar o analisador sintático, precisa-se realizar três particularidades: primeiro, um analisador tem como entrada uma sentença e como resultado produz a análise; segundo, uma gramática é obtida a partir de um conjunto de regras que o analisador pode utilizar; e terceiro, um léxico, definido por um dicionário de palavras corretas e *parts of speech* (por exemplo, verbo, advérbio, adjetivo, etc.).

Nesta dissertação, o analisador sintático escolhido foi o *Spacy*. O analisador possui boa relação com a linguagem de programação *Python* (BIRD; KLEIN; LOPER, 2009), que foi a linguagem selecionada para o desenvolvimento do protótipo desta abordagem. As características do *Spacy* são: *tokenization*, dado que, as sentenças são separadas em um vetor de objetos; *part of speech tagging* e *dependency parsing*, uma vez que as regras de mapeamento são baseadas em tais características; *lemmatization*, utilizado para verificar rótulos dos elementos de processo identificados no contexto desta abordagem. A Figura 2.6, mostra relação de dependência entre as palavras na sentença gerada pelo *Spacy*, em que "The boys" representa o sujeito nominal do verbo "Play" e "ball" equivale ao objeto direto da sentença.

Figura 2.6: Relação de dependência da ferramenta *Spacy*



Fonte: Desenvolvido por displaCy<sup>17</sup>

A escolha desta ferramenta, baseia-se na acurácia do analisador realizado por Choi, Tetreault and Stent (2015). O autor mostra que a ferramenta obteve uma acurácia (*overall accuracy*) extremamente satisfatória em relação a testes<sup>18</sup> realizados com outros analisadores (*parsers*). Em média, a ferramenta obteve uma acurácia de 90,53%

<sup>18</sup>As principais métricas relatadas são: *labeled attachment score* (LAS), *unlabeled attachment score* (UAS), *label accuracy score* (LS) e *exact match* (EM)

de exatidão. Os testes de velocidades (*Overall Speed*), mostram que a ferramenta obteve a maior performance dentre as outras. Portanto, o *Spacy* suporta todos os requisitos e condições (por exemplo, identificação de *parts of speech* e dependência entre palavras) para o desenvolvimento do protótipo apresentado nesta dissertação, além de ser código aberto<sup>19</sup>.

## 2.4 Trabalhos Relacionados

Ao longo da dissertação diversos trabalhos que contribuam para esse trabalho foram citados, porém esta seção discute trabalhos específicos do que vêm sendo propostos em áreas relacionadas a abordada nesta dissertação. Portanto, o estado da arte pode ser dividido em duas categorias: a extração de modelos de processo a partir de texto em linguagem natural e geração de texto a partir de modelos de processo. A Tabela 2.2, provê uma visão geral do estado da arte.

Com base na análise de trabalhos explorando a extração de modelos de processo a partir de texto em linguagem natural se pode destacar dois aspectos principais que devem ser considerados no contexto desta dissertação. Primeiro, deve ser considerada a origem da informação do texto em linguagem natural. Em Friedrich, Mendling and Puhlmann (2011), é proposta a extração de modelos de processo a partir de descrições textuais. A abordagem proposta pelos autores, considera três perspectivas: análise sintática, isto é, determinação de uma árvore sintática e de relações gramaticais entre as particularidades de sentenças; análise semântica, extração do significado de palavras ou sentenças; e resolução da anáfora, identificação de conceitos que são referências usando pronomes (*we, he e it*) e artigos (*this, that*). As limitações desta abordagem, partem do pressuposto que os textos analisados para a geração de modelos de processo, estão gramaticalmente corretos na língua inglesa, isto é, retira-se e/ou corrige manualmente palavras ou frases que estão incorretas gramaticalmente, tornando o texto a ser analisado pela ferramenta, gramaticalmente correto. Além disso, o texto necessita ser descrito sequencialmente. Isto é, o texto deve conter uma ordem cronológica dos fatos.

O segundo aspecto se refere a modelos de processos extraídos de texto em linguagem natural. Em Santoro, Gonçalves and Baiao (2009) e Gonçalves, Santoro and Baiao (2011), são descritas abordagens que exploram a técnica narrativa associada à mineração de texto e interpretação de linguagem natural para a geração de modelos de processo. O

---

<sup>19</sup><https://github.com/explosion/spaCy>; último acesso em 06-01-2017

trabalho mostra, que possíveis problemas de comunicação podem ocorrer, por exemplo, individualmente os participantes podem apontar suas visões gerais do processo dentro de uma estória. Desta forma, existe a possibilidade de múltiplos fluxos de trabalho (*workflows*) relatando o mesmo processo de negócio.

Em Jiexun et al. (2010), é proposta uma ferramenta de mineração denominada *policy-based process mining (PBPM)*, cujo objetivo é realizar a descoberta automática de modelos de processo baseados em políticas empresariais. Considerando que políticas empresariais é um tópico novo na pesquisa de BPM e de mineração de texto, a abordagem exige esforços adicionais para ser validada inteiramente e produzir soluções práticas. Todavia, existem aspectos positivos, tal como a identificação de modelos de processo em texto através de técnicas de classificações de texto. Além disso, essa abordagem utilizou métricas de recuperação de informação para avaliar o desempenho da ferramenta proposta no trabalho, assim como a utilizada nesta dissertação (WANG; ZHAO; ZHANG, 2009).

Em relação à geração de texto a partir de modelos de processo, foram identificadas três abordagens principais. Em Leopold (2013), é apresentada uma abordagem para extração de texto em linguagens natural a partir de modelos de processo. Tal abordagem, retrata diversos desafios para obter textos a partir de modelos de processo, tais como: planejamento do texto; planejamento de sentença; realização de superfície<sup>20</sup> e flexibilidade. Para a proposta apresentada nesta dissertação, é de especial interesse a abordagem do planejamento de texto. Dado que tal planejamento está dividida em três etapas: extração linguística de informação, onde ocorre a extração de componentes linguísticos dos rótulos (*labels*) dos elementos dos modelos de processo; geração *Annotated rpST*, faz a linearização do modelo de processo através da geração de uma estrutura de árvore; estruturação do texto, onde ocorre a aplicação de técnicas de estrutura de texto, como a inserção de *n* pontos com base na estrutura de árvore computadorizada.

As limitações presentes neste trabalho mostram evidências de que as frases geradas pela ferramenta são comparativamente curtas e simples. Outra limitação, ocorre em garantir um nível estável de complexidade entre os textos criados manualmente, assim, seria necessário treinar os escritores de texto. Além disso, tais textos apresentam uma linguagem extremamente técnica, por exemplo, *control flow*, *parallel gateway*, *exclusive gateway*. Desta forma, os analistas iniciantes apresentam dificuldade para extrair modelos ou fragmentos de processo a partir do texto gerado pela abordagem. Por fim, se a modelagem de processos não está bem documentada, o texto gerado também não estará. A

---

<sup>20</sup>Utilização de uma ferramenta denominada RealPro (LAVOIE; RAMBOW, 1997), na qual tem como entrada uma árvore sintática de sentenças e retorna as mesmas, gramaticalmente corretas.

fim de reduzir o tempo e esforço necessários para definir modelos de processo a partir de descrições textuais, (HAN; LEOPOLD; REIJERS, 2015) e (HAN; HENRIK; A., 2016), discutem uma abordagem para identificar inconsistências entre um modelo de processo e uma descrição textual correspondente ao modelo. Tal abordagem pode ser utilizada para identificar modelos de processo em um conjunto de descrições textuais que divergem de tais modelos.

Tabela 2.2: Estado da Arte para Identificar Elementos de Processo em Texto em Linguagem Natural

<b>Categorias</b>
<p><b>Extração de Modelos de Processo a Partir de Texto</b></p> <ul style="list-style-type: none"> <li>- Gerar Modelos de Processo a Partir de Texto.           <ul style="list-style-type: none"> <li>- (<i>FRIEDRICH; MENDLING; PUHLMANN, 2011</i>)</li> <li>- (<i>CHUENG; KOLIADIS; GHOSE, 2007</i>)</li> </ul> </li> <li>- Mineração de Processos a Partir de Texto em Linguagem Natural           <ul style="list-style-type: none"> <li>- (<i>SANTORO; GONÇALVES; BAIAO, 2009</i>)</li> <li>- (<i>JIEXUN et al., 2010</i>)</li> <li>- (<i>GONÇALVES; SANTORO; BAIAO, 2011</i>)</li> </ul> </li> </ul>
<p><b>Geração de Texto a Partir de Modelos de Processo</b></p> <ul style="list-style-type: none"> <li>- Gerar Texto a Partir de Modelos de Processo           <ul style="list-style-type: none"> <li>- (<i>LEOPOLD; MENDLING; POLYVYANY, 2014</i>)</li> <li>- (<i>MEITZ; LEOPOLD; MENDLING, 2013</i>)</li> <li>- (<i>LEOPOLD, 2013</i>)</li> </ul> </li> <li>- Inconsistências entre Modelos de Processo e Textos           <ul style="list-style-type: none"> <li>- (<i>HAN; LEOPOLD; REIJERS, 2015</i>)</li> <li>- (<i>HAN; HENRIK; A., 2016</i>)</li> </ul> </li> </ul>

Fonte: Os autores

Com o objetivo de extrair modelos de processos e garantir a qualidade do processo extraído a partir de documentos textuais, BORDIGNON (2016) realizou uma revisão sistemática cujo objetivo é apresentar uma revisão sobre técnicas e ferramentas de PLN para preparar e processar texto em linguagem natural. Foram encontrados 49 trabalhos que utilizam técnicas e/ou ferramentas de PLN nas etapas de identificação de processo, descoberta e análise do ciclo de vida de BPM. A maioria das ferramentas na etapa de descoberta de processo aplicaram *parsers* e *taggers* como técnica e as mais utilizadas foram *Trigram Tagger*, *Stanford Parser* e *Tagger*.

## 2.5 Considerações Finais

O capítulo 2 apresentou uma visão geral de conceitos utilizados nesta dissertação, tais como BPM, BPMN e PLN.

Em relação ao contexto de BPM, a modelagem de processos pode ser considerada uma das mais importantes e duradouras etapas do ciclo de vida de BPM. Tal etapa tem como propósito documentar, melhorar e padronizar processos de negócio, além de automatizar processos com sistemas de *Workflow* (por exemplo, BPMS). A abordagem apresentada nesta dissertação, se refere a esta etapa do ciclo de vida de BPM. Dado que, tal abordagem possa auxiliar os analistas em modelar o processo inicial (*as-is*) do ciclo de vida.

Além disso, este capítulo apresentou conceitos fundamentais da BPMN utilizados ao longo desta dissertação. A maioria destes conceitos são padronizados pela OMG e têm sido utilizados não apenas pela academia, mas também em toda a indústria (OMG, 2013). No contexto desta dissertação, foram selecionados quatro elementos de processo baseados na BPMN para serem identificados conforme as regras de mapeamento, tais como: atividades, eventos, desvio paralelo (AND), desvio exclusivo (XOR) e raias.

Em PLN, foram abordadas as três análises de processamento de linguagem natural, tais como: morfológica, sintática e semântica. Nesta dissertação, será conduzida a análise sintática, na qual visou gerar uma estrutura de árvore que represente a estrutura sintática da sentença analisada. Além disso, neste capítulo, foram abordadas as diferentes categorias léxicas presentes na análise sintática. Dado que, tais categorias juntamente com a dependência de palavras definiram as regras de mapeamento propostas no capítulo 3 desta dissertação.

No contexto dos trabalhos relacionados, existem abordagens que realizam a geração de textos a partir de modelos de processo. Tais textos, apresentam diversas palavras-chave extremamente técnicas em relação a experiência em modelagem de processos dos analistas novatos. Desta forma, tais analistas apresentam dificuldade para representar os modelos de processo extraídos de tal texto. Além disso, outras abordagens exigem que os textos sejam organizados seguindo uma ordem cronológica dos fatos.

A fim de minimizar este problema, em um trabalho anterior (FERREIRA; THOM, 2016), propuseram uma abordagem introdutória para gerar texto orientado a processo a partir de texto em linguagem natural. A partir desta abordagem, foi observado que o texto em linguagem natural deve ser processado antes da extração de modelos de processo. Por-

Tabela 2.3: Comparação dos trabalhos relacionados com a abordagem apresentada nesta dissertação.

<b>Autores</b>	RI	EM	GT	GD	ME
SANTORO; GONCALVES; BAIÃO, 2009		X			
FRIEDRICH; MENDLING; PUHLMANN, 2011		X			
CHUENG; KOLIADIS; GHOSE, 2007	X	X			
LEOPOLD, 2013			X	X	
JIEXUN et al., 2010		X			
<b>Abordagem desta dissertação</b>	X		X	X	X

Fonte: Os autores. Siglas: RI – Regras para Identificação de Modelos de Processo, EM – Extração de Modelos, GT – Geração de Texto, GD – Regras de Modelagem, ME – Marcação de Elementos de Processo.

tanto, a abordagem apresentada nesta dissertação é o primeiro passo para gerar um texto orientado a processo, realizando a marcação (identificação) de elementos de processo em texto para posteriormente gerar tal texto. A Tabela 2.3 mostra uma visão comparativa dos trabalhos relacionados com a abordagem desta dissertação, em que tal abordagem avança o conhecimento justamente na marcação dos elementos de processo em texto.

Com base nas abordagens apresentadas, este trabalho provê um caminho através das regras de mapeamento, para que seja possível identificar elementos de processo de negócio em texto e minimizar o trabalho do analista na etapa de modelagem de processos. Portanto, o próximo capítulo desta dissertação, apresenta a abordagem proposta.

### **3 UMA ABORDAGEM PARA IDENTIFICAÇÃO DE ELEMENTOS DE PROCESSO DE NEGÓCIO EM TEXTO EM LINGUAGEM NATURAL**

Neste capítulo é apresentada uma abordagem para identificação de elementos de processo de negócio em texto em linguagem natural. Em um trabalho anterior (FERREIRA; THOM, 2016), foi realizada uma abordagem introdutória para gerar texto orientado a processo a partir de texto em linguagem natural. Tal pesquisa caracteriza-se como base para a abordagem apresentada nesta dissertação.

A Figura 3.1 mostra a abordagem para identificar elementos de processo de negócio em texto em linguagem natural. Tal abordagem consiste em 4 etapas. São elas:

1. Entrada de dados
2. Análise sintática de texto
3. Análise lógica de texto
4. Saída de texto

A entrada de dados (texto em linguagem natural) se faz necessária, pois a partir destes textos é realizada a análise sintática. Tal análise tem função de classificar todas as sentenças do texto pelas categorias gramaticais (*parts of speech*) e relações de dependência entre as palavras.

A partir desta análise, são empregadas regras de mapeamento, para que se possa identificar elementos de processo em texto em linguagem natural.

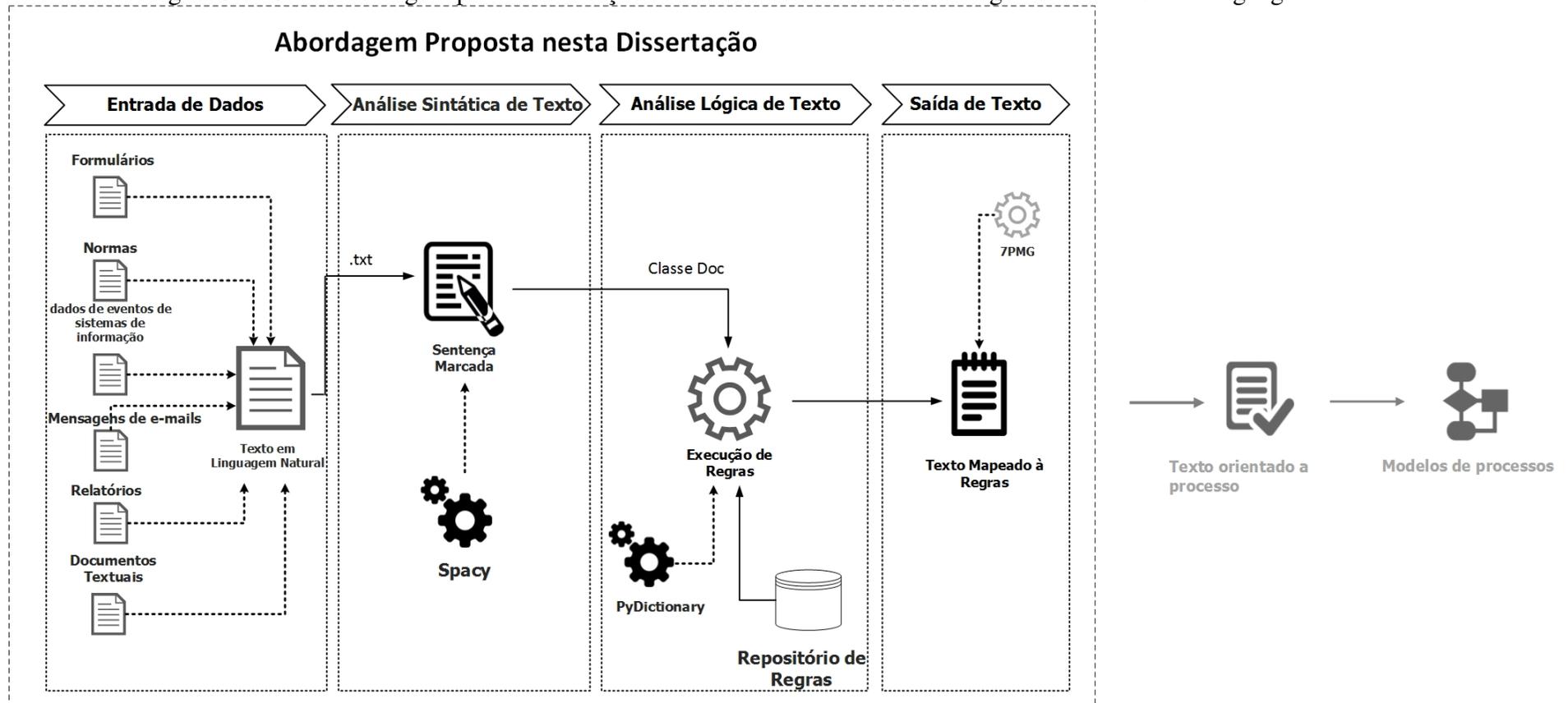
Por fim, a partir do momento que as regras de mapeamento são aplicadas em texto e os elementos de processo são identificados, tais sentenças são possíveis candidatas à modelagem de processos a partir do texto analisado.

Nas seções seguintes, serão apresentadas e demonstradas tais etapas em detalhes.

#### **3.1 Entrada de Dados**

Na língua inglesa existem diversas classificações de textos. Cada classificação apresenta características diferentes, tais como palavras, sentenças e questões relacionadas com cada particularidade do texto. Exemplos de classificações incluem: textos descritivos; comparativos e contraste; ordem de importância; problema e solução; causa e efeito; sequenciais. Em textos sequenciais, a informação é organizada em etapas ou um processo é descrito na ordem em que ocorre (WORKSHEETS, 2014).

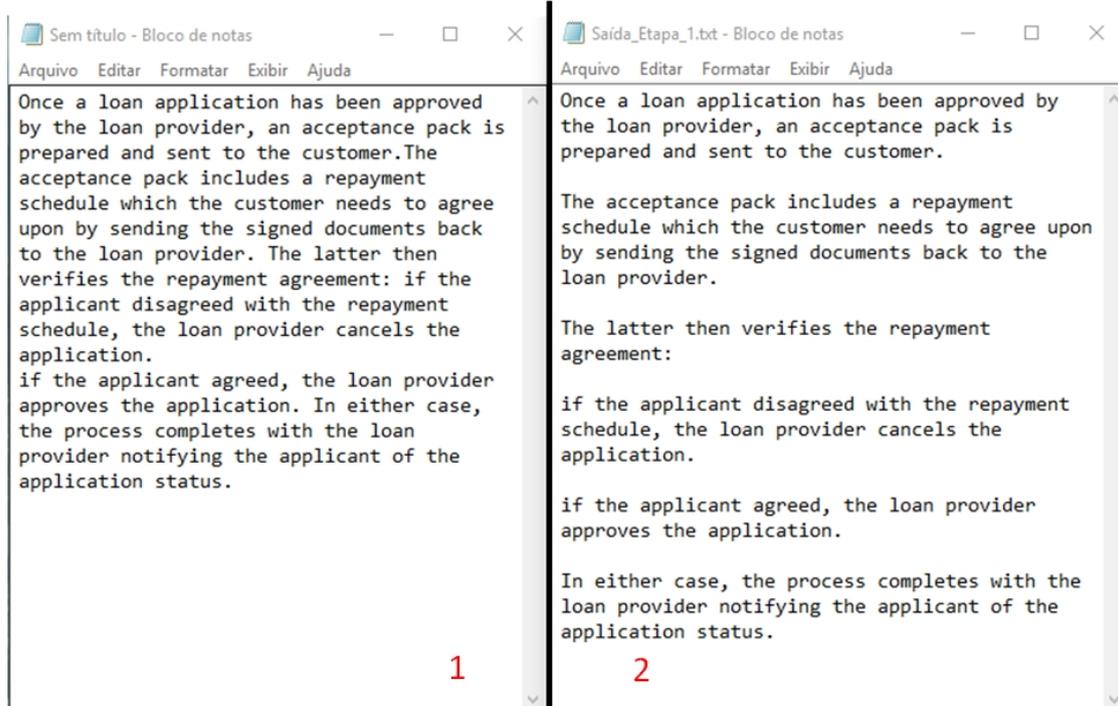
Figura 3.1: Uma Abordagem para Identificação de Elementos de Processo de Negócio em Texto em Linguagem Natural



Fonte: Os autores

As características presentes nos textos de classificação sequencial, há particularidades presentes em modelos de processo de negócio, ou seja, existem palavras-chave que são recorrentes nestes textos, tais como: *primeiro*, *segundo*, *próximo*, *então*, *finalmente*, *seguinte*, *agora*, *depois*, entre outras. Tais palavras apresentam possíveis relações (correlações) com elementos de modelagem da BPMN, tais como atividades, desvios, piscinas, raias, etc. Por outro lado, Jiexun et al. (2010) descreve que a extração de modelos ou fragmentos de processo a partir de textos, pode ser através de textos heterogêneos e não apenas textos homogêneos (por exemplo, uma ferramenta de extração de modelos de processo a partir de textos sequenciais).

Figura 3.2: Saída de dados separados por sentenças.



Fonte: Os autores

Neste contexto, tais textos podem ter diversas origens, como: formulários, normas, dados de eventos de sistemas de informação (*logs*), mensagens de e-mails, etc. Conforme Dumas et al. (2013), tais origens são também chamadas de textos em linguagem natural, incluindo sentenças não estruturadas que não correspondem aos textos sequenciais (por exemplo, textos descritivos) e, portanto, tornar-se difícil a extração de modelos de processo a partir dele.

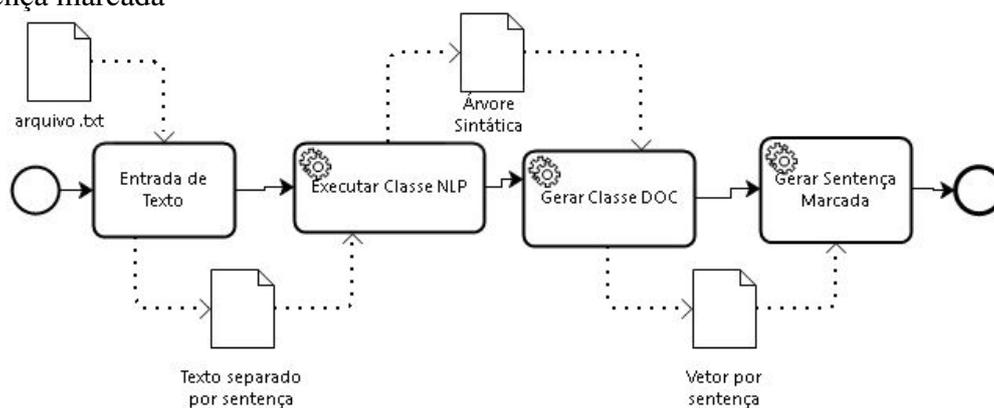
Portanto, esta etapa visa separar o texto por sentenças manualmente. Uma vez que, o analisador sintático realiza a análise sintática através de sentenças. Para isto, a Figura 3.2 descreve um exemplo de um documento textual (1) e a saída do mesmo documento

desta etapa com formato *.txt* separados por sentenças (2).

### 3.2 Análise Sintática de Texto

A fim de obter uma sentença marcada, dado que, tal sentença será analisada na próxima etapa pelas regras de mapeamento, é necessário considerar a análise sintática do documento textual. A proposta da análise sintática é determinar a estrutura (sintática) de textos de entrada. Conforme descrito no capítulo 2, o analisador sintático considera diversas características, incluindo: *tokenization*, *lemmatization*, *part of speech tagging*, etc. O *Part of speech* provê informações relevantes a respeito das regras gramaticais de uma palavra em seu contexto, além de fornecer informações sobre a inflexão<sup>1</sup> de uma palavra.

Figura 3.3: Visão geral estrutural da etapa de análise sintática de texto para gerar uma sentença marcada



Fonte: Os autores

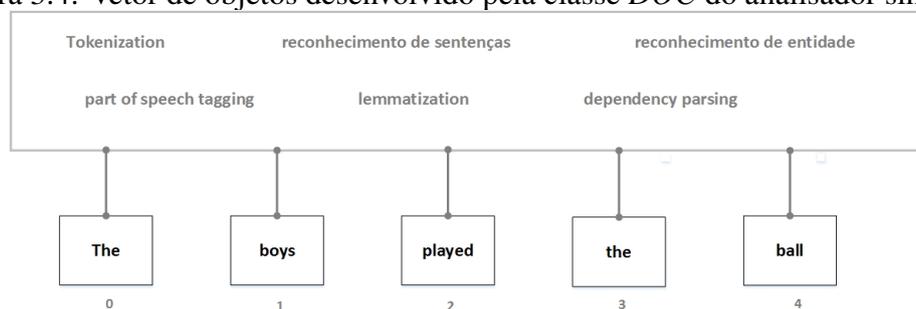
Nesta dissertação, as relações de dependência e o *part of speech tagging* são importantes componentes para identificar elementos de processo em sentenças. A fim de identificar o *parts of speech tagging* de sentenças no analisador sintático, é necessário dividir todas as sentenças do texto analisado. O resultado é um arquivo *.txt* separado por sentenças, como mostrado na tarefa "Entrada de Texto" da Figura 3.3.

Como o objetivo de gerar automaticamente uma sentença marcada, o algoritmo 1 representa tal objetivo. Tal algoritmo, descreve uma estrutura de repetição (*for* - linha 3), que percorre todas as sentenças (variável: *sent*) do texto. Em seguida, para realizar a análise sintática da sentença, é necessário executar a classe *NLP*. Tal classe tem como

<sup>1</sup>Inflexão é a modificação de uma palavra para expressar diferentes categorias gramaticais, tais como modo, tempo, voz, aspecto, pessoa, número, gênero e caso (BRINTON, 2000).

função identificar e relacionar todas as palavras em uma árvore sintática (essa classe é descrita no algoritmo 1 na linha 4). A árvore contém todas relações morfológicas das palavras. Essa etapa é descrita na tarefa de serviço "*Executar Classe NLP*" na Figura 3.3. Posteriormente, o analisador gera uma classe denominada como *DOC*, ilustrada na tarefa "*Gerar classe DOC*" da Figura 3.3. Esta classe é um vetor de objetos, cujo tamanho é equivalente ao número de palavras na sentença, em que cada posição é uma palavra da sentença que será manipulada na próxima etapa (análise lógica de texto). A Linha 5 do algoritmo 1 descreve a execução da classe.

Figura 3.4: Vetor de objetos desenvolvido pela classe *DOC* do analisador sintático.



Fonte: Os autores

Além disso, cada posição do vetor contém todas as características (por exemplo, *tokenization*, reconhecimento de sentenças, *part of speech tagging*, *lemmatization*, *dependency parsing*, e reconhecimento de entidade) da palavra em uma sentença. A Figura 3.4 descreve tais características do vetor conforme o exemplo: "*The Boys played the ball*".

---

#### Algoritmo 1 Algoritmo de Execução da Análise Sintática de Texto

---

**Entrada:** Arquivo *.txt*

**Saída:** Vetor de objetos (DOC)

- 1: **procedure** EXECUTESYNTACTICANALYSIS(*.txtFile*)
- 2:     *nlp* ← *English*()
- 3:     **for**  $\forall$  *sent* **in** *sentences* **do**
- 4:         *ExecuteNLPClass* ← *nlp*(*sent*)
- 5:         *Doc* ← *ExecuteNLPClass*(*sent*)
- 6:     **return** *Doc*

Fonte: Os autores

---

Finalmente, gera-se uma sentença marcada com todas as sentenças analisadas pelo analisador sintático. A Figura 3.3 mostra uma visão geral desta etapa.

### 3.3 Análise Lógica de Texto

A fim de auxiliar e minimizar o esforço do analista de processos na etapa de modelagem, foi desenvolvido, no contexto desta dissertação, **um conjunto de regras de mapeamento para identificação de elementos de processo de negócio em texto em linguagem natural**.

As regras se originam de um diversificado conjunto de classes gramaticais (por exemplo, *verbo*, *pronome*, *artigo*, *substantivo* etc) e relações de dependências (por exemplo, *sujeito*, *objeto direto*, *objeto indireto* etc). Com base no estudo sobre classes gramaticais realizado no contexto desta dissertação, não existe um padrão em que as classes gramaticais devem ser apresentadas no texto. Sendo assim, elas estão relacionadas entre si e representam um elemento de processo. Por exemplo, as sentenças que contém *Sujeito*, *Verbo* e *Objeto* representam um elemento de processo, tal como: Tarefa Manual e não dependem da ordem em que aparecem no texto. Em algumas sentenças, as classes gramaticais mesclam entre si, por exemplo: sentenças que contém a ordem de *Verbo*, seguindo de um *Sujeito* e posteriormente um *Objeto*, representam uma tarefa manual do processo e estão posicionadas de maneiras diferentes na sentença.

#### 3.3.1 Definição de Regras de Mapeamento

As regras foram definidas manualmente, dado que, foram analisados diversos textos, para que se observa-se padrões de possíveis elementos de processo a partir de classes gramaticais e dependências entre palavras. Os textos analisados foram de: Dumas et al. (2013) e Weske (2012). Desta forma, cada regra é classificada conforme a categoria do conjunto de elementos reduzidos da BPMN (por exemplo, objetos de fluxo e divisões) proposto pela OMG (2013) e pelo fato deste conjunto ser recorrentes em processos de negócio. Tais elementos são: atividades, eventos, desvio paralelo (AND), desvio exclusivo (XOR) e raias (*swimlanes*). Nesta dissertação, não foi incluído o desvio inclusivo (OR), pois existem diversas pesquisas que apresentam problemas estruturais e semânticos com tal elemento Mendling, Neumann and Aalst (2007), Kindler (2006), Figl, Recker and Mendling (2013), Kossak, Illibauer and Geist (2012). Pesquisas também apontam ambiguidades na semântica do desvio inclusivo (*OR-join*) (KINDLER, 2006). Além disso, modelos de processo que possuem somente desvio paralelo (AND) e desvio exclu-

Tabela 3.1: Regras para identificação de atividades.

<b>Atividades</b>		
<b>Regras</b>	<b>Descrição</b>	<b>Exemplo de Sentença</b>
Regra 1	<sujeito> +<verbo> +<objeto>	<i>The Support Officer</i> <sujeito> <i>updates</i> <verbo> <i>all group calendars</i> <objeto>
Regra 2	<sujeito>+<aux> +<verbo>+<objeto><(no futuro)>	<i>The secretary</i> <sujeito> <i>will</i> <aux> <i>send</i> <verbo> <i>to dispatch</i> <objeto.>
Regra 3	<verbo>+<artigo>+<objeto>	<ul style="list-style-type: none"> <li>- <i>choose</i>&lt;verbo&gt; <i>a</i>&lt;artigo&gt; <i>document</i>&lt;objeto&gt;.</li> <li>- <i>it do</i>&lt;verbo&gt; <i>a</i>&lt;artigo&gt; <i>order</i>&lt;objeto&gt;.</li> </ul>
Regra 4	<sujeito>+<verbo>+ <objeto>+ <conjunção>+ <verbo> +<objeto>	<i>A client</i> <sujeito> <i>calls</i> <verbo> <i>the help desk</i> <objeto> <i>and</i> <conjunção> <i>makes</i> <verbo> <i>a request</i> <objeto>.
Regra 5	<objeto>+<sujeito>+ <verbo>	<i>The severity</i> <objeto> <i>of the claimant</i> <sujeito> <i>is evaluated</i> <verbo>.
Regra 6	<verbo>+<conjunção> +<verbo> +<objeto>	<i>The first activity is to check</i> <verbo> <i>and</i> <conjunção> <i>repair</i> <verbo> <i>the hardware</i> <objeto>.
Regra 7	<sujeito>+<sujeito>+ <verbo>+<sujeito>	<i>Department of engineering</i> <sujeito> <i>and sell</i> <sujeito> <i>are informed</i> <verbo> <i>by Manager</i> <sujeito>.
Regra 8	<verbo>+ <sujeito>+ <objeto>	<ul style="list-style-type: none"> <li>- ... <i>eat</i>&lt;verbo&gt;<i>your</i>&lt;sujeito&gt;<i>pizza</i> &lt;objeto&gt;</li> <li>- ... <i>wait</i>&lt;verbo&gt;<i>until your</i>&lt;sujeito&gt;<i>pizza</i>&lt;objeto&gt; <i>comes</i>.</li> </ul>
Regra 9	<objeto>+<verbo>+<objeto indireto>	<i>...the forms</i> <objeto> <i>are</i> <to-be> <i>sent</i> <verbo> <i>to the claimant</i> <objeto indireto> ...

Fonte: Os autores

sivo (XOR) são menos propensos a problemas estruturais<sup>2</sup>(MENDLING; NEUMANN; AALST, 2007). O conjunto completo inclui 32 regras de mapeamento. Sendo que, 9 refere-se a atividades e eventos, 4 a desvio paralelo (AND), 7 a desvio exclusivo (XOR) e 3 a raias.

Nesta dissertação, os rótulos (*labels*) de elementos de processo são importantes fatores para identificar elementos de processo. Por exemplo, os rótulos das tarefas, partem do pressuposto que na sentença, existirá no mínimo um *Verbo* e *Objeto*. Conforme Mendling, Reijers and Aalst (2010), Mendling, Reijers and Recker (2010) e Gassen et al. (2014), os *labels* das atividades são representados pelo *Verbo* e em seguida o *Objeto*. Por exemplo, na sentença, "*O cliente envia o formulário para o analista*", o *label* da atividade

<sup>2</sup>problemas estruturais são problemas que podem ocorrer durante a execução de um processo de negócio.

se tornaria: *Enviar Formulário* e o participante: *Cliente*. Portanto, o uso dos *labels* é uma boa prática de modelagem, pois impacta diretamente na clareza e compreensão do processo (MENDLING; REIJERS; RECKER, 2010).

Tabela 3.2: Regras para identificação de eventos.

<b>Eventos</b>		
<b>Regras</b>	<b>Descrição</b>	<b>Exemplo de Sentença</b>
Regra 1	<sujeito> + <verbo> + <objeto>	After the agente <sujeito> has confirmed <verbo> the claim<objeto> to the clerk.
Regra 2	<sujeito>+<verbo> + <agente> + <objeto>	The SCT physical <sujeito> file was stored <verbo no passado> by <agente> the Back office <objeto>. (voz passiva)
Regra 3	<sujeito> + <verbo> + <objeto> + <conjunção> + <objeto>	After the customer <sujeito> has chosen <verbo> several products and <conjunção> services <objeto>.
Regra 4	<objeto>+<verbo (passado)> + <sujeito>	...a message <objeto> was generated <verbo> to the customer<sujeito>.
Regra 5	<objeto>+<verbo>+ <adjetivo>	Once the forms <objeto> are returned, they were checked<verbo> for completeness<adjetivo>.
Regra 6	evento temporário - <tempo> + <preposição> + <tempo>	...it may take thirty minutes<tempo> to <preposição> an hour <tempo>
Regra 7	<palavra-chave>+ <tempo>	If we are not contacted within <palavra-chave> 2 weeks <tempo> ...
Regra 8	<sujeito>+ <objeto> + <verbo <passado>>	The claimant <sujeito> was informed that claim <objeto> must be rejected <verbo>.
Regra 9	<objeto> + <verbo (presente perfeito)>+<sujeito>	...Urgent document <objeto> has been received <verbo> by the Manager <sujeito>

Fonte: Os autores

Neste contexto, para identificar elementos de processo em sentenças, conforme Mendling, Reijers and Aalst (2010), Leopold (2013), Mendling, Reijers and Recker (2010), as sentenças que contém tempos verbais no presente e/ou futuro representam atividades. Por outro lado, as sentenças que retratam tempos verbais no passado e/ou presente perfeito da língua inglesa são classificados como eventos. Uma outra diferença é a elaboração de *labels* para eventos. Em uma sentença, os *labels* de eventos são representados através do *Objeto* e em seguida o *Verbo* no particípio, tal como: "*Invoice Created*" (MENDLING, 2013; LEOPOLD, 2013).

Por exemplo, a regra 5 da Tabela 3.1, mostra um exemplo de identificação de atividades em sentenças. A regra contempla a sequência de um *Objeto*, seguido de um *Sujeito* e posteriormente um *Verbo*. A modelagem desta sentença se tornaria uma tarefa com a união do *Verbo* e *Objeto*. Portanto, tal sentença seria uma boa candidata para a

Tabela 3.3: Regras para identificação de desvio exclusivo (XOR).

<b>Desvio Exclusivo (XOR)</b>		
<b>Regras</b>	<b>Descrição</b>	<b>Exemplo de Sentença</b>
Regra 1	<verbo>+ <palavra-chave>+ <sujeito>+ <objeto>	It first <i>checked</i> <verbo> whether <palavra-chave> the claimant <sujeito> is insured <objeto> by the organization.
Regra 2	<palavra-chave> + <condição> + < tarefa/evento>	If <palavra-chave> the claimant requires two or more forms <condição>, the Department of customer select the forms <tarefa>.
Regra 3	<tarefa/evento> + <palavra-chave> + <condição>	After that they enter into a firm commitment to buy the stock and then offer it to the public <tarefa>, when <palavra-chave> they haven't still found any reason not to do it <condição>.
Regra 4	<tarefa>+ <palavra-chave>+ <condição> + <palavra-chave alternativa> + <tarefa>	The clerk checks <tarefa> whether <palavra-chave> the beneficiary's policy was valid at the time of the accident <condição>. If not <palavra-chave alternativa>, it send to Department of the intelligence <tarefa>.
Regra 5	<palavra-chave> + <condição> + <conjunção>+ <condição> + <tarefa/evento>	If <palavra-chave> Department of sell send the document <condição> and <conjunção> notify the department of engineering <condição>, the document is processed <tarefa/evento>.
Regra 6	<palavra-chave>+ <tarefa/evento>+ <tarefa/evento>	If <palavra-chave> the agent has completed his additional support <tarefa/evento>, the claim is closed <tarefa/evento>.
Regra 7	<palavra-chave alternativa>+ <tarefa/evento>	If not <palavra-chave alternativa>, it send to Department of the intelligence <tarefa/evento>.

Fonte: Os autores

modelagem.

As regras 5, 7 e 9 da Tabela 3.1, são exceções no contexto deste elemento de processo. Por exemplo, a regra 9, considera uma sequência *Objeto, Verbo e Objeto Indireto*. Tal regra apresenta o Verbo no passado. Esta regra apresenta um sequência de *Verbo TO BE* (no presente) e *Verbo* (verbo principal no passado). Isto é, se a sentença conter um *Verbo TO BE* no presente, dado que o *Verbo TO BE* indica uma ação pontual na sentença, em seguida o verbo principal no passado, portanto, essa sequência representa uma atividade no contexto desta dissertação. O restante das regras para este elemento de processo (atividades), segue o mesmo padrão de identificação (ver Tabela 3.1).

A regra 2 da Tabela 3.2, ilustra um exemplo de identificação de eventos em sentenças. A regra considera a sequência de um *Sujeito*, seguido de um *Verbo*, posteriormente um *Agente*<sup>3</sup> e finalmente por um *Objeto*. Assim, tal sentença também seria uma candidata

<sup>3</sup>Um agente é o complemento de um verbo passivo que é o sujeito superficial de sua forma ativa. Nesta dissertação, a preposição "by" é incluída como parte do agente

à modelagem de processos. O restante das regras para este elemento de processo, segue o mesmo padrão de identificação (ver Tabela 3.2).

Tabela 3.4: Regras para identificação de desvio paralelo (AND).

<b>Desvio Paralelo(AND)</b>		
<b>Regras</b>	<b>Descrição</b>	<b>Exemplo de Sentença</b>
Regra 1	<b>&lt; tarefa/evento &gt;+ &lt; palavra-chave &gt;+ &lt; tarefa/ evento &gt;</b>	<i>Forward the document &lt; tarefa/evento &gt;, In parallel with this &lt; palavra-chave &gt;, the RCC shall also notify the Executive Board &lt; tarefa/evento &gt;.</i>
Regra 2	<b>&lt; palavra-chave &gt;+ &lt; tarefa/evento &gt; + &lt; conjunção &gt;+ &lt; tarefa/evento &gt; + &lt; tarefa/evento &gt;</b>	<i>In parallel with this &lt; palavra-chave &gt; The supervisor dispatch the report &lt; tarefa/evento &gt; and &lt; conjunção &gt; inform the manager &lt; tarefa/evento &gt;, then the report is sent &lt; tarefa/evento &gt;.</i>
Regra 3	<b>&lt; palavra-chave &gt; + &lt; tarefa/evento &gt;</b>	<i>In the meantime &lt; palavra-chave &gt;, the engineering department prepares everything for the assembling of the ordered bicycle &lt; tarefa/evento &gt;.</i>
Regra 4	<b>&lt; palavra-chave &gt; + &lt; tarefa/evento &gt; + &lt; palavra-chave alternativa &gt; + &lt; tarefa/evento &gt;</b>	<i>The ongoing repair consists of two activities, which are executed, in an parallel order &lt; palavra-chave &gt;. The first activity is to check and repair the hardware &lt; tarefa/evento &gt;, whereas &lt; palavra-chave alternativa &gt; the second activity checks and configures the software. &lt; tarefa/evento &gt;.</i>

Fonte: Os autores

As regras de mapeamento da Tabela 3.3 e Tabela 3.4, apresentam palavras que denotam fluxo de controle (XOR e AND). Tais palavras são denominadas, nesta dissertação, como *palavras-chave* e *palavras-chave alternativas*. Essas palavras refere-se a uma condição (cláusula condicional) na sentença. Nesta dissertação, tais condições são definidas como tarefas e eventos na sentença. As palavras são divididas em 2 grupos:

Tabela 3.5: Regras para identificação de raias.

<b>Raias</b>		
<b>Regras</b>	<b>Descrição</b>	<b>Exemplo de Sentença</b>
Regra 1	<b>O sujeito da sentença</b>	<i>&lt; sujeito &gt;realiza uma &lt; tarefa/evento &gt;</i>
Regra 2	<b>&lt; tarefa &gt; + &lt; objeto indireto &gt;</b>	<i>She then submits an order &lt; tarefa &gt; to the customer &lt; objeto indireto &gt;.</i>
Regra 3	<b>&lt; evento &gt; + &lt; objeto indireto &gt;</b>	<i>The Manager forwarded the form &lt; evento &gt; to Official &lt; objeto indireto &gt;.</i>

Fonte: Os autores

- Desvio paralelo (AND): palavras que denotam paralelismo na sentença.

- palavras-chave: *while, meanwhile, in parallel, concurrently, meantime, in the meantime, in parallel with this, in addition to, simultaneously, at the same time e whereas.*

- Desvio Exclusivo (XOR): palavras que referem-se à exclusão em uma sentença.
  - *if, whether, if not, or, in case [of], otherwise, either, only, till, until, unless, when, only if.*
  - palavras-chave alternativas: *but, then, else, or, unless, without, either, otherwise, other, if its is not.*

A fim de identificar os sinônimos das *palavras-chave* e *palavras-chave alternativas*, nesta dissertação, foi atribuído um dicionário da linguagem de programação *Python*, para obter sinônimos de palavras. Tal dicionário é denominado como *PyDictionary*<sup>4</sup>.

A regra 2 da Tabela 3.3, ilustra um exemplo para identificar desvio exclusivo (XOR) em sentenças. A regra considera a sequência de uma *Palavra-Chave*, seguida por uma *Condição*, posteriormente uma *Tarefa* ou *Evento*. Por esta razão, tal sentença é uma candidata à modelagem de processos. O restante das regras para este elemento de processo (XOR), seguem mesmo padrão de identificação (ver Tabela 3.3).

A regra 2 da Tabela 3.4, descreve um exemplo de identificação de desvio paralelo (AND) em sentenças. A regra considera uma sequência de *Palavra-Chave*, seguido de uma *Tarefa*, então por uma *Conjunção*<sup>5</sup>, posteriormente por uma *Tarefa* e finalmente uma outra *Tarefa*. Portanto, tal sentença também seria uma candidata à modelagem de processos. As regras 2, 3 e 4 da Tabela 3.4, apresentam, em ambas, uma sequência de *Palavra-Chave* e *tarefa/evento*. A importância desta sequência, se dá ao fato de terem diferentes padrões de classes gramaticas após tal sequência (exceto regra 3). Por exemplo, a regra 4, considera a sequência de *Palavra-Chave*, seguido de uma *tarefa/evento*, posteriormente por uma *Palavra-Chave Alternativa* e finalmente por uma *tarefa/evento*. Portanto, a disparidade das regras 2 e 4 é determinada pela sequência após a *Palavra-Chave* e *tarefa/evento*. Neste caso, tal sequência é uma particularidade importante para identificar desvio paralelo (AND) em sentenças.

Finalmente, as regras de mapeamento da Tabela 3.5, mostram um exemplo de identificação de raias (*swimlanes*) em sentenças. Conforme as considerações realizadas nos textos descritos nesta seção, foi observado que o *Sujeito* ou *Objeto Indireto*, geralmente serão as raias da sentença. Desta forma, o *Sujeito* pode ser um ser humano, equipamento eletrônico ou algo que pratica uma ação na sentença (DUMAS et al., 2013; WESKE, 2012; FRIEDRICH; MENDLING; PUHLMANN, 2011).

<sup>4</sup><http://github.com/geekpradd/PyDictionary> ; último acesso em 11-01-2017

<sup>5</sup>conjunção: é a palavra invariável que liga duas sentenças ou dois termos semelhantes de uma mesma sentença. Na língua inglesa, geralmente é dado pelas palavras "and" e "or".

Em suma, nesta etapa são aplicadas as regras de mapeamento nas sentenças a fim de identificar elementos do processo de negócio.

A próxima seção descreve detalhes da identificação de elementos de processos de negócio utilizando algoritmos.

### 3.3.2 Aplicação das Regras de Mapeamento para Identificação de Elementos de Processo de Negócio em Texto em Linguagem Natural

Esta seção descreve a análise lógica de texto utilizando algoritmos, dado que, tais algoritmos apresentam lógicas para realizar as identificações dos elementos de processo em texto. Os algoritmos se originam das regras de mapeamento propostas nesta dissertação e foram utilizados para a implementação do protótipo desenvolvido no contexto desta abordagem. Tal protótipo é apresentado na seção 4.2 desta dissertação.

O processo de identificação através de algoritmos consiste em diversas etapas que são ilustradas como um modelo de processo BPMN, conforme a Figura 3.5. O principal requisito para executar as regras de mapeamento, é o vetor de objetos (*Doc*) da etapa de análise sintática do texto, como fonte de informação para as próximas tarefas da representação. As tarefas são descritas, como segue:

- **Verificar Palavras-Chave:** nesta tarefa é realizada a verificação de cada sentença do texto, buscando palavras-chave de desvio paralelo (AND) e exclusivo (XOR). Tal tarefa está descrita nos algoritmos 2 e 3 respectivamente, uma vez que ambos algoritmos recebem as mesmas variáveis de entrada (*indice* e *Doc*).

A variável *Doc* (linha 1) tem como origem, a análise sintática da sentença, descrita no algoritmo 1 (linha 6). A primeira condição do algoritmo 2 (linha 3) refere-se a verificação de dependência da palavra analisada de acordo com o contexto da sentença. Isto é, verifica-se se palavra analisada pertence a categoria léxica denominada como "*mark*". Tal categoria tem como função identificar um complemento causal (ver Tabela 2.1). Além de pertencer a esta categoria, tal palavra (variável *content* na linha 2) pode ser representada por *palavras-chave* classificadas no conjunto de palavras-chave do elemento de processo: desvio exclusivo (XOR) (ver seção 3.3). Esta lógica é seguida nas linhas 5, 7 e 9. Caso não ocorra a identificação da palavra analisada nas 4 primeiras condições (linhas 3, 5, 7 e 9), a palavra é enviada para os algoritmos 4 e 6. Por outro lado, caso não ocorram as identificações

---

**Algoritmo 2** Algoritmo de Verificação de Palavras-Chave do Desvio Exclusivo (XOR)
 

---

**Entrada:** índice e vetor com a análise sintática da sentença

**Saída:** Retorna variável booleana (se existe ou não a palavra na sentença)

```

1: procedure ISSIGNALWORDSEXOR(indice, Doc)
2:   content ← Doc[indice]
3:   if Doc[indice].dep ∈ 'mark' and ("if" = content or "whether" = content or "otherwise" =
   content or "either" = content or "only" = content or "till" = content or "until" =
   content or "when" = content) then
4:     return True
5:   else if ("if not" ∈ Sentence) then
6:     return True
7:   else if ("in case" ∈ Sentence) then
8:     return True
9:   else if ("in case of" ∈ Sentence) then
10:    return True
11:  isSynonyms, Content ← isSynonymsXOR(content)
12:  if isSynonyms = True then
13:    return True
14:  else
15:    return False

```

Fonte: Os autores

---

das palavras, incluindo as condições mencionadas e no algoritmo 4, a sentença não representa o desvio exclusivo (XOR). O algoritmo 3, segue a mesma lógica do algoritmo 2, porém este algoritmo apresenta um conjunto de palavras-chave divergente, uma vez que o algoritmo 3 representa o desvio paralelo (AND).

- **Verificar Palavras Sinônimas:** a fim de obter um maior número de palavras que representam paralelismo ou exclusividade, foi implementada uma tarefa, para que este objetivo fosse realizado. Os algoritmos 4 e 6, descrevem uma lógica para identificação de sinônimos baseada nos grupos de palavras-chave dos elementos de processo: desvio exclusivo (XOR) e desvio paralelo (AND), respectivamente. A entrada de dados para estes algoritmos, depende dos resultados dos algoritmos 4 e 6 (linha 12), uma vez que estes resultados foram descritos na tarefa "*Verificar Palavras-Chave*". O dado de entrada dos algoritmos 4 e 6, é representado por uma sentença a ser analisada verificando sinônimos de palavras-chave. Nesta dissertação, para verificar sinônimos de palavras, foi utilizado o dicionário da linguagem de programação *Python*, denominado como *Pydictionary*<sup>6</sup>. O dicionário considera diversas características, tais como: identificar significados, traduções, sinônimos e antônimos de palavras. Além disso, utiliza como fontes de informações *WordNet*<sup>7</sup> Miller (1995) para obter significados, *Google Translate*<sup>8</sup> para traduções e *Thesau-*

---

<sup>6</sup><https://pypi.python.org/pypi/PyDictionary/1.3.4>; último acesso em 11-01-2017

<sup>7</sup><http://wordnet.princeton.edu/>; último acesso em 11-01-2017

<sup>8</sup><https://translate.google.com.br/>; último acesso em 11-01-2017

rus Chapman (1977) para obter sinônimos e antônimos. O algoritmo 4, apresenta as palavras-chave que representam exclusividade em uma sentença (linha 4).

---

### Algoritmo 3 Algoritmo de Verificação de Palavras-Chave do Desvio Paralelo (AND)

---

**Entrada:** índice e o vetor com a análise sintática da sentença

**Saída:** Retorna variável booleana (se existe ou não a palavra na sentença)

```

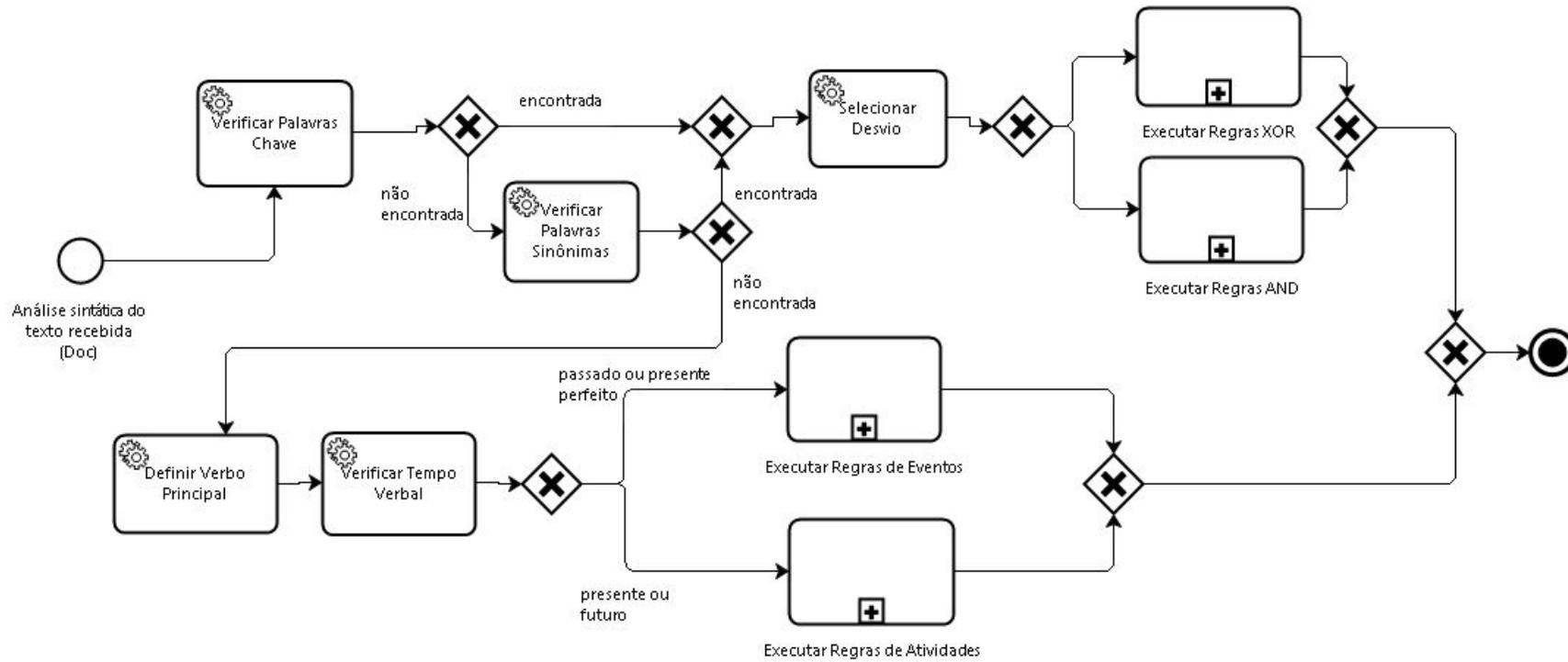
1: procedure ISSIGNALWORDSAND(indice, Doc)
2:   content ← Doc[indice]
3:   if “while” = content or “meanwhile” = content or “concurrently” =
   content or “meantime” = content or “simultaneously” = content or “whereas” = content
   then
4:     return True
5:   else if (“in parallel with this” ∈ Sentence) then
6:     return True
7:   else if (“in parallel” ∈ Sentence) then
8:     return True
9:   else if (“in the meantime” ∈ Sentence) then
10:    return True
11:  else if (“in addition to” ∈ Sentence) then
12:    return True
13:  else if (“at the same time” ∈ Sentence) then
14:    return True
15:  isSynonyms, Content ← isSynonymsAND(content)
16:  if isSynonyms = True then
17:    return True
18:  else
19:    return False

```

Fonte: Os autores

---

Figura 3.5: Representação estrutural da etapa de análise lógica de texto.



Fonte: Os autores

Posteriormente, na linha 5 apresentada no algoritmo 4, é utilizada uma estrutura de repetição (*For*) para percorrer cada palavra na sentença e verificar se tal palavra é sinônima do conjunto de palavras-chave do elemento de processo (linha 7). Caso houver alguma palavra sinônima, o algoritmo retorna uma variável booleana, além da palavra que foi identificada (linha 10). Caso não houver palavras-chave sinônimas, o algoritmo é finalizado, e então dirigido para a próxima tarefa ("*Definir Verbo Principal*") da Figura 3.5. O algoritmo 6, possui a mesma lógica apresentada no algoritmo 4. Porém, o conjunto de palavras-chave denotam paralelismos e representam o elemento de processo desvio paralelo (AND).

- **Executar Regras de Desvio Exclusivo (XOR):** Após serem identificadas as *palavras-chave* no texto, uma vez que tais palavras possibilitam a identificação de paralelismo e exclusividade, é necessário selecionar as regras de mapeamento de desvios (AND e XOR) que serão executadas. Nesta dissertação, será descrito o algoritmo da regra 3 do desvio exclusivo (XOR) presente na Tabela 3.3. O conjunto completo de algoritmos (atividades, eventos, desvio paralelo e desvio exclusivo) estão disponíveis em: <[www.github.com/renatoborgesss/python\\_parser](http://www.github.com/renatoborgesss/python_parser)>. O algoritmo 5, tem como objetivo identificar a existência de uma sequência formada por uma *tarefa* ou um *evento*, *palavra-chave* e uma *condição* em uma sentença. Para isto, é necessário identificar o índice e a palavra na sentença (linha 4), dado que, conforme a regra 3, tal palavra se apresenta entre uma ação e uma condição.

---

#### **Algoritmo 4** Algoritmo de Verificação de Palavras Sinônimas do Desvio Exclusivo (XOR)

---

**Entrada:** Sentença para ser analisada

**Saída:** Retorna variável booleana (se existe ou não uma palavra sinônimo na sentença) e a palavra identificada

```

1: procedure ISSYNONYMSXOR(String)
2:   isSynonyms ← False
3:   Content ← None
4:   ArrayPalavrasChave ← "if", "if not", "whether", "simultaneously", "whereas", "or", "in case of", "otherwise", "either", "only", "till", "until", "unless"
5:   for  $\forall$  palavra  $\in$  ArrayPalavrasChave do
6:     StringResult ← dictionary.synonym(palavra)
7:     if (String  $\in$  StringResult) then
8:       Content ← String
9:       isSynonyms ← True
10:    return isSynonyms, Content
11:  return isSynonyms, Content

```

---

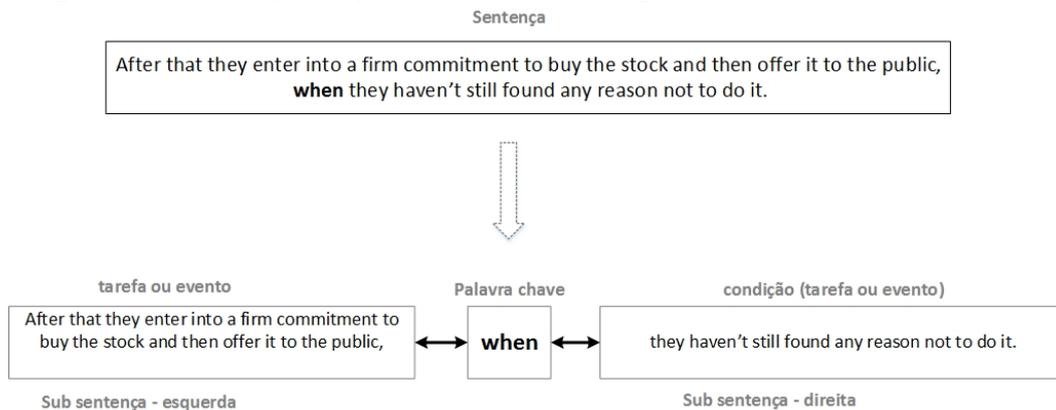
Fonte: Os autores

---

Após, é necessário separar a palavra e armazenar o restante da sentença em duas *strings* (linhas 7 e 8), conforme ilustrado na Figura 3.6. Nas linhas 11 e 13, respec-

tivamente, foi verificado se existem uma ação (tarefa ou evento) e uma condição, antes e depois da palavra-chave identificada. Desta forma, se a sentença representar a sequência definida, tal sentença é identificada como um desvio exclusivo (XOR) do modelo de processo. Se a sentença não for identificada, conforme a regra 3 de desvio exclusivo, tal sentença prossegue para regra 4, e assim sucessivamente.

Figura 3.6: Separação de *palavras-chave* das regras de desvio exclusivo (XOR)



Fonte: Os autores

---

### Algoritmo 5 Algoritmo de Identificação de desvio exclusivo (Regra 3)

---

**Entrada:** Vetor com a análise sintática da sentença

**Saída:** Retorna um vetor uma variável do tipo booleana

```

1: procedure GETRULE2XOR(Doc)
2:   Left ← False
3:   Right ← False
4:   IndiceWord, Content, Separate ← GetWordIndiceContentDivide(Doc)
5:   if (Separate = True) then
6:     array ← sent.split(Content)
7:     StringArrayLeft ← array[0]
8:     StringArrayRight ← array[1]
9:     DocLeft ← nlp(StringArrayLeft)
10:    DocRight ← nlp(StringArrayRight)
11:    if (ExecuteRulesActivities(DocLeft) = True or ExecuteRulesEvents(DocLeft) =
12:    True) then
13:      Left ← True
14:      if (ExecuteRulesActivities(DocRight) = True or ExecuteRulesEvents(DocRight) =
15:      True) then
16:        Right ← True
17:        if (Left = True and Right = True) then
18:          return True
19:        else
20:          return False
21:      else
22:        return False

```

Fonte: Os autores

---

- **Executar Regras de Desvio Paralelo (AND):** Assim como a tarefa "Executar Regras de Desvio Exclusivo (XOR)", a tarefa "Executar Regras de Desvio Paralelo

(AND)", é necessária para identificar *palavras-chave* que denotam paralelismo em uma sentença, além de executar as regras definidas para o elemento de processo desvio paralelo (AND). Nesta dissertação, será descrito o algoritmo da regra 1 do desvio paralelo presente na Tabela 3.4. Tal regra descreve uma sequência de *tarefa* ou *evento*, *palavra-chave* seguido de uma *tarefa* ou *evento* novamente. Para isto, é necessário identificar o índice e a palavra na sentença (linha 4), uma vez que tal palavra é posicionada entre duas ações.

---

**Algoritmo 6** Algoritmo de Verificação de Palavras Sinônimas de Desvio Paralelo (AND)

---

**Entrada:** Sentença para ser analisada

**Saída:** Retorna variável booleana (se existe ou não uma palavra sinônimo na sentença) e a palavra identificada

```

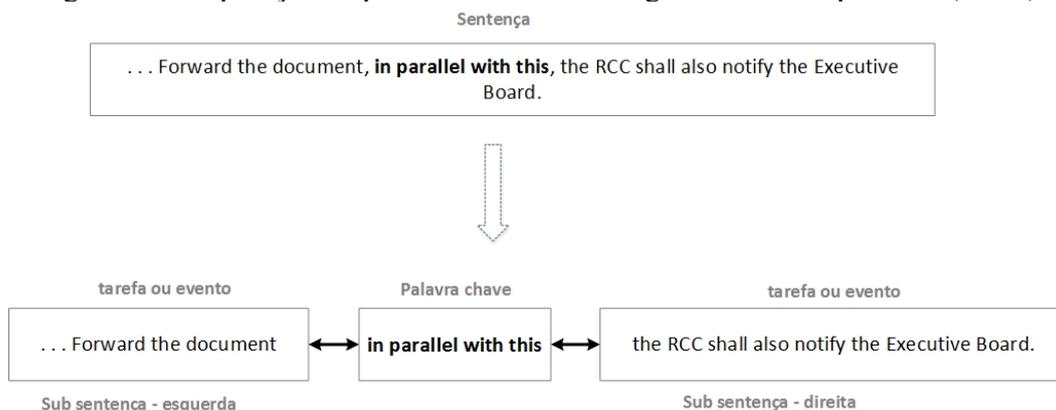
1: procedure ISSYNONYMSAND(String)
2:   isSynonyms ← False
3:   Content ← None
4:   ArrayPalavrasChave ← "while", "meanwhile", "meantime", "simultaneously", "whereas", "concurrently", "until", "unless", "when"
5:   for  $\forall$  palavra  $\in$  ArrayPalavrasChave do
6:     StringResult ← dictionary.synonym(palavra)
7:     if (String  $\in$  StringResult) then
8:       Content ← String
9:       isSynonyms ← True
10:    return isSynonyms, Content
11:  return isSynonyms, Content

```

Fonte: Os autores

---

Figura 3.7: Separação de *palavras-chave* das regras de desvio paralelo (AND)



Fonte: Os autores

Após, é considerado a necessidade de separar a palavra e armazenar o restante da sentença em duas *strings* (linhas 7 e 8). Então, a classe *NLP* é executada novamente, com o objetivo de gerar uma nova árvore sintática da sentença de origem. Esta árvore irá definir se a sentença é uma ação do lado esquerdo e direito da *palavra-chave* (linhas 11 e 13). Para ilustrar a execução, considere a Figura 3.7. Se a sentença re-

presentar a sequência definida da regra 1, neste caso, a sentença é identificada como um desvio paralelo (AND) do modelo de processo. Se a sentença não for identificada, conforme a regra 1 de desvio paralelo, tal sentença prossegue para regra 2, e assim sucessivamente.

- **Definir Verbo Principal:** Se a sentença não for identificada como desvio exclusivo (XOR) ou desvio paralelo (AND), é possível que esta possa ser representada por uma tarefa ou evento no modelo de processo (ver Figura 3.5). Para isso, tem-se como objetivo, identificar o verbo principal da sentença e posteriormente definir o seu tempo verbal. Sendo assim, foi desenvolvido uma tarefa para realizar tal função.

---

#### Algoritmo 7 Algoritmo de Definição de Verbo Principal da Sentença

---

**Entrada:** Vetor com a análise sintática da sentença

**Saída:** Retorna o verbo principal da sentença

```

1: procedure DEFINITIONMAINVERBACTIVITYOREVENT(Doc)
2:   for  $\forall$  word < Doc do
3:     if word.dep  $\in$  ('ROOT') then
4:       MainVerb  $\leftarrow$  word
5:     return MainVerb

```

Fonte: Os autores

---

O algoritmo 7, representa a tarefa "*Definir o Verbo Principal*". É importante identificar o verbo principal da sentença, uma vez que todas as palavras na sentença são dependentes dele. Isto é, o verbo principal é a raiz ("*ROOT*") da árvore sintática da sentença. O dado de entrada do algoritmo 7, é representado pela variável *Doc* (definida no algoritmo 1 linha 6). A estrutura de repetição (*For*) (linha 2) percorre cada palavra da sentença, a fim de identificar se a relação de dependência da palavra é representada por "*ROOT*" (linha 3). Se a palavra analisada, for a raiz da árvore sintática determinada pelo algoritmo 1, esta é o verbo principal da sentença (linha 4).

- **Verificar Tempo Verbal:** Após ser identificado o verbo principal da sentença na tarefa anterior, é necessário verificar o seu tempo verbal, isto é, identificar qual tempo verbal o verbo se apresenta, para que a sentença seja classificada como uma tarefa ou um evento. A fim que tal sentença seja representada como tarefa, o tempo verbal do verbo principal, deverá ser classificado como presente ou futuro. Por outro lado, se o tempo verbal do verbo principal for identificado como passado ou presente perfeito da língua inglesa, este será relacionado a um evento (ver Figura 3.5) (MENDLING; REIJERS; AALST, 2010; LEOPOLD, 2013; MENDLING; REIJERS; RECKER, 2010). O algoritmo 9 apresenta tal tarefa. Os dados de entrada do

---

**Algoritmo 8** Algoritmo de Identificação de desvio paralelo (Regra 1)
 

---

**Entrada:** Vetor com a análise sintática da sentença

**Saída:** Retorna um vetor uma variável do tipo booleana

```

1: procedure GETRULE1AND(Doc)
2:   Left ← False
3:   Right ← False
4:   IndiceWord, Content, Separate ← GetWordIndiceContentDivideAND(Doc)
5:   if (Separate = True) then
6:     array ← sent.split(Content)
7:     StringArrayLeft ← array[0]
8:     StringArrayRight ← array[1]
9:     DocLeft ← nlp(StringArrayLeft)
10:    DocRight ← nlp(StringArrayRight)
11:    if (ExecuteRulesActivities(DocLeft) = True or ExecuteRulesEvents(DocLeft) =
12:    True) then
13:      Left ← True
14:    if (ExecuteRulesActivities(DocRight) = True or ExecuteRulesEvents(DocRight) =
15:    True) then
16:      Right ← True
17:    if (Left = True and Right = True) then
18:      return True
19:    else
20:      return False

```

Fonte: Os autores

---

algoritmo 9, são o vetor de objetos *Doc* e o verbo principal da sentença (definida no algoritmo 7). A fim de identificar o tempo verbal da sentença, é necessário utilizar o *part of speech tagging*. Para isso, tem-se como "tag":

- **VB** – verbo em sua forma base (infinitivo)
- **VBD** – verbo no passado
- **VBG** – verbo no gerúndio
- **VBN** – verbo no passado do particípio
- **VBZ** – terceira pessoa do singular, tais como: *He, she, it* (verbo no presente)
- **VBP** – primeira e segunda pessoa do singular, tais como: *I e You* (verbo no presente)

A linha 2 descrita no algoritmo 9, apresenta uma condição para classificar o verbo principal da sentença como uma tarefa. Isto é, é verificado se a "tag" do verbo principal (variável: *MainVerb*), representa *VBP* ou *VBZ* ou *VB* ou *VBG*. Portanto, neste caso, a sentença é classificada como uma tarefa (linha 3). Por outro lado, a linha 4, apresenta outra condição para determinar o verbo principal como um evento, ou seja, verificando se a "tag" do verbo principal representa *VBD* ou *VBN*.

---

**Algoritmo 9** Algoritmo de Verificação do Tempo Verbal
 

---

**Entrada:** Vetor com a análise sintática e verbo principal da sentença

**Saída:** Retorna o tipo de ação (atividade ou evento)

```

1: procedure DEFINITIONVERBTENSEMAINVERB(MainVerb, Doc)
2:   if (MainVerb.tag = "VBP" or MainVerb.tag = "VBZ" or MainVerb.tag =
   "VB" or MainVerb.tag = "VBG") then
3:     return TypeSentenceActivity
4:   else if (MainVerb.tag = "VBD" or MainVerb.tag = "VBN") then
5:     return TypeSentenceEvent

```

Fonte: Os autores

---

O algoritmo 9 tem como saída, a classificação da sentença como uma tarefa ou um evento (linhas 3 e 5, respectivamente).

- **Executar Regras de Atividades:** Após determinar o tempo verbal do verbo principal da sentença, é necessário executar as regras de atividades ou eventos para que a sentença seja identificada. Para execução destas regras, é necessário definir o *Sujeito* e o *Objeto* (direto ou indireto) da sentença, uma vez que o sujeito representa a raia do modelo de processo e os objetos equivalem ao objeto do modelo.

---

**Algoritmo 10** Algoritmo de Identificação do Sujeito
 

---

**Entrada:** Vetor com a análise sintática da sentença

**Saída:** Retorna o índice e o conteúdo do sujeito da sentença.

```

1: procedure GETCONTENTINDICESUBJECT(Doc)
2:   for  $\forall$  indice < (Doc) do
3:     if Doc[indice].dep  $\in$  ('nsubj' or 'csubj' or 'nsubjpass' or 'xsubj' or 'agent' or 'csubjpass')
   then
4:       subject  $\leftarrow$  Doc[indice].subtree
5:       indiceSubject  $\leftarrow$  indice
6:   return subject, indiceSubject

```

Fonte: Os autores

---

Os algoritmos 10 e 11 respectivamente, apresentam esta definição (identificar o sujeito e objeto da sentença). A primeira condição do algoritmo 10 (linha 3), determina qual tipo de sujeito a sentença é constituída, baseado nas categorias gramaticais propostas por Choi and Palmer (2012). Após a definição do tipo de sujeito, o algoritmo armazena o sujeito e o índice (posição da palavra na sentença) do mesmo, uma vez que é necessário para executar as regras de mapeamento da próxima tarefa. O algoritmo 11, segue a mesma lógica do algoritmo 10. Porém, a condição da linha 3, tem como função determinar qual tipo de objeto tal sentença estabelece. Desta forma, as saídas dos algoritmos 10 e 11 respectivamente, são representados pelo sujeito, objeto e seus índices. Após isso, são executadas as regras de atividades. Nesta dissertação, será apresentado o algoritmo da regra 1 de atividades. O algoritmo 12, foi desenvolvido baseado na regra de mapeamento 1 da Tabela 3.1. Este

---

**Algoritmo 11** Algoritmo de Identificação do Objeto
 

---

**Entrada:** Vetor com a análise sintática da sentença

**Saída:** Retorna o índice o conteúdo do objeto.

```

1: procedure GETCONTENTINDICEOBJECT(Doc)
2:   for  $\forall$  indice < (Doc) do
3:     if Doc[indice].dep  $\in$  ('dobj' or 'iobj' or 'advcl' or 'advmod' or 'pobj' or 'oprd') then
4:       Object  $\leftarrow$  Doc[indice].subtree
5:       indiceObjeto  $\leftarrow$  indice
6:   return Object, indiceObjeto

```

Fonte: Os autores

---

algoritmo tem como objetivo identificar a sequência *Sujeito*, *Verbo* e *Objeto* em uma sentença. Para a identificação, é necessário executar os algoritmos 7, 10 e 11 (linhas 2, 3 e 4), uma vez que exista a necessidade da palavra (por exemplo, palavra *client* – sujeito) e seu índice na sentença. Na linha 6 e 9 do algoritmo 12, é verificado se há ocorrências de conjunções e auxiliares de futuro (por exemplo, "*will*" e "*going to*") na sentença, dado que há outras regras para tratar destas ocorrências (por exemplo, regra 2 e 4 da Tabela 3.1). Posteriormente, é verificado se a sentença tem *Sujeito*, *Verbo* e *Objeto* (linha 11). Se sim, é verificado se a sentença representa a sequência *Sujeito*, *Verbo* e *Objeto* (linha 14), conforme a definição da regra 1 de atividades da Tabela 3.1. Por fim, caso a sentença representar a sequência definida, neste caso, tal sentença é identificada como uma atividade do modelo de processo. Caso a sentença não for identificada, conforme a regra 1 de atividades, tal sentença prossegue para regra 2, e assim sucessivamente. Caso não ocorra a identificação da sentença pelas regras de atividades, eventos, desvio exclusivo ou desvio paralelo, o analista de processos detém a tomada de decisão em modelar a sentença conforme seu conhecimento.

- **Executar Regras de Eventos:** Assim como a tarefa "*Executar Regras de Atividades*", a tarefa "*Executar Regras de Eventos*", é necessária também para definir o *Sujeito* e *Objeto* (direto ou indireto) da sentença, descritos nos algoritmos 10 e 11. Desta forma, são executadas as regras de eventos. Nesta dissertação, será apresentado o algoritmo da regra 2 de eventos presente na Tabela 3.2. O algoritmo 13, tem como objetivo identificar a sequência de *Sujeito*, *Verbo*, *Agente* e *Objeto* em uma sentença. Para a identificação, se faz necessário executar os algoritmos 10, 7 e 11 (linhas 2,3 e 4), uma vez que exista a necessidade dos índices do *Sujeito* e *Verbo* e *Agente* na sentença. Na linha 7 do algoritmo 13, é verificado se há ocorrências de verbos no tempo verbal presente perfeito da língua inglesa, dado que há outras regras para tratar deste caso (por exemplo, regra 9 da Tabela 3.2). Desta forma, na

---

**Algoritmo 12** Algoritmo de Identificação de Atividades (Regra 1)
 

---

**Entrada:** Vetor com a análise sintática da sentença

**Saída:** Retorna um vetor contendo os índices do verbo, sujeito e objeto e uma variável booleana.

```

1: procedure GETRULE1ACTIVITY(Doc)
2:   Subject, IndiceSubject ← GetContentIndiceSubject(Doc)
3:   Verb, IndiceVerb ← GetContentIndiceVerb(Doc)
4:   Object, IndiceObject ← GetContentIndiceObject(Doc)
5:   OtherRule ← False
6:   if (isConjunction(Doc) = True) then                                ▷ verificar conjunções
7:     OtherRule ← True
8:   for  $\forall$  Indice < (Doc) do
9:     if InTheFuture = content then                                    ▷ verificar auxiliares
10:      OtherRule ← True
11:    if (IndiceVerb = Null or IndiceSubject = Null or IndiceObject = Null or OtherRule =
12:      True) then
13:      return ERROR
14:    else
15:      if IndiceSubject < IndiceVerb and IndiceVerb < IndiceObject then
16:        return OK

```

Fonte: Os autores

---

linha 9 do algoritmo, é verificado se a sentença tem *Sujeito*, *Verbo*, *Agente* e *Objeto*. Em caso afirmativo, é verificado se a sentença representa a sequência *Sujeito*, *Verbo*, *Agente* e *Objeto* (linha 12), conforme a definição da regra 2 de eventos. Portanto, caso a sentença representar a sequência definida, neste caso, tal sentença é identificada como um evento do modelo de processo. Se a sentença não for identificada tal como a regra 2 de eventos, a sentença prossegue para regra 3, e assim sucessivamente.

---

**Algoritmo 13** Algoritmo de Identificação de Eventos (Regra 2)
 

---

**Entrada:** Vetor com a análise sintática da sentença

**Saída:** Retorna um vetor contendo os índices do verbo, sujeito, objeto e uma variável booleana.

```

1: procedure GETRULE2EVENT(Doc)
2:   Subject, IndiceSubject ← GetContentIndiceSubject(Doc)
3:   Verb, IndiceVerb ← GetContentIndiceVerb(Doc)
4:   Object, IndiceObject ← GetContentIndiceObject(Doc)
5:   IndiceAgent, isAgent ← isAgente(Doc)
6:   OtherRule ← False
7:   if (isPresentperfect(Doc) = True) then
8:     OtherRule ← True
9:   if (IndiceVerb = Null or IndiceSubject = Null or IndiceObject = Null or OtherRule =
10:     True or isAgent = False) then
11:     return ERROR
12:   else
13:     if (IndiceSubject < IndiceVerb and IndiceVerb < IndiceAgent and IndiceAgent <
14:       IndiceObject) then
15:       return OK

```

Fonte: Os autores

---

### 3.4 Saída de Texto

A fim de obter um texto com as regras mapeadas, é necessário analisar todas as sentenças do texto, buscando identificar a dependência entre palavras e as classes gramaticais das mesmas. Tais sentenças identificadas podem ser entendidas como candidatas à extração de elementos processo, baseado nas regras de mapeamento identificadas. Após analisar todas as sentenças do texto, foram empregadas diretrizes de modelagem, com o objetivo de indicar tais diretrizes nos textos em linguagem natural, uma vez que as diretrizes (D) possibilitam um controle maior na modelagem e execução dos processos de negócio. Contudo, essas diretrizes são aplicadas a nível de modelagem. Foram consideradas as diretrizes de modelagem de processos apresentadas na pesquisa de Mendling, Reijers and Aalst (2010), que visam realizar a estruturação dos modelos de processo e diminuir a probabilidade de erros de entendimento durante a modelagem de processos. Como resultado, os autores propõem um conjunto de 7 diretrizes sobre modelagem de processos de negócio (*7PMG*). As diretrizes são:

- D1: Utilize o mínimo possível de elementos no modelo;
- D2: Minimize os caminhos de roteamento por elemento;
- D3: Utilize apenas um evento de início e um de fim;
- D4: Modele o processo o mais estruturado possível;
- D5: Evite elementos de roteamento OR;
- D6: Utilize o estilo verbo-objeto para os rótulos de atividades;
- D7: Decomponha modelos se este apresentar mais de 50 elementos.

Nesta dissertação, foram utilizadas as seguintes diretrizes:

- **D1:** O número de elementos de processo do modelo de processo depende exclusivamente, da quantidade de sentenças presentes no texto, uma vez que tais sentenças podem ser candidatas à elementos de processo, conforme as regras de mapeamento ou não. A tomada de decisão em modelar tais sentenças é em particular, do analista de processos.
- **D3:** a fim de diminuir a probabilidade de erros e permitir maior compreensão da modelagem (MENDLING; NEUMANN; AALST, 2007), foi verificado se os textos analisados, nesta dissertação, contêm eventos iniciais e finais no início e fim das

sentenças, respectivamente. Tal verificação foi realizada de forma automática<sup>9</sup>.

- **D5:** não foi incluído o desvio inclusivo (OR), devido a problemas estruturais e semânticos com tal elemento (MENDLING; NEUMANN; AALST, 2007; KINDLER, 2006; FIGL; RECKER; MENDLING, 2013; KOSSAK; ILLIBAUER; GEIST, 2012). Além de existir ambiguidades na semântica do desvio inclusivo (*OR-join*) (KINDLER, 2006).
- **D6:** a fim de permitir maior clareza e compreensão de modelos de processo, foram utilizados rótulos (*labels*) em elementos de processo. Isto é, todos os elementos de processo identificados no texto, tais como atividades e eventos, os rótulos foram definidos tal como verbo-objeto para atividades e objeto-participio para eventos. A definição dos rótulos foi realizada de forma automática.
- **D7:** com a finalidade de decompor o modelo de processo, foram verificadas e apontadas as quantidades de elementos de processo identificados em cada texto, para que o analista de processos utilize as decisões apropriadas, baseadas em seu conhecimento.

Em síntese, para descrever elementos do processo de negócio em texto em linguagem natural esta etapa inclui os seguintes passos: identificar eventos iniciais, eventos finais, raias (*swimlanes*), ações, tarefas, rótulos de tarefas, rótulos de eventos, além de apresentar o número de elementos do processo do texto analisado, bem como identificar possíveis elementos de processo omitidos pelo analista de processos.

A Figura 3.1 ilustra a próxima etapa desta dissertação, descrita como geração de texto orientado a processo. A abordagem aqui apresentada, servirá como dado de entrada para a preparação do texto orientado a processo. Tal texto está definido como uma estrutura que permite identificar: papel associado a uma atividade, piscinas relacionadas com raias, interação entre piscinas (fluxo de mensagem), identificar eventos (início, intermediário e fim) e apontar controles de fluxo (AND e XOR). Está fora do contexto desta dissertação, gerar um modelo (*template*) de como o texto deve ser estruturado para extração de modelos de processo a partir de texto (FERREIRA; THOM, 2016). Porém, os resultados desta dissertação são pré-requisitos para gerar texto orientado a processo.

---

<sup>9</sup>disponível em: <[www.github.com/renatoborgess/python\\_parser](http://www.github.com/renatoborgess/python_parser)>

### 3.5 Considerações Finais

Neste capítulo foram descritas as quatro etapas para identificação de elementos de processos de negócio em texto em linguagem natural de forma semiautomática.

Cada etapa da abordagem, mostra um objeto de saída que permite que a próxima etapa seja executada. Isto é, todas as etapas são dependentes da etapa anterior.

A etapa de entrada de dados visou separar (manualmente) todas as sentenças presentes no texto em um arquivo *.txt*. Para que a próxima etapa, denominada análise sintática de texto, gere um vetor de objetos (*Doc*), incluindo todas as sentenças do texto marcadas sintaticamente. A etapa de análise lógica de texto, mostra que a aplicação das regras de mapeamento definidas para atividades, eventos, desvio paralelo e exclusivo em sentenças, possibilita que tais sentenças sejam candidatas à extração de elementos de processo a partir delas. Por fim, na etapa de saída de texto, são aplicadas as diretrizes de modelagem propostas por Mendling, Reijers and Aalst (2010), nos textos analisados pela abordagem.

Como principais limitações da abordagem proposta, destaca-se que apesar do número de regras de mapeamento ser relativamente expressivo, é possível que ocorra omissão na identificação de elementos de processo em sentenças. Por esta razão, o analista de processo poderá utilizar os resultados da identificação de elementos de processo no texto em linguagem natural como ponto de partida para a modelagem. Outro ponto identificado é que, considerando o uso do dicionário de sinônimos (*PyDictionary*) para analisar as sentenças, é possível que tal dicionário, não identifique *palavras-chave* que denotam exclusividade ou paralelismo. Logo, o analista novamente, se encarregaria de realizar tal decisão. Portanto, o trabalho proposto nesta dissertação, não visa substituir o analista, e sim auxiliá-lo, para que as tomadas de decisões sejam exclusivamente dele.

A partir do conjunto de regras de mapeamento desenvolvidos, no próximo capítulo são apresentados os resultados obtidos e avaliação dos mesmos através de questionários e métricas de desempenho.

## 4 AVALIAÇÃO DA ABORDAGEM PROPOSTA E RESULTADOS OBTIDOS

Este capítulo descreve as avaliações da abordagem e os resultados obtidos através de questionários, verificando a usabilidade junto aos usuários e capturando a respectiva aceitação. Além disso, para medir o desempenho do protótipo desenvolvido, o presente trabalho faz o uso de métricas de recuperação de informação tais como precisão, revocação (*recall*) e medida-F (*F-measure*). Tal protótipo foi desenvolvido na linguagem de programação *Python*, além da utilização do analisador sintático denominado *Spacy* para realizar a análise sintática do texto.

Nas próximas seções serão detalhadas e discutidas as avaliações realizadas no contexto desta dissertação.

### 4.1 Avaliação Através de Questionário

Foram realizadas duas avaliações através de questionários junto aos usuários. O objetivo das avaliações é verificar a aceitabilidade, visando identificar se o uso da abordagem minimizará o tempo e o esforço da tarefa de modelagem de processos.

Conforme Pezzè and Young (2009), a validação baseada em participação de usuários é particularmente importante para validar a usabilidade do *software*. A aplicação pode ocorrer em diversos estágios do desenvolvimento e pode ser utilizado com diferentes propósitos, como por exemplo: identificar o modelo mental dos usuários finais, avaliar alternativas de projeto e testar a usabilidade.

O teste de aceitação proposto por Sommerville (2011), dispõe que os clientes experimentem um sistema para decidir se este está ou não pronto para ser aceito pelos desenvolvedores de sistemas e implantado no ambiente do cliente.

Portanto, foi aplicado o teste de aceitação por ser mais adequado a este trabalho, uma vez que o objetivo é avaliar a usabilidade da abordagem proposta, visando identificar dificuldades e obstáculos, medidos através da interação do usuário e regras de mapeamento, além de medir características, como grau de dificuldade ao identificá-las.

Para isto, aplicou-se a validação utilizando uma estrutura de questionário, objetivando capturar estatísticas com base na opinião dos usuários, principalmente analistas de processo (por exemplo, engenheiros de processo, diretores de processo, *process designer*, participantes de processo, donos do processo etc). Além disso, os questionários objetivaram obter experiências em modelagem de processos e conhecimentos em BPMN, além

de ser possível obter a opinião sobre o uso das regras de mapeamento e sua usabilidade. Os principais requisitos para responder os questionários, foram a experiência em modelagem de processos e experiência em BPMN, uma vez que as regras de mapeamento foram baseadas nesta notação. A estrutura de questionário utilizada foi a do *Google Forms*<sup>1</sup>, que possibilita uma fácil construção de questionários.

Nas próximas seções, serão descritos detalhes dos questionários e suas particularidades.

#### 4.1.1 Primeiro Questionário Aplicado

O primeiro questionário aplicado visava identificar elementos de processos de negócio em sentenças. Tais sentenças foram retiradas de textos<sup>2</sup> em linguagem natural. O período que o questionário ficou disponível foi de 20 de outubro de 2016 a 10 de novembro de 2016. Tal questionário foi amplamente divulgado em redes sociais (*Facebook* e *LinkedIn*), sites, listas de e-mails, dentre outros meios. Portanto, foram analisados dados de 43 participantes, sendo eles: analistas de processo, desenvolvedores de *softwares*, estudantes, dentre outros. O questionário foi dividido em 3 etapas.

A primeira etapa visava levantar informações sobre as experiências dos participantes. A Figura 4.1 ilustra o formulário desta etapa, cujos campos são os seguintes:

1. *Profissão*: Visava identificar a profissão do participante. Como opções disponíveis tinha-se: Estudante, professor, desenvolvedor de *software*, especialista de processo e outros.
2. *Escolaridade*: Visava identificar o nível de escolaridade do participante. Como opções disponíveis tinha-se: Ensino Médio, Superior Incompleto, Superior Completo, Pós-graduação (mestrado ou doutorado).
3. *Possui conhecimentos em modelagem de processos de negócio*: Pretendia identificar se o participante possuía algum conhecimento de modelagem de processos de negócio. Como opções disponíveis tinha-se: Sim ou Não.
4. A resposta dessa opção era condicionada a resposta da pergunta 3. Ao responder sim, o participante informava o tempo de experiência em modelagem de processos de negócio. Opções disponíveis: Menos de 2 anos, de 2 a 4 anos, mais de 4 anos.

---

<sup>1</sup><https://www.google.com/forms/about>; último acesso em 17-01-2017

<sup>2</sup>Os textos foram retirados da *Humboldt University of Berlin (HUB)*. Os textos da HU-Berlim representam exercícios de modelagem que são utilizados em tutoriais de BPMN.

5. *Possui conhecimentos em BPMN*: Principal questão da primeira etapa, uma vez que tal pergunta visava identificar se o participante possuía algum conhecimento da BPMN. Como opções disponíveis tinha-se: Sim ou Não.
6. Resposta condicionada à resposta da pergunta 5. Semelhante a questão 4, ao responder sim, o participante informava o tempo de experiência em BPMN. Opções disponíveis: Menos de 2 anos, de 2 a 4 anos, mais de 4 anos.
7. *Possui conhecimentos nas práticas de modelagem proposto por Jan Mendling*: Pergunta com objetivo de identificar se o participante possuía algum conhecimento nas 7 diretrizes de modelagem de processo (7PMG). Uma vez que tais diretrizes são utilizadas nesta abordagem, por exemplo, utilização do verbo-objeto em rótulos dos elementos de processo. Como opções disponíveis tinha-se: Sim ou Não.

Figura 4.1: Primeira etapa do questionário.

**Identificação de Elementos de Processos em Textos em Linguagem Natural**

Pesquisa que visa o estudo de Identificação de Elementos de Processo em Textos em Linguagem Natural e a criação de uma ferramenta que identifique elementos de processo em textos.

**\*Obrigatório**

**Parte 1 - Sobre Você**

**1 Profissão \***

- Estudante
- Professor
- Desenvolvedor de software
- Especialista de Processo (analista, diretor, engenheiro, etc...)
- Outro

**2 Escolaridade \***

- Ensino Médio
- Superior Incompleto
- Superior Completo
- Pós-Graduação (Mestrado e Doutorado)

**3 Possui conhecimento em modelagem de processos de negócio? \***

- Sim
- Não

**4 Se sim, quanto tempo? \***

- Menos de 2 anos
- Entre 2 e 4 anos
- Mais de 4 anos

**5 Possui conhecimento em BPMN (Notação e Modelo de Processos de Negócio) \***

- Sim
- Não

**6 Se sim, quanto tempo?**

- Menos de 2 anos
- Entre 2 e 4 anos
- Mais de 4 anos

**7 Possui conhecimento nas práticas de modelagem proposto por Jan Mendling et al. ? \***

Link: <http://www.wis.win.tue.nl/~wvdaalst/publications/p574.pdf>

- Sim
- Não

**8 Qual seu nível de conhecimento da gramática da Língua Inglesa. \***

Sem nenhum  1  2  3  4  5  Alto conhecimento

Fonte: Os autores

8. *Qual seu nível de conhecimento da gramática da Língua Inglesa*: visava identificar

o nível de conhecimento gramatical da língua inglesa do participante. A resposta foi baseada na escala Likert (LIKERT, 1932; EDWARDS, 1957; ROBERTSON, 2012) de cinco pontos, variando de sem nenhum conhecimento até alto conhecimento.

Todos os campos eram obrigatórios, exceto aqueles em que eram condicionada a resposta anterior. Por se tratar de um questionário que visava capturar a aceitação do usuário com a utilização das regras de mapeamento, objetivou-se identificar sua experiência com modelagem de processos (questões 3 e 4) e experiência da notação BPMN (questões 5 e 6). Esses conhecimentos influenciam na identificação de elementos de processo, baseadas nas regras de mapeamento, sendo que a experiência de modelagem de processo foca na capacidade de entender qual elemento notacional esta sendo identificado na sentença, e o conhecimento em BPMN foca em entender a análise semântica e sintaxe do elemento notacional identificado.

Figura 4.2: Segunda etapa do questionário.

**Parte 2 - Identificação de elementos de processo em Textos.**

**Abaixo são apresentadas sentenças que contém elementos de processo. Você pode deixar sua opinião sobre os elementos extraídos das mesmas.**

**Primeira Sentença**

A customer brings a defective computer and the CRS checks the defect and hands out a repair cost calculation back.

**Quais elementos de processo são mais adequados para representar esta sentença?**

\*  
 Três Tarefas e duas Piscinas (Customer e CRS)  
 Três Eventos e duas piscinas (Customer e CRS)  
 1 Gateway do tipo XOR  
 1 Gateway do tipo AND

**Justifique sua resposta. \***  
 Sua resposta

**Segunda Sentença**

If the customer decides that the costs are acceptable, the process continues, otherwise, she takes her computer home unrepaired.

**Quais elementos de processo são mais adequados para representar esta sentença?**

\*  
 1 Gateway do tipo XOR + 2 tarefas  
 Dois Eventos + 3 tarefas  
 1 Gateway do tipo AND + 3 tarefas  
 Duas Tarefas

**Terceira Sentença**

The ongoing repair consists of two activities, which are executed, in an parallel order. The first activity is to check and repair the hardware, whereas the second activity checks the software and configures the hardware.

**Quais elementos de processo são mais adequados para representar esta sentença?**

\*  
 1 Gateway do tipo AND + 4 tarefas  
 Um Evento + 2 tarefas  
 1 Gateway do tipo AND + 4 Eventos  
 1 Gateway do tipo XOR + 2 tarefas

**Justifique sua resposta. \***  
 Sua resposta

Além destas, buscou-se identificar nas questões 1 e 2 a origem do participante, a partir de sua escolaridade e profissão, uma vez que a profissão, tal como especialista de processos, é o público alvo deste questionário, pois são eles que modelam processos (por exemplo, modelo *as-is*) em um projeto BPM.

Figura 4.3: Terceira etapa do questionário.

**Parte 3 - Opinião da Identificação de elementos de processos em Textos.**

Sua opinião é importante para mim.

**Primeira Sentença**

A customer brings a defective computer and the CRS checks the defect and hands out a repair cost calculation back.

**Modelagem da Primeira Sentença**

Você concorda com a modelagem da primeira sentença ? \*

Concordo plenamente

Concordo

Concordo parcialmente

Discordo Parcialmente

Discordo

Se discorda, justifique sua resposta.

Sua resposta

**Segunda Sentença**

The SCT physical file was stored by the Back office. When the report is received, the respective SCT file is retrieved.

**Modelagem da Segunda Sentença**

Você concorda com a modelagem da segunda sentença ? \*

Concordo plenamente

Concordo

Concordo parcialmente

Discordo Parcialmente

Discordo

Se discorda, justifique sua resposta.

Sua resposta

**Terceira Sentença**

If the customer decides that the costs are acceptable, the process continues, otherwise, she takes her computer home unrepaired.

**Modelagem da Terceira Sentença**

Você concorda com a modelagem da terceira sentença ? \*

Concordo plenamente

Concordo

Concordo parcialmente

Discordo Parcialmente

Discordo

Fonte: Os autores

Após o preenchimento dos dados, o participante passava para a segunda etapa, conforme apresenta a Figura 4.2. O objetivo desta etapa é permitir a conformidade das respostas dos participantes de acordo com as regras de mapeamento criadas para o ele-

mento de processo mostrado nas sentenças. Para a primeira sentença, foi utilizada a regra 4 da Tabela 3.1. Para a segunda sentença, foi utilizada a regra 2 da Tabela 3.3. Por fim, na terceira sentença, foi utilizada a regra 4 da Tabela 3.4.

A estrutura da segunda etapa do questionário foi realizada da seguinte forma: foi apresentado uma sentença. A partir desta sentença, o participante realizava a modelagem da sentença conforme seus conhecimentos. Posteriormente, responderia a questão se concordava com a modelagem baseada nas regras de mapeamento. Por fim, o participante justificava sua resposta. O objetivo desta questão dissertativa, é obter maiores informações dos motivos de assinalar tal alternativa (por exemplo, *1 Gateway do tipo AND*). Assim como na primeira etapa, conforme a Figura 4.2, os campos "*Quais elementos de processo são mais adequados para representar esta sentença*" e "*Justifique sua resposta*" eram obrigatórios.

A terceira etapa do questionário visava obter a opinião dos participantes sobre a modelagem de processos mostrada no questionário. Assim, seis modelos de processo foram criados por um analista a partir de sentenças, e apenas dois dos modelos de processo foram modelados de acordo com as regras de mapeamento, tais como, segunda e quarta sentença, ilustradas nas Figuras 4.3 e 4.4. O restante dos modelos de processo (quatro modelos) foram propositalmente modelados incorretamente, uma vez que verifica-se, a conformidade das respostas dos participantes de acordo com as regras de mapeamento para tal sentença. Para a segunda modelagem, foi utilizada a regra 2 da Tabela 3.2. Para a quarta modelagem, foram utilizadas as regras 1 da Tabela 3.2 e 3.1, respectivamente. As Figuras 4.3 e 4.4, ilustram esta etapa.

A estrutura da terceira etapa do questionário foi realizada da seguinte forma: foi apresentada uma sentença. A partir desta sentença, o modelo de processo é mostrado. Posteriormente, o participante responderia, se concordava com a modelagem. Tal questão, foi baseada na escala Likert de cinco pontos, variando de concordo plenamente até discordo. Por fim, o participante justificava a resposta, uma vez que ao discordar dos modelos de processo corretos (dois e quatro), o objetivo desta questão dissertativa é obter maiores informações dos motivos de assinalar tal alternativa. Conforme as Figuras 4.3 e 4.4, os campos "*Você concorda com a modelagem desta sentença*" e "*Se discorda, justifique sua resposta*" eram obrigatórios, dado que o campo "*Se discorda, justifique sua resposta*" era condicionado a questão anterior.

Figura 4.4: Continuação da terceira etapa do questionário.

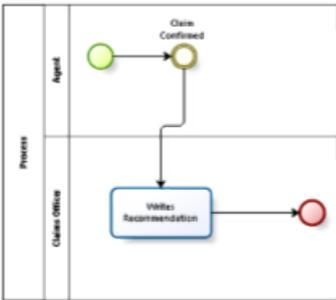
**Parte 3 (continuação) - Opinião da Identificação de elementos de processos em Textos.**

Sua opinião é importante para mim.

**Quarta Sentença**

After the agent has confirmed the claim to the clerk. The claims officer then writes a "settlement recommendation".

**Modelagem da Quarta Sentença**



Você concorda com a modelagem da quarta sentença ? \*

Concordo plenamente

Concordo

Concordo parcialmente

Discordo Parcialmente

Discordo

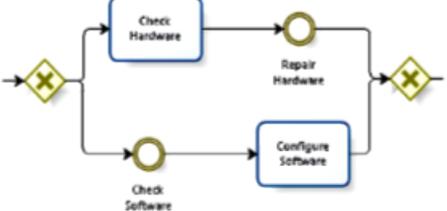
Se discorda, justifique sua resposta.

Sua resposta

**Quinta Sentença**

The ongoing repair consists of two activities, which are executed, in a parallel order. The first activity is to check and repair the hardware, whereas the second activity checks the software and configures the hardware.

**Modelagem da Quinta Sentença**



Você concorda com a modelagem da quinta sentença ? \*

Concordo plenamente

Concordo

Concordo parcialmente

Discordo Parcialmente

Discordo

Se discorda, justifique sua resposta.

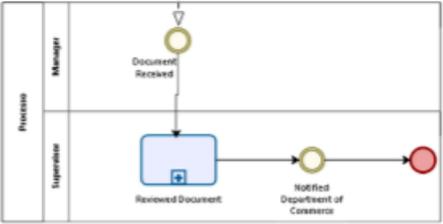
Sua resposta

---

**Sexta Sentença**

After the document has been received by the case manager, the supervisor has reviewed. Then, the Department of Commerce and Sell are notified.

**Modelagem da Sexta Sentença**



Você concorda com a modelagem da quinta sentença ? \*

Concordo plenamente

Concordo

Concordo parcialmente

Discordo Parcialmente

Discordo

Se discorda, justifique sua resposta.

Sua resposta

#### 4.1.2 Resultados Obtidos Através da Aplicação do Primeiro Questionário

A partir da aplicação do questionário, foi obtida a participação dos usuários, que foram distribuídos entre estudantes de graduação e pós-graduação de cursos da área da computação, especialistas de processo, professores, dentre outros.

As Figuras 4.8, 4.9 e 4.10, mostram os dados (43 participantes) relacionados de cada etapa, respectivamente.

A fim de obter as informações dadas pelos especialistas de processo, uma vez que tal profissão era o público alvo deste questionário, foram analisadas as respostas de 22 participantes. Foram selecionados as seguintes características dos participantes:

- Especialista de processos;
- Experiência em BPMN;
- Mais de dois anos de experiência em BPMN;
- Experiência em modelagem de processos;
- Mais de dois anos de experiência em modelagem de processos.

Estas características foram selecionadas, pois os principais requisitos para responder o questionário eram experiência em BPMN e experiência em modelagem de processos. Conforme Mendling, Strembeck and Recker (2012), Mendling, Reijers and Cardoso (2007), Schrepfer et al. (2009), Reijers and Mendling (2008), tais características influenciam nos resultados das modelagens. Portanto, tais participantes selecionados representam 51,16% do total de participantes do questionário.

A avaliação conduzida nesta dissertação demonstrou resultados promissores. A Figura 4.5, ilustra os resultados da segunda etapa do questionário, sendo que 90% dos participantes selecionados, concordam com o modelo apresentado na primeira sentença do questionário. Quanto a segunda sentença, 100% dos participantes concordaram com a modelagem, enquanto 90% aprovou a modelagem na terceira sentença. Importante observar que todas as sentenças foram modeladas conforme as regras de mapeamento propostas nesta dissertação.

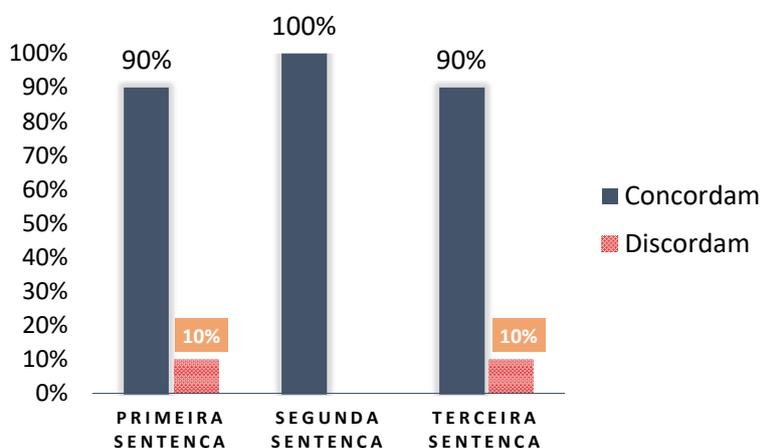
A terceira etapa do questionário, também demonstra resultados promissores. Os resultados obtidos nas modelagens de processo, que representam as modelagens propositalmente incorreta de acordo com as regras de mapeamento propostas nesta dissertação, mostram que a maioria dos participantes selecionados discordam (90%, 100%, 95% e 77%, respectivamente) com as modelagens apresentadas. A Figura 4.6 ilustra os resulta-

dos obtidos.

Por outro lado, a segunda e quarta modelagens, que representam as modelagens corretas de acordo com as regras de mapeamento, mostram que a maioria dos participantes concordaram (68% e 81%, respectivamente) com as modelagens apresentadas. A Figura 4.7 ilustra os resultados obtidos.

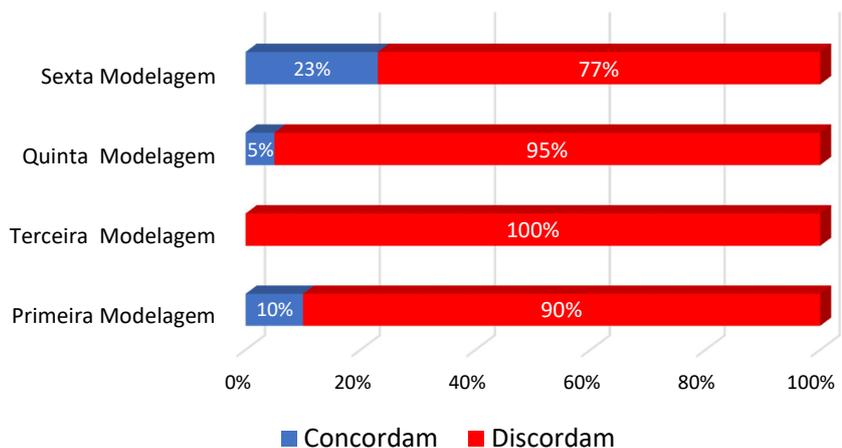
Com base na aceitação dos usuários em relação ao uso das regras de mapeamento para modelagem de processos a partir de texto em linguagem natural, pôde-se alcançar um dos objetivos específicos deste trabalho, que era "avaliar, através de pesquisa de opinião, as considerações de possíveis usuários quanto ao uso das regras de mapeamento definidas nesta dissertação e aplicabilidade delas em texto".

Figura 4.5: Resultados da segunda etapa do primeiro questionário através dos 22 participantes selecionados.



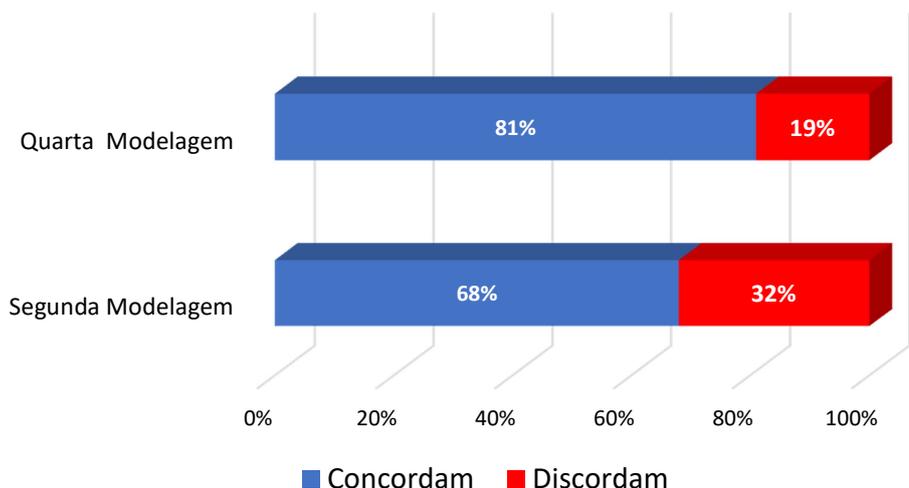
Fonte: Os autores

Figura 4.6: Resultados da terceira etapa do primeiro questionário através dos 22 participantes selecionados.



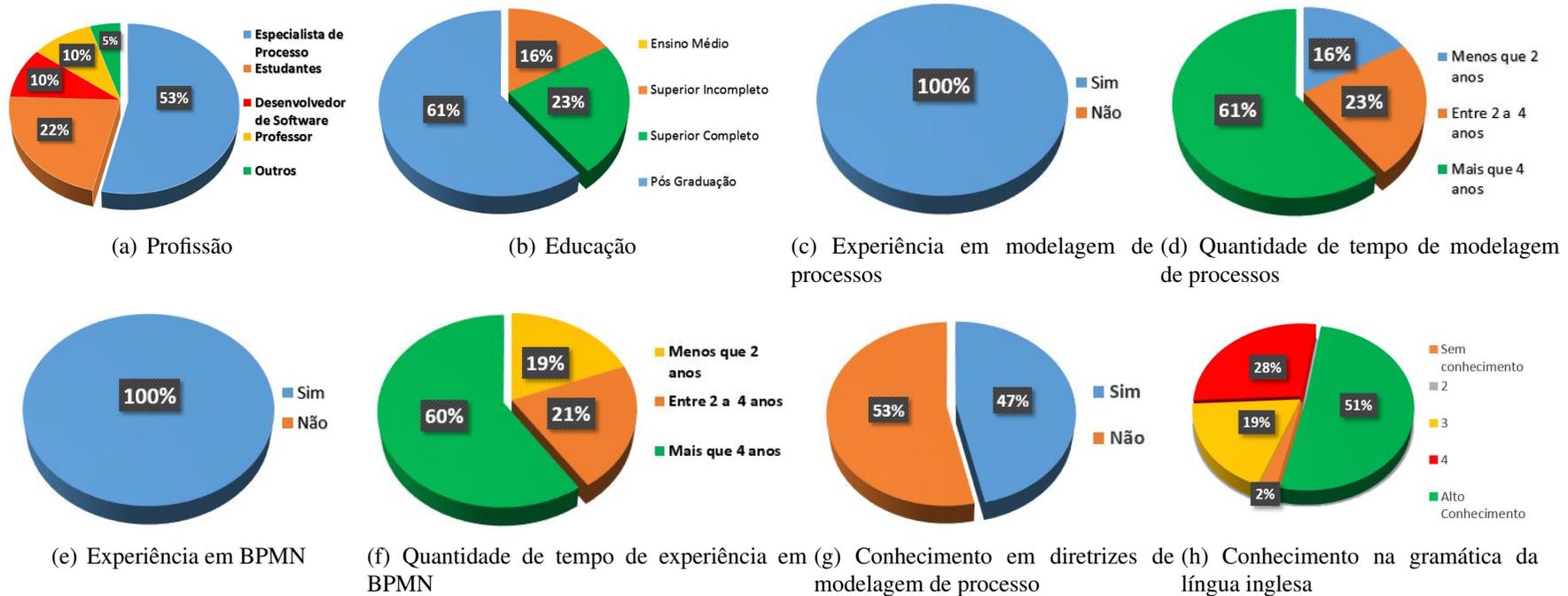
Fonte: Os autores

Figura 4.7: Resultados da terceira etapa do primeiro questionário através dos 22 participantes selecionados.



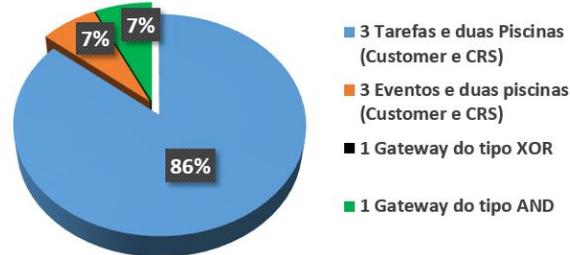
Fonte: Os autores

Figura 4.8: Resultados obtidos a partir da primeira parte do primeiro questionário.

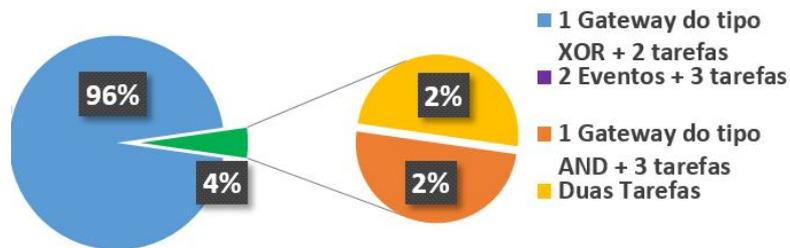


Fonte: Os autores. A Figura 4.8(a), mostra que mais da metade dos participantes são especialistas de processo (53%), 22% são estudantes, 10% são desenvolvedores de *softwares* e professores e, finalmente, 5% são outras profissões. A Figura 4.8(b), ilustra que 61% dos participantes detém pós-graduação (mestrado e/ou doutorado), 23% detém graduação e 16% detém curso de ensino superior incompleto. A Figura 4.8(c), descreve que 100% dos participantes detém experiência em modelagem de processos. A Figura 4.8(d), mostra que 61% dos participantes possui mais de quatro anos de experiência em modelagem de processos, 23% possui entre dois e quatro anos e 16% possui menos de dois anos de experiência. A Figura 4.8(e), mostra que 100% dos participantes detém experiência em BPMN. A Figura 4.8(f), ilustra que 60% dos participantes possui mais de quatro anos de experiência em BPMN, 21% possui entre dois e quatro anos e 19% detém menos de dois anos. A Figura 4.8(g), mostra que 53% dos participantes possui conhecimentos nas diretrizes de modelagem de processos e 47% sem conhecimento. Finalmente, a Figura 4.8(h), mostra que 51% dos participantes possui um alto conhecimento em inglês, 28% detém grande conhecimento, 19% possui bom conhecimento e apenas 2% não tem algum conhecimento da língua.

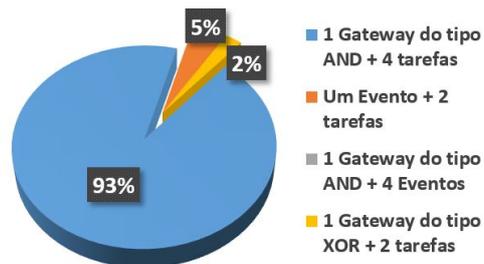
Figura 4.9: Resultados obtidos a partir da segunda parte do primeiro questionário.



(a) Primeira Sentença



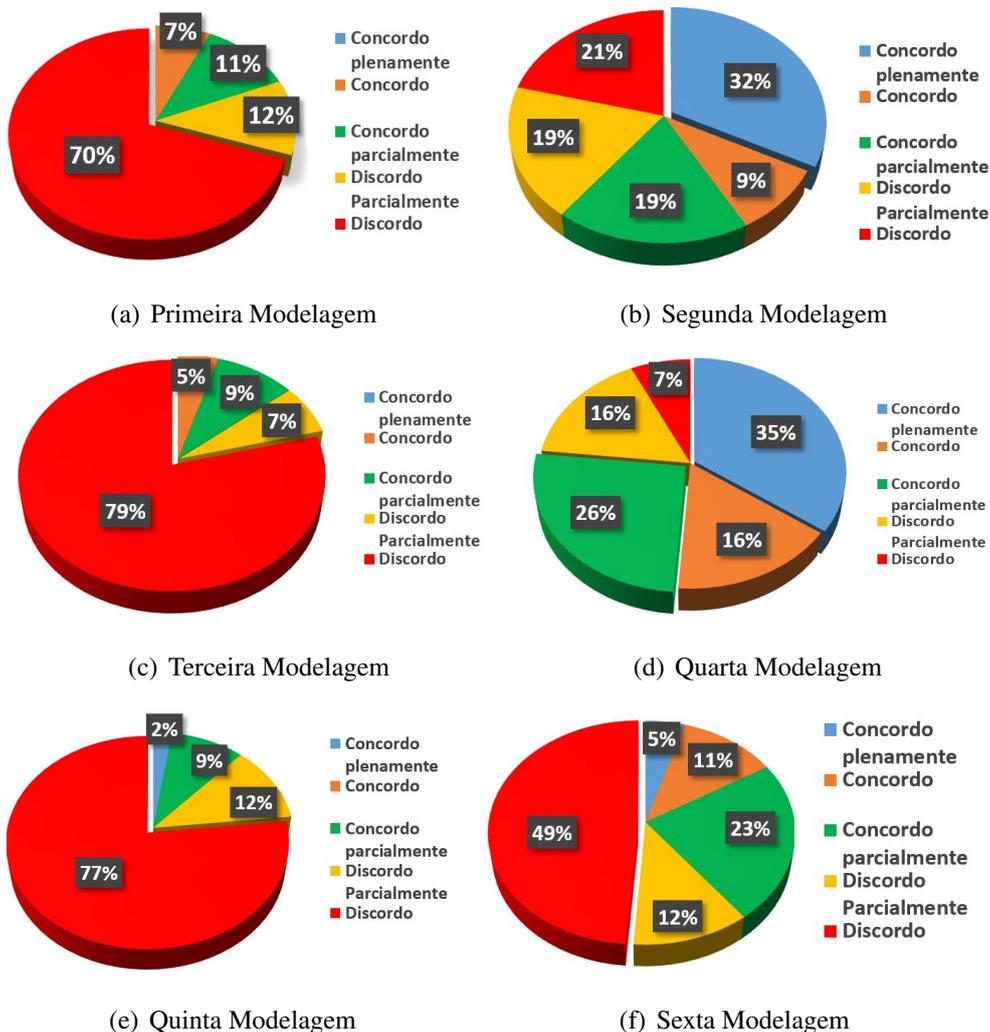
(b) Segunda Sentença



(c) Terceira Sentença

Fonte: Os autores. Nesta figura são apresentadas as avaliações dos 43 participantes do questionário. Os resultados da primeira sentença mostrada na Figura 4.9(a), descrevem que 86% dos participantes concordam com a modelagem. A Figura 4.9(b), ilustra que 96% dos participantes, também concordam com a modelagem apresentada. Finalmente, a Figura 4.9(c), descreve que 93% dos participantes concordam com a modelagem. Nesta etapa do questionário, todas as sentenças foram modeladas de acordo com as regras de mapeamento propostas nesta dissertação. Portanto, os resultados apresentam, em média, 91,66% de aceitação dos participantes.

Figura 4.10: Resultados obtidos da terceira parte do primeiro questionário.



Fonte: Os autores. Nesta figura são apresentadas as avaliações dos 43 participantes do questionário. As Figuras 4.10(a), 4.10(c), 4.10(e) e 4.10(f), que representam os modelos de processo propositalmente incorretos de acordo com as regras de mapeamento propostas nesta dissertação, mostram que a maioria dos participantes discordam (70%, 79%, 77% e 49%, respectivamente) da modelagem apresentada. Por outro lado, os modelos de processo das Figuras 4.10(b) e 4.10(d), nos quais são representados de acordo com as regras de mapeamento propostas nesta dissertação, mostram que a maioria dos participantes concordam (60% e 77%, respectivamente) com as modelagens propostas.

### 4.1.3 Segundo Questionário Aplicado

O segundo questionário a ser respondido visava controlar o tempo de modelagem com e sem o uso das regras de mapeamento, bem como avaliá-las em relação a aceitação dos participantes. O período em que ficou disponível foi de 11 de janeiro de 2017 à 30

de janeiro de 2017. Tal questionário foi amplamente divulgado em redes sociais, sites, listas de e-mails, dentre outros meios. Portanto foram analisados dados de 21 participantes, incluindo: especialistas de processo, desenvolvedores de *softwares*, estudantes, dentre outros. O questionário foi dividido em quatro etapas. A primeira etapa deste questionário visava levantar informações sobre a experiências dos participantes. Por esta razão, a primeira parte é idêntica ao primeiro questionário, apresentado na seção 4.1.1, exceto as questões 7 e 8, uma vez que tais questões não são adequadas ao objetivo deste questionário.

Figura 4.11: Segunda e terceira etapa do questionário.

Parte 2 - Modelagem de processo de negócio a partir de textos em linguagem natural	Parte 3 - Modelagem de processo de negócio a partir de textos em linguagem natural
Abaixo é apresentado um texto em linguagem natural retirado de Dumas et al 2013.	Abaixo é apresentado um texto em linguagem natural retirado de Dumas et al 2013. Porém, tal texto é marcado com elementos de processos de negócio conforme a BPMN.
Atividade 1 - Modele o texto conforme o seu conhecimento.	Atividade 1 - Modele o texto conforme o seu conhecimento.
Atividade 2 - Anote a quantidade de tempo (em minutos) gasto para realizar esta modelagem.	Atividade 2 - Anote a quantidade de tempo (em minutos) gasto para realizar tal modelagem.
<p>Once a loan application has been approved by the loan provider, an acceptance pack is prepared and sent to the customer.</p> <p>The acceptance pack includes a repayment schedule which the customer needs to agree upon by sending the signed documents back to the loan provider.</p> <p>The latter then verifies the repayment agreement:</p> <p>if the applicant disagreed with the repayment schedule, the loan provider cancels the application.</p> <p>if the applicant agreed, the loan provider approves the application.</p> <p>In either case, the process completes with the loan provider notifying the applicant of the application status.</p>	<p>Once a loan application has been approved by the loan provider (Evento de inicio), an acceptance pack is prepared (Tarefa: Prepare Acceptance Pack) and sent to the customer (Tarefa: Sent Acceptance Pack).</p> <p>The acceptance pack includes a repayment schedule which the customer needs to agree upon by sending the signed documents back (Tarefa: Receive Signed Documents) to the loan provider.</p> <p>The latter then verifies the repayment agreement (Tarefa: Verify Repayment):</p> <p>if the applicant disagreed with the repayment schedule (Desvio XOR), the loan provider cancels the application (Tarefa: Cancel Application).</p> <p>if the applicant agreed, the loan provider approves the application (Tarefa: Approve Application).</p> <p>The process completes with the loan provider notifying the applicant of the application status (Tarefa: Notify Application).</p>
<p>Quantidade de tempo gasto para realizar tal modelagem *</p> <p>Sua resposta</p>	<p>Quantidade de tempo gasto para realizar tal modelagem ? *</p> <p>Sua resposta</p>
<p>Grau de Dificuldade *</p> <p>1 2 3 4 5</p> <p>Fácil <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> Dificil</p>	<p>Grau de Dificuldade *</p> <p>1 2 3 4 5</p> <p>Fácil <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> Dificil</p>

Fonte: Os autores

Após o preenchimento dos dados na primeira etapa, o participante passava para a segunda etapa, conforme apresentada na Figura 4.11 (1). A fim de obter um controle de tempo de modelagem, o objetivo desta etapa era controlar o tempo estimado da modelagem e mensurar o nível de dificuldade de tal modelagem.

Nesta etapa, o texto de exemplo não estava identificado conforme as regras de

mapeamento. Portanto, o participante modelaria tal texto conforme seu conhecimento em modelagem de processos.

Figura 4.12: Quarta etapa do questionário.

**Parte 4 - Identificação de Elementos de Processo de Negócio a partir de Texto em Linguagem Natural**

**Sobre a Modelagem**

**1 Qual das opções há mais detalhes de modelagem de processo?**

\*

- Texto 1 - Sem identificação
- Texto 2 - Com Identificação

**2 Justifique sua resposta**

Sua resposta

---

**3 A Identificação de elementos de processos de negócio no Texto 2 ajudou a modelar tais elementos descrito no texto? \***

- Concordo totalmente
- Concordo parcialmente
- Indiferente
- Discordo parcialmente
- Discordo totalmente

**4 Você concorda que a identificação de elementos de processos de negócio em textos em linguagem natural auxiliará e minimizará o esforço necessário dos analistas de processos na etapa de modelagem? \***

- Concordo totalmente
- Concordo parcialmente
- Indiferente
- Discordo parcialmente
- Discordo totalmente

Fonte: Os autores

A estrutura da segunda etapa do questionário foi realizada da seguinte forma: inicialmente é apresentada a origem do texto e as atividades a serem desenvolvidas. Logo após, é apresentado o texto a ser modelado. Conforme Dumas et al. (2013), tal texto contém 9 sentenças a serem modeladas. Posteriormente, o participante responderia a quantidade de tempo gasto para realizar tal modelagem, uma vez que o objetivo do questionário é controlar o tempo de modelagem do participante. Por fim, o participante responderia o grau de dificuldade da modelagem, que era baseada na escala Likert de cinco pontos, va-

riando de fácil até difícil. Conforme a Figura 4.11 (1), os campos "*Quantidade de tempo gasto para realizar tal modelagem*" e "*Grau de Dificuldade*" eram obrigatórios.

A terceira etapa do questionário, semelhante a segunda etapa, visava obter um controle de tempo (em minutos) de modelagem e avaliar o nível de dificuldade de tal modelagem.

Nesta etapa, o texto de exemplo estava marcado conforme as regras de mapeamento.

A estrutura da terceira etapa do questionário foi realizada da seguinte forma: semelhante a segunda etapa, é apresentado a origem do texto e as atividades a serem implementadas. Em seguida, é apresentado o texto a ser modelado. Tal texto é similar ao da segunda etapa. Posteriormente, o participante responderia a quantidade de tempo consumido para realizar a modelagem do texto apresentado. Dado que, tais respostas iriam ser comparadas com as respostas da segunda etapa. Por fim, o participante responderia o grau de dificuldade da modelagem realizada, que era baseada na escala Likert de cinco pontos, variando de fácil até difícil. Conforme a Figura 4.11 (2), os campos "*Quantidade de tempo gasto para realizar tal modelagem*" e "*Grau de Dificuldade*" eram obrigatórios.

Por fim, a última etapa visava capturar a satisfação do usuário com a identificação dos elementos de processos no texto. Para isso, a Figura 4.12 apresenta os campos presentes no formulário:

1. *Qual das opções inclui mais detalhes de modelagem de processos?:* nesta questão, o participante informava quais das opções fornecia mais detalhes de modelagem de processos, conforme as seguintes opções: Texto 1 - Sem identificação ou Texto 2 - Com Identificação.
2. *Justifique sua resposta:* questão condicionada a questão 1. Visava obter as razões e motivos de selecionar qual texto fornecia mais detalhes de modelagem de processos.
3. *A identificação de elementos de processos de negócio no Texto 2 ajudou a modelar tais elementos descritos no texto?:* a partir da experiência com o texto marcado conforme as regras de mapeamento, o participante informava se com o seu uso, a identificação de elementos de processo, auxiliou a modelagem de processos do texto apresentado. Tal questão foi baseada na escala Likert de cinco pontos, variando de concordo totalmente até discordo totalmente.
4. *Você concorda que a identificação de elementos de processos de negócio em texto em linguagem natural auxiliará e minimizará o esforço necessário dos analistas de processos na etapa de modelagem?:* esta questão, visava obter a opinião do

participante em relação a identificação de elementos de processos de negócio em texto, e se tal identificação auxiliará e minimizará o esforço do analista na etapa de modelagem de processos em um projeto BPM. Esta questão foi baseada na escala Likert de cinco pontos, variando de concordo totalmente até discordo totalmente.

Ao finalizar essa etapa, o participante concluía a participação no questionário.

#### **4.1.4 Resultados Obtidos Através da Aplicação do Segundo Questionário**

A partir da aplicação do segundo questionário, foi obtida a participação dos usuários, que foram distribuídos entre estudantes de graduação e pós-graduação de cursos da área da computação, especialistas de processo, professores dentre outros.

As Figuras 4.13, 4.14 e 4.15, mostram os dados (21 participantes) relacionados das etapas do segundo questionário.

A fim de obter as informações dadas pelos especialistas de processo, uma vez que tal profissão era o público alvo deste questionário, foram analisadas as respostas de 12 participantes. Foram selecionados as seguintes características dos participantes:

- Especialista de processos;
- Experiência em BPMN;
- Mais de dois anos de experiência em BPMN;
- Experiência em modelagem de processos;
- Mais de dois anos de experiência em modelagem de processos.

Pelos mesmos motivos mencionados no primeiro questionário (ver seção 4.1.2), tais características influenciam nos resultados das modelagens. Portanto, tais participantes selecionados representam 57,14% do total de participantes do questionário.

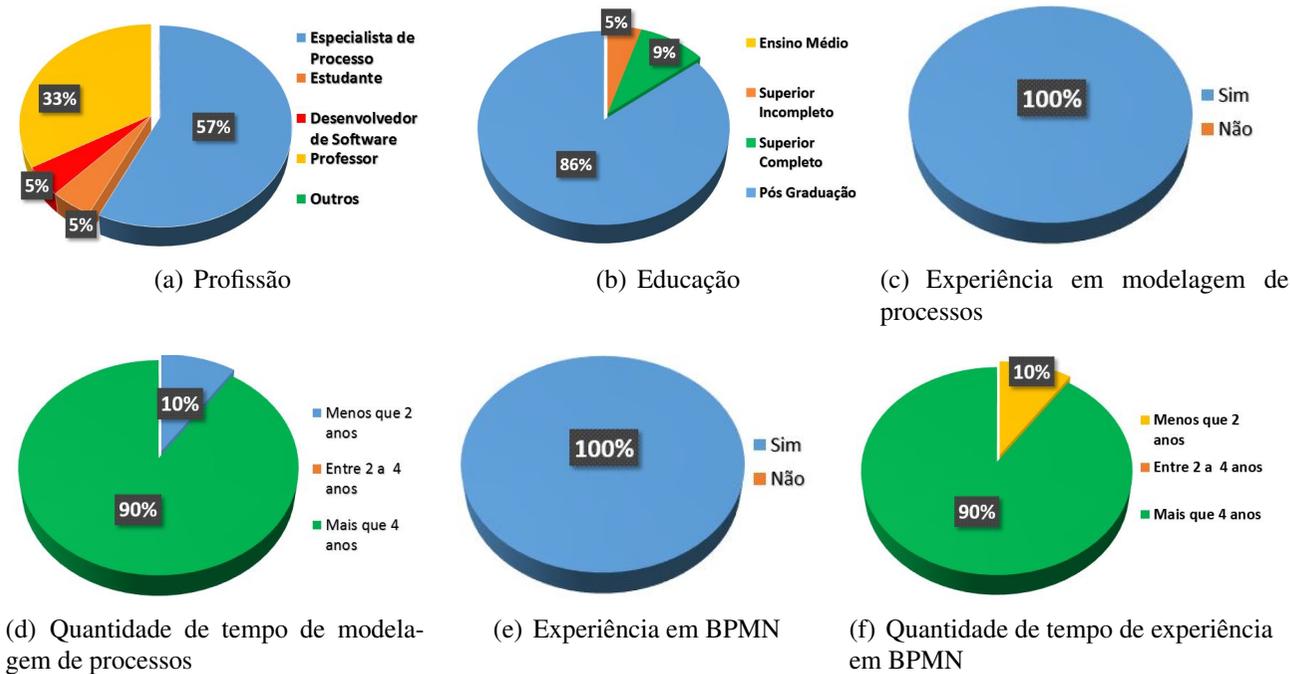
A avaliação do segundo questionário conduzida nesta dissertação, demonstrou resultados promissores. A fim de obter uma comparação de tempo de modelagem juntamente com o grau de dificuldade dos textos apresentados no questionário, a segunda etapa apresentou, em média, 17 minutos para a modelagem a partir do texto. Por outro lado, a terceira etapa apresentou, em média, 10 minutos para a modelagem. Dado que, tal etapa apresentava um texto com identificação de elementos de processo, conforme as regras de mapeamento propostas nesta dissertação. Portanto, a identificação de elementos de processo no texto minimizou 41% do tempo dos participantes selecionados. Na

segunda etapa, os participantes selecionados, relataram que o texto 1 (Figura 4.11/1) possuía, em média, 3,25 de nível de dificuldade. Contudo, a terceira etapa, os participantes, demonstraram que o texto 2 (Figura 4.11/2) possuía, em média, 1,33 de nível de dificuldade. Sendo que tais níveis foram baseados na escala Likert de cinco pontos, variando de fácil (1) até difícil (5).

A quarta etapa do questionário, também demonstra resultados promissores. Na primeira questão, que visava obter as opiniões dos participantes de quais dos textos incluíam mais detalhes na modelagem, mostra que 100% dos participantes selecionados, concordam que o texto 2 possuía mais detalhes para modelagem de processos. Na terceira questão, que visava se a identificação de elementos de processo no texto, auxiliou a modelagem de processos, 91,66% dos participantes concordam que identificação de elementos de processo nos textos realmente auxiliou a modelagem. Na quarta questão, que visava se a identificação de elementos de processo auxiliará e minimizará o esforço do analista na etapa de modelagem, 100% dos participantes concordaram que a identificação de elementos de processo realmente minimizará o esforço dos analistas.

Com base na aceitação dos usuários e as devidas opiniões em relação a identificação de elementos de processo em texto, pôde-se comprovar uma das hipóteses deste trabalho, que era "Tal abordagem auxiliará e minimizará o esforço necessário dos analistas de processo na etapa de modelagem do ciclo de vida de BPM".

Figura 4.13: Resultados obtidos a partir da primeira parte do segundo questionário.

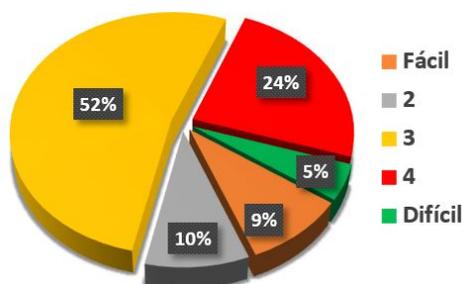


Fonte: Os autores. Nesta figura são apresentadas as avaliações dos 21 participantes do segundo questionário. A Figura 4.13(a), mostra que mais da metade dos participantes são especialistas de processo (57%), 5% são estudantes e desenvolvedores de *softwares* e, finalmente, 33% são professores. A Figura 4.13(b), ilustra que 86% dos participantes detém pós-graduação (mestrado e/ou doutorado), 9% detém graduação e 5% detém curso de ensino superior incompleto. A Figura 4.13(c), descreve que 100% dos participantes detém experiência em modelagem de processos. A Figura 4.13(d), mostra que 90% dos participantes possui mais de quatro anos de experiência em modelagem de processos e 10% possui menos de dois anos de experiência. A Figura 4.13(e), mostra que 100% dos participantes detém experiência em BPMN. A Figura 4.13(f), ilustra que 90% dos participantes possui mais de quatro anos de experiência em BPMN e 10% detém menos de dois anos.

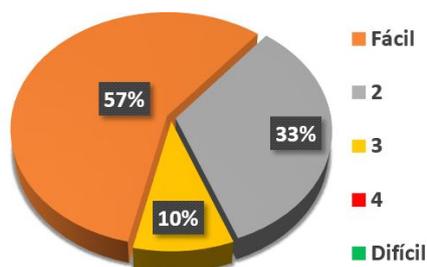
Figura 4.14: Resultados obtidos a partir da segunda e terceira parte do segundo questionário.

	Texto 1	Texto 2
Média (em minutos)	18	10

(a) Média de tempo de modelagem do Texto 1 e 2



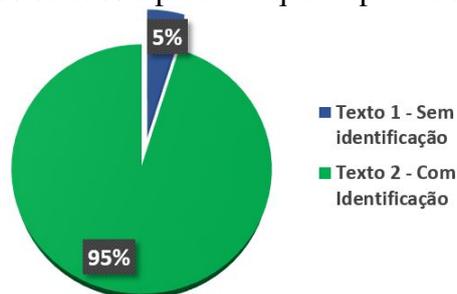
(b) Grau de dificuldade - Texto 1



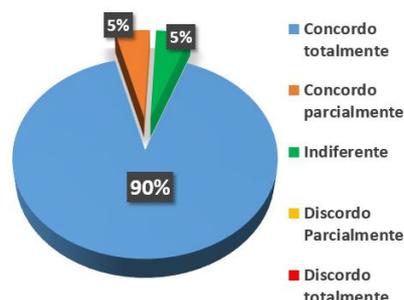
(c) Grau de dificuldade - Texto 2

Fonte: Os autores. Nesta figura são apresentadas as avaliações dos 21 participantes do segundo questionário. Os resultados da quantidade de tempo consumida para a modelagem dos textos mostrada na Figura 4.14(a), relatam que a modelagem do primeiro texto (sem identificação), consumiu 18 minutos dos participantes. Por outro lado, o segundo texto (com identificação), consumiu apenas 10 minutos dos esforços dos participantes. Portanto, houve 44,45% de economia de tempo da modelagem do segundo texto ao primeiro texto. Os resultados de grau de dificuldade do texto 1 mostrada na Figura 4.14(b), descrevem que 52% dos participantes assumiram que o texto 1 possuía nível 3 de dificuldade, 24% relataram que tal texto possuía nível 4, 10% assumiram que o texto 1 possuía nível 2, 9% mostraram que o nível do texto 1 era fácil e, finalmente, 5% disseram que o nível do texto 1 era difícil. A Figura 4.14(c) mostra que 57% dos participantes assumiram que o nível do texto 2 era fácil, 33% relataram que o texto 2 possuía nível 2 e, finalmente, 10% mostraram que o texto 2 possuía nível 3 de dificuldade.

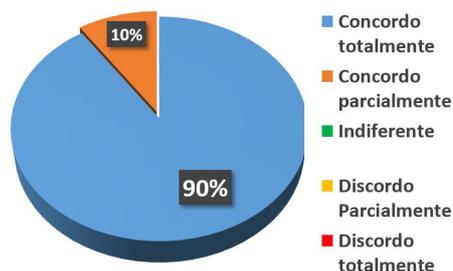
Figura 4.15: Resultados obtidos a partir da quarta parte do segundo questionário.



(a) Primeira Questão



(b) Terceira Questão



(c) Quarta Questão

Fonte: Os autores. Nesta figura são apresentadas as avaliações dos 21 participantes do segundo questionário. Os resultados da primeira da primeira questão mostrada na Figura 4.15(a), descrevem que 95% dos participantes concordam que o texto 2 possuía mais detalhes de modelagem de processos e 5% relatam que o texto 1 possuía mais detalhes. Sendo que o texto 2 apresentava a identificação de elementos de processo baseado nas regras de mapeamento propostas nesta dissertação. A Figura 4.15(b), ilustra que 90% dos participantes concordam que a identificação dos elementos de processo auxiliou a modelar os elementos descritos no texto 2 e, 5% concordam parcialmente e indiferente. A Figura 4.15(c), mostra que 90% dos participantes concordam que a identificação de elementos de processo de negócio em texto em linguagem natural auxiliará e minimizará o esforço dos analistas na etapa de modelagem de um projeto BPM e, 10% concordam parcialmente. Portanto, os resultados apresentam, em média, 91,66% de aceitação dos participantes.

## 4.2 Resultados Obtidos a partir do Desenvolvimento do Protótipo

Esta seção tem por intuito apresentar o protótipo desenvolvido para implementar a abordagem aqui apresentada.

A fim de identificar elementos de processos de negócio em texto em linguagem natural, foi implementado um protótipo para identificar tais elementos de forma semiautomática baseados nas regras de mapeamento propostas nesta dissertação.

Tal protótipo foi desenvolvido através da linguagem de programação *Python*. As razões pelas quais a linguagem foi selecionada foram apresentadas na seção 2.3.2. Foi utilizado um analisador sintático denominado *Spacy*, para identificar as relações de dependência e *parts of speech* das sentenças. A escolha do analisador baseia-se na acurácia realizada por Choi, Tetreault and Stent (2015). Os autores apresentam que o analisador obteve uma acurácia (*overall accuracy*) extremamente satisfatória em relação a testes realizados com outros analisadores. Em média, o analisador obteve uma acurácia de 90,53% de exatidão. Portanto, o *Spacy* suporta todos os requisitos e condições (por exemplo, identificação de *parts of speech* e dependência entre palavras) para o desenvolvimento do protótipo apresentado nesta dissertação.

Portanto, tal protótipo tem as seguintes características: identificar eventos iniciais, eventos finais, raias (*swimlanes*), ações, tarefas, rótulos de tarefas, rótulos de eventos, além de apresentar o número de elementos do processo do texto analisado.

O texto 1 a seguir, retirado de Dumas et al. (2013), foi selecionado como entrada de dados do protótipo. Este texto contém 9 sentenças. Conforme Dumas et al. (2013), a primeira sentença (1), representa um evento, a segunda (2), terceira (3), quarta (4) e quinta (5) sentenças equivalem a uma tarefa. A sexta (6) sentença, representa um desvio exclusivo (XOR) no modelo de processo. Em seguida, a sétima (7), oitava (8) e nona (9) sentenças, representam tarefas no modelo de processo extraído do texto.

Texto 1: *Once a loan application has been approved by the loan provider 1, an acceptance pack is prepared 2 and sent to the customer 3. The acceptance pack includes a repayment schedule which the customer needs to agree upon by sending the signed documents back to the loan provider 4. The latter then verifies the repayment agreement 5: if the applicant disagreed with the repayment schedule 6, the loan provider cancels the application 7. if the applicant agreed, the loan provider approves the application 8. In either case, the process completes with the loan provider notifying the applicant of the application status 9.*

Conforme a abordagem proposta nesta dissertação, o texto é separado manualmente por sentenças em um arquivo de formato *.txt*.

Posteriormente, tal texto é analisado sintaticamente pelo analisador. Desta forma, geram-se diversas sentenças marcadas. A partir destas sentenças, as regras de mapeamento são executadas, com o objetivo de identificar possíveis elementos de processo.

A primeira sentença do Texto 1, corresponde a regra 2 de identificação de eventos da Tabela 3.2. Tal regra resulta na sequência *Sujeito, Verbo, Agente e Objeto*. Desta forma, a saída do protótipo é ilustrada na Figura 4.16(a), no qual, o protótipo é capaz de identificar a raia (*lane*) e o rótulo (*label*) do elemento de processo.

A segunda e terceira sentenças identificadas no Texto 1, equivalem a regra 9 de identificação de atividades da Tabela 3.1. Tal regra, corresponde a sequência *Objeto, Verbo e Sujeito*. Deste modo, as saídas do protótipo são mostradas nas Figuras 4.16(b) e 4.17(a) respectivamente, no qual o protótipo identifica raia (*Lane*) e o rótulo (*label*) do elemento de processo.

Figura 4.16: Primeira e Segunda Sentenças Identificadas no Texto 1

```
Indicação de Lane: the loan provider
Indicação de Label:a loan application approved
-----
Indica-se um Evento
-----
```

(a) Primeira Sentença Identificada.

```
Sentença:an acceptance pack is prepared
-----
Indicação de Lane: the loan provider
Indicação de Label: Prepare Acceptance Pack
Indica-se uma Tarefa
-----
```

(b) Segunda Sentença Identificada.

Fonte: Os autores

A quarta e quinta sentenças identificadas no Texto 1, condizem com a regra 1 de identificação de atividades da Tabela 3.1. Tal regra, corresponde a sequência *Sujeito, Verbo e Objeto*. Desta forma, as saídas do protótipo são ilustradas nas Figuras 4.17(b) e 4.17(c) respectivamente, em que o protótipo identifica raia e o rótulo do elemento de processo.

Figura 4.17: Terceira, Quarta e Quinta Sentenças Identificadas no Texto 1

```
Sentença:sent to the customer
-----
Indicação de Lane: the loan provider
Indicação de Label: Sent Acceptance Pack
Indica-se uma Tarefa
-----
```

(a) Terceira Sentença Identificada.

```
Sentença:The acceptance pack includes a repayment
schedule which the customer needs to agree upon by sending
the signed documents back to the loan provider.
-----
Indicação de Lane: Customer
Indicação de Label: : Send Signed Documents
Indica-se uma Tarefa
-----
```

(b) Quarta Sentença Identificada.

```
Sentença:The latter then verifies the repayment agreement.
-----
Indicação de Lane: The latter
Indicação de Label: : Verify Repayment
Indica-se uma Tarefa
-----
```

(c) Quinta Sentença Identificada.

Fonte: Os autores

A sexta e sétima sentenças identificadas no Texto 1, correspondem a regra 2 de desvio exclusivo (XOR) da Tabela 3.3. Tal regra corresponde a sequência de *palavra-chave*, seguido de uma condição e uma ação (tarefa e evento). A saída do protótipo é apresentada na Figura 4.18, no qual, o protótipo é capaz de identificar a condição e a ação (tarefa ou evento) dos elementos de processo.

Figura 4.18: Sexta e Sétima Sentença Identificada

```
Sentença:if the applicant disagreed with the repayment schedule,
the loan provider cancels the application.
-----
Condicao: if the applicant disagreed with the repayment schedule
Tarefa/Evento: the loan provider cancels the application.
Indica-se um XOR Split
-----
```

Fonte: Os autores

A oitava e nona sentenças do Texto 1, corresponde, respectivamente as regras 2 e 7 de identificação de desvios exclusivos (XOR) da Tabela 3.3. As saídas do protótipo são ilustradas nas Figura 4.19(a) e 4.19(b) respectivamente, em que o protótipo identifica as raias e os rótulos dos elementos de processo.

Figura 4.19: Oitava e Nona Sentenças Identificadas no Texto 1

```

Sentença:if the applicant agreed, the loan
provider approves the application
-----
Indicação de Lane: the loan provider
Indicação de Label: : Approve Application
Indica-se uma Tarefa
-----

```

(a) Oitava Sentença Identificada.

```

Sentença:The process completes with the loan
provider notifying the applicant of the application status.
-----
Indicação de Lane: the loan provider
Indicação de Label: Notify Application
Indica-se uma Tarefa
-----

```

(b) Nona e última Sentença Identificada.

Fonte: Os autores

Portanto, o objetivo do protótipo foi permitir o estudo e viabilizar a validação da abordagem, uma vez que o mesmo foi baseado nas regras de mapeamento.

Na próxima seção, serão apresentados os resultados que foram obtidos com o protótipo baseados em métricas de desempenho.

#### 4.2.1 Avaliação de Desempenho

A fim de avaliar a eficiência do protótipo desenvolvido nesta dissertação, a seção 4.2.1.1, apresenta as métricas que foram selecionadas para medir o desempenho. Em seguida, a seção 4.2.1.2, mostra o conjunto de dados escolhidos como entradas de informações do protótipo. Por fim, são apresentados os resultados de desempenho do protótipo na seção 4.2.1.3.

##### 4.2.1.1 Métricas Selecionadas

Para definir um processo de avaliação para sistemas de recuperação de informação é fundamental um esforço organizacional, além de reunir as partes interessadas e desenvolver os recursos utilizados durante a avaliação. Além disso, é também necessário, estabelecer o que se deseja medir e, por conseguinte, as métricas a utilizar e a metodologia de acordo com essas métricas, tais como: como as medições são feitas, se a forma de julgamento será manual, automática ou ambas (AIRES et al., 2003; BAILEY; CRASWELL; HAWKING, 2003).

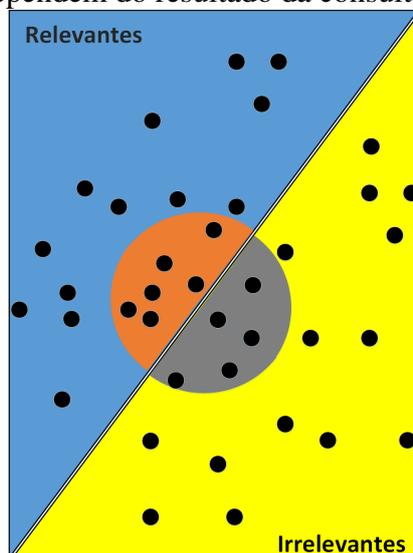
A Figura 4.20 representa o conjunto de informações de uma coleção de documentos sobre a doença de malária, discriminadas em informações relevantes (esquerda/azul e

laranja) e informações irrelevantes (direita/amarelo e cinza). O formato circular (laranja e cinza) significa o resultado da consulta sobre informações da doença.

A consulta ideal e desejável é quando for recuperada toda a parte esquerda da figura (azul e laranja), tendo assim 100% de precisão e revocação.

A matriz de confusão resume o desempenho de classificação de um classificador em relação ao conjunto de dados (*data sets*) (FORBES, 1995). Além disso, apresenta uma medida concreta do modelo de classificação, ao revelar o resultado de classificações corretas e preditas para cada classe em um conjunto de dados (TING, 2010). Nesta dissertação tais classes foram definidas como: atividades, eventos, desvio exclusivo (XOR) e desvio paralelo (AND).

Figura 4.20: Medidas dependem do resultado da consulta (circular/laranja e cinza)



Fonte: Os autores

A Figura 4.21(a) é apresentada um exemplo de matriz de confusão, contendo 3 classes: *A*, *B*, e *C*. Tais classes são fictícias para ilustrar este exemplo.

A primeira linha da matriz indica que 8 objetos foram classificados na classe *A*, sendo que 5 foram classificados corretamente como classe *A*, três foram identificados incorretamente como *B* e nenhum foi classificado como *C*.

A matriz de confusão é geralmente utilizada com duas classes, sendo que uma para designar a classe positiva e a outra para a classe negativa. Neste contexto, as quatro células da matriz são identificadas como Verdadeiro Positivos (TP), Falso Positivos (FP), Verdadeiro Negativos (TN), e Falso Negativos (FN), como indicado na Figura 4.21(b).

O resultado da matriz de confusão para a classe *A* do exemplo apresentado na Figura 4.21(a), é mostrado na Figura 4.21(c).

Figura 4.21: Exemplos de Matrizes de Confusão

		Classe Predita		
		A	B	C
Classe Atual	A	5	3	0
	B	2	3	1
	C	0	2	11

(a) Matriz de Confusão de 3 Classes: A, B, e C

		Classe Predita	
		Positivo	Negativo
Classe Atual	Positivo	TP	TN
	Negativo	FP	FN

(b) Matriz de Confusão com classes positivas e negativas.

<b>5 - Verdadeiro Positivos</b>	<b>17 - Verdadeiro Negativos</b>
<b>2 - Falso Positivos</b>	<b>3 - Falso Negativos</b>

(c) Resultado da matriz de confusão da classe A.

Fonte: Os autores

A Figura mostra que 5 objetos da classe A, foram classificados corretamente como A. Na célula de falso negativos, 3 objetos foram classificados incorretamente como B. Em seguida, 2 objetos (B), foram rotulados incorretamente como A. Por fim, na célula de verdadeiro negativos, mostra que os demais objetos, foram classificados corretamente como *não-A* (17 objetos).

Existem diversas métricas de classificações de desempenho que são definidas em termos destas quatro saídas, tais como precisão, revocação, medida-F, acurácia etc. Nesta dissertação, foram utilizadas as seguintes métricas:

- **Precisão:** A quantidade com que todos os exemplos classificados como positivos são realmente positivos. Além disso, nenhum exemplo negativo é inserido. A equação 4.1 ilustra esta métrica.
- **Revocação (*recall*):** Classifica como positivo todos os exemplos que são positivos. Nenhum exemplo positivo é desperdiçado. Tal métrica, mostra a quantidade de dados relevantes que foram recuperados. A equação 4.2 apresenta esta métrica.
- **Medida-F (*F-measure*):** Média harmônica ponderada da precisão e revocação. A equação 4.3 apresenta esta métrica.

$$Precisão = \frac{TP}{TP + FP} \quad (4.1)$$

$$Revocação = \frac{TP}{TP + FN} \quad (4.2)$$

$$Medida - F = \frac{2 \times Precisão(i) \times Recall(i)}{Precisão(i) + Recall(i)} \quad (4.3)$$

Portanto, os resultados do exemplo da Figura 4.21(a) para a classe A, são apresentados nas equações 4.4, 4.5 e 4.6.

$$Precisão(A) = \frac{TP}{TP + FP} = \frac{5}{5 + 2} = 0,714 = 71\% \quad (4.4)$$

$$Revocação(A) = \frac{TP}{TP + FN} = \frac{5}{5 + 3} = 0,625 = 63\% \quad (4.5)$$

$$Medida - F = \frac{2 \times Precisão(A) \times Recall(A)}{Precisão(A) + Recall(A)} = \frac{2 \times 0,714 \times 0,625}{0,714 + 0,625} = 66\% \quad (4.6)$$

A equação 4.4, mostra que a classe A obteve 71% de precisão, isto é, os objetos que foram classificados como positivos foram realmente positivos. A equação 4.5, ilustra que classe A, atingiu 63% de revocação. Isto significa que a classe A, mostra uma completude de 63%. Por fim, a equação 4.6, apresenta que a classe A, obteve 66% da medida-F. Visto que tal métrica atinge seu melhor valor em 100% e o pior em 0%. Por comodidade os resultados das métricas, serão apresentados como porcentagens.

Portanto, a precisão e revocação são métricas utilizadas para avaliar a qualidade dos resultados em diversas áreas de pesquisas. Desta forma, tais métricas são padrões na área de recuperação de informação (RI), e são utilizadas para contribuir com a avaliação de sistemas de RI que tem o objetivo de recuperar documentos relevantes a partir da consulta de um usuário, porém diversas outras áreas, como PLN, IA (JAPKOWICZ; SHAH, 2011) incluindo aprendizagem de máquina e extração de informação, utilizam destas métricas para avaliação (MANNING; RAGHAVAN; SCHÜTZ, 2008).

#### 4.2.1.2 Conjunto de Dados Selecionados

A fim de obter um maior conjunto de informações para entrada de dados do protótipo desenvolvido no contexto desta dissertação, foram selecionados 70 textos em linguagem natural. Dado que tais textos variam com relação a diversas dimensões, tais como:



(2010) e Wang, Zhao and Zhang (2009).

Inicialmente, as sentenças analisadas manualmente, foram classificadas conforme as classes definidas (atividades, eventos, desvio exclusivo e desvio paralelo) para medição. Tais classificações foram realizadas por dois especialistas de processos. A princípio, o primeiro especialista classificava cada sentença conforme sua experiência em modelagem de processos. Posteriormente, o segundo especialista, analisava se tal sentença (classificada anteriormente pelo primeiro especialista), representa realmente aquela classe definida. Tal procedimento é chamado de classes preditas na matriz de confusão, como indicado na Figura 4.21(a).

Posteriormente, o protótipo analisa cada sentença (instância) conforme as regras de mapeamento. Sendo assim, tais sentenças são analisadas e classificadas por classes.

Por fim, gera-se a matriz de confusão de cada classe e seus respectivos critérios, tais como: Verdadeiro Positivos (TP), Falso Positivos (FP), Verdadeiro Negativos (TN) e Falso Negativos (FN).

A partir destes critérios, pode-se medir a precisão, revocação e medida-F de cada classe, conforme as equações 4.1, 4.2 e 4.3, respectivamente.

As Tabelas 4.1, 4.2, 4.3 e 4.4, apresentam os resultados das métricas de cada classe.

Tabela 4.1: Resultados de desempenho do protótipo para a classe de atividades.

<b>Classe</b>	<b>Encontradas</b>	<b>Precisão</b>	<b>Revocação</b>	<b>Medida-F</b>
Atividades	140	75,48%	70%	73,12%

Fonte: Os autores

Para a classe de atividades, a Tabela 4.1, apresenta os resultados de desempenho. Tal classe atingiu 75,48% de precisão, 70% de revocação e 73,12% de medida-F.

Tabela 4.2: Resultados de desempenho do protótipo para a classe de eventos.

<b>Classe</b>	<b>Encontradas</b>	<b>Precisão</b>	<b>Revocação</b>	<b>Medida-F</b>
Eventos	106	81,81%	72,26%	76,74%

Fonte: Os autores

Tabela 4.3: Resultados de desempenho do protótipo para a classe de desvio exclusivo (XOR).

<b>Classe</b>	<b>Encontradas</b>	<b>Precisão</b>	<b>Revocação</b>	<b>Medida-F</b>
Desvio Exclusivo (XOR)	98	65,47%	73,95%	69,60%

Fonte: Os autores

Tabela 4.4: Resultados de desempenho do protótipo para a classe de desvio paralelo (AND).

Classe	Encontradas	Precisão	Revocação	Medida-F
Desvio Paralelo (AND)	43	71,69%	64,40%	67,85%

Fonte: Os autores

Para a classe de eventos, a Tabela 4.2, ilustra os resultados de desempenho. Ao analisar os resultados para tal classe, observa-se que obteve 81,81% de precisão, 72,26% de revocação e 76,74 de medida-F.

Para a classe de desvio exclusivo, a Tabela 4.2, mostra os resultados de desempenho. Tal classe atingiu 65,47% de precisão, 73,95% de revocação e 69,60% de medida-F.

Por fim, a classe de desvio paralelo, a Tabela 4.4, ilustra os resultados de desempenho. Sendo que, tal classe obteve 71,69% de precisão, 64,40% de revocação e 67,85% de medida-F.

### 4.3 Discussão dos Resultados

Este capítulo teve como foco aplicar técnicas que pudessem verificar e validar o trabalho desenvolvido. Buscou-se através destes, verificar se os objetivos específicos deste trabalho foram alcançados. Para isso, foi feito o uso de estrutura de questionário para proporcionar aos participantes interação com o trabalho desenvolvido, visando obter retorno quanto a satisfação com os mesmos.

A estrutura de questionário utilizada permitiu que os participantes pudessem aplicar as regras de mapeamento, bem como capturar satisfação com o uso.

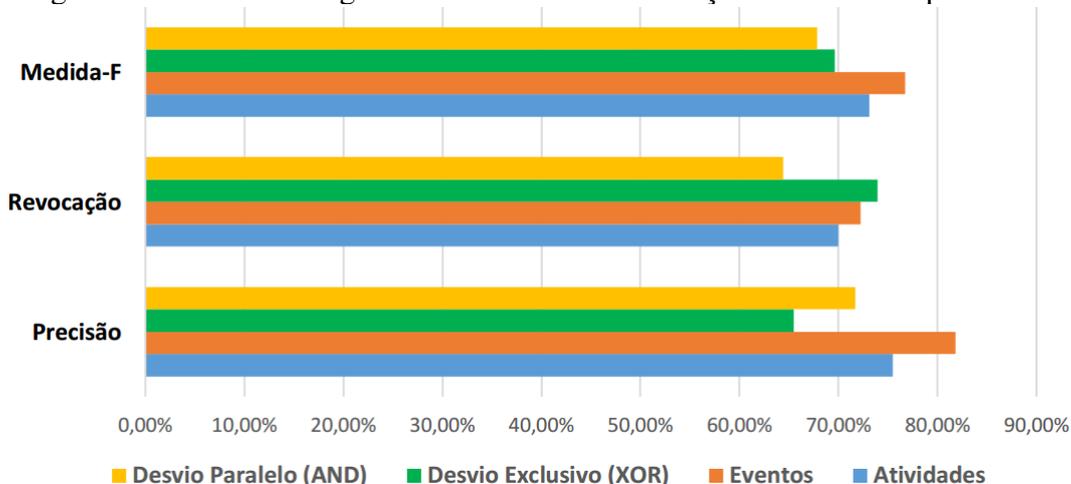
Dos resultados obtidos através do primeiro questionário, destacam-se que as modelagens realizadas incorretas, isto é, fragmentos de processo que não foram modelados conforme as regras de mapeamento, obtiveram uma reprovação significativa. Por outro lado, as modelagens realizadas conforme as regras de mapeamento, atingiram uma boa aceitação dos participantes. Portanto, considera-se plausível a utilização das regras de mapeamento para identificar elementos de processo em texto.

Dos resultados obtidos através do segundo questionário, destaca-se que as regras de mapeamento tiveram aceitação com o público, sendo que 90% acharam que auxiliou na modelagem. Além disso, os resultados mostram que a identificação de elementos de processo em texto minimizou 44,45% do tempo dos participantes em relação ao texto sem identificação.

Como principal limitação dos questionários propostos, destaca-se a quantidade de regras de mapeamento apresentadas aos participantes (9 regras). Isto equivale a 28% das regras de mapeamento propostas nesta dissertação. Em trabalhos futuros, considera-se a possibilidade de apresentar as demais regras de mapeamento aos participantes através de questionários, bem como obter as devidas opiniões e satisfações das mesmas. Outra possibilidade de trabalhos futuros é dividir um grupo de pessoas em duas turmas com o mesmo texto a ser modelado, sendo que a primeira turma modele processos sem o uso das regras de mapeamento e a segunda turma modele processos com o uso das regras de mapeamento. Neste caso, tal procedimento poderia minimizar a influência do uso das regras de mapeamento no mesmo grupo de pessoas.

O protótipo desenvolvido no contexto desta dissertação, permitiu avaliar o desempenho através de métricas de recuperação de informação. Apesar da quantidade de sentenças analisadas serem baixas, foi observado através da avaliação das métricas, resultados muito promissores, em termos de precisão, revocação e medida-F. Em média, obteve 73,61%, 70,15% e 71,82% respectivamente.

Figura 4.22: Resultados gerais das métricas de avaliação classificadas por classes



Fonte: Os autores

Portanto, os melhores resultados, em termos de precisão, foram as classes de eventos e atividades, respectivamente. Isto significa que, as sentenças que foram classificadas com tais classes foram realmente representadas por essas classes. Já em termos de revocação, as classes que obtiveram melhores resultados, foram desvio exclusivo e eventos, respectivamente. Isto é, tais classes apresentam uma boa completude em relação as demais. Com base na medida-F, as classes que atingiram resultados superiores foram eventos e atividades, respectivamente. A Figura 4.22, apresenta uma visão geral dos resultados.

## 5 CONCLUSÃO

Neste capítulo, as considerações finais são apresentadas, com ênfase nos objetivos alcançados e resultados obtidos através do desenvolvimento desta dissertação. Além disso, são discutidas as contribuições científicas geradas por este trabalho. Por fim, são expostas as perspectivas de trabalhos futuros e publicações científicas realizadas durante o mestrado.

Este trabalho traz contribuições para a identificação de elementos de processos em texto em linguagem natural. Tal identificação pode auxiliar o analista de processo na tarefa de modelagem de processos em um projeto BPM. Sendo assim, foram levantados os problemas motivadores deste trabalho, a grande variedade de documentos textuais existentes nas organizações e documentos que contém palavras-chave e sentenças que denotam elementos de processo, desta forma dificultam a extração por parte dos analistas iniciantes. Com base nos problemas levantados, foi apresentada uma abordagem para identificar elementos de processo em texto em linguagem natural de forma semiautomática. Para avaliar a abordagem foram utilizados questionários e métricas de desempenho.

Desta forma, os seguintes objetivos específicos foram alcançados:

- *Propor regras de mapeamento para identificar elementos de processo de negócio em texto em linguagem natural:* o desenvolvimento das regras de mapeamento foi o caminho utilizado para identificar elementos de processo de negócio em texto em linguagem natural. Tais regras de mapeamento foram baseadas na BPMN versão 2.0.2.
- *Implementar um protótipo fundamentado em técnicas e ferramentas de PLN e baseado nas regras de mapeamento para identificar elementos de processo de negócio de forma semiautomática:* a implementação de um protótipo viabilizou a identificação de elementos de processo de negócio em texto em linguagem natural de forma semiautomática. A implementação foi desenvolvida pela linguagem de programação *Python*. Além disso, a análise sintática dos textos foi realizada através do analisador sintático denominado *Spacy*, e a verificação de sinônimos de palavras foi realizada pelo dicionário *PyDictionary*.
- *Avaliar através de pesquisa de opinião de possíveis usuários as considerações quanto as regras de mapeamento criadas e aplicabilidade delas em texto:* A aplicação dos questionários visava obter a satisfação dos participantes em relação as regras de mapeamento, bem como a aplicabilidade delas em texto. Os resultados

do primeiro e segundo questionário apresentam, em média, 91,66% de aceitação dos participantes.

- *Medir o desempenho do protótipo implementado por meio de métricas de recuperação de informação, tais como a precisão, revocação e medida-F:* A avaliação do protótipo desenvolvido, visava medir o desempenho através das métricas de precisão, revocação e medida-F. Em média, tal protótipo obteve 73,61% de precisão, 70,15% de revocação e 71,82% de medida-F.

As principais contribuições deste trabalho, são:

- Definição de regras de mapeamento entre texto em linguagem natural e elementos de processo de negócio baseados na BPMN. Foi possível identificar elementos de processo através de classes gramaticais e dependência de palavras (por exemplo, sujeito, objeto indireto, objeto direto etc).
- Implementação de um protótipo, cuja finalidade era identificar elementos de processo de negócio em texto em linguagem natural de forma semiautomática. A intenção de desenvolver um protótipo no contexto desta dissertação, foi permitir o estudo e viabilizar a validação da abordagem.
- Geração de texto orientado a processo. Foi observado que este trabalho, é o primeiro passo para gerar tal texto, possibilitando assim, a extração de processos a partir dele.

Apesar dos resultados promissores, esta abordagem apresenta limitações, tal como as regras de mapeamento, que não abrangem todos os elementos de processo do conjunto de elementos reduzidos da BPMN. Sendo assim, apresenta oportunidades para trabalhos futuros. Outra limitação se refere a possibilidade de omissão na identificação dos elementos de processo, devido ao extenso número de palavras da língua inglesa. Em vista disso, o analista se encarregaria de realizar a decisão do elemento de processo na sentença. Uma outra questão se relaciona a ambiguidades presentes nos textos, sendo que, nesta abordagem, tal texto é tratado por sentenças, e conseqüentemente poderá haver viés de interpretação na identificação dos elementos. Por outro lado, existem diversos trabalhos que realizam tal análise por sentenças, por exemplo, Jiexun et al. (2010) e Wang, Zhao and Zhang (2009). Além disso, existem diversos trabalhos que tratam de inconsistências de modelos de processo a partir de descrições textuais, por exemplo, Han, Leopold and Reijers (2015) e Han, Henrik and A. (2016).

As principais vantagens deste trabalho são que a partir do momento em que se tem

um texto identificado com os devidos elementos de processo, é possível que os analistas de processo, possam ter um ponto de partida para o desenvolvimento da modelagem de processos. Porém o trabalho aqui apresentado, não visa substituir o analista, e sim assisti-lo, para que as tomadas de decisões sejam exclusivamente dele. Portanto, tal abordagem auxiliará o analista da tarefa de modelagem de processos.

Como trabalhos futuros foram identificadas as seguintes possibilidades:

- Para minimizar possível viés de interpretação (sentenças que denotam o mesmo elemento de processo), faz-se necessário utilizar análise semântica de PLN. Em vista disso, uma alternativa para minimizar ambiguidades no texto, é apresentada em Yarowsky (1992) e Sussna (1993). Outra possibilidade é a utilização do analisador semântico *WordNet* (MILLER, 1995).
- Desenvolver meios de gerar regras de mapeamento para identificar elementos de processo de forma automática. Utilizar fundamentos de inteligência artificial (reconhecimento de padrões), para identificar possíveis padrões de identificação de elementos de processo em texto em linguagem natural. Desta forma, poderão ser desenvolvidas quantidades relevantes de regras de mapeamento a partir dos padrões identificados, e conseqüentemente, possíveis omissões na identificação serão minimizadas.
- Com base no protótipo desenvolvido neste trabalho, pode-se implementar outros módulos de identificação, incluindo principalmente, uma quantidade de regras de mapeamento significativas para identificação dos elementos de processo em texto.
- A partir da abordagem apresentada nesta dissertação, observa-se o primeiro passo para o desenvolvimento de uma abordagem para gerar texto orientado a processo a partir de texto em linguagem natural. Espera-se gerar um modelo (*template*) de como o texto deve ser estruturado para extração de modelos de processo a partir dele.
- Apesar do uso de particularidades da língua inglesa neste trabalho, pode-se realizar uma possível integração para outros idiomas. A ferramenta de PLN, *Stanford Parser* pode ser utilizada, uma vez que tal ferramenta suporte idiomas, tais como alemão (RAFFERTY; MANNING, 2008), chinês (LEVY; MANNING, 2003) e árabe (GREEN; MANNING, 2010).

## 5.1 Publicações

Esta seção apresenta as produções científicas relacionadas a esta dissertação, em ordem cronológica. O primeiro artigo publicado discutiu a geração de texto orientado a processo a partir de texto em linguagem natural. O artigo apresentou uma visão inicial da abordagem apresentada nesta dissertação.

- FERREIRA, RENATO CÉSAR BORGES e THOM, LUCINÉIA HELOISA. (2016), **Uma Abordagem para Gerar Texto Orientado a Processo a partir de Texto em Linguagem Natural**, XII Brazilian Symposium on Information Systems, pp-77.

Além disso, um artigo foi publicado para a conferência *International Conference on Enterprise Information Systems*. O artigo discute a identificação de elementos de processo de negócio em texto em linguagem natural, relacionando a discussão apresentada no capítulo 3 desta dissertação.

- FERREIRA, RENATO CÉSAR BORGES e THOM, LUCINÉIA HELOISA e FANTINATO, MARCELO. (2017), **A Semi-Automatic Approach to Identify Business Process Elements in Natural Language Texts**, International Conference on Enterprise Information Systems.

Esta dissertação também originou um trabalho de conclusão de curso:

- BORDIGNON, ANA CLÁUDIA DE ALMEIDA. (2016), **A Systematic Literature Review on Natural Language Processing in Business Process Identification and Modeling**. Instituto de Informática. UFRGS. Lucinéia H. Thom (orientadora) e Renato César Borges Ferreira (co-orientador).

## REFERÊNCIAS

- AALST van der. Business process management: A comprehensive survey. **ISRN Software Engineering**, Hindawi Publishing Corporation, v. 2013, 2013.
- AALST, W. M. P. van der; HOFSTEDÉ, A. H. M. ter. Yawl: Yet another workflow language. **Inf. Syst.**, Elsevier Science Ltd., Oxford, UK, UK, v. 30, n. 4, p. 245–275, jun. 2005. Available from Internet: <<http://dx.doi.org/10.1016/j.is.2004.02.002>>.
- AIRES, R. et al. An initial proposal for cooperative evaluation on information retrieval in portuguese. In: \_\_\_\_\_. **Computational Processing of the Portuguese Language: 6th International Workshop, PROPOR 2003 Faro, Portugal, June 26–27, 2003 Proceedings**. Berlin Heidelberg: Springer Berlin Heidelberg, 2003. p. 227–234.
- ALLEN, J. **Natural Language Understanding**. Redwood City, CA, USA: Benjamin-Cummings Publishing Co., Inc., 1995.
- BAILEY, P.; CRASWELL, N.; HAWKING, D. Engineering a multi-purpose test collection for web retrieval experiments. **Inf. Process. Manage.**, Pergamon Press, Inc., Tarrytown, NY, USA, v. 39, n. 6, p. 853–871, nov. 2003. Available from Internet: <[http://dx.doi.org/10.1016/S0306-4573\(02\)00084-5](http://dx.doi.org/10.1016/S0306-4573(02)00084-5)>.
- BIRD, S.; KLEIN, E.; LOPER, E. **Natural Language Processing with Python**. 1st. ed. [S.l.]: O'Reilly Media, Inc., 2009.
- BLUMBERG, R.; ATRE., S. The problem with unstructured data. In: . [S.l.]: DM Review, 2003.
- BORDIGNON, A. C. D. A. **A systematic literature review on Natural Language Processing in Business Process Identification and Modeling**. 62 p. Monografia (Graduação) — Universidade Federal do Rio Grande do Sul, 2016.
- BRILL, E. A simple rule-based part of speech tagger. In: **Proceedings of the Third Conference on Applied Natural Language Processing**. Stroudsburg, PA, USA: Association for Computational Linguistics, 1992. (ANLC '92), p. 152–155. Available from Internet: <<http://dx.doi.org/10.3115/974499.974526>>.
- BRINTON, L. **The Structure of Modern English: A Linguistic Introduction**. [S.l.]: John Benjamins Pub, 2000. (The Structure of Modern English: A Linguistic Introduction, v. 1).
- BRISCOE, T.; CARROLL, J.; WATSON, R. The second release of the rasp system. In: **Proceedings of the COLING/ACL on Interactive Presentation Sessions**. Stroudsburg, PA, USA: Association for Computational Linguistics, 2006. (COLING-ACL '06), p. 77–80. Available from Internet: <<http://dx.doi.org/10.3115/1225403.1225423>>.
- BROCKE, J. vom; ROSEMANN, M. **Handbook on Business Process Management 1: Introduction, Methods, and Information Systems**. [S.l.: s.n.], 2010.
- CHAPMAN, R. **Roget's International Thesaurus**. [S.l.]: Crowell, 1977. (Harper Colophon Books).

CHOI, J. D.; PALMER, M. **Guidelines for the Clear Style Constituent to Dependency Conversion**. [S.l.], 2012.

CHOI, J. D.; TETREAUULT, J. R.; STENT, A. It depends: Dependency parser comparison using A web-based evaluation tool. In: **Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, ACL 2015, July 26-31, 2015, Beijing, China, Volume 1: Long Papers**. [S.l.: s.n.], 2015.

CHUENG, A.; KOLIADIS, G.; GHOSE, A. Process discovery from model and text artefacts. **2007 IEEE Congress on Services**, IEEE Computer Society, Los Alamitos, CA, USA, v. 00, n. undefined, p. 167–174, 2007.

DAVIS, R.; BRABANDER, E. **ARIS Design Platform: Getting Started with BPM**. Springer London, 2007. Available from Internet: <<https://books.google.com.br/books?id=yZbExAf6jU0C>>.

DUMAS, M. et al. **Fundamentals of Business Process Management**. [S.l.]: Springer, 2013. I-XXVII 1-399 p.

DUMAS, M. et al. **Understanding Business Process Models: The Costs and Benefits of Structuredness**. 2012.

EDWARDS, A. **Techniques of attitude scale construction**. Appleton-Century-Crofts, 1957. (Century psychology series). Available from Internet: <[https://books.google.pt/books?id=V\\\_EDAQAIAAJ](https://books.google.pt/books?id=V\_EDAQAIAAJ)>.

FANTINATO, M. et al. A survey on reuse in the business process management domain. **International Journal of Business Process Integration and Management**, v. 6, n. 1, p. 52–76, 2012.

FERREIRA, R. C. B.; THOM, L. H. Uma abordagem para gerar texto orientado a processo a partir de texto em linguagem natural. In: . [S.l.]: XII Brazilian Symposium on Information Systems, 2016. p. 77.

FIGL, K.; RECKER, J.; MENDLING, J. A study on the effects of routing symbol design on process model comprehension. **Decision Support Systems**, v. 54, n. 2, p. 1104–1118, 2013. Available from Internet: <<http://dx.doi.org/10.1016/j.dss.2012.10.037>>.

FORBES, A. D. Classification-algorithm evaluation: Five performance measures based on confusion matrices. **Journal of Clinical Monitoring**, v. 11, n. 3, p. 189–206, 1995. Available from Internet: <<http://dx.doi.org/10.1007/BF01617722>>.

FRIEDRICH, F.; MENDLING, J.; PUHLMANN, F. Process model generation from natural language text. *Advanced Information Systems Engineering*, p. 482–496, 2011.

GASSEN, J. B. et al. Business process modeling: Vocabulary problem and requirements specification. In: **Proceedings of the 32Nd ACM International Conference on The Design of Communication CD-ROM**. New York, NY, USA: ACM, 2014. (SIGDOC '14), p. 2:1–2:10. Available from Internet: <<http://doi.acm.org/10.1145/2666216.2666217>>.

GONÇALVES, J. C. A.; SANTORO, F. M.; BAIAO, F. A. Let me tell you a story - on how to build process models. In: . [S.l.]: Journal of Universal Computer Science, 2011. v. 17, n. 2, p. 276–295.

GREEN, S.; MANNING, C. D. Better arabic parsing: Baselines, evaluations, and analysis. In: **Proceedings of the 23rd International Conference on Computational Linguistics**. Stroudsburg, PA, USA: Association for Computational Linguistics, 2010. (COLING '10), p. 394–402. Available from Internet: <<http://dl.acm.org/citation.cfm?id=1873781.1873826>>.

HAN, v. d. A.; LEOPOLD, H.; REIJERS, H. A. Detecting inconsistencies between process models and textual descriptions. In: SPRINGER. **International Conference on Business Process Management**. [S.l.], 2015. p. 90–105.

HAN van der A.; HENRIK, L.; A., R. H. Dealing with behavioral ambiguity in textual process descriptions. In: \_\_\_\_\_. **Business Process Management: 14th International Conference -BPM 2016- Rio de Janeiro- Brazil- September 18-22- 2016. Proceedings**. Cham: Springer International Publishing, 2016.

HERBST, J. An inductive approach to the acquisition and adaptation of workflow models. In: **Proceedings of the IJCAI'99 Workshop on Intelligent Workflow and Process Management: The New Frontier for AI in Business**. [S.l.: s.n.], 1999. p. 52–57.

INDURKHYA, N.; DAMERAU, F. J. **Handbook of Natural Language Processing**. 2nd. ed. [S.l.]: Chapman & Hall/CRC, 2010.

JAPKOWICZ, N.; SHAH, M. **Evaluating Learning Algorithms: A Classification Perspective**. New York, NY, USA: Cambridge University Press, 2011.

JIEXUN, L. et al. A policy-based process mining framework: mining business policy texts for discovering process models. **Information Systems and E-Business Management**, Springer, 2010.

JOHN, L. Language. *Linguistic Society of America*, v. 69, n. 4, p. 825–828, 1993.

JURAFSKY, D.; MARTIN, J. H. **Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition**. 1st. ed. Upper Saddle River, NJ, USA: Prentice Hall PTR, 2000.

JURAFSKY, D.; MARTIN, J. H. **Speech and Language Processing**. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2009.

KATE, S.; GORDON, L. **Webster's New World English Grammar Handbook**. 2nd. ed. [S.l.: s.n.], 2009.

KINDLER, E. On the semantics of eps: Resolving the vicious circle. **Data Knowl. Eng.**, Elsevier Science Publishers B. V., Amsterdam, The Netherlands, The Netherlands, v. 56, n. 1, p. 23–40, jan. 2006. Available from Internet: <<http://dx.doi.org/10.1016/j.datak.2005.02.005>>.

KOK, D. de; BROUWER, H. **Natural Language Processing for the Working Programmer**. [s.n.], 2011. Available from Internet: <<http://nlpwp.org/book/>>.

KOSSAK, F.; ILLIBAUER, C.; GEIST, V. Event-based gateways: Open questions and inconsistencies. In: **Business Process Model and Notation - 4th International Workshop, BPMN 2012, Vienna, Austria, September 12-13, 2012. Proceedings**. [S.l.: s.n.], 2012. p. 53–67.

LAVOIE, B.; RAMBOW, O. A fast and portable realizer for text generation systems. In: **Proceedings of the Fifth Conference on Applied Natural Language Processing**. Stroudsburg, PA, USA: Association for Computational Linguistics, 1997. (ANLC '97), p. 265–268. Available from Internet: <<http://dx.doi.org/10.3115/974557.974596>>.

LAWRIE, D. et al. What's in a name? a study of identifiers. In: **Proceedings of the 14th IEEE International Conference on Program Comprehension**. Washington, DC, USA: IEEE Computer Society, 2006. (ICPC '06), p. 3–12. Available from Internet: <<http://dx.doi.org/10.1109/ICPC.2006.51>>.

LEOPOLD, H. **Natural language in business process models**. [S.l.]: Springer, 2013.

LEOPOLD, H.; MENDLING, J.; POLYVYANY, A. Supporting process model validation through natural language generation. In: . [S.l.]: IEEE Transactions on Software Engineering, 2014. v. 40.

LEVY, R.; MANNING, C. Is it harder to parse chinese, or the chinese treebank? In: **Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - Volume 1**. Stroudsburg, PA, USA: Association for Computational Linguistics, 2003. (ACL '03), p. 439–446. Available from Internet: <<http://dx.doi.org/10.3115/1075096.1075152>>.

LIKERT, R. **A Technique for the Measurement of Attitudes**. publisher not identified, 1932. (A Technique for the Measurement of Attitudes, N° 136-165). Available from Internet: <<https://books.google.com.br/books?id=9rotAAAAYAAJ>>.

MANNING, C. D.; RAGHAVAN, P.; SCHÜTZE, H. **Introduction to Information Retrieval**. New York, NY, USA: Cambridge University Press, 2008.

MARNEE, M.-C. de; MANNING, C. D. Stanford typed dependencies manual. 2009. Available from Internet: <<http://nlp.stanford.edu/software/dependenciesmanual.pdf>>.

MEITZ, M.; LEOPOLD, H.; MENDLING, J. An approach to support process model validation based on text generation. In: . [S.l.: s.n.], 2013. v. 33, n. 2, p. 7–20.

MENDLING, J. **Metrics for Process Models: Empirical Foundations of Verification, Error Prediction, and Guidelines for Correctness**. 1. ed. [S.l.]: Springer Publishing Company, Incorporated, 2008.

MENDLING, J. Managing structural and textual quality of business process models. In: \_\_\_\_\_. **Data-Driven Process Discovery and Analysis: Second IFIP WG 2.6, 2.12 International Symposium, SIMPDA 2012, Campione d'Italia, Italy, June 18-20, 2012, Revised Selected Papers**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013. p. 100–111. Available from Internet: <[http://dx.doi.org/10.1007/978-3-642-40919-6\\_6](http://dx.doi.org/10.1007/978-3-642-40919-6_6)>.

MENDLING, J.; NEUMANN, G.; AALST, W. van der. Understanding the occurrence of errors in process models based on metrics. In: \_\_\_\_\_. **On the Move to Meaningful**

**Internet Systems 2007: CoopIS, DOA, ODBASE, GADA, and IS: OTM Confederated International Conferences CoopIS, DOA, ODBASE, GADA, and IS 2007, Vilamoura, Portugal, November 25-30, 2007, Proceedings, Part I.** Berlin, Heidelberg: Springer Berlin Heidelberg, 2007. p. 113–130. Available from Internet: <[http://dx.doi.org/10.1007/978-3-540-76848-7\\_9](http://dx.doi.org/10.1007/978-3-540-76848-7_9)>.

MENDLING, J.; REIJERS, H. A.; AALST, W. M. P. van der. Seven process modeling guidelines (7pmg). **Inf. Softw. Technol.**, Butterworth-Heinemann, Newton, MA, USA, v. 52, n. 2, p. 127–136, feb. 2010. Available from Internet: <<http://dx.doi.org/10.1016/j.infsof.2009.08.004>>.

MENDLING, J.; REIJERS, H. A.; CARDOSO, J. What makes process models understandable? In: \_\_\_\_\_. **Business Process Management: 5th International Conference, BPM 2007, Brisbane, Australia, September 24-28, 2007. Proceedings.** Berlin, Heidelberg: Springer Berlin Heidelberg, 2007. p. 48–63. Available from Internet: <[http://dx.doi.org/10.1007/978-3-540-75183-0\\_4](http://dx.doi.org/10.1007/978-3-540-75183-0_4)>.

MENDLING, J.; REIJERS, H. A.; RECKER, J. Activity labeling in process modeling: Empirical insights and recommendations. **Inf. Syst.**, Elsevier Science Ltd., Oxford, UK, UK, v. 35, n. 4, p. 467–482, jun. 2010. Available from Internet: <<http://dx.doi.org/10.1016/j.is.2009.03.009>>.

MENDLING, J.; STREMBECK, M.; RECKER, J. C. Factors of process model comprehension : findings from a series of experiments. **Decision Support Systems**, Elsevier, v. 53, n. 1, p. 195–206, April 2012. Available from Internet: <<http://eprints.qut.edu.au/49413/>>.

MILLER, G. A. Wordnet: A lexical database for english. **Commun. ACM**, ACM, New York, NY, USA, v. 38, n. 11, p. 39–41, nov. 1995. Available from Internet: <<http://doi.acm.org/10.1145/219717.219748>>.

MUEHLEN, M. Z.; ROSEMAN, M. M.: Multi-paradigm process management. In: **In: Proc. of the Fifth Workshop on Business Process Modeling, Development, and Support - CAISE Workshops.** [S.l.: s.n.], 2004. p. 169–175.

OMG. **Business Process Model and Notation (BPMN) Version 2.0.** [S.l.], 2011. Available from Internet: <<http://taval.de/publications/BPMN20>>.

OMG. Business process modeling notation (bpmn). versão 2.0.2, 2013. Available from Internet: <<http://www.omg.org/spec/BPMN/2.0.2/>>.

PETERSON, J. L. **Petri Net Theory and the Modeling of Systems.** Upper Saddle River, NJ, USA: Prentice Hall PTR, 1981.

PEZZÈ, M.; YOUNG, M. **Teste e Análise de Software: Processos, Princípios e Técnicas.** Bookman, 2009. Available from Internet: <<https://books.google.com.br/books?id=aldOmo1oF2AC>>.

RAFFERTY, A. N.; MANNING, C. D. Parsing three german treebanks: Lexicalized and unlexicalized baselines. In: **Proceedings of the Workshop on Parsing German.** Stroudsburg, PA, USA: Association for Computational Linguistics, 2008. (PaGe '08), p. 40–46. Available from Internet: <<http://dl.acm.org/citation.cfm?id=1621401.1621407>>.

REIJERS, H.; MENDLING, J. Modularity in process models: Review and effects. In: **Proceedings of the 6th International Conference on Business Process Management**. Berlin, Heidelberg: Springer-Verlag, 2008. (BPM '08), p. 20–35. Available from Internet: <[http://dx.doi.org/10.1007/978-3-540-85758-7\\_5](http://dx.doi.org/10.1007/978-3-540-85758-7_5)>.

RITTGEN, P. Negotiating models. In: \_\_\_\_\_. **Advanced Information Systems Engineering: 19th International Conference, CAiSE 2007, Trondheim, Norway, June 11-15, 2007. Proceedings**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007. p. 561–573. Available from Internet: <[http://dx.doi.org/10.1007/978-3-540-72988-4\\_39](http://dx.doi.org/10.1007/978-3-540-72988-4_39)>.

ROBERTSON, J. Likert-type scales, statistical methods, and effect sizes. **Commun. ACM**, ACM, New York, NY, USA, v. 55, n. 5, p. 6–7, may 2012. Available from Internet: <<http://doi.acm.org/10.1145/2160718.2160721>>.

ROSA, M. L. et al. Managing process model complexity via concrete syntax modifications. **IEEE Transactions on Industrial Informatics**, v. 7, n. 2, p. 255–265, May 2011.

ROSEMANN, M. Potential pitfalls of process modeling: part A. **Business Process Management Journal**, Emerald Group Publishing Limited, v. 12, n. 2, p. 249–254, mar. 2006. Available from Internet: <<http://dx.doi.org/10.1108/14637150610657567>>.

ROTH, D. Learning to resolve natural language ambiguities: A unified approach. **AAAI-98 Proceedings**, 1998.

RUMBAUGH, J.; JACOBSON, I.; BOOCH, G. **Unified Modeling Language Reference Manual, The (2Nd Edition)**. [S.l.]: Pearson Higher Education, 2004.

SANTORO, F. M.; GONÇALVES, J. C. A.; BAIAO, F. A. Business process mining from group stories. **International Conference on Computer Supported Cooperative Work in Design**, IEEE Computer Society, Los Alamitos, CA, USA, p. 161–166, 2009.

SCHREPFER, M. et al. The impact of secondary notation on process model understanding. In: \_\_\_\_\_. **The Practice of Enterprise Modeling: Second IFIP WG 8.1 Working Conference, PoEM 2009, Stockholm, Sweden, November 18-19, 2009. Proceedings**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009. p. 161–175. Available from Internet: <[http://dx.doi.org/10.1007/978-3-642-05352-8\\_13](http://dx.doi.org/10.1007/978-3-642-05352-8_13)>.

SHARP, A.; MCDERMOTT, P. **Workflow Modeling: Tools for Process Improvement and Application Development**. 1st. ed. [S.l.]: Artech House, Inc., 2001.

SOMMERVILLE, I. **Engenharia de software**. Pearson Prentice Hall, 2011. Available from Internet: <<https://books.google.com.br/books?id=H4u5ygAACAAJ>>.

SUSSNA, M. Word sense disambiguation for free-text indexing using a massive semantic network. In: **Proceedings of the Second International Conference on Information and Knowledge Management**. New York, NY, USA: ACM, 1993. (CIKM '93), p. 67–74. Available from Internet: <<http://doi.acm.org/10.1145/170088.170106>>.

THOM, L.; IOCHPE, H. Bpmn poster. **Berliner BPM-Offensive**, 2010. Available from Internet: <<https://goo.gl/kQnvh9>>.

THOM, L. H. **Gerenciamento de Processos de Negócio e Aplicabilidade na Saúde e na Robótica**. [S.l.]: Biblioteca Digital Brasileira de Computação, 2012.

THOM, L. H.; REICHERT, M.; IOCHPE, C. Activity patterns in process-aware information systems: Basic concepts and empirical evidence. **International Journal of Business Process Integration and Management (IJBPIIM)**, Inderscience Publishers, 2009. Available from Internet: <<http://dbis.eprints.uni-ulm.de/473/>>.

TING, K. M. Confusion matrix. In: \_\_\_\_\_. **Encyclopedia of Machine Learning**. Boston, MA: Springer US, 2010. p. 209–209. Available from Internet: <[http://dx.doi.org/10.1007/978-0-387-30164-8\\_157](http://dx.doi.org/10.1007/978-0-387-30164-8_157)>.

WANG, H. J.; ZHAO, J. L.; ZHANG, L.-J. Policy-driven process mapping (pdp): Discovering process models from business policies. **Decis. Support Syst.**, Elsevier Science Publishers B. V., Amsterdam, The Netherlands, The Netherlands, v. 48, n. 1, p. 267–281, dec. 2009. Available from Internet: <<http://dx.doi.org/10.1016/j.dss.2009.08.006>>.

WEBER, B.; SADIQ, S.; REICHERT, M. Beyond rigidity–dynamic process lifecycle support. **Computer Science-Research and Development**, Springer, 2009.

WESKE, M. **Business Process Management: Concepts, Languages, Architectures**. [S.l.]: Springer-Verlag, Berlin, 2007.

WESKE, M. **Business Process Management**. Berlin Heidelberg: Springer Berlin Heidelberg, 2012.

WHITE, M. Information overlook. EContent, p. 26:7, 2003.

WHITE, S.; MIERS, D. **BPMN Modeling and Reference Guide: Understanding and Using BPMN**. Future Strategies Incorporated, 2008. Available from Internet: <<https://books.google.com.br/books?id=0Z2Td3bCYW8C>>.

WORKSHEETS, R. Text structure. <http://www.ereadingworksheets.com/text-structure/>, p. 1, 2014.

YAROWSKY, D. Word-sense disambiguation using statistical models of roget's categories trained on large corpora. In: **Proceedings of the 14th Conference on Computational Linguistics - Volume 2**. Stroudsburg, PA, USA: Association for Computational Linguistics, 1992. (COLING '92), p. 454–460. Available from Internet: <<http://dx.doi.org/10.3115/992133.992140>>.