UNIVERSIDADE DE SÃO PAULO
ESCOLA DE ARTES, CIÊNCIAS E HUMANIDADES
PROGRAMA DE PÓS-GRADUAÇÃO EM SISTEMAS DE INFORMAÇÃO

RICELLI MOREIRA SILVA RAMOS

**Determinação de conteúdo para geração de língua natural baseada em
personalidade**

São Paulo

2018

RICELLI MOREIRA SILVA RAMOS

**Determinação de conteúdo para geração de língua natural baseada em
personalidade**

Dissertação apresentada à Escola de Artes, Ciências e Humanidades da Universidade de São Paulo para obtenção do título de Mestre em Ciências pelo Programa de Pós-graduação em Sistemas de Informação.

Área de concentração: Processamento de Linguagem Natural: Geração de Língua Natural

Versão corrigida contendo as alterações solicitadas pela comissão julgadora em 25 de junho de 2018. A versão original encontra-se em acervo reservado na Biblioteca da EACH-USP e na Biblioteca Digital de Teses e Dissertações da USP (BDTD), de acordo com a Resolução CoPGr 6018, de 13 de outubro de 2011.

Orientador: Prof. Dr. Ivandré Paraboni

São Paulo

2018

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

CATALOGAÇÃO-NA-PUBLICAÇÃO

(Universidade de São Paulo. Escola de Artes, Ciências e Humanidades. Biblioteca)
CRB-8 4625

Ramos, Ricelli Moreira Silva

Determinação de conteúdo para geração de língua natural
baseada em personalidade / Ricelli Moreira Silva Ramos ;
orientador, Ivandré Paraboni. – 2018.
98 f.

Dissertação (Mestrado em Ciências) - Programa de Pós-
Graduação em Sistemas de Informação, Escola de Artes,
Ciências e Humanidades, Universidade de São Paulo.
Versão corrigida

1. Inteligência artificial. 2. Linguagem natural. 3. Análise de
texto. 4. Personalidade. I. Paraboni, Ivandré, orient. II. Título.

CDD 22.ed.– 006.35

Dissertação de autoria de Ricelli Moreira Silva Ramos, sob o título “**Determinação de conteúdo para geração de língua natural baseada em personalidade**”, apresentada à Escola de Artes, Ciências e Humanidades da Universidade de São Paulo, para obtenção do título de Mestre em Ciências pelo Programa de Pós-graduação em Sistemas de Informação, na área de concentração Metodologia e Técnicas da Computação, aprovada em 25 de junho de 2018 pela comissão julgadora constituída pelos doutores:

Prof. Dr. Ivandré Paraboni

Instituição: Universidade de São Paulo

Presidente

Prof. Dr. Luciano Antonio Digiampietri

Instituição: Universidade de São Paulo

Prof. Dr. Camilo Rodrigues Neto

Instituição: Universidade de São Paulo

Profa. Dra. Valéria Delisandra Feltrim

Instituição: Universidade Estadual de Maringá

Agradecimentos

Se existe uma palavra que defina como eu me sinto com relação a esse trabalho, essa palavra é gratidão! Sou grata pela oportunidade de pesquisar um tema tão interessante e desafiador. Sou grata pelo enorme aprendizado nesse período, e por cada resultado dessa pesquisa. Agradeço a Deus, à família, aos amigos, ao orientador, aos colegas e aos professores. A todos os que me apoiaram e me ajudaram das mais diversas formas o meu muito obrigada!

*“Understand well as I may, my comprehension can only be an infinitesimal fraction of all
I want to understand.”
(Ada Lovelace)*

Resumo

RAMOS, Ricelli Moreira Silva. **Determinação de conteúdo para geração de língua natural baseada em personalidade**. 2018. 98 f. Dissertação (Mestrado em Ciências) – Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo, 2018.

O presente trabalho aborda a determinação de conteúdo na fase de planejamento do documento no *pipeline* de Geração de Língua Natural (GLN) usando fatores de personalidade do modelo dos Cinco Grandes Fatores (CGF). O principal objetivo deste trabalho é gerar modelos computacionais de determinação de conteúdo baseados nos fatores de personalidade CGF. O trabalho aplicará técnicas já existentes de GLN para determinação de conteúdo, levando em conta os fatores de personalidade mapeados pelo modelo CGF. São utilizados os conceitos descritos por substantivos e os conceitos descritos por adjetivos relacionados aos substantivos na tarefa de descrição de cenas para a determinação de conteúdo. As principais contribuições desse trabalho são uma investigação de se e como a determinação de conteúdo de descrições textuais é influenciada pela personalidade do autor, além de entregar um modelo de determinação de conteúdo baseado em personalidade para os conceitos em que o trabalho foi aplicado, entre outras entregas. É apresentado o embasamento teórico com os conceitos fundamentais de GLN, e mais especificamente de determinação de conteúdo, foco dessa pesquisa. Além disso, são apresentados os modelos de personalidade humana, com destaque ao modelo CGF e inventários CGF, utilizados para a coleta de dados e execução dessa proposta. São apresentados também os principais trabalhos relacionados à GLN e modelo CGF, mesmo que não tratem especificamente da influência dos CGF na determinação de conteúdo. Um experimento para coleta do corpus utilizado na pesquisa é descrito, e também os modelos para determinação de conteúdo no âmbito de conceitos representando entidades visuais e seus predicados, assim como a avaliação desses modelos. Por fim, são apresentadas as conclusões obtidas com os modelos desenvolvidos e experimentos realizados.

Palavras-chaves: Geração de língua natural. Determinação de conteúdo. Fatores de personalidade. Modelo CGF.

Abstract

Ramos, Ricelli Moreira Silva. **Content planning for natural language generation based on personality**. 2018. 98 p. Dissertation (Master of Science) – School of Arts, Sciences and Humanities, University of São Paulo, São Paulo, 2018.

The present research approaches the content determination in the document planning phase of the Natural Language Generation (NLG) pipeline using personality factors of the Big Five Factor (BFF) model. The main objective of this research is to generate computational models of content determination based on the BFF personality factors. This research will apply existing NLG models to the content determination, taking into account the personality factors mapped by the BFF model. The concepts described by nouns and the concepts described by adjectives related to nouns in the task of describing scenes for content determination are used. The main contributions of this research are an investigation of if and how the content determination of textual descriptions is influenced by the personality of the author, in addition to providing a personality-based content determination model for the concepts in which the research was applied, among others deliveries. This document presents the theoretical basis and the fundamental NLG concepts, and more specifically the concept of content determination, which is the focus of this research. In addition, human personality models are presented, with emphasis on the BFF model and BFF inventories, used both for data collection and development of this proposal. The main studies related to NLG and the BFF model, even if they do not specifically address the influence of BFF in content determination, are also presented. An experiment for collecting the corpus used in the research is described, also the models to determine content in the scope of concepts representing visual entities and their predicates, as well as an evaluation of these models. Finally, the conclusions obtained with the developed models and experiments are presented.

Keywords: Natural language generation. Content determination. Big Five Factors. BFF model.

Lista de figuras

Figura 1 – <i>Pipeline</i> de sistemas de GLN	17
Figura 2 – Um exemplo de domínio visual	19
Figura 3 – Sistema de geração de língua natural	22
Figura 4 – Arquitetura de um sistema de geração de língua natural	23
Figura 5 – Exemplos de mensagens geradas na determinação de conteúdo	27
Figura 6 – Cálculo do coeficiente de Sorensen-Dice	30
Figura 7 – Arquitetura do gerador PERSONAGE	46
Figura 8 – <i>Pipeline</i> da arquitetura de um sistema de geração de língua natural, com os componentes do sistema PERSONAGE nas caixas cinzas	53
Figura 9 – Exemplo de estímulo visual para coleta do corpus b5-text	62
Figura 10 – Distribuição de idade dos sujeitos	63
Figura 11 – Sujeitos por fatores do modelo CGF	64
Figura 12 – Cena para descrição textual	65

Lista de quadros

Quadro 1 – Módulos e tarefas de um sistema de GLN	25
Quadro 2 – Adjetivos associados aos extremos de cada fator do modelo CGF . . .	34

Lista de tabelas

Tabela 1 – Distribuição de conceitos por fator do modelo CGF na descrição da figura 12	66
Tabela 2 – Número de conceitos Entidade por cena no corpus b5-text	72
Tabela 3 – Número de conceitos Predicado por cena no corpus b5-text	73
Tabela 4 – Exemplo de classes positivas e negativas para uma descrição da cena da figura 12	74
Tabela 5 – Resultados médios obtidos utilizando os conceitos Entidade do corpus b5-text relacionados a cada cena	75
Tabela 6 – Desempenho dos conceitos Entidade mais frequentes por cena	76
Tabela 7 – Resultados médios obtidos utilizando os conceitos Predicado do corpus b5-text relacionados a cada cena	77
Tabela 8 – Desempenho dos conceitos Predicado mais frequentes por cena	78
Tabela 9 – Média do coeficiente Dice	83
Tabela 10 – Média do coeficiente Dice para os conceitos Entidade por cena	83
Tabela 11 – Média do coeficiente Dice para os conceitos Predicado por cena	84
Tabela 12 – Média do coeficiente Dice para a combinação dos conceitos Entidade e Predicado por cena	84
Tabela 13 – Resultados gerais obtidos pelos modelos	85

Lista de abreviaturas e siglas

AM	Aprendizagem de Máquina
BFF	<i>Big Five Factors</i>
BFI	<i>Big Five Inventory</i>
CGF	Cinco Grandes Fatores
DC	Determinação de Conteúdo
EACH	Escola de Artes, Ciências e Humanidades
FFI	<i>Five Factor Inventory</i>
FL	Forma Lógica
GAPED	<i>The Geneva Affective Picture Database</i>
GLN	Geração de Linguagem Natural
HTML	<i>HyperText Markup Language</i>
IGFP-5	Inventário dos Cinco Grandes Fatores de Personalidade
LREC-2018	<i>Language Resources and Evaluation Conference</i>
NEO-FFI	<i>NEO-Five Factor Inventory</i>
NEO-PI	<i>NEO Personality Inventory</i>
NEO-PI-R	<i>NEO Personality Inventory Revised</i>
NLG	<i>Natural Language Generation</i>
PHP	<i>Hypertext Preprocessor</i>
PPgSI	Programa de Pós-graduação em Sistemas de Informação
RST	Relações retóricas
SARGS	Sistema de geração de relatório afetivo simples
SMOTE	<i>Synthetic Minority Over-sampling Technique</i>

SVM	<i>Support Vector Machine</i>
USP	Universidade de São Paulo
VC	Vapnik-Chervonenkis
XAMPP	<i>Apache + MariaBD + PHP + Perl</i>

Sumário

1	Introdução	16
1.1	<i>Objetivo</i>	20
1.2	<i>Hipótese</i>	20
1.3	<i>Organização da monografia</i>	20
2	Conceitos fundamentais	22
2.1	<i>Geração de língua natural</i>	22
2.2	<i>Determinação do conteúdo</i>	26
2.3	<i>Aprendizagem de máquina</i>	29
2.3.1	Árvores de decisão binárias	29
2.3.2	<i>Support Vector Machine (SVM)</i>	29
2.3.3	Métricas	30
2.4	<i>Modelos de personalidade humana</i>	31
2.4.1	Os cinco grandes fatores	32
2.4.2	Inventários CGF	34
3	Revisão bibliográfica	37
3.1	<i>O sistema LetterGen</i>	37
3.2	<i>Modelos de avatares autônomos</i>	39
3.3	<i>O sistema PERSONAGE</i>	41
3.3.1	A determinação de conteúdo no sistema PERSONAGE	45
3.4	<i>Avaliação e generalização do sistema PERSONAGE</i>	50
3.5	<i>Variação estilística contínua no sistema PERSONAGE</i>	51
3.6	<i>Evolução do sistema PERSONAGE</i>	54
3.7	<i>O sistema PERSONAGE em diferentes domínios</i>	55
3.8	<i>Outras ferramentas e aplicações</i>	58
3.9	<i>Discussão</i>	59
4	Construção do corpus b5	61
4.1	<i>Materiais</i>	61
4.1.1	Inventário de personalidade	61

4.1.2	Imagens de estímulo	62
4.2	<i>Sujeitos</i>	63
4.3	<i>Procedimento</i>	63
4.4	<i>Dados obtidos</i>	64
4.5	<i>Análise preliminar</i>	64
4.5.1	Resultados da análise preliminar	66
5	Método	68
5.1	<i>Objetivo</i>	68
5.2	<i>Hipótese</i>	69
5.3	<i>Atividades previstas</i>	69
6	Determinação de conteúdo baseada em personalidade	71
6.1	<i>Visão geral</i>	71
6.2	<i>Dados</i>	71
6.3	<i>Modelos propostos</i>	73
6.4	<i>Testes</i>	74
6.5	<i>Considerações</i>	77
6.6	<i>Outros experimentos</i>	80
7	Avaliação complementar	82
7.1	<i>Procedimento</i>	82
7.2	<i>Resultados obtidos</i>	82
7.3	<i>Considerações</i>	85
8	Conclusão	87
8.1	<i>Limitações</i>	88
8.2	<i>Trabalhos futuros</i>	88
	Referências¹	89
	Anexo A – Protocolo de aplicação	97
A.1	<i>Condições</i>	97
A.2	<i>Início do experimento</i>	97

¹ De acordo com a Associação Brasileira de Normas Técnicas. NBR 6023.

<i>A.3</i>	<i>Inventário</i>	98
A.3.1	Facebook	98
A.3.2	Inventário presencial	98

1 Introdução

Muitos sistemas computacionais apresentam informação ou resultados em forma textual, frequentemente com uso de expressões pré-gravadas e projetadas para retratar uma personalidade específica ou estilo linguístico. Embora esta abordagem ofereça total controle autoral e produza expressões de alta qualidade, ela sofre de problemas de portabilidade e escalabilidade, além de não ser sensível ao contexto em que está inserida. Para mitigar esses problemas, são aplicadas técnicas de geração de língua natural (GLN) (WALKER; RAMBOW, 2002).

A GLN trata da produção de formas linguísticas (e.g., texto) a partir de uma entrada não linguística (geração dados-para-texto) ou linguística (texto-para-texto). Aplicações de GLN incluem boletins de previsão do tempo baseados em dados de satélites (BELZ, 2008; SRIPADA et al., 2014), diagnósticos e relatórios médicos produzidos em função da leitura de sensores de equipamentos hospitalares (JORDAN et al., 2014; EUGENIO et al., 2014; SCHNEIDER et al., 2015), escrita criativa poética seguindo as métricas dos diferentes estilos (ZHANG; LAPATA, 2014; GABRIEL; CHEN; NICHOLS, 2015), descrição de imagens (KARPATHY; FEI-FEI, 2015; GILBERT et al., 2015), comunicação com crianças e adolescentes com necessidades especiais (TINTAREV et al., 2016), relatórios financeiros (PLACHOURAS et al., 2016), comunicação e *feedback* para ciência cidadã (WAL et al., 2016), relatos de notícias (LEPPÄNEN et al., 2017), entre outras.

Na geração de um texto, um sistema de GLN toma decisões baseadas em diversos níveis de conhecimento linguístico. Estas decisões tendem a ser modeladas na forma de uma arquitetura do tipo *pipeline*, composta de 3 etapas, visualizáveis na figura 1 (REITER; DALE; FENG, 2000).

A geração de texto segundo a arquitetura de *pipeline* de GLN inicia-se com a etapa de Planejamento do documento. Esta etapa recebe como entrada o objetivo da comunicação, e é composta pela determinação do conteúdo (DC) e estruturação do documento. A etapa seguinte, de Microplanejamento, é composta pela lexicalização, geração de expressões de referência, e agregação. Por fim, a etapa de Realização superficial é composta pela realização linguística e realização estrutural. Ao concluir essas etapas, o texto superficial é gerado, ou seja, a representação abstrata do texto produzida nas etapas anteriores é transformada em uma representação linear em língua natural (REITER; DALE; FENG,

Figura 1 – *Pipeline* de sistemas de GLN

Fonte: (REITER; DALE; FENG, 2000)

2000; REITER, 2007). Nesse trabalho aborda-se especificamente a etapa de planejamento de documento, focado na determinação de conteúdo.

Sistemas de GLN procuram gerar texto de forma correta e com naturalidade, aproximando-se tanto quanto possível do desempenho humano. Dada uma entrada representando o sentido a ser comunicado, um sistema de GLN pode apresentar uma saída fixa (i.e., gerando sempre o mesmo texto) ou considerar variações estilísticas ditadas por um perfil de interesse (e.g., um modelo de autor).

Uma forma de gerar texto automaticamente com variação estilística controlada, e que tem se tornado popular em anos recentes, é o uso de fatores de *personalidade humana* para modelar o texto de um autor-alvo que possua determinada personalidade de interesse. Sistemas deste tipo tomam como entrada, além do conteúdo a ser comunicado, informações sobre a personalidade que o sistema deve impor ao texto de modo que este tenha maior realismo e proximidade ao desempenho humano. Algumas aplicações com esse tipo de característica estão nos trabalhos que apresentam o sistema PERSONAGE de recomendação de restaurantes (MAIRESSE, 2008), e aplicações do mesmo sistema PERSONAGE em outros domínios como *story-telling* (LUKIN; RYAN; WALKER, 2014),

jogos (LUKIN; REED; WALKER, 2015), interação humano-robô (ALY; TAPUS, 2016), e atendimento aos clientes (HERZIG et al., 2017), ou diferentes modelos utilizando GLN como escrita criativa (GABRIEL; CHEN; NICHOLS, 2015) e comentários afetivos (STRAUSS; KIPP, 2008), entre outros.

De modo geral, sistemas de GLN baseados em personalidade consideram o modelo dos cinco grandes fatores (CGF). Este modelo divide a personalidade humana em cinco dimensões - “Neuroticismo”, “Extroversão”, “Amabilidade”, “Conscienciosidade” e “Abertura à experiência” - e é amplamente aceito pela Psicologia como a base adequada para representação da personalidade humana (NORMAN, 1963; GOLDBERG, 1990; PERVIN; JOHN, 2004; ANDRADE, 2008).

De especial interesse para a presente proposta, observa-se que a literatura em GLN de modo geral trata principalmente da natureza superficial da comunicação, correspondendo à escolha lexical e forma sintática, e trata tanto da língua oral quanto escrita. Em ambas são encontradas evidências da variação na forma de comunicação por indivíduos com diferentes personalidades. Os estudos nesse tema não tratam da questão de como a personalidade humana é refletida na determinação de conteúdo, muito embora seja razoável supor que tal relação exista. Por exemplo, um indivíduo com determinada personalidade escolhe falar de alguns itens, enquanto outro dá preferência para outros (MAIRESSE, 2008).

O tipo de aplicação de GLN a ser considerado neste trabalho, e usado para ilustrar esta questão, é o caso de sistemas que geram descrições de cenas visuais (e.g., imagens) em língua natural escrita. Por exemplo, considere o objetivo de produzir uma descrição textual da cena da figura 2.

Dado o objetivo de descrever o conteúdo da cena da figura 2, um sistema de GLN poderia produzir diferentes textos descrevendo a mesma figura baseado nos diferentes fatores de personalidade. Por exemplo, uma descrição textual como

“Uma menina e um menino em uma estrada de terra. Eles usam roupas antigas e desgastadas, têm cabelos escuros e lisos (...)”

pode representar um indivíduo com “Extroversão” mais alta. Já para um indivíduo com “Extroversão” mais baixa poderia produzir um texto como

“Irmãos em um terreno baldio. A garota tem aparência séria e usa vestido marrom com babados e o garoto parece curioso e usa blusa de moletom (...)”.

Figura 2 – Um exemplo de domínio visual

Fonte: Extraída da base GAPED (DAN-GLAUSER; SCHERER, 2011)

Neste exemplo há uma variação no conteúdo utilizado, por exemplo, o indivíduo extrovertido usou “menino” e “menina” separadamente e o introvertido usou “irmãos”, da mesma forma o indivíduo extrovertido abordou o cabelo e suas características, enquanto o outro falou da aparência das crianças, entre outras variações. Ou seja, diferentes indivíduos (ou indivíduos com diferentes personalidades) podem escolher falar de aspectos distintos da mesma cena.

Estas decisões baseadas em personalidade também poderiam ser incorporadas à tarefa de DC em um sistema de GLN para produção de língua natural de forma mais próxima ao desempenho humano, e também com maior variação estilística. No entanto, além do pouco conhecimento disponível sobre a determinação de conteúdo baseada em personalidade, observa-se que, no caso específico do Português brasileiro, estudos desta natureza são ainda mais escassos. Diante deste cenário, evidencia-se a oportunidade para um estudo de como os fatores do modelo CGF influenciam a determinação de conteúdo para o desenvolvimento de sistemas de GLN baseados em personalidade. Uma pesquisa em nível de mestrado acadêmico abordando esta questão é o tema desta monografia.

1.1 *Objetivo*

O objetivo da pesquisa é desenvolver modelos computacionais de determinação de conteúdo baseados nos fatores de personalidade CGF, de tal modo que esta determinação seja o mais próxima possível das escolhas humanas na tarefa de descrição de imagens.

O estudo consiste na coleta de um *cópus* de textos e inventários de personalidade, e no uso de técnicas supervisionadas de aprendizagem de máquina (AM) para determinação de conteúdo com base na personalidade de um autor-alvo de interesse. Em função das dificuldades naturais em coletar um *cópus* de grandes proporções deste tipo, entretanto, o modelo de determinação de conteúdo a ser desenvolvido não objetiva produzir um planejamento completo do que expressar em um documento (por exemplo, com relações retóricas (WILLIAM; THOMPSON, 1988)). Ao invés disso, a proposta será limitada a dois tipos de estimativa que seriam as condições iniciais para um futuro planejamento de documento completo: o modelo deve ser capaz de estimar *quais* objetos discursivos um autor-alvo com certa personalidade escolheria para serem mencionados em um texto, e *quais* os predicados - representados por adjetivos - seriam relacionados a esses objetos.

1.2 *Hipótese*

A hipótese investigada neste estudo é a de que o uso de informação de personalidade humana na tarefa de DC permite obter resultados mais próximos ao desempenho humano do que os resultados obtidos com o uso de modelos baseados nas seleções de conceitos mais frequentes no domínio de interesse. Essa hipótese será testada utilizando o *cópus* coletado nessa pesquisa para treinar e testar modelos de AM que tomem as melhores decisões de conteúdo com e sem informação de personalidade. Os modelos serão avaliados e comparados para demonstrar se existe e qual é a influência da personalidade na determinação de conteúdo na tarefa de descrição de imagens.

1.3 *Organização da monografia*

O restante deste documento é organizado da seguinte forma. O capítulo 2 apresenta os conceitos fundamentais utilizados nesse trabalho: sistemas de GLN, a tarefa de DC,

aprendizagem de máquina, o conceito de personalidade, o modelo dos cinco fatores, e os inventários de personalidade validados para o português brasileiro. O capítulo 3 apresenta a revisão bibliográfica, com a análise dos trabalhos que utilizam GLN e fatores de personalidade. O capítulo 4 apresenta o trabalho de coleta de dados para construção do corpus a ser utilizado nessa pesquisa. O capítulo 5 apresenta o método de pesquisa. O capítulo 6 apresenta os modelos desenvolvidos e testes realizados para a verificação da hipótese investigada nesse estudo. O capítulo 7 apresenta a avaliação complementar dos modelos desenvolvidos. Por fim, o capítulo 8 apresenta as conclusões obtidas com essa pesquisa.

2 Conceitos fundamentais

Nesse capítulo é apresentada uma visão geral dos principais conceitos utilizados nesse trabalho: sistemas de geração de língua natural, com aprofundamento da tarefa de determinação de conteúdo, aprendizado de máquina, o conceito de personalidade, o modelo dos cinco fatores de personalidade (CGF), os inventários de personalidade validados para o contexto brasileiro, e o inventário utilizado no presente trabalho.

2.1 Geração de língua natural

A conversação realizada por meio da língua natural é um componente básico do comportamento social, e um dos principais meios pelos quais os seres humanos expressam intenções, crenças, emoções, atitudes e personalidade (MAIRESSE; WALKER, 2009). Geração de língua natural, um subcampo da inteligência artificial e linguística computacional, foca-se em sistemas computacionais que consigam produzir linguagem humana, ou seja, sistemas que permitam a conversação, ou comunicação, entre indivíduo e máquina (REITER; DALE; FENG, 2000).

A figura 3 demonstra o funcionamento de um sistema de geração de língua natural, proposta por Reiter (1994) como um consenso de arquitetura para sistemas deste tipo, com base em pesquisas e aplicações anteriores. Essa arquitetura é detalhada em Reiter, Dale e Feng (2000).

Figura 3 – Sistema de geração de língua natural

Fonte: Traduzido e adaptado de Reiter, Dale e Feng (2000)

Na figura 3 é possível visualizar os módulos de um sistema de GLN típico, iniciando com o planejamento do documento dividido em duas tarefas principais: a primeira é a determinação (também chamada seleção) do conteúdo, e a segunda é a ordenação e estruturação (estruturação do documento). Em seguida, há o planejamento de sentenças e, por fim, a realização superficial.

Um exemplo prático de arquitetura de sistema de geração de língua natural é apresentado na figura 4, específico para gerar diálogos em uma interação entre indivíduo e um sistema GLN. Além dos módulos apresentados na figura 3, são apresentadas na figura 4 as entradas necessárias ao sistema de GLN, assim como quais bases de dados cada uma das tarefas utiliza em sua execução. Esse modelo foi desenhado para um sistema de GLN no domínio de voos de um aeroporto, porém pode ser aplicado a qualquer domínio de sistema de GLN (WHITE; CLARK; MOORE, 2010).

Figura 4 – Arquitetura de um sistema de geração de língua natural

Fonte: Traduzido e adaptado de White, Clark e Moore (2010)

No exemplo da figura 4 verifica-se que as entradas fundamentais do sistema são o histórico do discurso, ou seja, o que já foi comunicado entre sistema e usuário, o modelo de usuário, que trata dos objetivos da comunicação, e o domínio a ser tratado. Além destas, tem-se o gerenciador de diálogo, que informa o objetivo da comunicação que será gerada pelo sistema (WHITE; CLARK; MOORE, 2010).

De forma geral, as entradas de um sistema de GLN podem ser consideradas uma tupla de quatro elementos (k, c, u, d) sendo k a fonte de conhecimento a ser usada, ou

seja, a informação sobre o domínio a ser trabalhado; c o objetivo de comunicação a ser atingido, isto é, o propósito do texto que será gerado pelo sistema de GLN, u o modelo de usuário, ou seja, a caracterização do ouvinte ou audiência para quem o texto será gerado, e d o histórico de discurso, um modelo do que foi dito no texto produzido até então, também chamado de histórico de diálogo (REITER; DALE; FENG, 2000; WHITE; CLARK; MOORE, 2010).

O sistema de GLN recebe como entrada um objetivo comunicativo abstrato do gerenciador de diálogo. Nos próximos exemplos, será utilizado o domínio de transporte ferroviário. Na fase de apresentação de informações do diálogo, o objetivo comunicativo é descrever as viagens disponíveis que melhor atendam às restrições e preferências do usuário. Dado este objetivo comunicativo, o planejador de conteúdo seleciona e organiza a informação para transmitir, aplicando os itens do plano que implementam sua estratégia de apresentação. Ao fazê-lo, utiliza três fontes de conhecimento adicionais: o modelo de usuário, o domínio e o histórico de diálogo (WHITE; CLARK; MOORE, 2010).

A primeira tarefa do planejamento do documento, ou seja, o que o planejador de conteúdo realizará, é a determinação de conteúdo. Nela decidem-se quais informações devem ser comunicadas no documento de saída, ou seja, é o aspecto de conteúdo do módulo de planejamento de documento. A tarefa seguinte, estruturação do documento, decide como o conteúdo deve ser agrupado e relacionado em termos retóricos. É o aspecto estrutural do planejamento de documento. Em suma, dado o objetivo, o planejador de conteúdo seleciona e organiza a informação, aplicando a estratégia de apresentação definida. (REITER; DALE; FENG, 2000; WHITE; CLARK; MOORE, 2010).

O módulo seguinte é o planejamento de sentenças. O plano de conteúdo é enviado para o planejador de sentenças, que usa modelos (*templates*) léxicos para executar as tarefas de agregação, lexicalização e geração de expressões de referência. A saída do planejamento de sentenças é uma sequência de formas lógicas (FLs). O uso de FLs representa uma maneira prática e flexível de lidar com a interação de decisões tomadas no nível de planejamento de sentenças. Cada FL é realizada como uma sentença usando uma gramática lexical do realizador superficial (WHITE; CLARK; MOORE, 2010).

A agregação é a tarefa que mapeia a estrutura do plano de conteúdo em estruturas linguísticas e elementos textuais. A lexicalização decide quais palavras específicas (ou outros recursos linguísticos) devem ser utilizadas para expressar o conteúdo selecionado e estruturado nas tarefas anteriores, isto é, é o aspecto de conteúdo do microplanejamento. O

próximo passo é a geração de expressões de referência, que seleciona quais expressões devem ser usadas para referenciar os diferentes objetos do discurso, sendo outra parte de conteúdo do microplanejamento. Tem-se a seguir a agregação, que define como as estruturas criadas anteriormente serão mapeadas em estruturas linguísticas (basicamente frases e parágrafos). Esta etapa é a parte estrutural do microplanejamento. Ou seja, o planejador de sentenças transforma a informação selecionada e ordenada em estruturas sentenciais, utilizando os *templates* de lexicalização (REITER; DALE; FENG, 2000; WHITE; CLARK; MOORE, 2010).

O último módulo é composto pela realização superficial, formada pela realização linguística e pela realização estrutural. A realização linguística converte as representações abstratas das sentenças em texto objetivo, sendo o aspecto de conteúdo da realização superficial. A realização estrutural transforma as estruturas de parágrafos e seções em codificação do tipo HTML (*HyperText Markup Language*) ou similar, compreensível para o componente de apresentação do documento. Esta etapa corresponde ao aspecto estrutural da realização superficial. Ou seja, o realizador superficial recebe as estruturas e as traduz em texto efetivo para compreensão humana (REITER; DALE; FENG, 2000; WHITE; CLARK; MOORE, 2010).

O quadro 1 apresenta, de forma resumida, as tarefas dos principais módulos de um sistema de GLN. Nesse quadro, as tarefas estão separadas entre tarefas de estrutura e de conteúdo, conforme seu relacionamento direto com cada item. Isso não significa que tarefas como a lexicalização ou a geração de expressão de referência determinem qual conteúdo será gerado pelo sistema de GLN, e sim que essas tarefas tratam o conteúdo determinado em tarefas anteriores, de forma a atingir os objetivos comunicativos definidos para o sistema de GLN (REITER; DALE; FENG, 2000).

Quadro 1 – Módulos e tarefas de um sistema de GLN

Módulo	Tarefa de Conteúdo	Tarefa de Estrutura
Planejamento de documento	Determinação do conteúdo	Estruturação do documento
Microplanejamento	Lexicalização e geração de expressão de referência	Agregação
Realização	Realização linguística	Realização estrutural

Fonte: Traduzido e adaptado de Reiter, Dale e Feng (2000)

É importante notar que, conforme indicado na figura 3, a tarefa de determinação de conteúdo é responsável por “o que dizer”, ou seja, qual conteúdo o sistema de GLN

deve tratar, e, por isso, também é dependente do domínio (REITER; DALE; FENG, 2000). Essa tarefa é explicada detalhadamente na seção seguinte (2.2), pois será a tarefa de GLN explorada na presente pesquisa.

As tarefas e módulos seguintes da figura 3 (ordenação e estruturação, planejamento de sentenças e realização superficial) são responsáveis por definir “como dizer” o conteúdo selecionado na tarefa de determinação de conteúdo. Ou seja, são as tarefas que determinam quais palavras serão utilizadas, qual a ordem em que o conteúdo deve aparecer, eventuais marcadores linguísticos etc. (REITER; DALE; FENG, 2000).

Além disso, na figura 3, é possível verificar a importância do domínio para a aplicação em que o sistema de GLN está inserido, e para a determinação do conteúdo. Porém, para as demais atividades e módulos, o domínio é menos relevante, enquanto a linguagem-alvo a ser usada, incluindo, por exemplo, nível de formalidade, polidez etc., é fundamental (REITER; DALE; FENG, 2000).

2.2 *Determinação do conteúdo*

Embora o desenvolvimento de um módulo de determinação de conteúdo (DC) não seja o foco do presente trabalho, esta seção descreve o funcionamento desta tarefa para melhor contextualizar o estudo realizado apresentado nos capítulos 5, 6 e 7.

Determinação do conteúdo é uma tarefa do módulo “Planejador de conteúdo” de um sistema de GLN, e consiste em definir quais informações devem ser incluídas em uma comunicação que utiliza língua natural, como falas de personagens em jogos, boletins meteorológicos, diagnósticos médicos, entre outros. Ou seja, é nessa fase que se especifica a estrutura da informação a ser veiculada (REITER; DALE; FENG, 2000; MAIRESSE; WALKER, 2011).

Determinação de conteúdo também pode ser descrita como o processo de criar um conjunto de mensagens a partir das entradas do sistema e outras fontes de informação, como o histórico de discurso, o modelo de usuário e o domínio (REITER; DALE, 1997). A saída desse processo é uma parcela do plano de conteúdo, uma estrutura de alto nível que reflete o objetivo comunicativo do enunciado a ser gerado (WALKER, 2009).

A tarefa de determinação de conteúdo é um processo de extrema importância nos sistemas de GLN. Na maioria das aplicações é preferível receber um texto superficial com

falhas, mas que trate do conteúdo adequado, do que um texto superficial correto com conteúdo inapropriado (REITER; DALE; FENG, 2000; SRIPADA et al., 2001).

Determinação de conteúdo, via de regra, consiste em filtrar e resumir os dados de entrada, criando mensagens que servirão de entrada às etapas seguintes de um sistema de GLN. As mensagens criadas são expressas em alguma linguagem formal que categoriza entidades, conceitos e relações no domínio da mensagem criada. Na figura 5 é possível visualizar um exemplo de uma mensagem criada para duas comunicações diferentes (1) e (2), no domínio de transporte ferroviário, sendo (a) cada mensagem gerada no processo de determinação de conteúdo e (b) a comunicação que deverá ser gerada como saída do sistema de GLN (REITER; DALE, 1997).

Figura 5 – Exemplos de mensagens geradas na determinação de conteúdo

Fonte: Traduzido e adaptado de Reiter e Dale (1997)

Na figura 5, cada mensagem é representada como uma matriz de parâmetros com seus respectivos valores. Cada parâmetro descreve alguma relação entre as entidades ou conceitos especificados como os argumentos dessa relação. Pode-se verificar que as mensagens não correspondem exatamente às frases. Da mesma forma, as entidades de domínio, relações e conceitos geradas no processo de DC não corresponderão exatamente aos itens lexicais no texto resultante. Essas mensagens servirão de entrada aos próximos módulos do sistema GLN, que construirão o “como” dizê-las, ou seja, transformarão essas mensagens em texto superficial.

A tarefa de determinação de conteúdo depende de uma variedade de fatores, incluindo pelo menos os objetivos comunicativos, o conteúdo, as restrições de saída, e as fontes de informações adjacentes. Cada um desses fatores influencia as decisões tomadas na tarefa.

Os objetivos comunicativos determinam o propósito da comunicação a ser gerada, ou seja, são a meta que será satisfeita se o texto for gerado apropriadamente. Dependendo do propósito da comunicação, informações diferentes devem ser expressadas. Utilizando como exemplo a comunicação no domínio de transporte ferroviário (conforme a figura 5), um relatório sobre atrasos dos trens do dia anterior provavelmente terá informações com uma granularidade mais fina do que seria incluído em um relatório sobre os atrasos de trens que contemplasse todos os atrasos do mês anterior.

O conteúdo pode depender também das características (conhecidas ou assumidas) do ouvinte ou leitor. Seguindo com o mesmo exemplo, se o relatório for direcionado a alguém com mais experiência e conhecimento no assunto, não são necessárias muitas informações técnicas. Porém, se for um novato no assunto, precisará de mais explicações e orientações.

As restrições acerca da saída também podem influenciar a determinação de conteúdo. Por exemplo, quando o texto produzido tem um limite de caracteres ou um limite de páginas.

Finalmente, dependendo da natureza e conteúdo de informações disponíveis, as fontes de informações adjacentes podem impactar a determinação de conteúdo. Por exemplo, utilizando ainda o relatório de atrasos de trens, essas fontes podem ser a informação de que esse é o mês com mais atrasos de determinado período, ou as causas dos atrasos, entre outros. Como essas informações não dependem apenas do período ao qual o relatório se refere, podem ou não estar disponíveis (REITER; DALE; FENG, 2000).

A decisão de quais informações devem ou não ser incluídas em um texto e de que forma devem ser incluídas depende ainda do tipo de aplicação proposta. Por isso, não é possível estabelecer uma regra única aplicável a todos os sistemas de GLN. Ainda assim, os fatores anteriormente citados podem ajudar a guiar a construção do mecanismo de tomada de decisão para determinação de conteúdo (REITER; DALE; FENG, 2000), e essas informações são os objetos que servirão de entrada para os módulos seguintes em um sistema de GLN (REITER; DALE, 1997).

2.3 *Aprendizagem de máquina*

Nessa seção serão apresentadas brevemente as técnicas de aprendizagem de máquina utilizadas nesse trabalho, árvores de decisão binárias e *Support Vector Machine* (SVM). Também são descritas as métricas aplicadas para medição do desempenho dos modelos utilizados na pesquisa apresentada nos capítulos 5, 6 e 7.

2.3.1 Árvores de decisão binárias

Árvore de decisão binária é uma estrutura de dados hierárquica recursiva em que cada nó folha representa o rótulo de uma classe e cada nó de decisão (nó não folha) contém um teste sobre algum atributo. Para cada resultado do teste existe uma aresta para uma subárvore. Cada subárvore tem a mesma estrutura que a árvore. Uma árvore de decisão utiliza uma estratégia de dividir-para-conquistar: Um problema complexo é decomposto em sub-problemas mais simples. Recursivamente a mesma estratégia é aplicada a cada subproblema (BREIMAN et al., 1984; MONARD; BARANAUSKAS, 2003; RUSSEL; NORVIG, 2004; MICHALSKI; CARBONELL; MITCHELL, 2013).

2.3.2 *Support Vector Machine* (SVM)

Support Vector Machine (SVM) é uma abordagem de aprendizado de máquina altamente competitiva, podendo ser aplicada tanto para problemas de classificação quanto de regressão. A estratégia de aprendizado baseada em vetores-suporte utiliza a teoria de aprendizado estatístico Vapnik-Chervonenkis (VC) (VAPNIK; CHERVONENKIS, 1974) para desenvolver técnicas de aprendizado de máquina que maximizem a capacidade de generalização (LIMA et al., 2004; STEINWART; CHRISTMANN, 2008).

SVMs são utilizadas para tratar problemas com dados linearmente separáveis, através da utilização de vetores. Quando os dados não são linearmente separáveis, vetores de suporte adicionais são mapeados em um espaço de dimensão mais elevada. Após o mapeamento do problema para uma dimensão suficientemente grande, é possível gerar um hiperplano composto de n -dimensões que separe os dados em classes (CRISTIANINI; SHAW-TAYLOR, 2000). O objetivo no treinamento das SVMs é encontrar um hiperplano

que separa os dados com a maior margem possível. Quanto maior for essa margem, maior é a generalização do classificador (LARRANAGA et al., 2006; LORENA; CARVALHO, 2007).

Uma vantagem na utilização de SVMs em relação a outros algoritmos de aprendizagem é que ela pode ser analisada usando conceitos da teoria da aprendizagem computacional e, ao mesmo tempo, pode alcançar um bom desempenho quando aplicado a problemas reais (HEARST et al., 1998). Segundo Steinwart e Christmann (2008), há pelo menos três razões para o sucesso obtido com o uso de SVM: sua capacidade de aprender bem com apenas um número muito pequeno de parâmetros, sua robustez contra vários tipos de violação e *outliers* do modelo e sua eficiência computacional em comparação com vários outros métodos de aprendizagem de máquina.

2.3.3 Métricas

As principais métricas utilizadas nesse trabalho são precisão, revocação, medida F (*F-Measure*), coeficiente de Sorensen-Dice (coeficiente Dice) e teste de Wilcoxon. A precisão de um classificador é a razão entre verdadeiros positivos e a soma dos verdadeiros positivos e falsos positivos. A revocação é a razão entre os verdadeiros positivos e a soma dos verdadeiros positivos e falsos negativos. A medida F é uma média harmônica entre a precisão e a revocação. A medida F é um indicativo de desempenho de um modelo que, de forma geral, aponta que quanto mais próximo a medida F estiver de 1, melhor é o modelo, e resultados mais próximos de 0 sugerem que o modelo tem desempenho baixo (COSTA et al., 2007). O coeficiente Dice é um método estatístico para medir a similaridade entre dois conjuntos, e é calculado conforme a figura 6.

Figura 6 – Cálculo do coeficiente de Sorensen-Dice

$$dice(A, B) = \frac{2 \times |A \cap B|}{|A| + |B|}$$

Fonte: Adaptado de Dice (1945) e Sørensen (1948)

O teste de Wilcoxon é um teste de hipóteses não paramétrico, usado para testar diferenças nas distribuições populacionais. Pode ser aplicado para comparar duas amostras relacionadas, amostras emparelhadas ou medidas repetidas em uma única amostra, e verificar se os postos médios populacionais diferem (WILCOXON, 1945).

2.4 Modelos de personalidade humana

A palavra “personalidade” é originária do latim “persona”. Persona significa máscara, e vem da máscara utilizada no teatro grego para representar as emoções dos atores (COSTA et al., 2016; HALL; LINDZEY; CAMPBELL, 2000).

Uma definição clássica de personalidade segundo Allport (1937) é “A personalidade é a organização dinâmica dos sistemas psicofísicos que determinam, no interior do indivíduo, seus ajustamentos únicos ao meio”. Ainda segundo Allport (1937), a personalidade é o que faz com que os indivíduos comportem-se de forma coerente ao longo do tempo, com características “notavelmente reconhecíveis”. E, mesmo em diferentes ambientes, o comportamento de um indivíduo pode mudar, mas sua personalidade, em essência, não se altera.

Personalidade é uma palavra ampla, utilizada em diferentes contextos e diferentes maneiras. Pode, por exemplo, definir habilidades sociais ou atração social. Pode também ser considerada como a impressão causada por um indivíduo, ou a eficiência em produzir sentimentos e emoções positivas ou negativas a outros indivíduos (ANDRADE, 2008; HALL; LINDZEY; CAMPBELL, 2000; LUNDIN, 1977).

A personalidade costuma ser estável no decorrer da vida de um indivíduo, ainda que modificações possam ocorrer. Ou seja, a personalidade é formada desde a infância e continua a desenvolver-se até a idade adulta sem grandes modificações, a não ser que haja algum tipo de gatilho para essa mudança, como um grande trauma (COSTA et al., 2016).

A definição adequada do termo personalidade depende da linha teórica a qual a pesquisa está relacionada (ANDRADE, 2008) e nenhuma definição pode ser generalizada (HALL; LINDZEY; CAMPBELL, 2000). Segundo Pasquali (2003), personalidade é um conceito tão amplo que cada psicólogo tem uma definição única para o termo.

Nesse trabalho, dentre as abordagens existentes para definir e referir-se à personalidade, foi selecionado o modelo de personalidade humana dos cinco grandes fatores (CGF). Esse modelo objetiva mapear a personalidade através da determinação dos grandes fatores - “Amabilidade”, “Neuroticismo”, “Extroversão”, “Abertura à experiência” e “Conscienciosidade” - e da combinação entre esses fatores (ANDRADE, 2008). Esses fatores não representam uma teoria específica, mas consolidam as análises dos termos em

língua natural que indivíduos usam para se descreverem e descreverem aos outros (JOHN; NAUMANN; SOTO, 2008).

O modelo CGF não tem como objetivo substituir os modelos já existentes, e sim ser um integrador entre as diferentes teorias de personalidade. Inclusive o “Grande” no nome do modelo não significa superioridade, e sim que todos os fatores de personalidade do modelo CGF são extremamente amplos (JOHN; NAUMANN; SOTO, 2008).

Há quatro pilares que justificam a amplitude do modelo dos cinco grandes fatores: (1) estudos demonstram que os fatores de personalidade manifestam-se em padrões de comportamento; (2) os fatores de personalidade são encontrados em diversas teorias da personalidade, e também em descrições linguísticas de personalidade; (3) os fatores de personalidade foram encontrados em diferentes idades, sexos, raças e nacionalidades; e (4) há evidências que sugerem que os fatores de personalidade possuem base biológica (ANDRADE, 2008; COSTA; MCCRAE, 1992a).

2.4.1 Os cinco grandes fatores

Nesta seção o modelo CGF é descrito, assim como os inventários de personalidade para esse modelo validados para o contexto brasileiro, destacando-se o Inventário dos Cinco Grandes Fatores de Personalidade (IGFP-5) a ser utilizado no presente trabalho. Inventários são testes que avaliam traços comportamentais, a fim de colher indícios sobre a personalidade de um indivíduo.

O modelo dos cinco grandes fatores tem sido considerado um padrão para avaliar personalidade, dadas as evidências de sua aplicabilidade nos mais diferentes contextos (MAIRESSE, 2008; NUNES et al., 2007). Esse modelo considera os seguintes fatores: “Abertura à Experiência”, “Conscienciosidade”, “Extroversão”, “Amabilidade” e “Neuroticismo” (ANDRADE, 2008). Esses fatores de personalidade são amplamente aceitos como as dimensões mais importantes da variação comportamental de seres humanos (NORMAN, 1963; GOLDBERG, 1990).

O fator “Abertura à Experiência”, também chamado de “Cultura”, “Imaginação” ou “Intelecto”, reflete a complexidade, abertura e profundidade da mente humana. Os indivíduos francos, artísticos, criativos, originais e espirituosos têm pontuação alta nesse fator. Indivíduos superficiais e conservadores têm pontuação baixa (ANDRADE, 2008).

O fator “Conscienciosidade”, também chamado de “Falta de impulsividade” ou “Vontade”, reflete o controle de impulsos, a disciplina, execução de deveres e cumprimento de obrigações e compromissos. Os indivíduos que possuem alta pontuação para esse fator são cuidadosos, pontuais, organizados e confiáveis. Os indivíduos com baixa pontuação são descuidados, desorganizados e pouco confiáveis (ANDRADE, 2008).

O fator “Extroversão”, também chamado “Expansão”, reflete energia, atividade, dominância e emoções positivas. Alta pontuação é comum aos indivíduos ativos, entusiasmados, sociáveis e falantes. Baixa pontuação é comum aos indivíduos retraídos, submissos e quietos (ANDRADE, 2008).

O fator “Amabilidade”, também chamado de “Agradabilidade” ou “Sociabilidade”, reflete altruísmo, modéstia e interesse pelo outro. Indivíduos com alta pontuação nesse fator costumam ser cooperativos, agradáveis e amáveis. Os com baixa pontuação podem ser frios e indelicados (ANDRADE, 2008).

Finalmente, o fator “Neuroticismo”, também chamado de “Instabilidade emocional”, reflete a estabilidade ou o controle emocional. Indivíduos com alta pontuação nesse fator podem ser nervosos, exageradamente sensíveis e preocupados. Indivíduos com baixa pontuação costumam ser estáveis, calmos e tranquilos. Enquanto a “Conscienciosidade” está relacionada à disciplina e controle de impulsos, de forma geral, esse fator está relacionado ao controle das emoções (ANDRADE, 2008).

Cada fator do modelo CGF é composto por um conjunto de traços. No quadro 2 é possível visualizar adjetivos representando esses traços associados aos extremos dos fatores do modelo CGF. É importante ressaltar que a personalidade do indivíduo é composta pelos cinco fatores, e esses fatores não necessariamente são extremos, portanto determinada pessoa pode ter várias das características associadas aos diferentes fatores de personalidade. Fatores de personalidade não podem ser mapeados como variáveis discretas. Ao contrário, para se aproximar à realidade, é necessário mapeá-los de forma contínua, e sobrepostos uns aos outros (MAIRESSE; WALKER, 2011).

Para medir a personalidade de um sujeito não há um “padrão de ouro”, apenas técnicas que dão *insights* de como é a personalidade deste. Algumas técnicas existentes incluem o autorrelato, ou seja, inventários que o indivíduo responde sobre sua própria personalidade, o relato de observador, onde o observador responde ao inventário sobre o observado, os testes não projetivos, em que o sujeito realiza um teste sem saber como seu resultado afeta a avaliação da personalidade (por exemplo, o teste da mancha de tinta

Quadro 2 – Adjetivos associados aos extremos de cada fator do modelo CGF

Fator	Alto	Baixo
Abertura	criativo, intelectual, imaginativo, curioso, culto, complexo	mente estreita, conservador, simples
Conscienciosidade	competente, disciplinado, obediente, cuidadoso, organizado	desorganizado, impulsivo, não confiável, descuidado, esquecido
Extroversão	moderado, assertivo, sociável, ativo, espontâneo, otimista, falador	tímido, quieto, conservador, passivo, solitário, temperamental, sem alegria
Amabilidade	confiável, amigável, atencioso, generoso, útil, altruísta	hostil, egoísta, suspeito, não cooperativo, malicioso
Neuroticismo	neurótico, ansioso, deprimido, consciente, sensível, vulnerável	calmo, temperado, de confiança, pacífico, confiante

Fonte: Traduzido e adaptado de Mairesse e Walker (2011)

Rorschach), os dados externos, onde a personalidade é inferida através de informações objetivas sobre o sujeito, como o histórico empregatício, por exemplo, e a indução emocional, em que emoções são provocadas no sujeito, e suas reações são analisadas. Os métodos mais utilizados são geralmente os dois primeiros (autorrelato e relato de observador), por sua simplicidade na aplicação (MAIRESSE, 2008).

2.4.2 Inventários CGF

Se, por um lado, o modelo CGF foi um consolidador entre diferentes teorias da personalidade, por outro, surgiram diversos inventários para determinar os fatores de personalidade de cada indivíduo (JOHN; NAUMANN; SOTO, 2008). Um inventário é uma espécie de questionário com itens a serem respondidos em uma escala, por exemplo, de 1 a 5. Conforme Eysenck (1991), foram desenvolvidos literalmente centenas de inventários incorporando milhares de traços, em grande parte sobrepostos, mas também contendo especificidades, onde cada achado empírico é relevante apenas para um fator de personalidade específico.

Essa realidade começou a mudar com o trabalho de Costa e McCrae (1985). Nele foi projetado um inventário chamado *NEO Personality Inventory* (NEO-PI). A sigla NEO identificava que era capaz de mapear as dimensões “Neuroticismo”, “Extroversão” e “Abertura” (do inglês *Openness*), com seis facetas conceituais derivadas de cada fator de personalidade.

Em seguida, no trabalho de Costa e McCrae (1992b), este inventário foi revisado e nomeado *NEO Personality Inventory Revised* (NEO-PI-R). Nessa versão, foi feita a inclusão

dos fatores “Amabilidade” e “Conscienciosidade”, que permitiu a medição diferenciada de cada fator de personalidade em termos de seis facetas mais específicas para cada um (JOHN; NAUMANN; SOTO, 2008), conforme a lista a seguir. Para cada faceta de cada fator de personalidade, foram listados 8 itens para auxiliar a classificação, resultando em um modelo de 240 itens (COSTA; MCCRAE, 1992b).

- **Neuroticismo:** Ansiedade, raiva/hostilidade, depressão, embaraço/constrangimento, impulsividade, vulnerabilidade;
- **Extroversão:** Acolhimento, gregarismo¹, assertividade, atividade, busca de sensações, emoções positivas;
- **Abertura:** Fantasia, estética, sentimentos, ações variadas, ideias, valores;
- **Amabilidade:** Confiança, franqueza, altruísmo, complacência, modéstia, sensibilidade;
- **Conscienciosidade:** Competência, ordem, senso de dever, esforço por realizações, autodisciplina, ponderação.

Para fornecer uma medida mais rápida para os fatores de personalidade, Costa e McCrae (1989) e Costa e MacCrae (1992) propõem uma versão abreviada baseada em análise fatorial de item por nível das versões do NEO-PI (COSTA; MCCRAE, 1985) e NEO-PI-R (COSTA; MCCRAE, 1992b) chamada NEO-FFI, de *NEO-Five Factor Inventory*. A escala de 12 itens por fator de personalidade do FFI consiste em itens que caracterizam fortemente cada um dos cinco fatores do modelo CGF. Com isso, o inventário foi reduzido de 240 itens para 60 itens (COSTA; MACCRAE, 1992). O conteúdo de cada item das escalas foi ajustado para garantir a cobertura adequada das facetas. Entretanto, essas escalas representam os elementos centrais de cada fator do modelo CGF, e, portanto, não representam igualmente cada uma das seis facetas que definem cada fator (JOHN; NAUMANN; SOTO, 2008). Ainda assim, o inventário NEO-FFI atingiu índices de 68% a 86% de confiança para caracterizar os domínios do modelo CGF (COSTA; MACCRAE, 1992). Esse inventário é utilizado em situações em que há pouco tempo disponível e para obter informações gerais sobre a personalidade do indivíduo examinado (ANDRADE, 2008).

Para atender à necessidade de um instrumento curto que medisse os fatores do modelo CGF em indivíduos, o Inventário dos Cinco Grandes Fatores (BFI - *Big Five*

¹ Nos seres humanos, a tendência a desejar sempre a companhia de outrem; sociabilidade.

Inventory) de 44 itens foi construído (JOHN; NAUMANN; SOTO, 2008). O BFI é um instrumento de autorrelato rápido, composto por afirmações simples, e respondido em uma escala *likert* de 5 pontos, variando de “Discordo totalmente” a “Concordo totalmente”, e que objetiva avaliar a personalidade baseado no modelo CGF. O BFI foi desenvolvido para ser uma representação canônica dos fatores de personalidade que pretende captar seus principais elementos baseado nos achados de estudos, amostras ou instrumentos anteriores. O objetivo na criação do BFI foi desenvolver um inventário breve que permita avaliação flexível e eficiente dos cinco fatores, sem a necessidade de medidas diferenciadas para cada uma das facetas (JOHN; NAUMANN; SOTO, 2008).

Esse inventário usa frases afirmativas, e não adjetivos simples, porque em Goldberg e Kilkowski (1985) foi identificado que as respostas aos inventários foram mais consistentes quando um item foi acompanhado de uma definição. Essas frases são elaboradas considerando um ou dois dos adjetivos relacionados a cada fator de personalidade, e são a informação central em torno da qual cada afirmação é construída. Por exemplo, para o fator “Abertura”, o adjetivo “criativo” tornou-se a frase “É inventivo, criativo.” e o adjetivo “perseverante” do fator “Conscienciosidade” é refletido na frase “Insiste até concluir a tarefa ou o trabalho.”. Com isso, o inventário BFI mantém as vantagens identificadas em trabalhos anteriores na utilização de adjetivos, como brevidade e clareza, e evita algumas fraquezas identificadas nesses trabalhos, como ambiguidade. Apesar do inventário BFI possuir apenas 8 a 10 itens por fator de personalidade, inclusive com mais de um item em uma única afirmação, ele não reduz a cobertura de conteúdo, nem perde propriedades psicométricas (JOHN; NAUMANN; SOTO, 2008).

O inventário BFI possui uma versão em português para o contexto brasileiro denominada o Inventário dos Cinco Grandes Fatores de Personalidade (IGFP-5). O instrumento foi proposto em Andrade (2008) e foi aplicado a 5.089 respondentes das cinco regiões brasileiras, mas apresenta restrições de variabilidade devido à maioria dos indivíduos respondentes serem do sexo feminino (66,9%) e estudantes do Ensino Superior (79,0%), entre outros itens, como faixa etária e estado civil.

3 Revisão bibliográfica

Este capítulo objetiva discutir estudos de GLN baseada em personalidade, com um foco especial na tarefa de DC. Dentre estes estudos, o sistema PERSONAGE (MAIRESSE, 2008), em suas diversas versões, é o único exemplo completo de sistema de GLN baseado em fatores de personalidade, e por isso a revisão bibliográfica foi organizada predominantemente em torno deste sistema.

O capítulo é dividido em nove seções: as seções 3.1 e 3.2 apresentam os primeiros trabalhos de GLN com algum embasamento em fatores de personalidade. A seção 3.3 apresenta o sistema PERSONAGE. As seções 3.4 a 3.7 apresentam o progresso do sistema PERSONAGE e suas diversas variações e aplicações. A seção 3.8 discorre sobre outras ferramentas e aplicações que utilizam geração de língua natural e fatores de personalidade. Por fim, a seção 3.9 apresenta o desenvolvimento da discussão baseada nessa revisão bibliográfica.

3.1 O sistema *LetterGen*

O trabalho de Pautler e Quilici (1998) apresenta um modelo para “perlocuções sociais”, ou seja, comunicação social que seja compreendida da mesma forma pelo locutor e receptor. O trabalho integra trabalhos anteriores em geração de língua natural, psicologia social e estudos de comunicação, e descreve como este modelo foi usado para construir um sistema automatizado chamado *LetterGen* para gerar mensagens de e-mail e cartas socialmente adequadas.

O sistema proposto trabalha com os objetivos comunicativos e sociais gerais do usuário, questiona o usuário sobre subobjetivos e informações de *background* e gera o texto de uma mensagem apropriada pelo planejamento de atos de fala individuais. Este modelo foi implementado em um sistema que gera um e-mail socialmente apropriado em resposta a metas comunicativas especificadas pelo usuário.

Uma das informações de *background* importantes para o sistema proposto por Pautler e Quilici (1998) são os fatores de personalidade (tanto para a personalidade do emissor quanto para a do receptor), pois podem afetar a forma do texto gerado. As

informações de perfis de personalidade são solicitadas pelo sistema e utilizadas para gerar as opções de comunicação que melhor atendam às necessidades do usuário.

Além das metas explícitas indicadas pelo usuário, o sistema trabalha com alguns padrões:

- Reduzir custos, evitando aspectos indesejados de uma situação atual ou incipiente, como percepções sociais indesejadas.
- A manutenção do *status quo* de um ato porque um de seus efeitos reforçaria um aspecto desejado da situação atual. Por exemplo, dispor-se a ajudar outra pessoa para reforçar a autoimagem de pessoa generosa.
- Hábito baseado em fatores de personalidade gerando um ato para expressar um fator de personalidade.

A implementação do sistema proposto contém um modelo detalhado de geração de falas de acordo com as restrições informadas e os efeitos desejados. O sistema é capaz de gerar diferentes tipos de mensagens, incluindo iniciar ou terminar uma amizade, aplicar ou renunciar a um trabalho, felicitar ou consolar alguém, aceitar ou recusar um convite, encorajar ou desencorajar alguém, agradecer, e pedir desculpas a alguém.

Cada um dos tipos de mensagens inclui um modelo organizacional que coloca as frases geradas em uma ordem apropriada. A interação com o usuário do sistema *LetterGen* é feita a partir de um tipo de mensagem selecionada, e o sistema sugere pelo menos três atos de fala para o usuário escolher. Essas mensagens são desenhadas para descrever traços de personalidade específicos (por exemplo, simpático, consciencioso, neurótico) e emoções (por exemplo, gratidão, simpatia, alegria, amizade, angústia), baseadas em estudos psicolinguísticos anteriores.

O trabalho de Pautler e Quilici (1998) foi um dos primeiros encontrados na literatura que tratam de geração de língua natural considerando fatores de personalidade. Porém o sistema *LetterGen* tem uma base limitada gerada manualmente, apesar de apresentar variações de fatores de personalidade baseadas nos estudos psicolinguísticos, depende da inserção de novos tipos de mensagem em sua base para conseguir gerar outras “perlocuções sociais”.

3.2 Modelos de avatares autônomos

No trabalho de Scheutz e Römmer (2001) é apresentada a arquitetura de um agente interativo crível com personalidade, e que pode atuar em nome de um usuário em vários contextos de jogo *multiplayer* quando o usuário não está on-line. Em uma primeira etapa, o agente coleta a informação sobre a personalidade do usuário através de um inventário e, em seguida, integra estas informações ao sistema reativo do agente, que também implementa um sistema afetivo primitivo. Os agentes podem interagir com seus usuários através de um sistema de geração de língua natural integrado ao sistema deliberativo do agente e podem contar ao jogador o que aconteceu com o agente no jogo enquanto o usuário não estava presente.

Para obter as características da personalidade do jogador, foi utilizado um inventário baseado no modelo CGF. Além disso, durante o decorrer do jogo, o comportamento do jogador é coletado para comparar os traços de personalidade diagnosticados com o comportamento do jogador, alterando se necessário seu perfil.

Para obter os fatores da personalidade de forma breve e eficiente, Scheutz e Römmer (2001) escolhem perguntas que correspondem diretamente às cinco dimensões do modelo CGF. No exemplo a seguir, uma pergunta relacionada ao fator “Extroversão”.

“Você acha que fez um bom trabalho escrevendo um relatório sumário da reunião da semana passada, mas seu chefe rasga-o em pedaços, porque ele não gosta do formato. Como você reage?”

- (a) Peço desculpas e sugiro reformatar o relatório imediatamente.
- (b) Eu digo a ele para não ser ridículo.

No exemplo apresentado, se um indivíduo responder (a) identifica-se uma personalidade de mais submissão, e se responder (b) trata-se de uma personalidade com características de dominância. Outro tipo de questão importante para esse trabalho eram as questões para entender as preferências do usuário durante o jogo, como no exemplo a seguir.

“Suponha que você desejou comer um bife suculento durante o dia inteiro. Quando você finalmente pode sentar em um restaurante para comê-lo, uma

linda moça com traços estrangeiros caminha até você e pergunta se você poderia recomendar quaisquer pontos turísticos para ela ver. O que você faria?”

- (a) Diz a ela que está ocupado e chama o garçom.
- (b) Sorri para ela e pergunta se ela quer se juntar a você.
- (c) Levanta-se imediatamente e guia-a pela cidade.

Nesse exemplo, as respostas do usuário são então usadas para adaptar o sistema afetivo do agente, compreendendo se ele prioriza a fome ou o relacionamento, em jogos em que há construção de personagens e relacionamentos.

Em Scheutz e Römmer (2001) o modelo proposto é uma solução pragmática para criação de avatares autônomos em jogos *multiplayer*, e permite uma grande flexibilidade, já que o agente pode ser facilmente ajustado a vários contextos de jogo e até modificado durante um jogo. Por exemplo, é possível iniciar um jogo com um agente de usuário (com base em inventários iniciais e apenas algumas perguntas) e, em seguida, melhorar o agente conforme mais informações sobre o usuário tornam-se disponíveis.

Na arquitetura proposta, o sistema deliberativo fornece a interface básica de língua natural para a interação do usuário e adiciona componentes de planejamento e raciocínio que permitem aos projetistas ampliar a capacidade dos agentes e adaptá-los a ambientes de jogos específicos. A interface de língua natural usa um sensor de texto adicional para ler comandos do usuário e responde através de um gerador de texto adicional usando o sistema de geração de relatório afetivo simples (SARGS). SARGS é um sistema de GLN rudimentar, integrado na camada deliberativa, que pode ser incorporado a um sistema GLN híbrido de simulação de jogo, que use modelos de texto para fornecer informações sobre o estado atual do jogo.

O sistema SARGS é um sistema de GLN afetivo, que utiliza os estados afetivos do agente para a determinação do conteúdo, ou seja, para decidir quais aspectos da descrição de percepções, emoções e ações (pretendidas e realizadas) serão utilizados para um usuário com base nos seus estados afetivos atuais. Quanto maior a ativação de um estado, ou seja, quanto mais forte o diagnóstico de um fator de personalidade ou preferência do usuário, mais provável que o aspecto correspondente faça parte de uma descrição ou ação no jogo.

Para o planejamento e realização de sentenças, o sistema SARGS usa uma base gramatical de *templates* de diferentes tipos de frases, como sentenças declarativas, perguntas, entre outras. Para cada evento em um jogo, um *template* é selecionado aleatoriamente,

porém levando em consideração o *template* usado para descrever o evento anterior. Por exemplo, um *template* para descrever a percepção de um evento pode ser representado por: $S \Rightarrow TAdv N V NP$, onde S é uma sentença, $TAdv$ um advérbio temporal, N um substantivo, V um verbo, e NP um substantivo complementar ao verbo (SCHEUTZ; RÖMMER, 2001).

Para descrever um evento particular, as categorias gramaticais nos *templates* de frases são preenchidas com palavras de uma base de dados de palavras, que contém uma variedade de palavras para cada modalidade sensorial e possível percepção, cada estado interno e diferentes ações possíveis. O banco de dados também contém vários adjetivos, advérbios e conjunções, que podem ser usados para conectar frases e descrever propriedades de entidades percebidas (por exemplo, “legal”), bem como graus de “sentimentos”.

Regras especiais são usadas para lidar com a combinação entre eventos. Ao acompanhar o que muda e o que permanece igual entre dois eventos, o sistema SARGS determina quais estados internos usar, escolhendo apenas aqueles que realmente mudaram dentro de um intervalo de tempo. Em seguida, o sistema SARGS usa advérbios e locuções adverbiais como “ainda”, “novamente”, “ainda não”, e afins, para descrever os eventos, sejam eles percepções, estados internos, intenções ou ações executadas, em uma sequência de eventos. Além disso, o sistema SARGS usa outros advérbios e locuções adverbiais como “muito”, “não muito”, e afins, para descrever a intensidade de um sentimento, ou seja, o grau de ativação de um estado interno. Por exemplo, se um agente tiver um nível de excitação de 0.8, onde 1.0 significa o máximo de animação, e esse nível estava em 0.3 no evento anterior, o sistema SARGS poderia adicionar a seguinte expressão à sua descrição do evento atual: “(...) ainda mais animado (...)”.

Scheutz e Römmmer (2001) propõem uma arquitetura para agentes que pode representar usuários em ambientes de jogo. No entanto, a arquitetura testada não é aplicada em nenhum ambiente, e portanto a avaliação da aplicabilidade dessa arquitetura em jogos reais permanece pendente.

3.3 O sistema PERSONAGE

Nesta seção é apresentado o sistema PERSONAGE, um sistema de geração de textos de recomendação de restaurantes. O sistema PERSONAGE é o único projeto de

um sistema de GLN baseado em personalidade encontrado na literatura com descrição completa e detalhada.

O trabalho de Mairesse e Walker (2007) apresentou a primeira tentativa de incorporar fatores de personalidade a um sistema de GLN utilizando o sistema PERSONAGE. Nesse trabalho o fator “Extroversão” e seu oposto, “Introversão”, são tratados de forma que seja possível ao receptor da mensagem em uma comunicação perceber a variação linguística causada pelo fator. Essa percepção é possível devido ao tratamento dos diferentes parâmetros que controlam a geração automática de texto.

Em Mairesse e Walker (2007), primeiramente foram mapeados os parâmetros relacionados ao fator “Extroversão” sugeridos pelas pesquisas em psicolinguística, sendo os principais a prolixidade, restauração, repetição, polaridade de conteúdo, polaridade de repetições, polaridade de reivindicações, concessões, polaridade de concessões, polarização, conteúdo positivo primeiro, autorreferências, complexidade das reivindicações, orações relativas, palavras de sinalização, conjunções, pontos finais (pausas), assunto implícito, negações, palavras que indicam opinião individual (e.g., “eu acho”, “parece-me”, e afins), confirmações, palavras de ênfase, pontos de interrogação, variação, repetição e frequência lexical.

Em seguida, cada parâmetro foi classificado de acordo com sua influência no fator de personalidade (“Extroversão” vs. “Introversão”) baseada na pesquisa psicolinguística, criando uma escala de variação para cada parâmetro. Essa escala foi composta de três pontos: alto, médio e baixo, sendo que a pontuação 0 significa “Extroversão” baixa (ou “Introversão” alta), 0,5 “Extroversão” média e 1 “Extroversão” alta para a influência de cada parâmetro na manifestação do fator de personalidade.

Para cada módulo específico do sistema de GLN, foram utilizados parâmetros relacionados às tarefas desse módulo. Para o módulo de determinação e ordenação do conteúdo foram utilizados as características e parâmetros relacionados a seguir:

- **Tamanho do conteúdo:** Pessoas extrovertidas falam mais do que introvertidas, (FURNHAM, 1990; PENNEBAKER; KING, 1999), apesar de não ser claro se geram mais conteúdo ou apenas conteúdo redundante. Para controlar esse item foram utilizados os parâmetros prolixidade, repetição e restauração.
- **Polaridade:** As pessoas extrovertidas costumam ser mais positivas. Os introvertidos engajam-se em mais “conversas problemáticas” e utilizam mais expressões de

insatisfação (THORNE, 1987). Para controlar essa característica, foi utilizado o conjunto de parâmetros de polaridade, definido como positivo ou negativo com base no valor escalar de cada parâmetro do conjunto. Esses parâmetros são a polaridade de conteúdo, de repetições, de reivindicações, de concessões e polarização.

- **Ordenação de conteúdo:** Apesar do conhecimento de que pessoas extrovertidas costumam ser mais positivas, não se sabe em que ordem o conteúdo positivo é gerado. O parâmetro conteúdo positivo primeiro controla em que momento o conteúdo positivo aparece e a que informação esse conteúdo é relacionado.

Após a determinação do conteúdo, o módulo seguinte trata do planejamento de sentenças, em que foram abordados os seguintes aspectos:

- **Seleção de modelo sintático:** Para tratar desse aspecto foi necessário utilizar os parâmetros de complexidade sintática, pois pessoas introvertidas tendem a utilizar construções sintáticas mais complexas. Também a autorreferência, pois extrovertidos fazem esse tipo de referência com mais frequência. E a polaridade, anteriormente explicada, também associada ao modelo sintático.
- **Operações de agregação:** Nos extrovertidos, a agregação se dá com frases mais longas, porém utilizando construções mais simples e termos informais. Para controlar esse aspecto são utilizados os parâmetros representando orações relativas, palavras de sinalização e conjunções.
- **Transformação pragmática:** Para diferenciar esse item foi necessário abordar os seguintes parâmetros: palavras que indicam opinião individual, ponto de interrogação, negações e assunto implícito. Introvertidos utilizam menos palavras implícitas, porém mais interrogações, e mais palavras que indicam opinião e negações.
- **Escolha lexical:** Os introvertidos tendem a usar um vocabulário mais rico, e isso é controlado através do parâmetro frequência lexical.

A principal hipótese em Mairesse e Walker (2007) foi a de que a língua gerada por parâmetros variáveis orientados pela pesquisa psicolinguística pode refletir “Extroversão” ou “Introversão”. Para testar essa hipótese, três juízes especialistas avaliaram um conjunto de enunciados gerados como se tivessem sido proferidos por um amigo respondendo a um pedido de recomendação de restaurantes. Estes enunciados foram gerados para manipular sistematicamente os parâmetros de “Extroversão” / “Introversão”. Os juízes classificaram cada enunciado para a “Extroversão” percebida, respondendo às duas perguntas que

medem esse fator no Inventário de Personalidade de Dez Itens (RAMMSTEDT; JOHN, 2007), uma versão simplificada do inventário BFI.

Para avaliar se o sistema PERSONAGE gera língua reconhecível como “Extroversão” ou “Introversão”, foi feito um teste t de amostra independente entre as classificações médias de 40 expressões de “Introversão” e 40 expressões de “Extroversão”. As expressões de “Introversão” têm uma classificação média de 2,96 em uma escala que vai de 0 a 7, enquanto as expressões de “Extroversão” têm uma classificação média de 5,98 na mesma escala. Além disso, ao dividir os dados em duas porções de igual tamanho em torno da classificação de valores neutros, os índices de enunciados do sistema PERSONAGE caem no intervalo previsto pelo conjunto de parâmetros em 89,2% das vezes. Expressões de “Extroversão” também são consideradas ligeiramente mais naturais do que as de “Introversão”.

Outra abordagem considerada em Mairesse e Walker (2007) é a sobregeração com parâmetros aleatórios, seguida de classificação por meio de um modelo estatístico treinado a partir do *feedback* de juízes humanos. Esta abordagem oferece suporte à geração de enunciados para qualquer valor de “Extroversão” / “Introversão” de entrada, e também consegue identificar quais parâmetros afetam a percepção dos juízes humanos. Todos os modelos de aprendizado de máquina utilizados (regressão linear, árvore de decisão, máquinas de vetores de suporte com kernel linear, e máquinas de vetores de suporte com kernel radial) superaram o *baseline* de 0,83 de erro absoluto médio ($p < 0,05$), mas o modelo linear obtém o melhor desempenho com um erro absoluto médio de 0,65. O melhor modelo produz um coeficiente de correlação de 0,59 com as avaliações dos juízes. Esta correlação é maior do que as correlações entre pares de juízes, sugerindo que o modelo proposto pode funcionar como um substituto ao julgamento humano.

O trabalho apresenta também os resultados de um experimento demonstrando que se pode treinar um modelo que corresponda ao desempenho humano na atribuição de “Extroversão” ou “Introversão” ao texto gerado por um sistema de GLN. O trabalho mostra que as expressões geradas no domínio de recomendação de restaurante podem manifestar personalidade, mas em Mairesse e Walker (2007) é enfatizada a necessidade de investigar quais atos de fala reconhecidamente manifestam personalidade em outros domínios restritos para garantir a efetividade do modelo proposto.

Nesse trabalho é demonstrado que o mapeamento dos parâmetros baseados nos estudos psicolinguísticos foi efetivo, porém podem haver parâmetros adicionais aos quais a compreensão psicolinguística pode ser mapeada. Além disso, esse é um dos poucos

trabalhos com uma aplicação completa de GLN com fator de personalidade, apresentando uma descrição detalhada de toda a arquitetura do sistema de GLN, incluindo o módulo de determinação de conteúdo, apresentado detalhadamente na seção a seguir.

3.3.1 A determinação de conteúdo no sistema PERSONAGE

No trabalho seguinte de Mairesse (2008), identificou-se que é possível aplicar as variações linguísticas causadas pelos diferentes fatores de personalidade a um sistema de GLN, e essa aplicação foi realizada no sistema PERSONAGE. Segundo o autor, a maioria dos estudos existentes não foca na geração automática de uma linguagem, e sim em entender as relações entre marcas linguísticas e diferentes personalidades dos indivíduos. Assim, nesse trabalho os parâmetros que afetam a personalidade foram aplicados em um sistema de geração de língua natural no domínio de recomendação de restaurantes. Esse é o primeiro trabalho do autor a trabalhar com todos os fatores de personalidade do modelo CGF.

Na figura 7 é apresentada a arquitetura detalhada do sistema PERSONAGE, e suas ferramentas complementares. Segundo Mairesse (2008), o sistema segue os padrões de arquitetura de sistemas de GLN propostos em Reiter, Dale e Feng (2000).

O primeiro componente na figura 7 é o módulo de determinação de conteúdo, que especifica a estrutura da informação a ser veiculada. A árvore de plano de conteúdo resultante desse módulo é processada pelo planejador de sentenças, que seleciona modelos sintáticos para expressar proposições individuais, e agrega-os para produzir a estrutura sintática completa da mensagem. O componente de inserção dos marcadores pragmáticos modifica a estrutura sintática para produzir os vários efeitos pragmáticos, dependendo das regras de inserção de marcadores. O próximo componente seleciona o termo mais apropriado para cada palavra de conteúdo, dados os parâmetros de escolha lexical. Finalmente, o realizador *RealPro* (LAVOIE; RAMBOW, 1997) converte a estrutura sintática final em uma sequência de caracteres aplicando regras gramaticais de superfície, como inflexão morfológica ¹ e inserção de palavra de função ².

O módulo de DC do sistema PERSONAGE visa a aprimorar os objetivos comunicativos determinados pelo usuário, e selecionar e estruturar o conteúdo proposicional de

¹ A modificação de uma palavra para expressar diferentes categorias gramaticais

² Palavras que expressam relações gramaticais com outras palavras dentro de uma frase

Figura 7 – Arquitetura do gerador PERSONAGE

Fonte: Traduzido e adaptado de Mairesse (2008)

forma a obter texto de recomendações de restaurantes gerados com os diferentes traços de personalidade reconhecíveis.

Em Mairesse (2008) fatores de personalidade foram mapeados em parâmetros aplicados ao módulo de determinação de conteúdo para selecionar o que é dito, e em que ordem é dito. Para fins de aprendizado, primeiro foram geradas mensagens abrangendo toda a gama de parâmetros. Em seguida, juízes humanos avaliaram o resultado do primeiro passo apoiados em um teste de personalidade padrão, e então foram calculados os valores de combinação de fatores para cada mensagem e eventuais outras características da mensagem. Com isso, foi possível treinar um modelo estatístico para prever as decisões de geração a partir das classificações dos juízes e gerar a personalidade-alvo usando os valores dos parâmetros previstos, através da estimação dos valores dos parâmetros.

A hipótese de pesquisa em Mairesse (2008) é a de que os marcadores linguísticos de emoções relacionadas a um fator de personalidade podem ser usados para transmitir esse fator no diálogo, para todos os fatores do modelo CGF. Esta hipótese é motivada pela associação de fatores de personalidade a uma tendência a expressar ou reprimir emoções específicas durante um longo período.

Para testar esta hipótese, foram mapeados sistematicamente os parâmetros que afetam as decisões de linguagem que podem ser implementadas em um sistema de GLN para cada fator do modelo CGF. Alguns fatores, por exemplo, “Extroversão” foram mais amplamente estudados do que outros, e seus mapeamentos são, portanto, mais abrangentes e mais propensos à generalização.

O mapeamento dos parâmetros para a determinação do conteúdo relacionados à “Extroversão” não mudaram em relação aos apresentados em Mairesse e Walker (2007). No entanto, o mesmo mapeamento foi feito para os demais fatores do modelo CGF. Foram definidos parâmetros para os diversos resultados encontrados nos estudos psicolinguísticos, de forma a refletir os resultados psicolinguísticos nos textos de recomendação de restaurante gerados pelo sistema PERSONAGE.

O fator “Neuroticismo” é amplamente associado à negatividade, ainda mais do que a introversão, e pode assim ser controlado pelos parâmetros de *polaridade de conteúdo*, de *concessão* e de *repetição*. Os neuróticos também produzem mais repetições lexicais, e um gerador neurótico teria um valor de parâmetro de *repetições* elevado.

Adicionalmente, sugere-se que neuróticos provavelmente apresentam conteúdo positivo primeiramente em seus enunciados, e por isso teriam um alto valor no parâmetro de *conteúdo positivo primeiro*, enquanto que os indivíduos mais estáveis terminariam seus enunciados com conteúdo mais positivo para ter um maior impacto argumentativo. Seguindo o mesmo pressuposto, o neuroticismo foi associado a um alto valor do parâmetro *polarização*, isto é, a produção de conteúdo mais extremo (independentemente de ser positivo ou negativo).

Indivíduos com pontuação mais alta no fator “Agradabilidade” produzem palavras de emoção mais positivas tanto em texto escrito quanto em conversa, e isso é refletido pelos parâmetros de *polaridade de concordância*, de *concessão* e de *repetição*. No domínio da apresentação de informações, o otimismo é associado à apresentação de informação positiva primeiro (parâmetro *conteúdo positivo primeiro*), enquanto a empatia é transmitida através do pedido de confirmações explícitas (parâmetro *pedido de confirmação*). A agressividade verbal é geralmente expressa através de ataques pessoais. Assim, a baixa agradabilidade foi associada à mitigação da competência do usuário, por exemplo, fazendo-se com que a resposta do sistema começasse com “todo mundo sabe disso...” (parâmetro *mitigação da competência*). Mairesse (2008) ainda explica que indivíduos com pontuação baixa no

fator “Agradabilidade” produzem mais rejeições, modeladas com um valor de parâmetro de *rejeição inicial* elevado.

Para o fator “Conscienciosidade”, a polaridade desempenha um papel importante, ainda que não esteja tão fortemente relacionada com esse fator quanto a outros. Indivíduos conscienciosos produzem palavras de emoção mais positivas em sua comunicação, enquanto os que têm pontuação mais baixa nesse fator utilizam mais palavras emocionais negativas. Por isso, os parâmetros relacionados à *polaridade* são utilizados para transmitir esta diferença ao planejamento do conteúdo.

A “Conscienciosidade” é associada também a solicitações de confirmações, contemplada no parâmetro de *confirmação de pedido pessoal*. Outra proposição é que os indivíduos com “Conscienciosidade” baixa não se incomodam em colocar a informação em perspectiva, refletida em um baixo valor do parâmetro *concessão*. Indivíduos conscientes também são diretos, evitando rejeições e repetições de informação, isto é, com valores de parâmetros *rejeição inicial*, *repetições* e *reaconfirmações* menores.

A relação entre o fator “Abertura” e a polaridade não é tão clara como em outros fatores de personalidade. Apesar disso, para a determinação de conteúdo em Mairesse (2008) foram levantados indícios de que indivíduos com pontuação alta no fator “Abertura” usam mais palavras de emoção positivas. Assim, foram modelados valores altos para todos os parâmetros de *polaridade* para esse fator.

O fator “Abertura” também foi associado a um valor do parâmetro *verbosidade* superior e um valor do parâmetro *rejeição inicial* inferior, com base no pressuposto de que a cultura e o intelecto se correlacionam com esse fator. Também foi levantada a hipótese de que os indivíduos com pontuação mais alta nesse fator tendem a colocar informações em perspectiva, interpretada no domínio de determinação de conteúdo do sistema PERSONAGE como a produção de concessões (ou seja, um alto valor de parâmetro *concessões*). Assume-se ainda que esses indivíduos usam mais formas de cortesia, resultando em confirmações de solicitação mais explícitas, ou seja, o parâmetro de *confirmações de solicitação* recebe valor mais alto.

Em Mairesse (2008) é testada também a hipótese de que a ansiedade pode ser projetada no domínio da apresentação de informações através de pedidos explícitos ou implícitos de confirmação, produzindo enunciados começando com “Eu não tenho certeza ... você disse X?”, por exemplo. A inserção destes marcadores é controlada, respectivamente, pelos parâmetros *pedido de confirmação* e *rejeição inicial*.

A tarefa de determinação de conteúdo foi implementada da seguinte forma. O primeiro passo foi converter os valores de atributos do objeto ao qual a mensagem fazia referência em um plano de conteúdo, isto é, em uma estrutura de alto nível que reflete o objetivo comunicativo geral da mensagem. No domínio de recomendação de restaurantes, em que o sistema PERSONAGE foi aplicado, esses atributos poderiam ser, por exemplo, qualidade do serviço, preço, etc. Este plano combina proposições que expressam informações sobre atributos individuais usando relações retóricas (RST) (WILLIAM; THOMPSON, 1988), ou seja, sempre que um intervalo de texto é mais essencial para transmitir a informação desejada, ele é referido como o núcleo da relação (N), enquanto o outro espaço de texto é definido como o satélite (S).

O sistema PERSONAGE trabalha com dois tipos de objetivos comunicativos: recomendação e comparação de restaurantes. Ele pode recomendar um restaurante de acordo com a solicitação do usuário. Por exemplo, o usuário deseja um restaurante italiano de preço médio localizado na Zona Sul, o sistema procurará na base de dados do domínio um restaurante que atenda a essas especificações e gerará o texto da recomendação refletindo a personalidade selecionada. Cada plano de conteúdo de recomendação contém um atributo (núcleo) sobre a qualidade geral do restaurante selecionado, e essa recomendação é justificada por um conjunto de proposições satélites descrevendo os atributos do restaurante. O plano de conteúdo é convertido automaticamente em uma estrutura de árvore, denominada árvore de plano de conteúdo.

O plano de conteúdo determina qual será o conteúdo do enunciado e a ordem desse conteúdo, por meio dos atributos do restaurante e dos parâmetros definidos para cada fator do modelo CGF. Os demais módulos do sistema PERSONAGE não serão descritos detalhadamente por não estarem diretamente relacionados ao objetivo da presente pesquisa.

Em Mairesse (2008) foram geradas 240 mensagens, sendo 80 de “Extroversão” alta ou baixa e 160 para a avaliação dos outros quatro fatores de personalidade. As 240 mensagens foram agrupadas em 20 conjuntos de mensagens geradas a partir do mesmo plano de conteúdo. Cada conjunto continha duas mensagens por fator de personalidade (quatro para “Extroversão”), geradas com configurações de parâmetros tanto para o extremo inferior quanto para o extremo superior de cada fator, e foram submetidos a três juízes humanos.

De forma geral, os resultados obtidos foram satisfatórios, pois na maioria dos casos os juízes reconheceram corretamente o extremo do fator de personalidade ao qual a mensagem gerada estava relacionada.

Um resultado interessante que emerge é que a “Extroversão” observada é a característica mais fácil de modelar a partir de conversas. Curiosamente, a “Amabilidade” é a característica mais difícil de modelar usando o mesmo conjunto de dados, produzindo os piores resultados. Modelos de “Abertura à experiência” funcionam melhor, com 63% de classificações corretas, 7% de melhoria em relação ao *baseline* da regressão e 61% de classificações corretas entre todos os pares no conjunto de teste. Modelos de outros fatores de personalidade funcionam apenas ligeiramente melhor do que o *baseline*, sugerindo que estes fatores de personalidade são geralmente de difícil detecção no texto.

Em Mairesse (2008) foi demonstrado que é possível, para juízes humanos, perceber a personalidade almejada através de uma frase gerada pelo sistema de GLN. Entretanto, é importante ressaltar que esse trabalho foi aplicado especificamente no domínio de recomendações de restaurantes. Supõe-se que seja possível estender a outros domínios obtendo resultados similares, porém não foi feita nenhuma variação de domínio neste trabalho. Além disso, o trabalho foi desenvolvido através da parametrização baseada em estudos psicolinguísticos, e não utilizou um corpus como o projeto aqui apresentado.

3.4 Avaliação e generalização do sistema PERSONAGE

Em Mairesse e Walker (2008a), é demonstrado que o gerador PERSONAGE pode produzir texto com personalidade reconhecível por avaliadores humanos para todas as características do modelo CGF. Também é testada a capacidade do sistema de variar a personalidade dos personagens em um sistema narrativo interativo existente, mostrando que algumas formas de variação podem ser obtidas automaticamente em um novo domínio, dependendo do nível de representação das expressões.

Para avaliar se o sistema gera texto que pode ser reconhecido como projeção de personalidade extrema, ou seja, pontuação extremamente alta ou extremamente baixa, para os diferentes fatores do modelo CGF, foram comparadas as classificações médias das 20 expressões que expressam o extremo inferior de cada fator do modelo CGF e as 20 expressões que expressam o superior. Testes t pareados mostram que os avaliadores podem

discriminar entre dois conjuntos de expressões extremas para cada fator ($p < 0,01$). Os resultados obtidos foram satisfatórios, pois os juízes humanos puderam reconhecer o fator de personalidade do enunciado gerado pelo sistema na maioria das vezes. Os resultados obtidos por Mairesse e Walker (2008a) mostram que o sistema PERSONAGE pode produzir variação de personalidade reconhecível para todos os fatores de personalidade do modelo CGF.

Em Mairesse e Walker (2008a) são fornecidos também exemplos que sugerem que o sistema pode projetar vários tipos de personalidade em um diálogo do sistema narrativo interativo *Façade* (MATEAS; STERN, 2003). Nesse experimento, os diálogos de exemplo do sistema *Façade* foram convertidos manualmente em uma sequência de estruturas sintáticas, processados em seguida pelo sistema PERSONAGE para inserção de marcadores pragmáticos e componentes de escolha lexical. Esse experimento, no entanto, não foi avaliado por juízes humanos, ou seja, não é possível garantir o reconhecimento dos fatores de personalidade na alteração dos diálogos do sistema *Façade*.

Segundo Mairesse e Walker (2008a), o sistema PERSONAGE é o primeiro sistema a demonstrar claramente que os cinco grandes fatores de personalidade podem ser manifestados por um gerador automático para gerar texto em aplicações de diálogo que requerem inteligência social.

3.5 *Variação estilística contínua no sistema PERSONAGE*

Essa seção discute os trabalhos de Mairesse e Walker (2008b), Mairesse e Walker (2009), Mairesse e Walker (2010) e Neff et al. (2010), extensões do sistema PERSONAGE que utilizam as técnicas de GLN baseadas em personalidade em uma escala contínua, permitindo assim a geração de mensagens que refletem múltiplos fatores de personalidade simultaneamente e de forma mais próxima ao desempenho humano.

No trabalho de Mairesse e Walker (2008b) é apresentada uma nova abordagem de GLN que aprende a produzir variações textuais reconhecíveis ao longo de uma dimensão estilística significativa, sem o custo computacional incorrido por técnicas de supergeração (*overgeneration*). Nesse trabalho é executada a primeira avaliação de um método de geração baseado em dados que projeta múltiplas personalidades simultaneamente e em escala

contínua. O método proposto é comparado a um gerador baseado em regras no mesmo domínio.

O sistema PERSONAGE é utilizado mais uma vez para a geração das variações do modelo CGF dentro do fator “Extroversão”, mas enquanto trabalhos anteriores trataram o fator de personalidade como um valor discreto, nesse caso o fator é tratado em uma escala contínua. O sistema PERSONAGE com a adição dessa funcionalidade foi chamado PERSONAGE-PE, que representa uma nova versão do sistema PERSONAGE em que as regras baseadas nas pesquisas psicolinguísticas foram substituídas por modelos estatísticos.

Para geração da base de dados utilizada foi necessário empregar um gerador de base para produzir múltiplas mensagens variando-se aleatoriamente seus parâmetros, coletar julgamentos humanos que classificassem a personalidade de cada mensagem, e projetar modelos estatísticos para prever os parâmetros dos julgamentos de personalidade utilizando técnicas de aprendizado de máquina.

Para a avaliação da proposta, foram usados os modelos com melhor desempenho para prever valores de parâmetros a partir das pontuações das personalidades-alvo, e foram geradas as mensagens de saída usando os valores previstos para os parâmetros. Essas mensagens foram apresentadas a 24 participantes, que deveriam avaliar a naturalidade e personalidade de cada uma. Os resultados obtidos foram comparados com a versão anterior do sistema PERSONAGE (baseada em regras) e, em média, tiveram um desempenho melhor, com mais naturalidade e maior reconhecimento pelos participantes leigos sobre os fatores de personalidade projetados.

A abordagem proposta em Mairesse e Walker (2008b) fornece uma maneira de projetar múltiplas dimensões simultaneamente, incluindo as dimensões relevantes como características nos dados e parâmetros de treinamento dos modelos. Ainda assim, este método não funcionou tão bem para o fator “Conscienciosidade” e são necessários outros estudos para identificar o porquê disso.

No estudo posterior de Mairesse e Walker (2009) é demonstrada a arquitetura detalhada do sistema PERSONAGE. Essa arquitetura pode ser visualizada na figura 8, em que se vê um modelo de padrão de sistema de GLN (caixas brancas) comparado à arquitetura do sistema PERSONAGE (caixas cinzas). Na figura 8 também é possível observar em quais módulos da arquitetura o sistema PERSONAGE utiliza parâmetros diferenciados para controlar a variação linguística baseada em personalidade.

Figura 8 – *Pipeline* da arquitetura de um sistema de geração de língua natural, com os componentes do sistema PERSONAGE nas caixas cinzas

Fonte: Traduzido de Mairesse e Walker (2009)

Os resultados obtidos em Mairesse (2008) são avaliados detalhadamente em Mairesse e Walker (2009), verificando quais parâmetros do sistema PERSONAGE realmente afetaram a percepção dos extremos dos fatores de personalidade e identificando as correlações para melhoria da calibração da ferramenta. Após essa calibração, a naturalidade das mensagens geradas pelo sistema PERSONAGE foi avaliada pelos juízes humanos, obtendo uma pontuação média de 4,59, em uma escala de 0 a 7. Embora a naturalidade possa ser melhorada adicionando-se restrições à ferramenta para evitar resultados inconsistentes, esses resultados podem ser considerados promissores dada a natureza extrema da personalidade modelada: mensagens produzidas por um ser humano com fatores de personalidade tão extremos também podem ser percebidos como antinaturais.

O trabalho seguinte de Mairesse e Walker (2010) explica como configurar a ferramenta de GLN PERSONAGE para os diferentes fatores de personalidade. Nele é utilizado novamente o domínio específico de recomendações de restaurantes. Embora a abordagem do sistema PERSONAGE possa ser trivialmente estendida a outros domínios de apresentação de informações modificando o dicionário de geração, estender a abordagem a novos objetivos comunicativos (por exemplo, solicitações) exige novas regras de transformação sintática. No entanto, acredita-se que tais regras podem ser implementadas de forma genérica para uma gama maior de objetivos comunicativos e reutilizadas em todas as aplicações.

Finalmente, em Neff et al. (2010), foca-se em um fator no modelo CGF, a “Extroversão”, e demonstra-se como os fatores verbais e não-verbais influenciam a percepção desse fator. Um experimento foi conduzido com um agente virtual que demonstra como a geração de língua, taxa gestual e o conjunto de parâmetros de desempenho de movimento podem variar para aumentar ou diminuir a “Extroversão” percebida.

Para testar a “Extroversão” verbal, foi utilizada a ferramenta PERSONAGE, e para a demonstração não-verbal foi utilizado o software de animação desenvolvido nos trabalhos anteriores dos autores (NEFF; KIM, 2009; NEFF et al., 2008).

O trabalho de Neff et al. (2010) também foi aplicado no domínio de recomendações de restaurantes. A geração verbal foi combinada com a não-verbal, e foram gravados vídeos (de animação) com os perfis retratados. A experiência foi feita com 40 participantes, que assistiram aos vídeos gerados via *web* e foram questionados sobre o nível de “Extroversão” e de naturalidade. Tanto a “Extroversão” de linguagem verbal, quanto a taxa gestual e desempenho gestual (indicadores da linguagem não-verbal) apresentaram correlação positiva com a “Extroversão” percebida. Cada um desses fatores tiveram impacto estatístico significativo na percepção dos participantes.

Com os resultados dos trabalhos de variação estilística do sistema PERSONAGE é possível identificar a importância de utilizar os fatores de personalidade variando em uma escala contínua, devido a sua maior proximidade às características de indivíduos reais e, por consequência, a obtenção de língua natural gerada pelo sistema com maior similaridade ao desempenho humano.

3.6 *Evolução do sistema PERSONAGE*

Em Mairesse e Walker (2011) o sistema PERSONAGE evolui para um gerador de língua natural altamente parametrizável. É apresentado um novo método para geração estatística de língua natural, utilizando modelos treinados para estimar parâmetros em dados sobre personalidade, e antecipando as decisões de geração necessárias para transmitir qualquer combinação de valores escalares ao longo dos cinco fatores de personalidade.

Dois pontos positivos encontrados são que, ao testar os modelos treinados, foi verificado que os avaliadores humanos, sem conhecimento prévio do sistema PERSONAGE, identificaram a personalidade proposta para a maioria dos fatores do modelo CGF. Além

disso, o método proposto para a estimação de parâmetros é um método de geração computacionalmente tratável e não requer nenhuma fase de supergeração.

O método de estimação de parâmetros proposto pode aprender a controlar a saída do gerador a partir dos dados, sem qualquer trabalho manual adicional ou ajuste de parâmetros. Outra vantagem desse método em relação a outras abordagens é que ele pode aprender a atingir quaisquer combinações de valores de pontuação escalar dos fatores do modelo CGF, enquanto que a manipulação manual de todas as formas possíveis de múltiplas dimensões estilísticas que podem afetar as expressões de saída requer um esforço considerável. Além disso, esse método pode facilmente modelar dimensões estilísticas contínuas, ao mesmo tempo que produzir tais modelos manualmente é inviável.

3.7 O sistema *PERSONAGE* em diferentes domínios

Nessa seção discute-se brevemente um conjunto de estudos que fazem aplicação do sistema *PERSONAGE* em domínios distintos ao de recomendações de restaurantes. Não houve modificações nas técnicas para aplicação dos fatores do modelo CGF, por isso, nessa seção são apresentadas as variações de domínio e os principais resultados obtidos.

Em Khosmood e Walker (2010) é apresentado o sistema *Grapevine*, um gerador de conversas estilo “fofocas”. A conversa de fofocas foi modelada com uma série de atos de fala controlados por um gerenciador de diálogo. Personagens foram modelados com fatores derivados do modelo CGF. A realização do texto usa o *RealPro* (LAVOIE; RAMBOW, 1997), um realizador estilisticamente aprimorado com o sistema *PERSONAGE* (MAIRESSE; WALKER, 2007). Os resultados mostraram que a linguagem gerada pelo sistema *Grapevine* foi suficientemente clara para que os participantes entendessem o conteúdo da fofoca. Porém, os resultados sobre a identificação das características dos fatores do modelo CGF indicam que há possivelmente espaço para a melhoria no motor de geração, pois nenhuma característica foi identificada corretamente sequer em 50% dos casos.

Em Walker et al. (2011a), é apresentado um método para a aprendizagem de modelos de estilo linguístico de personagens a partir de um corpus de diálogos cinematográficos, e esse método é testado em uma experiência de percepção. O método foi aplicado no contexto do *SpyFeet*, um protótipo de jogo de RPG. Os resultados obtidos foram positivos,

porém demonstram que ainda há muito trabalho a fazer para que seja possível aplicar o sistema de GLN PERSONAGE no jogo *SpyFeet*.

Em Walker et al. (2011b), dois modelos de geração de falas de personagem são comparados: um baseado em fatores de personalidade do modelo CGF utilizando o sistema PERSONAGE, e o outro derivado de um corpus de personagens e filmes do arquivo IMDb³. Esses modelos de estilo foram aplicados a falas de personagens para um jogo piloto de narrativa baseado em realidade aumentada chamado *Murder in the Arboretum*.

Utilizou-se uma métrica quantitativa para estimar a qualidade de um modelo de personagem com o objetivo de prever a qualidade do modelo sem utilizar experimentos de percepção. Identificou-se que modelos de personagem baseados em corpus derivados de personagens individuais são frequentemente mais detalhados e específicos do que modelos baseados em personalidade, mas existe uma forte correlação entre a avaliação de personalidade de diálogo do personagem original e de mensagens geradas para o jogo *Murder in the Arboretum* que usam os modelos de personalidade derivados. Além disso, identificou-se que a combinação de vários personagens com o mesmo arquétipo não resultou em ganhos em relação à utilização de um único personagem. Ou seja, tanto o modelo baseado nas falas de um personagem quanto o modelo baseado em fatores de personalidade aplicados às falas do jogo *Murder in the Arboretum* possuem vantagens e desvantagens, e podem ser selecionados dependendo do objetivo da comunicação (WALKER et al., 2011b).

Em McCormick (2012) é proposta a aplicação do sistema PERSONAGE em um jogo de computador com o objetivo de atribuir naturalidade e diferentes fatores de personalidade aos personagens do jogo. Identificou-se que, nas expressões geradas pelo sistema PERSONAGE, a combinação de fatores de personalidade que pode-se transmitir consistentemente e com naturalidade é muito pequena para que o sistema seja um componente de GLN suficientemente sofisticado em um sistema de diálogo de personagem em um jogo. No entanto, se a capacidade do sistema PERSONAGE de gerar enunciados com naturalidade for melhorada, mantendo-se a manifestação dos fatores de personalidade reconhecíveis, o sistema poderia então ser aplicado a este domínio.

Em Lukin, Ryan e Walker (2014) o sistema PERSONAGE foi utilizado para a aplicação de variações estilísticas na geração de língua natural como uma maneira de ajudar os autores a reduzir a carga autoral de escrever o diálogo diversas vezes, ou passar para roteiristas reescreverem de diversas formas, sem garantias de que a essência pretendida de

³ *The Internet Movie Script Database*, disponível em <http://www.imsdb.com.br>

alguma fala não seria perdida. Essa primeira avaliação mostrou que é possível perceber diferenças entre as características dos personagens utilizando-se o modelo proposto. Além disso, a aplicação permite que o modelo proposto seja estendido a outros personagens de forma simples.

Em Lukin, Reed e Walker (2015), é criado um planejador de frases parametrizado para a geração de histórias interativas (do termo em inglês *storytelling*) que fornece operações de agregação, variações no discurso e no ponto de vista. Além disso, é apresentada uma avaliação por parte do usuário de diferentes narrativas geradas, no domínio de informações turísticas. Para as variações de personalidade, o sistema PERSONAGE é utilizado. O experimento indicou que o usuário prefere receber as informações originais (escritas por um humano) do que qualquer uma das variações geradas pelo planejador. Porém, as variações geradas obtiveram notas altas em naturalidade, e foram considerados compreensíveis pelos participantes.

Em Bowden et al. (2016) é apresentada a ferramenta *M2D*, de *storytelling*, que transforma monólogos em diálogos levando em consideração a personalidade dos personagens do diálogo que será gerado. Foram realizadas várias experiências para identificar se as narrativas dialógicas são mais envolventes e se as variantes geradas automaticamente em forma linguística que correspondem a diferenças de personalidade podem ser reconhecidas em um diálogo. Para a geração da personalidade dos personagens do diálogo, a estrutura do sistema PERSONAGE foi utilizada. As conclusões são que esse tipo de narrativa com diálogo realmente gera maior engajamento dos ouvintes e que, usando o fator de personalidade “Extroversão”, foi possível a juízes humanos perceberem a personalidade do personagem no diálogo.

Em Aly e Tapus (2016), o objetivo foi o de mapear o comportamento verbal humano para um robô com comportamento verbal e não verbal baseado na dimensão da personalidade “Extroversão”. A correspondência da personalidade robótica e do princípio de atração de similaridade é explorada (REEVES; NASS, 1996; NASS; LEE, 2001; TAPUS; MATARIC, 2008; PARK; JIN; POBIL, 2012), além dos diferentes efeitos do comportamento robótico expresso pela fala e pelos gestos do robô. A fala foi adaptada para a dimensão “Extroversão” utilizando-se o sistema PERSONAGE. Nos experimentos, foi identificada a preferência dos participantes pelo robô com pontuação no fator “Extroversão” similar à sua e, de forma geral, as falas e expressões do robô foram avaliadas como adequadas. Apesar dos resultados positivos, esse trabalho deve ser considerado experimental, pois teve apenas

20 participantes. Ainda assim, ele demonstra a importância de levar em consideração a personalidade na interação humano-robô, tanto na geração de língua natural quanto nos gestos.

Em Herzig et al. (2017), o sistema PERSONAGE é integrado ao modelo SEQ2SEQ (XU et al., 2017) de geração de respostas a solicitações de clientes em mídias sociais. Foram geradas respostas às solicitações reais de clientes, com diferentes fatores de personalidade. Essas respostas foram avaliadas por juízes humanos tanto em naturalidade quanto na identificação da personalidade projetada e tiveram resultados positivos em pouco mais de 60% das vezes, o que foi considerado um resultado satisfatório, porém insuficiente para aplicação com clientes reais.

3.8 Outras ferramentas e aplicações

Em Strauss e Kipp (2008) é apresentado o sistema ERIC, um agente incorporado para comentários afetivos em tempo real. Além do raciocínio sobre eventos que mudam dinamicamente, o sistema pode produzir uma língua natural coerente e comportamento não-verbal baseado em um modelo em camadas de afeto (personalidade, humor, emoção). O estudo mostra como o raciocínio, a geração de língua natural baseada em modelos e a avaliação afetiva podem ser implementados dentro do mesmo paradigma baseado em regras.

O sistema ERIC foi aplicado em dois domínios diferentes: uma corrida de cavalos virtual e um jogo de batalha de tanques para vários jogadores. O grau de generalidade do sistema foi avaliado de forma empírica, medindo-se o esforço necessário para mudar de um domínio para o outro. As variações de personalidade foram mapeadas usando-se o modelo CGF, e geradas por um módulo de inferência. De forma geral, o sistema ERIC demonstrou que é possível utilizar um agente incorporado para comentar esportes e competições, porém melhorias ainda são necessárias para que o agente tenha a mesma naturalidade e expressão de um comentarista humano.

Em Gabriel, Chen e Nichols (2015) é apresentada a ferramenta *InkWell*, um programa de GLN desenvolvido com o objetivo de ajudar os escritores a obter variações estilísticas em seus textos com base em escrita criativa e imitando aspectos de escritores especificados, e seus fatores de personalidade. A ferramenta é construída com base em um

processo de otimização que produz variações do texto fornecido, avalia essas variações quantitativamente e seleciona as que melhor mimetizam o escritor-alvo.

A ferramenta *InkWell* usa um *template* - uma especificação do texto original anotado no qual as palavras e as características das palavras que a ferramenta considera são variáveis - e produz revisões. Além disso, a ferramenta pode usar restrições como palavras curtas, palavras conservadoras, ou soar de forma similar a determinados escritores.

Por meio do estudo com uma escritora que trabalha com escrita criativa, verificou-se que a ferramenta *InkWell* pode ser usada para explorar palavras alternativas e metáforas, bem como para auxiliar a uma análise mais profunda do significado e da trajetória do projeto de um escritor. Além disso, com um experimento utilizando técnicas de Haiku⁴, observou-se que a ferramenta conseguiu assumir o papel de um escritor escolhendo suas próprias palavras (não simplesmente encontrando sinônimos) e usando um senso comum para trabalhar em direção à coerência.

Apesar dos bons resultados obtidos nos testes executados, um desafio restante são as transformações sintáticas. Por exemplo, o escritor Hemingway tende a usar frases curtas e muitos “e”s, e a ferramenta *InkWell* atualmente não pode transformar a estrutura da frase para coincidir o estilo deste autor, mas apenas fazer transformações triviais após a otimização.

3.9 Discussão

Os estudos que relacionam GLN e personalidade iniciaram há mais de vinte anos, porém os resultados mais expressivos nesse tema foram obtidos a partir do trabalho de Mairesse e Walker (2007), em que foi apresentado o sistema PERSONAGE. Grande parte dos estudos sobre o tema nos anos subsequentes é relacionado a esse sistema, com exceção dos trabalhos de Strauss e Kipp (2008) e de Gabriel, Chen e Nichols (2015).

Com a revisão bibliográfica foi possível identificar que há poucos estudos sobre o tema, mesmo analisando-se as publicações dos últimos vinte anos. A maioria dos estudos encontrados, além de serem vinculados ao mesmo sistema (PERSONAGE), trabalham com o modelo CGF e com a língua inglesa, e uma parcela considerável é vinculada ao domínio de recomendação de restaurantes. Há, entretanto, vários trabalhos aplicando a ferramenta em outros domínios, como fofocas em Khosmood e Walker (2010), personagens de filmes

⁴ Tipo de poesia oriental

em Walker et al. (2011a), personagens de jogos de computador em Walker et al. (2011b) e McCormick (2012), escrita criativa em Lukin, Ryan e Walker (2014), informações turísticas em Lukin, Reed e Walker (2015), *storytelling* em Bowden et al. (2016), personalidade na GLN combinada a gestos de um robô em Aly e Tapus (2016) e respostas a clientes em mídias sociais em Herzig et al. (2017).

O fator de personalidade com o maior número de estudos em GLN é a “Extroversão”, utilizada exclusivamente em Mairesse e Walker (2007), Neff et al. (2010) e Aly e Tapus (2016) e em todos os demais trabalhos dessa revisão, em conjunto aos demais fatores do modelo CGF. Além de, na maioria dos trabalhos estudados, foi o fator que obteve melhores resultados. Supõe-se que isso é devido a maior facilidade em mapeá-la e identificá-la na língua natural do que os demais fatores (MAIRESSE; WALKER, 2007; MAIRESSE, 2008).

Dos estudos levantados, poucos tratam da fase da GLN que executa a determinação de conteúdo. Os que a tratam não se aprofundam nesse tema em específico. Há pouca descrição sobre o assunto nos trabalhos aplicados, e a maioria deles utiliza o mesmo modelo de arquitetura detalhado no trabalho de Reiter, Dale e Feng (2000) e apresentado na figura 3 anterior. Além disso, não foram encontrados estudos de GLN para o Português do Brasil que tratam desta questão.

4 Construção do corpus b5

Neste capítulo será descrito o trabalho de coleta de dados para aplicação na presente pesquisa. De modo mais específico, foi realizada uma coleta de dados, para obter um corpus de textos descritivos de imagens acompanhado dos inventários de personalidade dos respectivos participantes. O corpus, denominado corpus b5 (RAMOS et al., 2018), foi coletado presencialmente de forma colaborativa com outros alunos do Programa de Pós-graduação em Sistemas de Informação (PPgSI) e da graduação em sistemas de informação da USP-EACH, e é composto de vários subcomponentes não aproveitados diretamente nesta pesquisa, e que não serão discutidos. Assim, o foco deste capítulo será uma porção específica do corpus b5, o subcorpus b5-text de descrições de imagens acompanhadas de informações de personalidade de seus autores.

4.1 Materiais

Foi desenvolvida uma ferramenta para coleta de dados na forma de experimentos presenciais com sujeitos humanos. Nela são solicitadas respostas aos 44 itens de um inventário de personalidade, e apresentada uma série de 10 imagens de estímulo a serem descritas em forma textual pelo participante, além de outras tarefas não relacionadas ao presente projeto. A ferramenta para coleta de dados foi desenvolvida na linguagem de programação PHP utilizando o servidor portátil Apache web livre XAMPP¹, e banco de dados MySQL.

Para cada uma das 10 imagens, há um *box* de texto para receber a descrição completa contendo algumas frases. A resposta do participante é armazenada em banco de dados, vinculada ao seu identificador e ao identificador de cada imagem.

4.1.1 Inventário de personalidade

O inventário de personalidade aplicado foi o IGFP-5, validado para o contexto brasileiro em Andrade (2008), e detalhado na subseção 2.4.2. Esse inventário possui 44 questões de múltipla escolha. Para cada questão, o participante deve indicar o quanto

¹ *Xampp: Apache + MariaDB + PHP + Perl*, disponível em <https://www.apachefriends.org/index.html>

discorda ou concorda da afirmação proposta através da seleção de 1 a 5 em uma escala *likert*.

As respostas das 44 questões do IGFP-5 foram utilizadas para calcular os valores das cinco dimensões de personalidade de cada indivíduo segundo o método padrão para este instrumento, detalhado em John, Donahue e Kentle (1991).

4.1.2 Imagens de estímulo

As imagens utilizadas para a descrição de cenas são originárias do banco de imagens GAPED (DAN-GLAUSER; SCHERER, 2011), uma base de imagens criada com o objetivo de aumentar a disponibilidade de estímulos emocionais visuais, para aplicação de pesquisa que tenha a necessidade desse tipo de estímulo. Foram selecionadas dez imagens com alta carga emocional, o que deve auxiliar a revelar possíveis diferenças entre a língua natural utilizada por indivíduos com os diferentes fatores de personalidade (DAN-GLAUSER; SCHERER, 2011).

Na figura 9 é possível visualizar um exemplo de imagem utilizada para essa tarefa. As imagens da base GAPED são classificadas por valência e significância normativa, e designadas de modo a despertar diferentes graus e tipos de emoção. As figuras selecionadas para a tarefa têm o grau de valência variando entre 3 e 54, sendo que o grau de valência da base GAPED varia entre 0 para imagens muito negativas a 100 para imagens muito positivas.

Figura 9 – Exemplo de estímulo visual para coleta do *cópus b5-text*

Fonte: Extraída da base GAPED (DAN-GLAUSER; SCHERER, 2011)

Para cada cena, foi solicitado ao participante escrever uma descrição detalhada da imagem com pelo menos 200 caracteres, em que o indivíduo deve relatar tudo o que está

vendo na imagem. Os participantes foram orientados a descrever tudo que observaram na cena, como se tivessem que contar o que está sendo visto para uma pessoa deficiente visual. A resposta do participante foi armazenada em banco de dados, vinculada ao seu identificador e ao identificador de cada imagem.

4.2 Sujeitos

As descrições de cenas foram feitas por 151 sujeitos. Desses sujeitos, 85 (56,3%) eram do sexo masculino. Na figura 10 é apresentado um gráfico com a distribuição dos indivíduos por idade. A maioria dos participantes (134) têm idade entre 18 e 30 anos.

Figura 10 – Distribuição de idade dos sujeitos

Fonte: Ricelli Moreira Silva Ramos, 2018

No gráfico apresentado na figura 11 é possível visualizar a distribuição dos participantes que tiveram pontuação acima da média para cada um dos fatores do modelo CGF no IGPF-5, e o número de indivíduos com pontuação acima da média para as diferentes combinações de fatores.

4.3 Procedimento

A coleta de dados foi feita de forma presencial por cinco alunos do PPgSI. A tarefa de descrições de imagens, acompanhada das demais atividades para construção do corpus levou, em média, uma hora e meia por participante.

A coleta de dados foi constituída dos seguintes passos: (1) voluntários foram contatados, (2) agendou-se um horário com cada um deles, (3) foi disponibilizado um computador com a interface desenvolvida para a coleta de dados, (4) o participante

Figura 11 – Sujeitos por fatores do modelo CGF

Fonte: Ricelli Moreira Silva Ramos, 2018

foi instruído conforme o protocolo disponível no anexo A, (5) após as orientações, o participante preencheu as descrições, e a aplicação da coleta foi finalizada.

4.4 Dados obtidos

Com a aplicação dessa coleta, foram obtidos 151 inventários de personalidade do tipo IGFP-5 (ANDRADE, 2008), e também foram coletadas 1510 descrições de imagens, tendo em média 24,5 palavras por descrição. Os textos coletados foram submetidos à correção ortográfica, pois o objetivo do presente trabalho é gerar texto minimamente correto, e posteriormente analisados com uso do *parser* PALAVRAS (BICK, 2000).

4.5 Análise preliminar

Como forma de obter uma visão geral das estratégias de determinação de conteúdo adotadas por indivíduos de diferentes personalidades, foi realizado um levantamento dos conceitos que representam as entidades visuais da imagem (pessoas, objetos etc.), e que são expressos na forma de substantivos que cada tipo de personalidade utilizou com maior

frequência para descrever as 10 imagens do corpus b5-text. Este levantamento foi realizado em duas etapas: primeiramente foi realizada a anotação semântica dos substantivos do corpus b5-text. Em seguida, foi computada a frequência com que cada conceito anotado ocorria no corpus.

Nesta tarefa de anotação semântica, substantivos empregados com o mesmo sentido (e.g., “menino” e “criança” na cena da figura 12) foram associados a um mesmo rótulo representando seu significado. Essa anotação foi feita com a colaboração de um segundo anotador, e posteriormente a tarefa foi validada por um terceiro. Além disso, no levantamento foi listada a ordem em que cada conceito foi utilizado para cada participante. Por exemplo, na descrição da figura 12, enquanto um participante começa a descrição com “Um bebê recém-nascido de pele morena e lábios inchados”, outro inicia com “Várias cobertas e mantas espalhadas pela cama. Enrolada em uma manta branca com estampas vermelha está uma criança (...)”, ou seja, enquanto um participante escreveu primeiramente sobre o conceito “bebê”, outro optou por falar do conceito “cobertas”. Com esses dados, será possível identificar e mapear no modelo computacional se há variações na ordenação do conteúdo de acordo com as diferentes combinações de fatores de personalidade.

Figura 12 – Cena para descrição textual

Fonte: Extraída da base GAPED (DAN-GLAUSER; SCHERER, 2011)

Este procedimento resultou em um mapeamento dos conceitos selecionados pelos indivíduos que produziram o corpus b5-text, e da ordem em que estes conceitos são apresentados em cada texto. Desse modo, os dados obtidos estão prontos para exploração (e.g., com uso de métodos de aprendizagem de máquina). No capítulo 6 é apresentado um estudo exploratório desses dados utilizando AM. No restante da presente seção, é realizada primeiramente uma análise preliminar dos dados obtidos.

4.5.1 Resultados da análise preliminar

Para exemplificar a diferença encontrada no conteúdo gerado pelos indivíduos com pontuações mais altas ou mais baixas em cada um dos fatores do modelo CGF, foram analisados os dez conceitos que representam as entidades visuais da imagem, expressos na forma de substantivos, mais frequentes observados nas descrições da figura 12.

Na tabela 1 são apresentados os conceitos mais utilizados para cada um dos fatores do modelo CGF, e suas respectivas frequências, com cada fator de personalidade dividido entre a pontuação alta (sim) e baixa (não), para que fosse possível visualizar se existe e qual é a variação no uso de um mesmo conceito por indivíduos com variações nos fatores de personalidade. Os valores alto e baixo da pontuação de personalidade representam os indivíduos que ficaram acima (sim) e abaixo (não) da média da pontuação do grupo de indivíduos que participaram da coleta de dados para cada um dos fatores. As frequências foram obtidas do corpus b5-text e refletem os conteúdos utilizados pelos indivíduos que participaram da coleta de dados, representados em diversas expressões similares. Por exemplo, as expressões “bebê”, “menino”, “criança”, “recém-nascido”, “neném” e “menina” representam um único conceito, pois descrevem a mesma entidade: o bebê no centro da figura 12.

Tabela 1 – Distribuição de conceitos por fator do modelo CGF na descrição da figura 12

conceito	freq	Extroversão		Amabilidade		Conscienciosidade		Neuroticismo		Abertura	
		sim	não	sim	não	sim	não	sim	não	sim	não
bebê	183	59%	41%	48%	52%	55%	45%	46%	54%	48%	52%
manta	140	61%	39%	50%	50%	56%	44%	51%	49%	49%	51%
lábios	98	61%	39%	51%	49%	54%	46%	50%	50%	52%	48%
olhos	46	54%	46%	37%	63%	59%	41%	57%	43%	46%	54%
mãe	42	55%	45%	64%	36%	52%	48%	43%	57%	50%	50%
mãos	39	51%	49%	51%	49%	51%	49%	46%	54%	41%	59%
braço	29	52%	48%	48%	52%	52%	48%	55%	45%	34%	66%
pele	25	72%	28%	52%	48%	52%	48%	48%	52%	56%	44%
cabelo	24	50%	50%	46%	54%	42%	58%	67%	33%	46%	54%
nariz	24	71%	29%	54%	46%	63%	38%	58%	42%	67%	33%

Fonte: Ricelli Moreira Silva Ramos, 2018

Apesar de a diferença entre extremos de personalidade ser muitas vezes pequena, o que pode indicar que a influência de personalidade sobre a determinação de conteúdo é sutil, observa-se em alguns casos um contraste mais acentuado. Este é o caso, por exemplo, dos conceitos *pele* e *nariz*, selecionados com mais frequência pelos indivíduos com pontuação alta no fator “Extroversão”. Esta análise preliminar, apoiada pelo levantamento

e tratamento dos dados descrito nas seções anteriores, forneceram subsídios iniciais para o desenvolvimento da pesquisa discutida nos próximos capítulos.

5 Método

Conforme apresentado na revisão bibliográfica no capítulo 3, há poucos estudos que relacionam geração de língua natural com fatores de personalidade, e nenhum que trate especificamente do processo de determinação de conteúdo nesse tema. Os estudos encontrados costumam manter o foco na forma superficial da comunicação (ou seja, “como” dizer, correspondendo à escolha lexical e à forma sintática). Isso faz sentido na medida que a forma superficial (e principalmente a escolha lexical) tende a refletir de forma direta os diferentes fatores do modelo CGF, mas o conteúdo comunicado (ou “o que” dizer) também pode em princípio ser dependente da personalidade em alguma escala. Além disso, a análise preliminar do *corpus b5-text* do capítulo anterior apresenta algumas evidências iniciais de que indivíduos com diferentes personalidades podem fazer escolhas de conteúdo distintas.

Com base nestas considerações, e tirando proveito da base de dados coletada, há a oportunidade de desenvolver um estudo sobre o papel do modelo CGF na determinação de conteúdo textual em Português brasileiro. Esse projeto explora essa oportunidade. Inicialmente, serão retomados os objetivos e hipótese já introduzidos no capítulo 1, e em seguida serão apresentadas as principais atividades do presente trabalho.

5.1 Objetivo

O objetivo da pesquisa proposta é desenvolver modelos computacionais de determinação de conteúdo baseados nos fatores de personalidade CGF, de tal modo que esta determinação seja o mais próxima possível das escolhas humanas na tarefa de descrição de imagens.

Em função das dificuldades naturais para coleta de um grande *corpus* textual com informações de personalidade, o modelo de DC a ser desenvolvido não objetiva produzir um planejamento completo do documento. Ao invés disso, o modelo deve ser capaz de estimar *quais* os objetos discursivos que um autor-alvo com certa personalidade escolheria para serem mencionados em um texto, e *quais* predicados seriam relacionados a esses objetos.

5.2 Hipótese

A hipótese investigada neste estudo é a de que o uso de informação de personalidade humana na tarefa de DC permite obter resultados mais próximos do desempenho humano do que seria possível com uso de modelos baseados nas seleções de conceitos mais frequentes no domínio de interesse.

Esta hipótese será testada comparando-se a um modelo de *baseline* que toma como entrada informações sobre os possíveis objetos discursivos de um domínio (neste caso, descrições de cenas visuais do corpus b5-text) e o modelo proposto, que inclui a informação de personalidade dos respectivos autores de cada texto. Na comparação serão utilizadas métricas tradicionais de aprendizagem de máquina, precisão, revocação e medida F. Espera-se que o modelo que utiliza conhecimento de personalidade apresente resultados superiores ao modelo que não faz uso deste tipo de conhecimento.

Comprovando-se que os modelos computacionais de determinação de conteúdo com base em fatores de personalidade têm mais acurácia se comparados ao modelo de *baseline*, demonstra-se que a personalidade do autor é refletida não apenas nas suas escolhas lexicais, como em Mairesse (2008), mas também nas escolhas do conteúdo (ou “o que” dizer), um resultado com consequências práticas potencialmente relevantes para a pesquisa em GLN.

5.3 Atividades previstas

Esse projeto tem como requisito básico a coleta de uma base de textos e inventário de fatores de personalidade de seus autores, e o tratamento deste dados, que resultou no corpus b5-text detalhado no capítulo 4. Estes dados são utilizados para o treinamento de modelos supervisionados de AM com o objetivo de aprender decisões de determinação de conteúdo (o que dizer) com e sem acesso a informações de personalidade, descrito no capítulo 6. O modelo proposto e sistemas de *baseline* pertinentes serão avaliados com base em métricas tradicionais da área de AM, precisão, revocação e medida F, e os resultados serão analisados e discutidos, e essa avaliação é apresentada no capítulo 7.

Considerando-se o domínio a ser utilizado, as principais decisões de determinação de conteúdo que serão tomadas são referentes aos elementos visuais das cenas do corpus b5-text, como pessoas, objetos etc. Assim, o modelo proposto será limitado aos conceitos

que representam as entidades visuais da imagem (pessoas, objetos etc.), expressas na forma de substantivos, e os predicados a elas associados, expressos na forma de adjetivos.

Serão criados dois tipos de modelos, com e sem informação de personalidade, descritos na seção 6.3 e esses modelos serão avaliados usando validação cruzada, conforme os resultados apresentados na seção 6.4. Em seguida será realizada uma avaliação complementar considerando uma tarefa de DC hipotética (e necessariamente incompleta pois não se aborda nesse estudo a seleção de todos tipos de conteúdo presentes nos textos, e sim dos conceitos que representam as entidades visuais da imagem (pessoas, objetos etc.), expressas na forma de substantivos, e os predicados a elas associados, expressos na forma de adjetivos), detalhada no capítulo 7.

6 Determinação de conteúdo baseada em personalidade

6.1 *Visão geral*

Este capítulo apresenta dois modelos de determinação de conteúdo: um modelo de seleção de conceitos que representam as entidades ou elementos visuais das cenas do *córpus* realizados na forma de substantivos, chamados aqui conceitos Entidade, e um modelo de seleção de predicados a eles associados realizados na forma de adjetivos, chamados aqui conceitos Predicado. A determinação de conteúdo, em ambos os casos, leva em conta informações de personalidade provenientes do modelo CGF para descrição de imagens no *córpus* b5-text, de tal modo que esta seleção seja o mais próxima possível das escolhas humanas na tarefa de descrição de imagens.

6.2 *Dados*

Os modelos de DC desenvolvidos são de caráter ilustrativo, considerando apenas os dois tipos de conteúdo mais frequentes no *córpus* b5-text: os conceitos que representam as entidades ou elementos visuais das cenas do *córpus* realizados na forma de substantivos (conceitos Entidade), e os predicados a eles associados na forma de adjetivos (conceitos Predicado).

Como o conjunto de dados oriundo do *córpus* b5-text é pequeno e necessariamente esparsos, não há sentido em investigar conceitos muito infrequentes, e nem aqueles que aparecem na grande maioria dos exemplos, pois a seleção desses independe de personalidade. Assim, os modelos desenvolvidos tratam apenas dos casos intermediários em que há um número suficiente de instâncias de uso do conceito pelos modelos de AM, porém não tão grande a ponto de ser utilizado pela grande maioria dos participantes.

Para o experimento de seleção de conceitos Entidade, foram utilizados os dados do *córpus* b5-text, selecionados e tratados conforme descrito na seção 4.5. Dentre os conceitos definidos para cada cena, foram eliminados os conceitos que apareciam em menos de 33% das descrições, ou seja, dados esparsos sobre o conceito, e também os que apareciam em mais de 90% das vezes, ou seja, conceitos que eram utilizados na maioria absoluta das descrições, independente de personalidade.

Na tabela 2 é apresentado o total de instâncias dos conceitos Entidade para cada uma das cenas, a quantidade de conceitos Entidade e, separadamente, quantos conceitos foram utilizados nesse experimento, após a eliminação das instâncias esparsas e muito frequentes. Para cada conceito utilizado foi implementado um classificador binário, discutido na seção 6.3.

Tabela 2 – Número de conceitos Entidade por cena no *cópus b5-text*

Cena	Instâncias	Entidades	Utilizados (classes)
01	2889	83	15
02	2383	48	7
03	1838	34	2
04	1811	43	4
05	2745	64	7
06	1929	35	6
07	2158	39	3
08	1875	42	7
09	2036	47	8
10	2899	63	14

Fonte: Ricelli Moreira Silva Ramos, 2018

Para cada conceito Entidade do passo anterior, foram identificados todos adjetivos a ele associado, e cada dupla <Entidade-adjetivo> forma um conceito Predicado. Detalhadamente, o levantamento das duplas <Entidade-adjetivo> do *cópus b5-text* foi feito em três etapas. A primeira, automatizada, extraiu todos os adjetivos classificados pelo *parser* PALAVRAS (BICK, 2000) e a palavra a qual esse adjetivo se referia, seja ela substantivo, pronome, advérbio ou verbo. Em seguida, todas as duplas foram analisadas pela autora, com a correção das classificações incorretas feitas pelo *parser* PALAVRAS através da leitura dos textos gerados pelos autores no *cópus b5-text*, assim como a substituição dos pronomes pelos substantivos aos quais se referiam. Por fim, as duplas <Entidade-adjetivo> que se referiam ao mesmo conceito na cena foram agrupadas. Por exemplo “sandálias vermelhas” e “chinelos vermelhos” na figura 2 referem-se à cor do calçado da menina, gerando assim a base apresentada na tabela 3. Foram consideradas duplas Entidade-adjetivo, e não adjetivos isolados, porque o adjetivo pode ser utilizado para a descrição de vários itens distintos, por exemplo, “camiseta branca”, “tigela branca” e “pessoa branca”, ou seja, é necessária a informação da dupla <Entidade-adjetivo> para que seja possível distinguir o conceito - cor de uma roupa, de um objeto ou da pele de pessoa - que o adjetivo “branca” pode representar em diferentes contextos. Da

mesma forma que para os conceitos Entidade, para cada conceito Predicado utilizado foi implementado um classificador binário, discutido na seção 6.3.

Tabela 3 – Número de conceitos Predicado por cena no corpus b5-text

Cena	Instâncias	Predicados	Utilizados (classes)
01	914	337	6
02	639	218	4
03	887	311	5
04	588	300	2
05	1040	334	8
06	892	245	8
07	435	200	0
08	565	194	5
09	959	282	9
10	959	393	4

Fonte: Ricelli Moreira Silva Ramos, 2018

Apesar do alto número de conceitos apresentado na tabela 3, poucos puderam ser utilizadas para esse experimento, pois a absoluta maioria é muito infrequente no corpus b5-text para uso de métodos de aprendizagem de máquina. Na tabela 3 também é apresentado o número total de instâncias dos conceitos Predicado para cada uma das cenas, a quantidade de conceitos Predicado e, separadamente, a quantidade de conceitos utilizados, que são os conceitos com mais de 20 aparições para cada cena, corte estabelecido de forma empírica para esta análise.

6.3 Modelos propostos

Foi criado um classificador binário para decidir quando selecionar ou não cada conceito Entidade e cada conceito Predicado definidos, selecionados e tratados nas etapas anteriores. Esse classificador teve por objetivo aprender quando os indivíduos utilizavam ou não determinado conceito baseado na pontuação do indivíduo nos cinco grandes fatores do modelo de personalidade utilizado. As características de aprendizado consideradas foram o identificador da cena, o identificador (arbitrário) definido para representar o conceito Entidade ou conceito Predicado, e os cinco valores escalares de personalidade dos autores. Os valores escalares de personalidade de todos os autores foram submetidos a todos os classificadores de conceitos, com a classe positiva (1) para os autores que mencionaram esse conceito e a classe negativa (0) para os autores que não mencionaram o mesmo conceito. Por exemplo, na cena da figura 12, os conceitos Entidade que cada autor poderia

utilizar para descrever a cena e seriam utilizados nessa pesquisa são “boca” e “cobertor”, e os conceitos Predicado são “grandes; lábios”, “branca; colcha”, “vermelhos; detalhes”, “moreno; bebê” e “vermelha; boca”. A descrição, extraída do *b5-text*,

“Bebê, aparentemente do sexo feminino, com lábios e mãos grandes, pele clara e cabelos escuros. Enrolada em um cobertor branco estampado com figuras vermelhas. O bebê se encontra aparentemente em uma cama que está coberta por uma colcha estampada de bolinhas azuis e figuras pretas em fundo bege.”

gerou os dados da tabela 4, com as instâncias que foram utilizadas para os classificadores binários dos conceitos da cena da figura 12.

Tabela 4 – Exemplo de classes positivas e negativas para uma descrição da cena da figura 12

Conceito	Instâncias
Boca	1
Cobertor	1
Grandes; lábios	1
Branca; colcha	1
Vermelhos; detalhes	1
Moreno; bebê	0
Vermelha; boca	0

Fonte: Ricelli Moreira Silva Ramos, 2018

6.4 Testes

Foi utilizada inicialmente uma árvore simples para visualização do aprendizado, e, em seguida, para melhoria do desempenho, foi utilizada a *support vector machine* (SVM) com kernel linear. Para validação cruzada da SVM foi utilizada a técnica *k-fold*, com $k=10$, e *grid search* para estimar os parâmetros mais adequados da SVM. Cada conceito foi submetido a um classificador binário com a informação de personalidade do autor, e outro classificador sem a informação de personalidade (*baseline*). Além disso, para o modelo com informação de personalidade, foi utilizada a *Synthetic Minority Over-sampling Technique* (SMOTE) (CHAWLA et al., 2002) com $k=5$ (*number of nearest neighbors*) para balanceamento de dados. Para o *baseline*, não foi utilizada SMOTE pois a técnica diminui o desempenho do modelo de *baseline*, uma vez que esse modelo é baseado apenas

na frequência em que os conceitos aparecem, sem a utilização de outras informações de personalidade dos autores.

Na tabela 5 são apresentados os resultados obtidos utilizando o modelo de SVM com *k-fold* para os conceitos que representam as entidades visuais da cena, expressas na forma de substantivos, de cada uma das imagens com e sem a informação de personalidade, e uma média geral dos resultados obtidos. São apresentados o número de classes, a média da precisão (P), da revocação (R) e da medida F (F) para cada cena. Os melhores valores obtidos para medida F estão destacados.

Tabela 5 – Resultados médios obtidos utilizando os conceitos Entidade do corpus b5-text relacionados a cada cena

Cena	Classes	<i>Baseline</i>			Personalidade		
		P	R	F	P	R	F
01	15	0,65	1,00	0,79	0,83	0,84	0,83
02	7	0,65	1,00	0,78	0,84	0,81	0,82
03	2	0,80	1,00	0,89	0,83	0,80	0,81
04	4	0,62	1,00	0,76	0,79	0,77	0,78
05	7	0,66	1,00	0,78	0,78	0,80	0,78
06	6	0,58	1,00	0,73	0,79	0,79	0,80
07	3	0,61	1,00	0,76	0,78	0,79	0,78
08	7	0,64	1,00	0,77	0,78	0,83	0,81
09	8	0,64	1,00	0,78	0,79	0,81	0,80
10	14	0,63	1,00	0,77	0,81	0,77	0,79
Média	7,3	0,65	1,00	0,78	0,80	0,80	0,80

Fonte: Ricelli Moreira Silva Ramos, 2018

Como forma de detalhar estes resultados, a tabela 6 apresenta os resultados para os conceitos Entidade mais frequentes em cada cena, dentre os que foram utilizados, com a informação de personalidade, em comparação ao *baseline* sem a informação de personalidade. Nessa tabela são apresentados o número de instâncias (Inst) de cada conceito no corpus b5-text, e a precisão (P), a revocação (R) e a medida F (F) para esses conceitos em todas as cenas e, em destaque, os melhores valores de medida F.

Na tabela 7 são apresentados os resultados para conceitos Predicado obtidos utilizando o modelo de SVM com *k-fold* para cada uma das cenas com e sem a informação de personalidade, e uma média geral dos resultados obtidos. São apresentados o número de classes, a média da precisão (P), da revocação (R) e da medida F (F) para cada cena. Os melhores valores obtidos para medida F estão destacados.

Como forma de detalhar estes resultados, a tabela 8 apresenta os resultados para os conceitos Predicado mais frequentes no texto descritivo do corpus b5-text para cada cena,

Tabela 6 – Desempenho dos conceitos Entidade mais frequentes por cena

Cena	Conceito Entidade	Inst	P	Baseline			Personalidade		
				R	F	P	R	F	
01	Estrada	125	0,86	1,00	0,93	0,76	0,96	0,85	
01	Menina	117	0,80	1,00	0,89	0,79	0,90	0,84	
01	Menino	117	0,80	1,00	0,89	0,82	0,96	0,88	
01	Rolos de arame	114	0,75	1,00	0,86	0,84	0,92	0,88	
02	Barraca	132	0,89	1,00	0,94	0,92	0,95	0,93	
02	Terra	125	0,82	1,00	0,90	0,86	0,86	0,86	
02	Lona	81	0,54	1,00	0,70	0,81	0,80	0,81	
02	Muro	77	0,51	1,00	0,58	0,84	0,76	0,80	
03	Cobertor	130	0,85	1,00	0,92	0,84	0,86	0,85	
03	Boca	114	0,76	1,00	0,86	0,83	0,74	0,79	
04	Imagem (metadescrição)	79	0,51	1,00	0,68	0,79	0,75	0,76	
04	Mão	52	0,68	1,00	0,81	0,88	0,77	0,82	
04	Semblante	52	0,63	1,00	0,71	0,77	0,77	0,77	
04	Braço	50	0,66	1,00	0,80	0,73	0,82	0,77	
05	Criança (indef.)	108	0,75	1,00	0,86	0,71	0,95	0,82	
05	Senhora	96	0,65	1,00	0,79	0,79	0,85	0,82	
05	Imagem (metadescrição)	91	0,89	1,00	0,94	0,78	0,85	0,81	
05	Casaco	72	0,51	1,00	0,68	0,81	0,80	0,81	
06	Feridas	109	0,71	1,00	0,83	0,83	0,81	0,82	
06	Pernas	93	0,64	1,00	0,78	0,77	0,83	0,80	
06	Lençol	81	0,54	1,00	0,70	0,86	0,80	0,83	
06	Cobertor	80	0,54	1,00	0,70	0,75	0,84	0,79	
07	Chão	102	0,68	1,00	0,81	0,74	0,88	0,81	
07	Pelúcia	87	0,57	1,00	0,72	0,79	0,84	0,81	
07	Imagem (metadescrição)	62	0,59	1,00	0,75	0,80	0,66	0,73	
08	Arame Farpado	131	0,86	1,00	0,92	0,76	0,97	0,86	
08	Madeira	129	0,85	1,00	0,92	0,90	0,98	0,93	
08	Camiseta	97	0,64	1,00	0,78	0,75	0,77	0,76	
08	Imagem (metadescrição)	76	0,49	1,00	0,65	0,73	0,73	0,73	
09	Chão	132	0,89	1,00	0,94	0,83	0,92	0,87	
09	Pessoas	106	0,69	1,00	0,82	0,77	0,73	0,80	
09	Camisa	88	0,58	1,00	0,74	0,80	0,75	0,78	
09	Touca	87	0,57	1,00	0,72	0,82	0,81	0,82	
10	Crianças	117	0,78	1,00	0,88	0,82	0,90	0,86	
10	Fogão	111	0,73	1,00	0,84	0,83	0,65	0,73	
10	Panela	109	0,73	1,00	0,84	0,79	0,89	0,84	
10	Senhora	106	0,70	1,00	0,83	0,76	0,87	0,81	
	Média	98	0,69	1,00	0,81	0,80	0,84	0,82	

Fonte: Ricelli Moreira Silva Ramos, 2018

Tabela 7 – Resultados médios obtidos utilizando os conceitos Predicado do corpus b5-text relacionados a cada cena

Cena	Classes	<i>Baseline</i>			Personalidade		
		P	R	F	P	R	F
01	6	0,81	1,00	0,89	0,83	0,99	0,90
02	4	0,67	1,00	0,80	0,82	0,99	0,90
03	5	0,69	1,00	0,81	0,79	0,98	0,88
04	2	0,57	1,00	0,72	0,79	0,99	0,87
05	8	0,82	1,00	0,90	0,82	1,00	0,90
06	8	0,67	1,00	0,79	0,79	0,98	0,88
07	-	-	-	-	-	-	-
08	5	0,80	1,00	0,89	0,81	0,99	0,89
09	9	0,76	1,00	0,86	0,78	0,99	0,87
10	4	0,80	1,00	0,89	0,81	0,99	0,89
Média	5,1	0,73	1,00	0,84	0,80	0,99	0,89

Fonte: Ricelli Moreira Silva Ramos, 2018

dentre as que foram utilizadas, com a informação de personalidade, em comparação ao *baseline* sem a informação de personalidade. Nessa tabela são apresentadas as instâncias (Inst) de cada conceito no corpus b5-text, e a precisão (P), a revocação (R) e a medida F (F) para esses conceitos em todas as cenas e, em destaque, os melhores valores de medida F obtidos para cada conceito.

6.5 Considerações

Os resultados apresentados na seção 6.4 sugerem que a variação de personalidade exerce influência na seleção de conceitos Entidade, na tarefa de descrição de imagens, porém de forma limitada. Alguns conceitos tiveram desempenho melhor quando o classificador tinha acesso às informações de personalidade, enquanto outros tiveram desempenho similar ao *baseline* sem informação de personalidade, e ainda, em alguns casos, o modelo de *baseline* apresentou o melhor desempenho, dado que, por não ter a informação sobre os sujeitos, selecionou a classe mais frequente, e manteve a revocação sempre em 1. Apesar disso, na média dos conceitos mais frequentes, o modelo com a informação de personalidade superou o *baseline*.

Os conceitos Entidade com melhor performance utilizando a informação de personalidade são relacionados tanto à descrição dos indivíduos presentes nas cenas quanto aos objetos ao redor, além de alguns casos em que a metadescrição - quando o participante escreveu “a imagem ...”, “a cena ...”, “a figura...” etc. - foi um dos conceitos Entidade com

Tabela 8 – Desempenho dos conceitos Predicado mais frequentes por cena

Cena	Conceito Predicado	Inst	P	Baseline			Personalidade	
				R	F	P	R	F
01	Vermelhas; sandálias	46	0,70	1,00	0,82	0,82	0,96	0,88
01	Marrom; vestido	37	0,76	1,00	0,86	0,86	0,99	0,92
01	Longo; vestido	26	0,83	1,00	0,91	0,84	0,98	0,90
02	Preta; lona	74	0,51	1,00	0,68	0,81	0,99	0,89
02	Branco; balde	65	0,57	1,00	0,73	0,76	0,98	0,86
02	Grande; balde	30	0,80	1,00	0,89	0,86	1,00	0,92
03	Grandes; lábios	89	0,56	1,00	0,71	0,77	0,99	0,87
03	Branca; colcha	59	0,62	1,00	0,76	0,74	0,99	0,85
03	Vermelhos; detalhes	56	0,54	1,00	0,78	0,75	0,99	0,85
04	Africanas; pessoas	75	0,52	1,00	0,69	0,84	0,97	0,90
04	Erguidos; braços	35	0,62	1,00	0,76	0,73	1,00	0,84
05	Sentadas; pessoas	38	0,75	1,00	0,86	0,86	0,96	0,81
05	Verde; agasalho	38	0,75	1,00	0,86	0,77	1,00	0,87
05	Branco; cabelos	28	0,81	1,00	0,90	0,77	1,00	0,87
06	Rosa; colcha	79	0,52	1,00	0,68	0,79	0,99	0,88
06	Azul; cobertor	74	0,51	1,00	0,68	0,83	0,97	0,89
06	Azul; agasalho	67	0,56	1,00	0,71	0,77	0,99	0,87
08	Negro; homem	111	0,72	1,00	0,84	0,86	0,96	0,91
08	Azul; boné	35	0,77	1,00	0,87	0,72	1,00	0,84
08	Farpados; arames	25	0,83	1,00	0,91	0,85	1,00	0,92
09	Amarela; bermuda	51	0,54	1,00	0,70	0,75	0,98	0,85
09	Xadrez; bermuda	45	0,70	1,00	0,82	0,80	0,99	0,89
09	Azul; camisa	37	0,75	1,00	0,86	0,77	0,97	0,86
10	Seco; campo	40	0,74	1,00	0,85	0,83	0,99	0,90
10	Idosa; mulher	30	0,80	1,00	0,89	0,83	0,99	0,90
10	Vermelhas; calças	30	0,80	1,00	0,89	0,75	1,00	0,86
Média		51	0,68	1,00	0,80	0,80	0,99	0,88

Fonte: Ricelli Moreira Silva Ramos, 2018

melhor desempenho. Da mesma forma, não foi identificada nenhuma relação específica entre os conceitos Entidade de pior desempenho. Isso sugere que a personalidade pode influenciar o uso de conceitos Entidade independente da relação com objetos específicos em uma cena.

Os resultados apresentados na seção 6.4 também sugerem que a variação de personalidade exerce influência na seleção de conceitos Predicado na descrição de imagens. A maioria dos conceitos tiveram desempenho melhor quando o classificador tinha acesso às informações de personalidade, enquanto outros tiveram desempenho similar ao *baseline* sem informação de personalidade, e ainda, em alguns poucos casos, o modelo de *baseline* apresentou o melhor desempenho, dado que, por não ter a informação sobre os sujeitos,

selecionou a classe mais frequente, e manteve a revocação sempre em 1, devido ao desbalançamento da base de dados. Nesses casos, verificou-se que o número de conceitos Predicado obtidos do *cópus b5-text* estava normalmente próximo ao corte de frequência estabelecido para esse experimento (20 aparições), e, por isso, acredita-se que se houvesse um número maior de dados, o desempenho do modelo proposto seria melhor. Além disso, na média dos conceitos Predicado mais frequentes, o modelo com a informação de personalidade superou com alguma distância os resultados do *baseline*.

A maioria dos conceitos Predicado utilizados nos modelos propostos relacionados diretamente à descrição física das imagens. Dos 26 conceitos apresentados na tabela 8, 12 são relacionados às cores de objetos da cena, e nenhuma é relacionado a um conceito que não seja visível na cena. Adjetivos relacionados a emoções como “triste”, “preocupado”, etc. não foram utilizados no *cópus b5-text* com uma frequência que tornasse possível aos modelos de aprendizado aplicados nesse experimento obterem conhecimento. Isso pode sugerir que a personalidade pode influenciar o uso de termos em descrições visuais independentes de emoções.

Utilizando a personalidade para determinar o conteúdo na tarefa de descrição de imagens, tanto para a descrição com os conceitos Entidade quanto para a descrição com os conceitos Predicado, diferentemente de quando trabalha-se com a escolha léxica, como em Mairesse (2008), Walker et al. (2011a), entre outros, não foi identificada uma relação direta dos conceitos com algum teor afetivo, e sim, as descrições independente desse teor tiveram um desempenho levemente melhor ao utilizar a informação de personalidade, possivelmente porque os conceitos utilizados para descrição, em sua absoluta maioria, não possuem esse teor.

É importante também destacar que, conforme apresentado na tabela 7, na média de todas as cenas utilizadas no experimento com os conceitos Predicado o modelo utilizando personalidade nunca teve desempenho inferior ao *baseline*. Isso sugere a influência da personalidade na assertividade da seleção de conteúdo, mesmo que de forma sutil, é constante para os conceitos Predicado na tarefa de descrição de imagens.

6.6 Outros experimentos

Além do teste com validação cruzada realizado na seção 6.4, foram executados outros experimentos na tentativa de obter maior aprendizado dos dados de descrição de imagens do *corpus b5-text*, porém sem resultados satisfatórios. Foram feitos experimentos para tratar a ordem dos conceitos Entidade e Predicado, assim como para identificar o primeiro termo a aparecer na descrição, também para determinar a quantidade de termos, tanto para conceitos Entidade, quanto para conceitos Predicado, e ainda para quantidade de palavras em geral, versus o fator “Extroversão”, e, além disso, determinar a influência do teor afetivo dos conceitos na determinação de conteúdo.

Para os experimentos que trataram da ordem dos conceitos Entidade e Predicado, havia a suspeita de que indivíduos com fatores de personalidade diferentes usariam conceitos em ordem diferentes, por exemplo, enquanto alguns indivíduos descrevem primeiro o ambiente da cena, outros preferem descrever primeiro os humanos. Os resultados desses experimentos foram insatisfatórios, mas não é possível concluir que isso significa que não há influência da personalidade nessa ordenação, pois esse resultado pode ter sido causado pelo tamanho limitado do *corpus b5-text*.

Foram feitos experimentos também para tentar identificar os três primeiros ou o primeiro termo (tanto para os conceitos Entidade quanto para os conceitos Predicado) que os indivíduos com diferentes personalidades utilizaram. Em Mairesse (2008) foi demonstrado que indivíduos com determinadas personalidades utilizaram conteúdo positivo ou negativo primeiro, porém ao tentar identificar esse e outros tipos de conteúdo primeiro, os experimentos não tiveram resultados expressivos, o que pode ser causado tanto pelo limitado tamanho do *corpus b5-text* quanto pelas características da tarefa de descrição de imagens, em que não é tão explícita a diferença do conteúdo positivo e negativo, enquanto que no domínio de recomendação de restaurantes de Mairesse (2008), essa diferenciação é bastante clara.

Um outro experimento tentou identificar a relação entre as pontuações no fator “Extroversão” e a quantidade de conceitos Entidade, conceitos Predicado ou palavras utilizadas. Para as três situações não foi possível desenvolver um modelo de regressão nem de classificação (ao dividir os termos em subclasses de 0 a 6, 6 a 10, etc.) que conseguisse mapear e aprender adequadamente alguma relação entre o fator e quantidade

de termos utilizados. Como na tarefa de descrição de cenas foi solicitada a descrição com o máximo de detalhes possível, conforme o protocolo no A, e indivíduos com “Extroversão” mais baixa costumam ser submissos e meticolosos na execução de tarefas (MAIRESSE, 2008; ANDRADE, 2008), é possível que esses indivíduos tenham usado tantos conceitos e termos quanto os indivíduos com “Extroversão” mais alta, apesar de um indivíduo com “Extroversão” mais alta normalmente ser mais falador. Mas o resultado insatisfatório desse experimento também pode ser causado pelo tamanho limitado do *corpus b5-text*, insuficiente para mapear essa relação.

Foi feita ainda uma análise dos conceitos utilizando o dicionário LIWC-BR (FILHO; PARDO; ALUÍSIO, 2013), buscando categorias desse dicionário em que houvesse maior influência dos fatores de personalidade CGF para determinação de conteúdo, porém não foi possível obter resultados satisfatórios, uma vez que a maioria dos conceitos com repetições suficientes para a aplicação de algum modelo de AM encontrados no *corpus b5-text* são descritivos e não fazem referências afetivas, ou seja, são independentes da categoria no dicionário. Conceitos como “triste”, “magoado” e “chateado”, mesmo quando se referem ao mesmo indivíduo em uma cena, não podem ser agrupados em um único, pois não necessariamente refletem o mesmo, diferente de conceitos como “menino”, “garoto” e “moleque”, nos quais é possível de fato determinar que se referem à mesma entidade na descrição de uma cena.

7 Avaliação complementar

Em complemento ao teste dos modelos individuais apresentado no capítulo anterior, neste capítulo é apresentada uma avaliação adicional destes modelos em uma tarefa de determinação de conteúdo. Nesta avaliação, os textos do *corpus b5-text* são divididos em conjuntos de treinamento e teste, e realiza-se a medição da proporção de conteúdos corretamente selecionados pelos modelos para cada texto de cada autor do *corpus*. Com base nessas informações, são apresentadas algumas considerações sobre os resultados desse trabalho.

7.1 *Procedimento*

Os dados do *corpus b5-text* foram divididos entre dois conjuntos, de treinamento e de teste. Os textos completos foram separados em dois conjuntos balanceados de modo que instâncias de todos os 151 autores apareceram em número semelhante em ambos os conjuntos. Aproximadamente 20% dos textos de cada cena foram utilizados para teste, enquanto os demais foram utilizados para o treinamento dos modelos.

Os modelos foram retreinados utilizando apenas o conjunto de treinamento. Em seguida, foram submetidos aos testes com o objetivo de computar o conteúdo a ser representado em cada texto de teste com base na personalidade do seu autor. Na execução dos testes, foram considerados cada um dos conceitos (Entidade ou Predicado) possíveis de cada cena, ou seja, se uma cena tem 27 conceitos possíveis, para cada documento de teste são criadas 27 instâncias de teste, e cada uma é submetida a um dos 27 classificadores. O resultado é, para cada documento de teste, um conjunto de conceitos (Entidade e Predicado) que deveriam figurar naquele texto dada a cena descrita e a personalidade do autor em questão.

7.2 *Resultados obtidos*

Após o treinamento dos modelos utilizando os conjuntos separados descritos na seção 7.1, a avaliação foi executada e os resultados obtidos foram consolidados. Os modelos foram testados para os conjuntos de conceitos Entidade, conceitos Predicado e para a combinação entre eles.

Na tabela 9 são apresentadas as médias do coeficiente Dice para o *baseline* e para o modelo com informação de personalidade (nomeado modelo b5), nas situações descritas anteriormente.

Tabela 9 – Média do coeficiente Dice

Situação	<i>baseline</i>	b5
Entidade	0,67	0,67
Predicado	0,14	0,28
Entidade + Predicado	0,60	0,61
Média geral	0,47	0,52

Fonte: Ricelli Moreira Silva Ramos, 2018

Na tabela 10 é apresentada a média do coeficiente Dice para os conceitos Entidade por cena para o *baseline* e o modelo b5, destacando o modelo com melhor resultado para cada cena.

Tabela 10 – Média do coeficiente Dice para os conceitos Entidade por cena

Cena	<i>baseline</i>	b5
01	0,70	0,70
02	0,70	0,72
03	0,88	0,74
04	0,37	0,62
05	0,54	0,63
06	0,63	0,61
07	0,61	0,59
08	0,69	0,67
09	0,68	0,68
10	0,60	0,64
Média geral	0,67	0,67

Fonte: Ricelli Moreira Silva Ramos, 2018

Na tabela 11 é apresentada a média do coeficiente Dice para os conceitos Predicado por cena para o *baseline* e o modelo b5, com destaque para o maior valor obtido em cada cena.

Na tabela 12 é apresentada a média do coeficiente Dice para a combinação dos conceitos Entidade e Predicado por cena para o *baseline* e o modelo b5, destacando o modelo com melhor resultado para cada cena.

Para verificar se a diferença entre o *baseline* e os modelos propostos é estatisticamente significativa, foi realizado o teste de Wilcoxon utilizando os resultados do coeficiente Dice apresentados anteriormente. Com relação à seleção de conceitos Predicado, observa-se que o valor médio de coeficientes Dice do modelo proposto é superior ao do *baseline*

Tabela 11 – Média do coeficiente Dice para os conceitos Predicado por cena

Cena	<i>baseline</i>	b5
01	0	0,27
02	0	0,16
03	0,42	0,42
04	0	0,30
05	0	0,14
06	0,25	0,32
07	-	-
08	0,56	0,34
09	0	0,35
10	0	0,21
Média geral	0,14	0,28

Fonte: Ricelli Moreira Silva Ramos, 2018

Tabela 12 – Média do coeficiente Dice para a combinação dos conceitos Entidade e Predicado por cena

Cena	<i>baseline</i>	b5
01	0,62	0,63
02	0,48	0,64
03	0,68	0,63
04	0,63	0,54
05	0,50	0,47
06	0,52	0,48
07	0,61	0,59
08	0,66	0,57
09	0,55	0,56
10	0,56	0,58
Média geral	0,60	0,61

Fonte: Ricelli Moreira Silva Ramos, 2018

considerado. Aplicando-se um teste de Wilcoxon é possível observar que esta diferença é significativa ($W = -5127$, $Z = -5.57$, $p < 0.001$). No caso da seleção de conceitos Entidade e da seleção combinada de conceitos Entidade e Predicado, as diferenças observadas não são estatisticamente significativas.

Finalmente, a tabela 13 apresenta um resumo do número de casos em que cada modelo apresentou o melhor desempenho nas três situações, ou seja, quando o valor do coeficiente Dice foi maior para o modelo b5 do que para o modelo de *baseline* (coluna Personalidade), quando houveram empates (coluna Empate), e quando o valor do *baseline* foi maior que o modelo b5 (coluna *Baseline*).

Tabela 13 – Resultados gerais obtidos pelos modelos

Situação	Personalidade	Empate	<i>Baseline</i>
Entidade	142	51	112
Predicado	103	137	35
Entidade + Predicado	143	27	135
Média	129	72	94

Fonte: Ricelli Moreira Silva Ramos, 2018

7.3 Considerações

Conforme os dados e resultados apresentados na seção 7.2, a diferença de desempenho entre o modelo b5 com informação de personalidade e o modelo *baseline* sem informação de personalidade é insignificante para os conceitos Entidade, e também para a combinação dos conceitos Entidade e Predicado. Essa diferença é realmente significativa apenas no caso dos conceitos Predicado. Uma possível explicação para isso é o fato de que os participantes foram instruídos a descrever tudo que havia na cena e, como resultado, a maioria realmente menciona todas as pessoas e coisas que viu. Ou seja, independentemente de personalidade, a maioria dos autores menciona aproximadamente o mesmo conjunto completo de objetos. Aparentemente a seleção de conceitos Entidade depende muito mais de percepção (e.g., enxergar direito) do que personalidade, ou a personalidade exerce uma influência muito limitada nessa seleção.

Por outro lado, quando trabalha-se com conceitos mais abstratos, como os representados por adjetivos (conceitos Predicado), começam a aparecer diferenças em função de personalidade, possivelmente porque há uma decisão pessoal de comentar ou não um determinado aspecto do que se está vendo. Quando os conceitos Predicado foram utilizados, a variação de termos é maior, e a gama diversificada pode ser causada, entre outros aspectos, pelas variações de personalidade dos indivíduos. Na tabela 9, o modelo com informação de personalidade para conceitos Predicado tem resultados significativamente maiores que o *baseline*, porém, no modelo para conceitos Entidade e, por consequência, a combinação dos modelos dos conceitos Entidade e Predicado, a diferença entre o modelo b5 e o *baseline* chega a ser insignificante.

Na tabela 10, que apresenta o resultado médio do coeficiente Dice por cena para os conceitos Entidade, apesar de haver desempenhos melhores de um ou outro modelo dependendo da cena, pode-se observar que a diferença é sutil na maioria dos casos, e há vários casos de empate entre o modelo b5 e o *baseline* dentro das cenas.

Na tabela 11, que apresenta o resultado médio do coeficiente Dice por cena para os conceitos Predicado, é importante ressaltar que o modelo b5 é superior ao *baseline* na maioria das cenas, porém os resultados não são tão altos devido à esparsidade do conjunto de conceitos Predicado em comparação ao conjunto de conceitos Entidade. Ou seja, muitas vezes o resultado do modelo b5 também foi 0, dada a pouca expressividade do conjunto de conceitos Predicado.

Na tabela 12, que apresenta o resultado médio do coeficiente Dice por cena para os a combinação dos conceitos Entidade e Predicado, apesar de haver desempenhos melhores mais vezes do *baseline* do que do modelo b5, pode-se observar que a diferença é sutil na maioria dos casos, sendo praticamente insignificante.

Na tabela 13 é possível verificar uma superioridade do desempenho do modelo b5 frente ao *baseline* em todas as situações propostas, conforme resultados também apresentados no capítulo 6, porém essa superioridade é limitada no caso dos conceitos Entidade e combinação dos conceitos Entidade e Predicado, de forma a não ter relevância estatística. Os modelos dos conceitos Predicado, por outro lado, tiveram um desempenho satisfatório, estatisticamente significativo.

8 Conclusão

Este trabalho descreveu um estudo cujo principal objetivo foi a geração de modelos computacionais de determinação de conteúdo baseados nos fatores de personalidade CGF, de tal modo que esta determinação seja o mais próxima possível das escolhas humanas na tarefa de descrição de imagem. A hipótese investigada neste estudo foi a de que o uso de informação de personalidade humana na tarefa de DC permitiria obter resultados mais próximos ao desempenho humano do que os resultados obtidos com o uso de modelos baseados nas seleções de conceitos mais frequentes no domínio de interesse. Essa hipótese foi testada utilizando o corpus coletado nessa pesquisa para treinar e testar modelos de AM que tomassem as melhores decisões de conteúdo com e sem informação de personalidade. Os modelos foram avaliados e comparados para demonstrar se existe e qual é a influência da personalidade na determinação de conteúdo na tarefa de descrição de imagens. Os resultados obtidos demonstram que pode existir uma influência sutil da personalidade na tarefa de DC.

Este trabalho também representa uma contribuição potencial para a área de GLN ao investigar se e como a determinação do conteúdo de descrições textuais é influenciada pela personalidade do autor. Os resultados obtidos podem assim ser levados em conta no futuro desenvolvimento de sistemas de GLN que considerem a personalidade do autor-alvo de interesse para geração de textos mais realistas e, de modo específico, representam uma contribuição para a pesquisa de sistemas de GLN com foco no Português brasileiro.

As principais contribuições desse estudo especificamente foram:

- um modelo de determinação de conteúdo baseado em personalidade na forma de conceitos que representam as entidades visuais da imagem (pessoas, objetos etc.), expressos na forma de substantivos, e os predicados a elas associados, e expressos na forma de adjetivos;
- um modelo de *baseline* de tomada de decisão mais provável na determinação de conteúdos para os conceitos que representam as entidades visuais da imagem (pessoas, objetos etc.), expressos na forma de substantivos, e os predicados a elas associados, e expressos na forma de adjetivos;
- o corpus b5 (RAMOS et al., 2018) de documentos textuais rotulados com informações de personalidade, construído em colaboração com outros pesquisadores;

- uma ferramenta para coleta de dados presenciais, composta pelo módulo de identificação de fatores de personalidade e módulo para descrição de cenas.

Os resultados iniciais deste trabalho - referentes à construção do corpus b5 - foram publicados na conferência *Language Resources and Evaluation Conference* (LREC-2018) em um artigo intitulado *Building a corpus for personality dependent natural language understanding and generation*.

8.1 Limitações

Essa pesquisa não objetivou o desenvolvimento de artefatos computacionais de GLN, mas sim um estudo sobre como a personalidade do locutor-alvo de interesse poderia ser explorada por modelos computacionais deste tipo.

Esse trabalho foi limitado a descrições em língua portuguesa, e provenientes unicamente do corpus b5-text obtido na fase de coleta de dados. Também foi limitado à fase inicial de determinação de conteúdo (quais objetos referenciar) de um sistema GLN, não abordando os demais passos exigidos para a construção de um planejamento completo do documento a ser gerado. Além disso, foram realizados outros experimentos com os dados do corpus b5-text nos quais não foi possível obter resultados satisfatórios para fases complementares da determinação de conteúdo de um sistema GLN.

8.2 Trabalhos futuros

Como trabalhos futuros, sugere-se realizar a coleta de corpora maiores com informação de personalidade em outros domínios, para que experimentos similares demonstrem se realmente há ou não influência da personalidade nessa etapa da determinação de conteúdo, ou se o resultado aqui obtido foi em função do tamanho do corpus ou do domínio selecionado. Também seria interessante que os resultados obtidos nos experimentos desse estudo fossem avaliados em sua naturalidade por juízes humanos, porém para tal avaliação é necessário que o avaliador humano conheça planejamento de documentos de GLN, pois o modelo aqui proposto não gera fala (escrita) humana, que poderia ser avaliada por qualquer indivíduo leigo, e sim uma representação abstrata de texto, parte do planejamento do documento.

Referências¹

- ALLPORT, G. W. *Personality: A psychological interpretation*. 2013. ed. New York: Henry Holt, 1937. Citado na página 31.
- ALY, A.; TAPUS, A. Towards an intelligent system for generating an adapted verbal and nonverbal combined behavior in human–robot interaction. *Autonomous Robots*, Springer, v. 40, n. 2, p. 193–209, 2016. Citado 3 vezes nas páginas 18, 57 e 60.
- ANDRADE, J. M. d. *Evidências de validade do inventário dos cinco grandes fatores de personalidade para o Brasil*. Tese (Doutorado) — Universidade de Brasília, Instituto de Psicologia, Departamento de Psicologia Social e do Trabalho, 2008. Citado 9 vezes nas páginas 18, 31, 32, 33, 35, 36, 61, 64 e 81.
- BELZ, A. Automatic generation of weather forecast texts using comprehensive probabilistic generation-space models. *Natural Language Engineering*, v. 14, n. 4, p. 431–455, 2008. Citado na página 16.
- BICK, E. *The Parsing System “Palavras”: Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework*. Århus: Aarhus Universitetsforlag, 2000. Citado 2 vezes nas páginas 64 e 72.
- BOWDEN, K. K. et al. M2d: Monolog to dialog generation for conversational story telling. In: *International Conference on Interactive Digital Storytelling*. Switzerland: Springer, 2016. p. 12–24. Citado 2 vezes nas páginas 57 e 60.
- BREIMAN, L. et al. *Classification and regression trees*. New York: CRC press, 1984. Citado na página 29.
- CHAWLA, N. V. et al. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Int. Res.*, AI Access Foundation, USA, v. 16, n. 1, p. 321–357, 2002. Citado na página 74.
- COSTA, E. et al. A review of performance evaluation measures for hierarchical classifiers. In: *Evaluation Methods for machine Learning II: papers from the AAAI-2007 Workshop*. Vancouver: AAAI, 2007. p. 1–6. Citado na página 30.
- COSTA, P. T.; MACCRAE, R. R. *Revised NEO personality inventory (NEO PI-R) and NEO five-factor inventory (NEO-FFI): Professional manual*. Odessa: Psychological Assessment Resources, Incorporated, 1992. Citado na página 35.
- COSTA, P. T.; MCCRAE, R. R. The NEO personality inventory. PAR Psychological Assessment Resources, 1985. Citado 2 vezes nas páginas 34 e 35.
- COSTA, P. T.; MCCRAE, R. R. The revised NEO-PI/NEO-FFI manual supplement. *Psychological Assessment Resources: Odessa, FL*, 1989. Citado na página 35.
- COSTA, P. T.; MCCRAE, R. R. Four ways five factors are basic. *Personality and individual differences*, Elsevier, v. 13, n. 6, p. 653–665, 1992. Citado na página 32.
- COSTA, P. T.; MCCRAE, R. R. Neo PI-R professional manual. Psychological assessment resources, 1992. Citado 2 vezes nas páginas 34 e 35.

¹ De acordo com a Associação Brasileira de Normas Técnicas. NBR 6023.

COSTA, S. B. et al. Psychometric properties of portuguese version of Big Five Inventory (BFI). *International Journal of Developmental and Educational Psychology. Revista INFAD de Psicología.*, v. 1, n. 2, p. 83–94, 2016. Citado na página 31.

CRISTIANINI, N.; SHAW-TAYLOR, J. *An introduction to support vector machines and other kernel-based learning methods*. UK: Cambridge University Press, 2000. Citado na página 29.

DAN-GLAUSER, E. S.; SCHERER, K. R. The Geneva affective picture database (GAPED): a new 730-picture database focusing on valence and normative significance. *Behavior research methods*, Springer, v. 43, n. 2, p. 468, 2011. Citado 3 vezes nas páginas 19, 62 e 65.

DICE, L. R. Measures of the amount of ecologic association between species. *Ecology*, Wiley Online Library, v. 26, n. 3, p. 297–302, 1945. Citado na página 30.

EUGENIO, B. di et al. PatientNarr: Towards generating patient-centric summaries of hospital stays. In: *Proceedings of the 8th International Natural Language Generation Conference (INLG-2014)*. Philadelphia, Pennsylvania, USA: Association for Computational Linguistics, 2014. p. 6–10. Citado na página 16.

EYSENCK, H. J. Dimensions of personality: 16, 5 or 3?—criteria for a taxonomic paradigm. *Personality and individual differences*, Elsevier, v. 12, n. 8, p. 773–790, 1991. Citado na página 34.

FILHO, P. P. B.; PARDO, T. A. S.; ALUÍSIO, S. M. An evaluation of the Brazilian Portuguese LIWC dictionary for sentiment analysis. In: *Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology*. Fortaleza, CE, Brasil: Sociedade Brasileira de Computação, 2013. Citado na página 81.

FURNHAM, A. *Language and personality*. John Wiley & Sons, 1990. Citado na página 42.

GABRIEL, R. P.; CHEN, J.; NICHOLS, J. Inkwell: A creative writer’s creative assistant. In: *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition*. Mountain View: ACM, 2015. p. 93–102. Citado 4 vezes nas páginas 16, 18, 58 e 59.

GILBERT, A. et al. Overview of the ImageCLEF 2015 Scalable Image Annotation, Localization and Sentence Generation task. In: *CLEF2015 Working Notes*. Toulouse, France: CEUR-WS.org, 2015. (CEUR Workshop Proceedings). Citado na página 16.

GOLDBERG, L. R. An alternative “description of personality”: the Big-Five factor structure. *Journal of personality and social psychology*, American Psychological Association, v. 59, n. 6, p. 1216, 1990. Citado 2 vezes nas páginas 18 e 32.

GOLDBERG, L. R.; KILKOWSKI, J. M. The prediction of semantic consistency in self-descriptions: Characteristics of persons and of terms that affect the consistency of responses to synonym and antonym pairs. *Journal of personality and social psychology*, American Psychological Association, v. 48, n. 1, p. 82, 1985. Citado na página 36.

HALL, C. S.; LINDZEY, G.; CAMPBELL, J. B. *Teorias da personalidade*. São Paulo: Artmed Editora, 2000. Citado na página 31.

- HEARST, M. A. et al. Support vector machines. *IEEE Intelligent Systems and their applications*, IEEE, v. 13, n. 4, p. 18–28, 1998. Citado na página 30.
- HERZIG, J. et al. Neural response generation for customer service based on personality traits. In: INLG. *Proceedings of the 10th International Conference on Natural Language Generation*. Santiago de Compostela, Spain, 2017. p. 252–256. Citado 3 vezes nas páginas 18, 58 e 60.
- JOHN, O. P.; DONAHUE, E.; KENTLE, R. *The Big Five Inventory - Versions 4a and 54*. Berkeley, CA, USA, 1991. Citado na página 62.
- JOHN, O. P.; NAUMANN, L. P.; SOTO, C. J. Paradigm shift to the integrative big five trait taxonomy. *Handbook of personality: Theory and research*, v. 3, p. 114–158, 2008. Citado 4 vezes nas páginas 32, 34, 35 e 36.
- JORDAN, P. et al. TBI-Doc: Generating patient & clinician reports from brain imaging data. In: *Proceedings of the 8th International Natural Language Generation Conference (INLG)-2014*. Philadelphia, Pennsylvania, USA: Association for Computational Linguistics, 2014. p. 143–146. Citado na página 16.
- KARPATHY, A.; FEI-FEI, L. Deep visual-semantic alignments for generating image descriptions. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Boston, MA, USA: IEEE, 2015. p. 3128–3137. Citado na página 16.
- KHOSMOOD, F.; WALKER, M. Grapevine: a gossip generation system. In: *Proceedings of the Fifth International Conference on the Foundations of Digital Games*. Santa Cruz: ACM, 2010. p. 92–99. Citado 2 vezes nas páginas 55 e 59.
- LARRANAGA, P. et al. Machine learning in bioinformatics. *Briefings in bioinformatics*, p. 86–112, 2006. Citado na página 30.
- LAVOIE, B.; RAMBOW, O. A fast and portable realizer for text generation systems. In: *Proceedings of the fifth conference on Applied natural language processing*. Ithaca: Association for Computational Linguistics, 1997. p. 265–268. Citado 2 vezes nas páginas 45 e 55.
- LEPPÄNEN, L. et al. Data-driven news generation for automated journalism. In: *Proceedings of the 10th International Conference on Natural Language Generation*. Santiago de Compostela, Spain: Association for Computational Linguistics, 2017. p. 188–197. Citado na página 16.
- LIMA, C. A. d. M. et al. *Comitê de Máquinas: uma abordagem unificada empregando máquinas de vetores-suporte*. Tese (Doutorado) — Unicamp, 2004. Citado na página 29.
- LORENA, A. C.; CARVALHO, A. C. de. Uma introdução às support vector machines. *Revista de Informática Teórica e Aplicada*, v. 14, n. 2, p. 43–67, 2007. Citado na página 30.
- LUKIN, S. M.; REED, L.; WALKER, M. A. Generating sentence planning variations for story telling. In: *16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. Prague, Czech Republic: Association for Computational Linguistics, 2015. p. 188–197. Citado 3 vezes nas páginas 18, 57 e 60.

LUKIN, S. M.; RYAN, J. O.; WALKER, M. A. Automating direct speech variations in stories and games. *Proceedings of the GAMNLP*, Santa Cruz, 2014. Citado 3 vezes nas páginas 17, 56 e 60.

LUNDIN, R. W. *Personalidade: uma análise do comportamento*. São Paulo: EPU- Editora Pedagógica Universitária, 1977. Citado na página 31.

MAIRESSE, F. *Learning to adapt in dialogue systems: Data-driven models for personality recognition and generation*. Tese (Doutorado) — University of Sheffield, United Kingdom, 2008. Citado 17 vezes nas páginas 17, 18, 32, 34, 37, 45, 46, 47, 48, 49, 50, 53, 60, 69, 79, 80 e 81.

MAIRESSE, F.; WALKER, M. PERSONAGE: Personality generation for dialogue. In: *Annual Meeting-Association For Computational Linguistics*. Sheffield: Association for Computational Linguistics (ACL), 2007. v. 45, n. 1, p. 496. Citado 7 vezes nas páginas 42, 43, 44, 47, 55, 59 e 60.

MAIRESSE, F.; WALKER, M. A. A personality-based framework for utterance generation in dialogue applications. In: *AAAI Spring Symposium: Emotion, Personality, and Social Behavior*. Sheffield: Stanford University, 2008. p. 80–87. Citado 2 vezes nas páginas 50 e 51.

MAIRESSE, F.; WALKER, M. A. Trainable generation of Big-Five personality styles through data-driven parameter estimation. In: . Cambridge: ACL, 2008. p. 165–173. Citado 2 vezes nas páginas 51 e 52.

MAIRESSE, F.; WALKER, M. A. *Can Conversational Agents Express Big Five Personality Traits through Language?: Evaluating a Psychologically-Informed Language Generator*. Cambridge: Cambridge University Engineering Department & Department of Computer Science, University of Sheffield, 2009. Citado 4 vezes nas páginas 22, 51, 52 e 53.

MAIRESSE, F.; WALKER, M. A. Towards personality-based user adaptation: psychologically informed stylistic language generation. *User Modeling and User-Adapted Interaction*, Kluwer Academic Publishers, v. 20, n. 3, p. 227–278, 2010. Citado 2 vezes nas páginas 51 e 53.

MAIRESSE, F.; WALKER, M. A. Controlling user perceptions of linguistic style: Trainable generation of personality traits. *Computational Linguistics*, MIT Press, v. 37, n. 3, p. 455–488, 2011. Citado 4 vezes nas páginas 26, 33, 34 e 54.

MATEAS, M.; STERN, A. Façade: An experiment in building a fully-realized interactive drama. In: *Game developers conference*. Georgia: GDC Vault, 2003. v. 2. Citado na página 51.

MCCORMICK, C. Evaluating the perception of personality and naturalness in computer generated utterances. 2012. Citado 2 vezes nas páginas 56 e 60.

MICHALSKI, R. S.; CARBONELL, J. G.; MITCHELL, T. M. *Machine learning: An artificial intelligence approach*. Palo Alto, California: Springer Science & Business Media, 2013. Citado na página 29.

MONARD, M. C.; BARANAUSKAS, J. A. Indução de regras e árvores de decisão. *Sistemas Inteligentes. Rezende, SO Editora Manole Ltda*, p. 115–140, 2003. Citado na página 29.

NASS, C.; LEE, K. M. Does computer-synthesized speech manifest personality? experimental tests of recognition, similarity-attraction, and consistency-attraction. *Journal of experimental psychology: applied*, American Psychological Association, v. 7, n. 3, p. 171, 2001. Citado na página 57.

NEFF, M.; KIM, Y. Interactive editing of motion style using drives and correlations. In: *Proceedings of the 2009 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. New Orleans: ACM, 2009. p. 103–112. Citado na página 54.

NEFF, M. et al. Gesture modeling and animation based on a probabilistic re-creation of speaker style. *ACM Transactions on Graphics (TOG)*, ACM, v. 27, n. 1, p. 5, 2008. Citado na página 54.

NEFF, M. et al. Evaluating the effect of gesture and language on personality perception in conversational agents. In: *International Conference on Intelligent Virtual Agents*. Berlin Heidelberg: Springer, 2010. p. 222–235. Citado 3 vezes nas páginas 51, 54 e 60.

NORMAN, W. T. Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *The Journal of Abnormal and Social Psychology*, American Psychological Association, v. 66, n. 6, p. 574, 1963. Citado 2 vezes nas páginas 18 e 32.

NUNES, C. et al. Criação de uma bateria brasileira de avaliação da personalidade pelo modelo dos cinco grandes fatores. In: *III Congresso Brasileiro de Avaliação Psicológica*. São Paulo: Vetor, 2007. Citado na página 32.

PARK, E.; JIN, D.; POBIL, A. P. del. The law of attraction in human-robot interaction. *International Journal of Advanced Robotic Systems*, SAGE Publications Sage UK: London, England, v. 9, n. 2, p. 35, 2012. Citado na página 57.

PASQUALI, L. *Os Tipos Humanos: A Teoria da Personalidade*. São Paulo: Vozes, 2003. Citado na página 31.

PAUTLER, D.; QUILICI, A. A computational model of social perlocutions. In: *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics-Volume 2*. Honolulu: ACL, 1998. p. 1020–1026. Citado 2 vezes nas páginas 37 e 38.

PENNEBAKER, J. W.; KING, L. A. Linguistic styles: language use as an individual difference. *Journal of personality and social psychology*, American Psychological Association, v. 77, n. 6, p. 1296, 1999. Citado na página 42.

PERVIN, L. A.; JOHN, O. P. *Personalidade: Teoria e Pesquisa*. 8. ed. São Paulo: Artmed, 2004. Citado na página 18.

PLACHOURAS, V. et al. Interacting with financial data using natural language. In: *ACM. Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. Pisa, Italy, 2016. p. 1121–1124. Citado na página 16.

RAMMSTEDT, B.; JOHN, O. P. Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of research in Personality*, Elsevier, v. 41, n. 1, p. 203–212, 2007. Citado na página 44.

RAMOS, R. M. S. et al. Building a corpus for personality-dependent natural language understanding and generation. *11th International Conference on Language Resources and Evaluation (LREC-2018)*, ELRA, Miyasaki, Japan, 2018. Citado 2 vezes nas páginas 61 e 87.

REEVES, B.; NASS, C. *The Media Equation*. Chicago: University of Chicago Press, 1996. Citado na página 57.

REITER, E. Has a consensus NL generation architecture appeared, and is it psycholinguistically plausible? In: *Proceedings of the Seventh International Workshop on Natural Language Generation*. Ithaca: Association for Computational Linguistics, 1994. p. 163–170. Citado na página 22.

REITER, E. An architecture for data-to-text systems. In: *Proceedings of the Eleventh European Workshop on Natural Language Generation*. Germany: Association for Computational Linguistics, 2007. p. 97–104. Citado na página 17.

REITER, E.; DALE, R. Building applied natural language generation systems. *Natural Language Engineering*, Cambridge Univ Press, v. 3, n. 01, p. 57–87, 1997. Citado 3 vezes nas páginas 26, 27 e 28.

REITER, E.; DALE, R.; FENG, Z. *Building natural language generation systems*. Cambridge: Cambridge University Press, 2000. v. 33. Citado 10 vezes nas páginas 16, 17, 22, 24, 25, 26, 27, 28, 45 e 60.

RUSSEL, S.; NORVIG, P. Inteligência artificial. *Editora Campus*, p. 26, 2004. Citado na página 29.

SCHEUTZ, M.; RÖMMER, B. Autonomous avatars? From users to agents and back. In: SPRINGER. *International Workshop on Intelligent Virtual Agents*. Berlin Heidelberg, 2001. p. 61–71. Citado 3 vezes nas páginas 39, 40 e 41.

SCHNEIDER, A. H. et al. Using technology to enhance rural resilience in pre-hospital emergencies. *Scottish Geographical Journal*, Taylor & Francis, v. 131, n. 3-4, p. 194–200, 2015. Citado na página 16.

SØRENSEN, T. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on danish commons. *Biol. Skr.*, v. 5, p. 1–34, 1948. Citado na página 30.

SRIPADA, S. et al. A case study: NLG meeting weather industry demand for quality and quantity of textual weather forecasts. In: *Proceedings of the 8th International Natural Language Generation Conference (INLG)-2014*. Philadelphia, Pennsylvania, USA: Association for Computational Linguistics, 2014. p. 1–5. Citado na página 16.

SRIPADA, S. G. et al. A two-stage model for content determination. In: *Proceedings of the 8th European workshop on Natural Language Generation-Volume 8*. Aberdeen: Association for Computational Linguistics, 2001. p. 1–8. Citado na página 27.

STEINWART, I.; CHRISTMANN, A. *Support vector machines*. New York: Springer Science & Business Media, 2008. Citado 2 vezes nas páginas 29 e 30.

STRAUSS, M.; KIPP, M. ERIC: a generic rule-based framework for an affective embodied commentary agent. In: *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 1*. Saarbrücken: International Foundation for Autonomous Agents and Multiagent Systems, 2008. p. 97–104. Citado 3 vezes nas páginas 18, 58 e 59.

TAPUS, A.; MATARIC, M. J. Socially assistive robots: The link between personality, empathy, physiological signals, and task performance. In: *AAAI Spring Symposium: Emotion, Personality, and Social Behavior*. Los Angeles: AAAC, 2008. p. 133–140. Citado na página 57.

THORNE, A. The press of personality: A study of conversations between introverts and extraverts. *Journal of Personality and Social Psychology*, American Psychological Association, v. 53, n. 4, p. 718, 1987. Citado na página 43.

TINTAREV, N. et al. Personal storytelling: Using natural language generation for children with complex communication needs, in the wild... *International Journal of Human-Computer Studies*, v. 92–93, p. 1–16, 2016. Citado na página 16.

VAPNIK, V. N.; CHERVONENKIS, A. J. Theory of pattern recognition. Nauka, 1974. Citado na página 29.

WAL, R. et al. The role of automated feedback in training and retaining biological recorders for citizen science. *Conservation Biology*, Wiley Online Library, v. 30, n. 3, p. 550–561, 2016. Citado na página 16.

WALKER, M. A. Endowing virtual characters with expressive conversational skills. In: SPRINGER. *International Workshop on Intelligent Virtual Agents*. Berlin Heidelberg, 2009. p. 1–2. Citado na página 26.

WALKER, M. A. et al. Perceived or not perceived: Film character models for expressive NLG. In: SI, M. et al. (Ed.). *Lecture Notes in Computer Science 7069*. Vancouver: Springer Verlag, 2011. p. 109–121. Citado 3 vezes nas páginas 55, 60 e 79.

WALKER, M. A. et al. Murder in the arboretum: Comparing character models to personality models. In: *Intelligent Narrative Technologies*. Santa Cruz: AAAI, 2011. Citado 2 vezes nas páginas 56 e 60.

WALKER, M. A.; RAMBOW, O. C. Spoken language generation. *Computer Speech & Language*, Academic Press, v. 16, n. 3, p. 273–281, 2002. Citado na página 16.

WHITE, M.; CLARK, R. A.; MOORE, J. D. Generating tailored, comparative descriptions with contextually appropriate intonation. *Computational Linguistics*, MIT Press, v. 36, n. 2, p. 159–201, 2010. Citado 3 vezes nas páginas 23, 24 e 25.

WILCOXON, F. Individual comparisons by ranking methods. *Biometrics bulletin*, JSTOR, v. 1, n. 6, p. 80–83, 1945. Citado na página 30.

WILLIAM, M.; THOMPSON, S. Rhetorical structure theory: towards a functional theory of text organization. *Text*, v. 8, n. 3, p. 243–281, 1988. Citado 2 vezes nas páginas 20 e 49.

XU, A. et al. A new chatbot for customer service on social media. In: ACM. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. Denver, CO, USA, 2017. p. 3506–3510. Citado na página 58.

ZHANG, X.; LAPATA, M. Chinese poetry generation with recurrent neural networks. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, 2014. p. 670–680. Citado na página 16.

Anexo A – Protocolo de aplicação

A.1 Condições

As condições definidas para a aplicação da coleta de dados são:

- Participantes maiores de 18 anos;
- Pessoas dispostas a ajudar;
- Preferencialmente em local calmo, silencioso;
- Responder individualmente e sem plateia;
- Responder com calma;
- Evitar interrupções, se possível.

O experimento de coleta de dados é composto de duas partes independentes: inventário de fatores de personalidade e redação com base em cenas. O inventário deve ser preferencialmente respondido via Facebook¹.

A.2 Início do experimento

Ao iniciar o experimento, o aplicador deve passar as seguintes orientações ao participante:

- Trata-se de um experimento de coleta de dados para uso no meu mestrado em computação.
- O resultado do meu trabalho depende de você me dar as melhores respostas possíveis, com toda atenção.
- As suas respostas serão usadas para fim de pesquisa científica, mas sempre de forma anônima.
- São duas partes: inventário e redação, se faltar tempo podemos interromper e continuar em outra ocasião.
- Sem contar o inventário, leva cerca de 20-25 minutos.
- Se você usa Facebook, pode primeiro preencher o inventário do Facebook e marcamos outro dia para continuar. Ou você faz agora mesmo e depois eu lhe ajudo com o resto.

¹ Inventário disponível no link <https://teste-personalidade.herokuapp.com/>

A.3 Inventário

A.3.1 Facebook

São 44 perguntas de marcar, sobre a sua personalidade. Por favor responda com atenção. No final a ferramenta lhe mostra seu perfil psicológico. Quando acabar, basta me avisar e marcamos a parte presencial.

A.3.2 Inventário presencial

Mas se você não tem Facebook, ou se não usa com frequência, ou se não deseja usar para essa pesquisa, então vamos continuar por aqui. São 44 perguntas de marcar, sobre a sua personalidade. Por favor responda com atenção. Ao final você me chama e então continuamos.