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

DANIELLE SAMPAIO MONTEIRO

Seleção de conteúdo referencial com base em traços de personalidade

São Paulo

2018

DANIELLE SAMPAIO MONTEIRO

Seleção de conteúdo referencial com base em traços de 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: Metodologia e Técnicas da Computação

Versão corrigida contendo as alterações solicitadas pela comissão julgadora em 28 de setembro 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 4936

Monteiro, Danielle Sampaio

Seleção de conteúdo referencial com base em traços de personalidade / Danielle Sampaio Monteiro ; orientador, Ivandré Paraboni. – 2018.

72 f. : il.

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. Personalidade - Análise. I. Paraboni, Ivandré, orient. II. Título.

CDD 22.ed.– 006.3

Dissertação de autoria de Danielle Sampaio Monteiro, sob o título “**Seleção de conteúdo referencial com base em traços de 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 28 de setembro de 2018 pela comissão julgadora constituída pelos doutores:

Prof. Dr. Ivandré Paraboni

Universidade de São Paulo

Presidente

Profa. Dra. Karina Valdivia Delgado

Universidade de São Paulo

Profa. Dra. Ana Amélia Benedito Silva

Universidade de São Paulo

Profa. Dra. Ariadne Maria B. Rizzoni Carvalho

Universidade Estadual de Campinas

*Dedico esta dissertação à minha família, em especial à minha avó querida Aldenora Maria
(in memoriam), a qual me ensinou tantos valores.*

Agradecimentos

Por todo esse caminho percorrido até o momento para fechar mais um ciclo de crescimento acadêmico, profissional e moral, aqui deixo meu total e sincero agradecimento...

A Deus, minha família, meus amigos e meus mestres por mais essa etapa de vida concluída! Embora seja um clichê, eu realmente devo muito a todos eles.

A Deus pela fé e força que tive nos momentos mais difíceis dessa caminhada para me reerguer, e olha que parece que foram muitos. À minha família, especialmente à minha mãe, Antonia, avós Aldenôra (*in memoriam*) e Genésio (*in memoriam*) e meu irmão Jailton pela educação, tanto acadêmica como moral. Ao meu esposo Malcon por ser meu alicerce e não ter deixado eu desistir em inúmeras ocasiões, por ser o grande responsável por minha estadia numa cidade tão longe, e claro por me completar na vida como um todo sendo parte do propósito de criar uma família com o nosso, já tão amado, filho Miguel e agregando minha sogra D Ceíça, cunhada Magda e sogro Mauro a minha família. E sem deixar de lembrar da família que estreitei os laços por meio do seu acolhimento em seu lar, em especial meus amores Carla e Larissa.

Aos meus amigos que com certeza são parte de tudo que me mantém viva, pelo apoio, puxões de orelha e experiências. Em especial aos conquistados no decorrer dessa caminhada e que eram cúmplices dos mesmos desafios (minha irmã de quarto Fernanda, a mãe de todos Duda, o comediante Paulo, e a socorre-todos Sahra). E claro que os antigos que já me conhecem de tantos outros desafios, em especial minha irmã de alma Lethycia, Emanuele, Yasmim e Teresa.

Aos meus professores pela oportunidade de enxergar mais longe. O conjunto deles me fez chegar até aqui, e vai fazer eu ir mais longe. E claro, que em especial ao meu orientador Ivandré por ter me guiado, incentivado e especialmente pela dedicação, auxílio e paciência.

“O que passou, passou, mas o que passou luzindo, resplandecera para sempre.”

(Johann Goethe)

Resumo

MONTEIRO, Danielle Sampaio. **Seleção de conteúdo referencial com base em traços de personalidade**. 2018. 71 f. Dissertação (Mestrado em Ciências) – Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo, 2018.

O presente trabalho traz um estudo no âmbito de Geração de Língua Natural, com ênfase na tarefa de Geração de Expressões de Referência (GER), a qual consiste em gerar expressões referenciais semelhantes às produzidas por humanos. Existem estudos que exploram o uso da variação individual do ser humano no aprendizado do padrão de seleção de conteúdo na construção de descrições, contudo, treinar tais conjuntos de dados é computacionalmente caro. O trabalho apresenta um modelo de seleção de conteúdo para GER, baseado em traços de personalidade, o qual generaliza padrões de comportamentos referenciais similares em cada perfil de personalidade. Na pesquisa também realizou-se um levantamento bibliográfico sobre o tema, e construiu-se um corpus com expressões de referência contendo informações de personalidade de cada participante, as quais foram anotadas tomando por base o modelo dos Cinco Grandes Fatores. Este corpus tem como finalidade ser utilizado como entrada tanto no modelo desenvolvido, como em outros estudos na área. Os resultados comprovam que modelos de GER dependentes da personalidade superam os algoritmos GER tradicionais, e que são uma alternativa viável em abordagens que dependam da variação de locutores.

Palavras-chaves: Geração de Expressões de Referência, Traços de Personalidade, Modelo dos Cinco Grandes Fatores.

Abstract

MONTEIRO, Danielle Sampaio. **Selection of referential content based on personality traits**. 2018. 71 p. Dissertation (Master of Science) – School of Arts, Sciences and Humanities, University of São Paulo, São Paulo, 2018.

The present work presents a study in the field of Generation of Natural Language, with emphasis on the task of Generation of Reference Expressions (GER), which is to generate reference expressions similar to those produced by humans. There are studies that explore the use of individual human variation in learning the pattern of content selection in the construction of descriptions, however, training such datasets is computationally expensive. The paper presents a content selection model for GER based on personality traits, which generalizes patterns of similar referential behavior in each personality profile. The research also carried out a bibliographic survey on the subject, and a corpus was constructed with reference expressions containing personality information of each participant, which were annotated based on the model of the Five Great Factors. This corpus is intended to be used as an input in both the developed model and other studies in the area. The results show that personality-dependent GER models exceed traditional GER algorithms, and are a viable alternative in approaches that depend on the variation of speakers.

Keywords: Generation of Reference Expressions, Personality Traits, Big Five.

Lista de figuras

Figura 1 – Contexto com três entidades $p1$, $p2$ e $p3$	20
Figura 2 – Exemplos de cenas exibidas aos participantes do cóp $\acute{u}$ pus TUNA Furniture e TUNA People, respectivamente.	23
Figura 3 – Exemplo de cena exibida aos participantes do cóp $\acute{u}$ pus GRE3D3.	24
Figura 4 – Exemplo de cena exibida aos participantes do cóp $\acute{u}$ pus GRE3D7.	25
Figura 5 – Exemplo de cena exibida aos participantes do cóp $\acute{u}$ pus <i>Stars</i>	26
Figura 6 – Exemplos de cenas exibidas aos participantes do cóp $\acute{u}$ pus <i>Stars2</i>	27
Figura 7 – Exemplo de cena de estímulo visual para o cóp $\acute{u}$ pus <i>b5-ref</i>	45
Figura 8 – Contexto com três entidades $p1$, $p2$ e $p3$	53

Lista de algoritmos

Algoritmo 1 – Algoritmo Incremental	35
Algoritmo 2 – Algoritmo Combinação das previsões	53

Lista de quadros

Quadro 1 – Resumo dos trabalhos apresentados que consideram a variação humana para a tarefa de GER	43
Quadro 2 – Quadro da base de conhecimento.	69

Lista de tabelas

Tabela 1 – Atributos mais frequentes.	46
Tabela 2 – Resultado Dice médio do experimento piloto	49
Tabela 3 – Resultado Dice médio e desvio padrão (DP) por polaridade de cada dimensão	49
Tabela 4 – Resultado dos classificadores na fase de treinamento	58
Tabela 5 – Coeficiente de Dice médio e Desvio Padrão (DP) dos modelos computa- cionais de GER.	59
Tabela 6 – Resultados das métricas Precisão, Revocação e Medida F dos classifica- dores dos modelos implementados	59

Lista de abreviaturas e siglas

BFI	Big Five Inventory
CGF	Cinco Grandes Fatores
GER	Geração de Expressões de Referência
GLN	Geração de Língua Natural
IGFP-5	Inventário dos Cinco Grandes Fatores de Personalidade
NEO-PI	Inventário de Personalidade NEO
NEO-PI-R	Inventário de Personalidade NEO Revisado
NEO-FFI	Inventário dos Cinco Fatores NEO
PLN	Processamento de Língua Natural

Sumário

1	Introdução	16
1.1	<i>Problema de pesquisa</i>	17
1.2	<i>Objetivos</i>	18
1.3	<i>Hipótese de pesquisa</i>	18
1.4	<i>Organização do Documento</i>	18
2	Conceitos Fundamentais	20
2.1	<i>Geração de Expressões de Referência</i>	20
2.2	<i>Córpus para GER</i>	22
2.2.1	O córpus TUNA	23
2.2.2	O córpus GRE3D3	23
2.2.3	O córpus GRE3D7	24
2.2.4	O córpus Stars	25
2.2.5	O córpus Stars2	26
2.3	<i>Avaliação em GER</i>	28
2.4	<i>O modelo de personalidade humana dos Cinco Grandes Fatores (CGF)</i>	28
2.5	<i>Inventários CGF</i>	30
2.6	<i>Geração de Língua Natural e Personalidade</i>	31
3	Trabalhos Relacionados	33
3.1	<i>Algoritmos Tradicionais de GER</i>	33
3.1.1	Full Brevity	33
3.1.2	Algoritmo Guloso (Greedy)	34
3.1.3	Algoritmo Incremental	35
3.1.4	Algoritmo Baseado em Grafos	37
3.2	<i>A Variação Humana em GER</i>	38
3.2.1	Extensões do algoritmo Incremental e Full Brevity com variação humana	38
3.2.2	Uso de exemplos de descrições recentes	39
3.2.3	Variação Humana nos córpus GRE3D3 e GRE3D7	39
3.2.4	Variação Humana usando algoritmos baseados em grafos	40

3.2.5	Variação Humana em múltiplos domínios	41
3.2.6	Considerações	43
4	Construção do corpus <i>b5-ref</i>	44
5	Experimento piloto	48
5.1	<i>O experimento</i>	48
5.2	<i>Resultados</i>	49
5.3	<i>Discussão</i>	50
6	Modelo computacional proposto	51
6.1	<i>Classificadores</i>	51
6.2	<i>Características de aprendizagem</i>	52
6.3	<i>Combinação das predições dos classificadores</i>	52
7	Experimentos para avaliação do modelo <i>PersonalityGER</i> . . .	55
7.1	<i>Modelos considerados</i>	55
7.2	<i>Hipóteses</i>	56
7.3	<i>Dados</i>	56
7.4	<i>Procedimento</i>	57
8	Resultados	58
8.1	<i>Resultados da tarefa de classificação</i>	58
8.2	<i>Resultados na tarefa de GER</i>	59
8.3	<i>Discussão</i>	60
8.4	<i>Considerações</i>	61
8.5	<i>Teste complementar com atributos afetivos</i>	62
9	Conclusão	63
	Referências¹	64
	Apêndice A – Base de conhecimento para o corpus <i>b5-ref</i> . .	69

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

1 Introdução

Sistemas de Geração de Língua Natural (GLN) são modelos que utilizam técnicas de Inteligência Artificial e Processamento de Língua Natural (PLN) para gerar automaticamente textos, isolados ou como parte de documentos multimídia, a partir de alguma representação, geralmente não linguística (REITER; DALE; FENG, 2000).

Com o aumento da disponibilização de dados e do poder da computação, aumentou também o interesse de pesquisadores em explorar novas possibilidades de aplicação da tarefa de GLN em uma variedade de domínios (GATT; KRAHMER, 2017). Alguns exemplos de aplicações são o uso na geração de relatórios clínicos de pacientes (EUGENIO et al., 2014; JORDAN et al., 2014); na geração de diagnósticos de equipamentos de alta tecnologia como turbinas de gás (REITER et al., 2005); na geração automática de poesia (ZHANG; LAPATA, 2014; OLIVEIRA, 2015); na criação de conteúdo para a plataforma *Twitter* e *chatbots* (DALE, 2016); na geração de notícias com base em dados de um jogo de futebol (AIRES,); dentre outros.

A arquitetura computacional da tarefa de GLN comumente é dividida em três módulos (REITER; DALE; FENG, 2000):

- **Planejamento do documento:** decide quais informações devem ser comunicadas no texto e como essa informação deve ser organizada.
- **Planejamento de sentenças:** decide como a informação é linguisticamente expressa no texto gerado (escolher quais palavras de conteúdo devem ser usadas para expressar conceitos e dados de domínio; escolha de expressões de referência para identificar entidades de domínio; escolha de estruturas sintáticas em sentenças geradas; escolha de quantas mensagens devem ser expressas em cada frase).
- **Realização Superficial:** gera um texto real (forma superficial), baseado na informação selecionada pela tarefa de planejamento de documentos e nas escolhas linguísticas feitas pelo planejamento de sentenças.

Os limites entre os módulos não são precisos, podendo as decisões serem tomadas em conjunto, não somente em módulos independentes (REITER, 2010).

De especial interesse para o presente trabalho, o módulo de planejamento da sentença inclui a tarefa de Geração de Expressão de Referência (GER) que se preocupa

com a geração de descrição de identificação única de um determinado objeto alvo, de forma que a descrição gerada se assemelhe às produzidas por seres humanos.

O desafio de aperfeiçoamento da tarefa de GER consiste exatamente em gerar expressões semelhantes a de locutores humanos, tendo em vista que pessoas diferentes sob circunstâncias idênticas produzem descrições diferentes (VIETHEN; DALE, 2010).

1.1 Problema de pesquisa

Os modelos tradicionais de GER comumente geram a mesma saída se executados com base na mesma entrada, ou seja, produzem saída fixa. Locutores humanos, por outro lado, podem apresentar grande variação linguística, variando assim as descrições produzidas por indivíduos diferentes, dado um mesmo contexto de entrada. Como alternativa para gerar expressões mais próximas das produzidas por humanos, muitos estudos desenvolveram modelos de GER que usam informações específicas a respeito de cada locutor, como por exemplo o uso de listas de atributos preferidos utilizados por cada indivíduo no estudo Bohnet (2008). Entretanto, estes modelos exigem grandes volumes de dados para treinamento de modelos específicos para cada indivíduo, tornando-se de difícil implementação e generalização.

Como solução intermediária entre modelos de GER fixos e os baseados em locutores individuais, consideramos neste trabalho o uso de modelos que usam informações mais genéricas de grupos de pessoas com comportamento referencial similar.

Uma maneira possível de implementar um modelo de GER mais geral e que ainda considere as diferenças entre os locutores, e que será a base do presente trabalho, é assumir que essas diferenças são influenciadas por traços de personalidade. Conhecendo os traços de personalidade de um locutor, seria em princípio possível gerar descrições semelhantes às produzidas por outros locutores com personalidade semelhante, podendo pelo menos superar modelos tradicionais de GER que produzem uma saída fixa, e sem os custos de coleta de exemplos linguísticos de cada locutor em questão.

O modelo dos Cinco Grandes Fatores (CGF) (GOLDBERG, 1990) tem sido considerado por muitos estudiosos como um modelo de representação dos traços de personalidade. O modelo abrange cinco fatores de personalidade que resumem aspectos de comportamento.

Com base nestas observações, este trabalho introduz um estudo sobre GER baseado em personalidade humana para representar uma solução intermediária entre modelos de GER fixos e os modelos baseados em locutor, sem as desvantagens de ambos.

1.2 Objetivos

O objetivo principal desse projeto é desenvolver um modelo de seleção de conteúdo para geração de expressões de referência que leva em consideração traços de personalidade de um locutor alvo de interesse, e que produza descrições próximas das produzidas por humanos.

Como objetivo secundário, pretende-se construir um corpus de GER com informações de personalidade, que auxiliará o desenvolvimento do modelo e que será disponibilizado para dar suporte a outros estudos voltados a geração de texto com base em traços de personalidade.

1.3 Hipótese de pesquisa

A partir do exposto, levanta-se a hipótese de que, conhecendo os traços de personalidade de um locutor, os modelos computacionais de GER baseados neste conhecimento podem gerar expressões de referência semelhantes às produzidas por outros locutores com personalidade semelhante e, ao fazê-lo, superar modelos tradicionais de GER que não utilizam informação de personalidade.

1.4 Organização do Documento

Este documento apresenta um projeto de pesquisa em nível de mestrado acadêmico enfocando o problema computacional da geração de expressões de referência baseada em personalidade humana. O trabalho realizado consiste na construção de um corpus paralelo de expressões de referência e inventários de personalidade, na implementação de modelos de GER para fins comparativos, e no desenvolvimento de um novo modelo de GER que leva em consideração informações de personalidade utilizando métodos supervisionados de aprendizado de máquina.

Ao propor um modelo inédito de GER baseado em personalidade, este estudo pretende avançar o conhecimento sobre o problema computacional de GER e sobre o uso de informação de personalidade em sistemas de GLN e, de modo mais específico, contribuir para o avanço da pesquisa em GLN desenvolvida para o português brasileiro.

Além desta Introdução, este trabalho está organizado da seguinte forma:

- O capítulo 2 apresenta os conceitos fundamentais utilizados como base para o desenvolvimento deste trabalho.
- O capítulo 3 contém uma revisão bibliográfica elucidando assuntos relativos aos métodos tradicionais utilizados na área de GER e trabalhos relacionados.
- O capítulo 4 apresenta a metodologia do trabalho, relatando o procedimento utilizado para definição do modelo de avaliação.
- O capítulo 4 contém a descrição sobre a coleta e anotação dos dados coletados para a criação do corpus *b5-ref*.
- O capítulo 5 apresenta um experimento piloto.
- O capítulo 6 apresenta o modelo proposto.
- Os capítulos 7 e 8 apresentam os métodos de avaliação e os resultados alcançados na pesquisa.
- O capítulo 9 refere-se as considerações finais a respeito da presente proposta.

2 Conceitos Fundamentais

Este capítulo trata de conceitos básicos da abordagem para GER, descreve alguns corpus dedicados a estudos deste tipo e as métricas de avaliação mais comumente utilizadas na área. Além disso, é apresentado um breve histórico do desenvolvimento do modelo dos Cinco Grandes Fatores (CGF) e alguns dos inventários usados para estimar suas dimensões.

2.1 Geração de Expressões de Referência

A Geração de Expressões de Referências (GER) é uma subtarefa fundamental de GLN e tem sido uma linha ativa de pesquisas (KRAHMER; DEEMTER, 2012). GER é a tarefa que dado um referente que se deseja descrever, fornece uma expressão referencial discriminatória que permite identificar este referente com êxito em um dado contexto (REITER; DALE; FENG, 2000).

A tarefa de GER pode ser dividida em dois componentes independentes: seleção de conteúdo (determinar quais propriedades incluir na descrição para distinguir um objeto-alvo de interesse dos demais objetos no mesmo contexto) e realização textual, ou superficial (determinar como representar as propriedades selecionadas em língua natural) (KRAHMER; DEEMTER, 2012). O presente estudo trata em específico da fase de seleção de conteúdo, que será referenciada como sinônimo de GER no decorrer do trabalho.

A entrada da tarefa computacional de GER é um contexto C contendo um objeto-alvo r que se deseja descrever, e demais objetos ditos distraidores D . Um exemplo de contexto deste tipo é ilustrado na Figura 1¹.

Figura 1 – Contexto com três entidades $p1$, $p2$ e $p3$

Fonte: Imagem construída a partir de exemplos da base Face Place (RIGHI; PEISSIG; TARR, 2012)

¹ O uso de imagens da base Face Place é uma cortesia de Michael J. Tarr, Center for the Neural Basis of Cognition and Department of Psychology, Carnegie Mellon University.

O contexto da Figura 1 pode ser representado por uma base de conhecimento composta de propriedades semânticas na forma de pares atributo-valor como segue.

Por exemplo, na Figura 1 temos o contexto $C = \{ p1, p2 \text{ e } p3 \}$ e considera-se os atributos *mf*, *race* e *emotion*, então tem-se que as propriedades de cada entidade de C podem ser representadas da seguinte forma:

- $p1 = \{ [mf, female], [race, caucasian], [emotion, neutral] \}$
- $p2 = \{ [mf, male], [race, caucasian], [emotion, positive] \}$
- $p3 = \{ [mf, female], [race, black], [emotion, positive] \}$

Tomando-se por entrada um contexto C deste tipo e um objeto-alvo r , o objetivo de um algoritmo de GER é selecionar um subconjunto de propriedades L que sejam verdadeiras para r , tal que L permita distinguir r dos outros objetos distraidores em C .

Seja o objeto-alvo $r = p1$ a ser descrito, iremos inicialmente selecionar seu atributo *mf* e comparar com os mesmos atributos dos distraidores ($D = \{p2, p3\}$). Sendo a propriedade de $p2$ igual a $[mf, male]$, ou seja, diferente da propriedade de $p1$, o distraidor $p2$ é excluído do conjunto D e a propriedade selecionada adicionada à descrição, $L = \{[mf, female]\}$. Como o conjunto D ainda não está vazio, é selecionado outro atributo para comparação (*race*). Sendo a propriedade de $p3$ igual a $[race, black]$ diferente da propriedade de $p1$, o distraidor $p3$ é excluído do conjunto D e a propriedade selecionada adicionada à L . Como o conjunto D agora está vazio, a tarefa de GER retornaria $L = \{[mf, female], [race, caucasian]\}$, que poderia ser realizada, por exemplo, como “A mulher caucasiana”.

A tarefa de referenciar objetos é realizada cotidianamente pelas pessoas de maneira trivial. Computacionalmente, entretanto, desenvolver essa tarefa de forma natural à percepção humana é um desafio que envolve analisar fatores ligados à psicolinguística para modelar como o locutor estrutura as propriedades para realizar a tarefa, e então desenvolver algoritmos eficientemente capazes de descrever o objeto-alvo (VIETHEN; DALE et al., 2009).

É esperado que uma expressão de referência seja facilmente compreendida e livre de ambiguidade; por isso, os primeiros algoritmos de GER evitavam o uso de mais atributos que o necessário, baseando-se na máxima de brevidade proposta em Grice (1975). Esta máxima sugere que a descrição ideal é a com menor número possível de propriedades. Por exemplo, no contexto da figura anterior, se o objeto-alvo fosse $p2$, apenas a propriedade

relacionada ao sexo seria selecionada, pois ela sozinha elimina todos os distraidores, e “O homem” seria gerado.

Em Engelhardt, Bailey e Ferreira (2006) foi observado que por outro lado, locutores tendem a incluir atributos redundantes às suas descrições de objetos-alvos, que por consequência, não são necessariamente descrições breves. Descrições que contêm informações que não são estritamente necessárias para desambiguação são ditas superespecificadas. Utilizando o exemplo anterior, a descrição do objeto-alvo *p2* poderia ser realizada de forma superespecificada como em “A pessoa caucasiana de expressão positiva que é homem”.

Resumidamente, pessoas não são breves, mas o objetivo dos modelos de GER é ser breve, para evitar que a expressão não seja ambígua (DALE; REITER, 1995).

2.2 *Córpus para GER*

A utilização de *córpus* é um recurso empregado em diversas áreas do Processamento de Língua Natural (PLN) e também para auxiliar os estudos de GER. Com o uso de *córpus* de GER pode-se entender como algoritmos podem gerar expressões de referência semelhantes às expressões de referência produzidas por seres humanos. Segundo em Krahmer e Deemter (2012), estes *córpus* podem ser separados em *córpus* de propósito geral e *córpus* dedicados.

Córpus de GER de propósito geral são aqueles em que a tarefa de descrição de objetos é secundária dentro de uma outra tarefa principal qualquer, tornando natural a realização da tarefa de GER. Como exemplos, estão os *córpus* *Map Task* (ANDERSON et al., 1991) e COCONUT (EUGENIO et al., 2000). *Córpus* de GER dedicados são aqueles em que a tarefa principal é referenciar entidades. Como exemplos, estão os *córpus* TUNA (GATT; SLUIS; DEEMTER, 2007), GRE3D3 (VIETHEN; DALE, 2008), GRE3D7 (VIETHEN; DALE, 2011), *Stars* (PARABONI et al., 2014) e *Stars2* (PARABONI; GALINDO; IACOVELLI, 2016). Como a proposta do trabalho inclui a construção de um *córpus* dedicado, alguns destes estudos são discutidos a seguir.

2.2.1 O córpus TUNA

O córpus TUNA (GATT; SLUIS; DEEMTER, 2007) é um conjunto de expressões de referência coletados em uma série de experimentos via *web*, nos quais era solicitado ao participante descrever objetos-alvos em dois domínios diferentes: TUNA Furniture: com imagens de itens de mobília; TUNA People: com fotografias de matemáticos famosos em preto e branco.

Exemplos de imagens de estímulo utilizadas no córpus TUNA são ilustradas na Figura 2. As descrições foram solicitadas com base em imagens contendo um ou dois objetos-alvos, os quais podiam ser do mesmo tipo ou não. Em cada imagem, os objetos-alvos eram identificados com uma borda vermelha e acompanhados de 6 distraidores, posicionados aleatoriamente em uma grade 3x5. Não foram utilizados relacionamentos entre os objetos, ou seja, as propriedades utilizadas eram atômicas.

Figura 2 – Exemplos de cenas exibidas aos participantes do córpus TUNA Furniture e TUNA People, respectivamente.

Fonte: Deemter et al. (2012)

No total, o córpus TUNA resultou em 2280 expressões de referência produzidas por 60 participantes, sendo 1200 descrições no domínio TUNA Furniture e 1080 no domínio TUNA People.

2.2.2 O córpus GRE3D3

O córpus GRE3D3 (VIETHEN; DALE, 2008) foi construído para analisar o uso de atributos relacionais na tarefa de GER em um contexto tridimensional com até 3 objetos, que podiam ser do tipo cubo ou esfera, nas cores amarela, azul, verde ou vermelha, e nos tamanhos pequeno ou grande. Em cada cena, o objeto-alvo era identificado com uma

seta. Os contextos em GRE3D foram planejados para que não fosse necessário o uso de atributos relacionais para descrever unicamente o objeto-alvo. Um exemplo de imagem de estímulo utilizada neste corpus é ilustrado na Figura 3.

Figura 3 – Exemplo de cena exibida aos participantes do corpus GRE3D3.

Fonte: Viethen e Dale (2008)

No total, o corpus GRE3D3 contém 630 descrições produzidas por 63 participantes (34 homens e 29 mulheres). Na análise de resultados em Viethen e Dale (2008), chegou-se à conclusão de que as relações espaciais eram utilizadas mesmo quando não havia necessidade, principalmente no começo do experimento, pois os participantes no decorrer do experimento percebiam não haver necessidade de utilizá-las. Ainda assim, alguns participantes utilizaram relações espaciais em todas as descrições, o que confirmou a hipótese de que as relações espaciais estavam relacionadas à preferência de cada participante. Concluiu-se também que, quando o objeto-alvo se encontrava no topo de outro objeto, o uso de relações espaciais era mais comum do que entre objetos lado-a-lado.

2.2.3 O corpus GRE3D7

O corpus GRE3D7 (VIETHEN; DALE, 2011) é uma extensão do corpus GRE3D3 com cenas tridimensionais e figuras geométricas, porém contendo cenas compostas por sete objetos divididos em dois grupos de cores iguais (um de 4 e outro de 3). Neste domínio, a relação entre o objeto-alvo e o objeto relacional nunca era única, ou seja, um dos dois outros pares de objetos em cada cena foi organizado de modo a exibir a mesma relação espacial que o par do objeto-alvo exibia. No entanto, os objetos no par com a mesma relação eram de tipos diferentes, de modo que uma descrição contendo o tipo do alvo e uma relação é totalmente distinta. Por exemplo, na figura 4 temos a cena de estímulo

utilizada no corpus GRE3D7, a qual possui dois relacionamentos lado-a-lado, ‘o quadrado vermelho ao lado da bola vermelha’ e ‘o quadrado amarelo ao lado do quadrado vermelho’.

Figura 4 – Exemplo de cena exibida aos participantes do corpus GRE3D7.

Fonte: Viethen e Dale (2011)

Assim como no corpus GRE3D3, o uso de atributos relacionais não era necessário para diferenciar o objeto-alvo. Para evitar a repetição deste padrão (como identificado no experimento anterior), as imagens de estímulo foram intercaladas com cenas conhecidas como *fillers*, que não representavam as condições do experimento.

No total, o corpus GRE3D7 contém 4480 descrições produzidas por 287 participantes. Na análise em Viethen e Dale (2011), concluiu-se que quando o objeto-alvo é do mesmo tamanho que outros objetos-relacionais na cena, é mais provável que relações espaciais sejam incluídas. Percebeu-se também que objetos-relacionais que possuem a mesma cor que o objeto-alvo têm maior probabilidade de serem selecionados, e que quando o objeto-alvo se encontrava no topo de outro objeto o uso desta relação espacial era novamente preferível ao uso de atributos relacionais laterais.

2.2.4 O corpus Stars

O corpus *Stars* (PARABONI et al., 2014) foi desenvolvido para auxiliar na análise de expressões referenciais produzidas por humanos em condições que favorecem a subespecificação do objeto tomado por ponto de referência em expressões relacionais como “a estrela ao lado da *caixa*” em um contexto com vários objetos do tipo caixa, porém apenas um ao lado de uma estrela. Além disso, o corpus *Stars* faz uso de cenas visuais mais complexas do que as utilizadas nos corpus GRE3D3/7, envolvendo a necessidade de descrever até três objetos (ou seja, um alvo, um ponto de referência e um segundo ponto de referência) como em “a estrela ao lado da caixa, próxima a um círculo”.

Um exemplo de estímulo utilizado neste córpus é ilustrado na figura 5.

Figura 5 – Exemplo de cena exibida aos participantes do córpus *Stars*.

Fonte: Paraboni et al. (2014)

O córpus *Stars* é representado por meio de 6 cenas intercaladas com 5 cenas *fillers* (para que os participantes não percebessem padrões de referência). As cenas eram compostas de estrelas, círculos e quadrados, onde o participante descreveria a estrela apontada por uma seta.

No total, o córpus *Stars* contém 704 descrições produzidas por 64 participantes. A análise em Paraboni et al. (2014) concluiu que, embora os participantes utilizem tanto descrições superespecificadas quanto subespecificadas, sendo que as subespecificadas foram menos frequentes quando era possível descrever o objeto-relacional pela cor, e que expressões de referência mínimas foram raras.

2.2.5 O córpus *Stars2*

O córpus *Stars2* (PARABONI; GALINDO; IACOVELLI, 2016) foi construído com o objetivo de investigar a questão da superespecificação de expressões de referência em contextos com tipos de objetos e oportunidades para produção de expressões de referência mais complexas.

Assim como no caso do córpus *Stars*, as expressões podiam envolver a necessidade de descrever até três objetos, e as situações de estímulos para a produção das descrições disponibilizavam uma variedade de alternativas para especificação referencial e o uso de propriedades altamente salientes que não eram estritamente necessárias para a desambiguação.

A coleta do córpus foi realizada com a participação de duplas de participantes (um locutor e um ouvinte) em um experimento com uso de uma ferramenta online chamada

Lausanne (IACOVELLI; GALINDO; PARABONI, 2014). A ferramenta apresentava o mesmo contexto de referência em duas telas, sendo uma para o locutor e outra para o ouvinte. Cabia ao locutor descrever o objeto-alvo indicado e enviar a descrição para o ouvinte, que então tinha que identificar o objeto em questão na sua própria tela (PARABONI; GALINDO; IACOVELLI, 2016).

O experimento *Stars2* consistiu no uso de 8 imagens em 2 modos de apresentação, ou seja, 16 imagens de estímulo, onde cada modo era ligeiramente diferenciado em grupos de objetos sobrepostos e não sobrepostos, como na figura 6. As imagens de estímulos foram elaboradas em duas condições: 4 imagens em que a descrição do objeto-alvo necessitava do uso de propriedade relacional, e 4 imagens em que o objeto-alvo poderia ser descrito só por propriedades atômicas. As 16 imagens eram apresentadas e escolhidas num conjunto de 64 opções de forma aleatória. Cada imagem possuía 15 objetos (cones, cubos e esferas).

Exemplos de estímulos utilizados neste cópua são ilustrados na figura 6.

Figura 6 – Exemplos de cenas exibidas aos participantes do cópua *Stars2*.

Fonte: Paraboni, Galindo e Iacovelli (2016)

No total, o cópua *Stars2* contém 884 descrições produzidas por 56 participantes. A análise em Paraboni, Galindo e Iacovelli (2016) concluiu que a propriedade que locutores humanos tendem a escolher para superespecificação é a mais saliente, observação esta que não é inteiramente considerada por modelos de GER existentes, como os algoritmos Greedy (DALE, 1989) e Incremental (DALE; REITER, 1995), a serem discutidos no próximo capítulo.

2.3 Avaliação em GER

Partindo do pressuposto de que os algoritmos de GER devam gerar expressões que sejam semelhantes às produzidas por humanos, vários estudos foram feitos para determinar métodos de avaliação desses algoritmos. O método dominante na área consiste em mensurar a similaridade entre as expressões geradas e as que estão em um *cópus* de expressões referenciais (KRAHMER; DEEMTER, 2012).

As expressões de referência geradas por algoritmos de GER podem ser comparadas a expressões de referência produzidas por humanos a partir de avaliação intrínseca por meio das métricas de Dice (DICE, 1945).

O coeficiente Dice (DICE, 1945) é calculado levando em consideração dois conjuntos de expressões de referência A e B, um gerado pelo algoritmo e o outro produzido por seres humanos (descrições do *cópus* escolhido). A fórmula é apresentada a seguir (DICE, 1945):

$$Dice(A, B) = \frac{2 \times |A \cap B|}{|A| + |B|} \quad (1)$$

O coeficiente de Dice varia entre 0 (nenhum elemento compartilhado entre os conjuntos) e 1 (total similaridade entre os dois conjuntos).

Finalmente, a Acurácia indica se os conjuntos A e B são idênticos atribuindo o valor 1, ou 0, caso contrário (binária), ou seja, verifica se há igualdade total entre a expressão gerada e a do *cópus*. Devido a sua baixa granularidade, essa métrica é geralmente interpretada com fins ilustrativos em complemento às métricas que detalham mais o comportamento do algoritmo, como a métrica Dice.

2.4 O modelo de personalidade humana dos Cinco Grandes Fatores (CGF)

Na área da Psicologia, a análise de personalidade é um instrumento de medida padronizada de uma amostra para se fazer inferências sobre o comportamento dos examinados. Estudos sobre personalidade sugerem que as pessoas possuem predisposições para responder a diferentes estímulos (dependendo do contexto), e que com testes objetivos de personalidade é possível estabelecer uma medida padronizada desse comportamento (ANDRADE, 2008). Em outras palavras, a personalidade pode ser usada para resumir, prever e explicar a conduta de uma pessoa (PERVIN; JOHN, 2009).

A Teoria dos Traços de Personalidade é uma das principais abordagens para o estudo da personalidade na Psicologia. Como descrito em Jaques e Nunes (2013), o primeiro estudo a trabalhar com o conceito de traços psicológicos para embasar a Teoria dos Traços de Personalidade foi Allport e Allport (1921). Esta teoria propõe que cada pessoa possui características (traços) que determinam padrões de comportamento, e que ao se identificar e mensurar o grau (intensidade) em que esses traços existem no indivíduo, determina-se então a personalidade do indivíduo. Em Jaques e Nunes (2013) é exemplificado o conceito da teoria da seguinte forma: duas pessoas podem ter um traço de personalidade “calmo”, e cada uma delas ter um nível de “calma” diferente devido às influências que cada uma passou ao longo de sua vida. No entanto, o traço latente seria o mesmo, e a personalidade dos indivíduos poderia, portanto, ser descrita com base em traços deste tipo.

O modelo dos Cinco Grandes Fatores (CGF) é uma versão moderna da Teoria dos Traços, que descreve dimensões humanas básicas de forma consistente e replicável (HUTZ et al., 1998). O modelo abrange cinco dimensões (ou fatores de personalidade): Abertura às experiências; Conscienciosidade; Extroversão; Amabilidade e Neuroticismo. Estas cinco dimensões são perspectivas derivadas de análises dos termos da linguagem que os indivíduos usam em seus ambientes naturais para descrever eles próprios e os outros (JOHN; SRIVASTAVA, 1999).

O estudo em John e Srivastava (1999) apresenta as etapas que originaram o modelo dos CGF. Inicialmente, são citados estudos que extraíram do dicionário termos associados a traços de personalidade por meio da abordagem lexical, a qual postula que as características relevantes da personalidade são codificadas na linguagem. O trabalho em Allport e Odbert (1936) apresenta um estudo léxico dos termos relevantes de personalidade a partir de um dicionário completo da língua inglesa, no qual foram selecionados 18.000 termos que poderiam ser usados para distinguir o comportamento de duas pessoas. Partindo desta lista, o trabalho em Cattell (1943), que utilizando procedimentos de agrupamento semântico e empírico, originou uma lista de 35 variáveis, as quais após várias análises de fatores oblíquos foram identificadas em 12 fatores que se tornaram parte do questionário de 16 Fatores de Personalidade. Este estudo estimulou outros pesquisadores a examinarem a estrutura dimensional dos traços de personalidade (ANDRADE, 2008).

Dentre os estudos subsequentes sobre este assunto está o estudo em Tupes e Christal (1961), que identificou cinco fatores recorrentes: Abertura às experiências que engloba características relacionadas a flexibilidade de pensamento, imaginação e interesses

culturais; Conscienciosidade características relacionadas a responsabilidade e honestidade; Extroversão relacionado a características sociáveis como expressividade, socialização, entusiasmo, etc.; Neuroticismo relacionado com características relacionadas a afeto positivo, ansiedade, etc.; e Amabilidade, relacionado a características sociáveis como cuidado, egoísmo, apoio emocional, amor, cooperação, etc. (HUTZ et al., 1998).

Estes fatores são denominados de forma diversa em outros estudos, mas os traços de que os constituem e seu agrupamento são equivalentes nas diferentes abordagens (HUTZ et al., 1998). Intitulados como os Cinco Grandes Fatores, essas cinco dimensões representam a personalidade no nível mais amplo de abstração, e cada dimensão resume um grande número de características de personalidade distintas e mais específicas (JOHN; SRIVASTAVA, 1999). As dimensões são bipolares, ou seja, o modelo defende que cada pessoa apresenta uma variação positiva ou negativa para cada dimensão.

2.5 Inventários CGF

Existem diferentes testes de personalidade e diferentes meios pelos quais pode-se extrair informações sobre as características de um ou mais indivíduos. A avaliação psicológica é um processo de coleta de dados, cuja realização inclui métodos e técnicas padronizadas, dentre eles estão os inventários que possuem como base o modelo CGF, utilizados em virtude de sua replicabilidade e generalidade.

Ao longo do tempo vários inventários foram desenvolvidos, mas surgiu a necessidade de padronizar esses instrumentos. O estudo em McCrae e Costa (1989) desenvolveu o *NEO Personality Inventory* (NEO-PI) para avaliar somente três dimensões: *Neuroticism* (neuroticismo), *Extraversion* (extroversão) e *Openness to experience* (abertura às experiências), por isso foi nomeado pelo acrônimo das três dimensões.

Posteriormente, o estudo em Costa e McCrae (1992) propôs o Inventário de Personalidade Revisado (NEO-PI-R) incluindo as dimensões *Conscienciosidade* e *Amabilidade*. Este inventário consiste em 240 itens que permitem mensurar seis facetas para cada uma das dimensões. Por ser considerado longo, os mesmos autores desenvolveram o *NEO Five Factor Inventory* (NEO-FFI), que é uma versão reduzida baseada em uma análise fatorial do NEO-PI com 60 itens (COSTA; MACCRAE, 1992).

Embora o inventário NEO tenha sido reduzido, ainda demandava muito tempo e gerava dúvidas devido à ambiguidade de seus itens. O estudo em John, Donahue e Kentle (1991) desenvolveu então o BFI (*Big Five Inventory*) composto por 44 itens. Esse inventário, ao contrário dos anteriores, não utiliza adjetivos únicos, mas sim frases curtas baseadas em adjetivos de traços que o item pretendia avaliar, evitando significados ambíguos ou múltiplos. Os 44 itens classificados em uma escala Likert de 5 passos, variam de 1 = “discordo totalmente” até 5 = “concordo totalmente”. Para gerar a pontuação de cada uma das dimensões do modelo CGF os itens são agrupados em itens de pesos positivos e pesos negativos, e calcula-se a média ponderada do total das respostas.

Vale ressaltar que devido ao fato de que as pontuações das dimensões de personalidade são valores escalares, estas devem ser interpretadas de forma relativa ao grupo de participantes analisados, e não de forma absoluta. Por exemplo, ao invés de classificar um participante como introvertido quando este tiver pontuado um valor x para a dimensão *Extroversão*, é preciso analisar o grau de Extroversão dos outros participantes e classificar relativamente a estes – dado dois graus de *Extroversão*, x_1 e x_2 , tais que $x_1 < x_2$, um participante de grau x_1 é menos extrovertido do que um participante de grau x_2 (RAMOS et al., 2017).

O inventário BFI tem se tornado bastante popular por ser mais breve, por não gerar muitas dúvidas, e por ter apresentado replicabilidade em vários países do mundo, inclusive no Brasil (ANDRADE, 2008). A adaptação para o português brasileiro considerada no presente trabalho é o Inventário dos Cinco Grandes Fatores de Personalidade (IGFP-5) validado em Andrade (2008) para 5.089 indivíduos das cinco regiões brasileiras.

2.6 Geração de Língua Natural e Personalidade

Estudos de GLN baseados em personalidade são escassos na literatura, e não foi possível identificar nenhum estudo específico de GER com estas características. Nesta seção descrevemos brevemente o trabalho em Mairesse (2008), que é um dos mais influentes na área de geração de texto com base no modelo CGF.

O trabalho em Mairesse (2008) propõe um modelo de geração de linguagem natural para projetar diferentes traços de personalidade com base no modelo dos CGF, e explora o reconhecimento de personalidade usando modelos treinados por meio de dados extraídos

de ensaios e conversas. A parte do estudo que trata do reconhecimento de personalidade não será tratada no presente trabalho, sendo discutida somente a parte sobre GLN, a qual também é estendida em Mairesse e Walker (2011) e Mairesse e Walker (2010).

Em conjunto, estes estudos descrevem o sistema PERSONAGE, que é um sistema de geração de linguagem natural, desenvolvido no domínio de recomendação e comparação de restaurantes em Nova York.

PERSONAGE implementa a arquitetura tradicional GLN proposta em Reiter, Dale e Feng (2000) vista no capítulo 1 deste trabalho, a qual consiste em componentes sequenciais que processam a saída por meio da correlação entre marcadores linguísticos (por exemplo: repetição, gagueira, pronominalização, verbosidade, etc.) e os traços de personalidade. O gerador usa modelos estatísticos treinados para prever os parâmetros da personalidade alvo ao longo das cinco grandes dimensões da personalidade.

Para avaliar os textos gerados, juízes especialistas avaliam o conjunto de enunciados gerados como se tivessem sido proferidos por um amigo respondendo em um diálogo sobre recomendação de restaurantes. Estes textos manipulam sistematicamente parâmetros relacionados com cada dimensão CGF. Por exemplo, um enunciado é gerado com o parâmetro neurótico, outro com os parâmetros emocionalmente estáveis (MAIRESSE; WALKER, 2008). Com as avaliações dos juízes humanos, os resultados indicaram que os juízes perceberam a personalidade dos enunciados do sistema.

O sistema PERSONAGE realiza um grande número de decisões de geração de baixo e alto nível baseadas em informações de personalidade, mas a questão do uso de informações de personalidade na seleção de conteúdo de GER não é tratada diretamente.

3 Trabalhos Relacionados

Neste capítulo será apresentado um levantamento bibliográfico de algoritmos tradicionais de GER e modelos de abordagens que utilizam a variação humana individual para a tarefa de GER. A análise dos estudos baseados em variação humana é útil para o presente trabalho, pois há o levantamento das possíveis técnicas de modelagem para personalização da seleção de conteúdo. Além disso, estudos deste tipo consistem no que há de mais próximo ao presente trabalho na literatura, já que não foi possível identificar iniciativas de uso de informação de personalidade em GER.

3.1 Algoritmos Tradicionais de GER

Nesta seção serão abordados os quatro algoritmos mais populares e consolidados em estudos como abordagens iniciais para GER.

3.1.1 Full Brevity

O algoritmo Full Brevity (DALE, 1989) propõe construir uma descrição distinta mínima selecionando sempre as propriedades mais discriminatórias disponíveis, baseando-se em dois dos princípios das *Máximas de Conversação de Grice* (GRICE, 1975): o princípio da adequação que exige que uma expressão de referência identifique sem ambiguidade o objeto-alvo e forneça informações suficientes para isso; e o princípio da eficiência que exige que a expressão de referência usada não contenha mais informações do que as necessárias.

Assim como em diversos outros algoritmos de GER, o algoritmo Full Brevity recebe como entrada um contexto de objetos e o objeto-alvo, e tenta gerar uma descrição para o objeto-alvo usando uma única propriedade. Se isso falhar, consideram-se todas as combinações possíveis de duas propriedades para verificar se alguma delas é suficiente para a geração de uma descrição livre de ambiguidade, e assim por diante. Ou seja, o objetivo do algoritmo é encontrar um número mínimo de propriedades possíveis de r que eliminem todos os distraidores no contexto.

Como exemplo de aplicação do algoritmo Full Brevity, considera-se como objeto-alvo a pessoa $p2$ do contexto apresentado na figura 1 da seção 2.1. O algoritmo receberia como

entrada o conjunto de todas as pessoas do contexto e os atributos utilizados seriam $\{mf, race \text{ e } emotion\}$. A descrição do objeto-alvo seria inicializada vazia e os distraidores seriam definidos como todos os objetos do contexto ($p1$ e $p3$) exceto o objeto-alvo. As possíveis descrições para r seriam $\{[mf, male]\}$, $\{[race, caucasian]\}$ e $\{[emotion, positive]\}$. Dentre estas, a propriedade mf , por si só, elimina todos os distraidores e, portanto, a saída do algoritmo seria $L = \{[mf, male]\}$, que poderia ser realizada como “O homem.”

Embora atenda ao que a tarefa de GER propõe, há algumas desvantagens no uso desta abordagem. Em primeiro lugar, o problema de encontrar uma descrição mínima tem complexidade NP-completo e, portanto, é computacionalmente caro. Em segundo lugar, é observado em Dale e Reiter (1995) que seres humanos nem sempre produzem descrições mínimas deste tipo.

3.1.2 Algoritmo Guloso (Greedy)

O algoritmo heurístico Greedy (DALE, 1989) seleciona iterativamente a propriedade que exclui a maior quantidade de distraidores em cada estágio e, como resultado, nem sempre é gerada a descrição mínima. A execução do algoritmo finaliza quando encontra um conjunto de propriedades que elimine todos os distraidores, ou quando todas as propriedades do objeto-alvo são verificadas.

Considere-se novamente o objetivo de descrever o objeto $p2$ na figura 1. Gerando uma descrição L , neste caso, o atributo mf teria o poder discriminatório 2, pois seu valor ($[mf, male]$) eliminaria os distraidores $p1$ e $p3$. Já os atributos $race$ e $emotion$ teriam o poder discriminatório igual a 1, pois seus valores, respectivamente *caucasian* e *negative*, eliminam apenas um distraidor em relação a cada propriedade. Sendo assim, a propriedade ($[mf, male]$) é adicionada a L , e como o atributo elimina dois distraidores existentes, o algoritmo é encerrado e retorna $L = \{[mf, male]\}$, que poderia ser realizada como “O homem”. Embora a saída neste caso seja a mesma do algoritmo Full Brevity na seção anterior, é facilmente observado que este nem sempre seria o caso.

Em Dale e Haddock (1991) é proposto uma extensão do algoritmo *Greedy* que gera expressões de referência relacionais, como em “A mulher à esquerda”. O algoritmo evita o uso de informação redundante, permitindo identificar o objeto-alvo por meio das relações com outros objetos, sem identificar cada um desses objetos separadamente, como por

exemplo “A mulher à esquerda do homem” em vez de “A mulher caucasiana à esquerda do homem sorrindo”. Além disso, trata do problema da recursividade infinita, como em “A mulher à esquerda do homem que está à direita da mulher que...”.

Segundo em Krahmer e Deemter (2012), o algoritmo Greedy é um algoritmo mais eficiente do que o Full Brevity, mas logo foi superado pelo algoritmo Incremental (DALE; REITER, 1995), que se revelou uma abordagem mais influente na área.

3.1.3 Algoritmo Incremental

Um dos algoritmos mais conhecidos na área, e que ajudou a definir o problema computacional de GER, é o Algoritmo Incremental (AI) apresentado em Dale e Reiter (1995). A ideia central deste algoritmo é a seleção de atributos do objeto-alvo de acordo com uma lista ordenada de preferências P . Os atributos são considerados para inclusão na descrição do objeto-alvo incrementalmente, de acordo com a ordem estabelecida na lista de preferências, desde que eliminem pelo menos um dos distraidores. O critério de parada do algoritmo é a obtenção de um conjunto único de propriedades do objeto-alvo que consiga distingui-lo dos distraidores do contexto, ou o término da lista de preferência, caso em que não teria sido possível distinguir o objeto-alvo dos distraidores.

Algoritmo 1 Algoritmo Incremental

```

1: function GETDESCRICAOC( $C, t, P$ )
2:    $L[t] \leftarrow \{\}$ 
3:    $D \leftarrow C - \{t\}$ 
4:   for  $A_i \in P$  do
5:      $V \leftarrow \text{valor}(t, A_i)$ 
6:     if  $\text{eliminaDistraidor}(D, (A_i, V)) > 0$  then
7:        $L[t] \leftarrow L[t] \cup (A_i, V)$ 
8:        $D \leftarrow D - \text{eliminaDistraidor}(D, (A_i, V))$ 
9:     if  $D == 0$  then
10:      return  $L$ 
11:   return  $L$ 
12: function ELIMINADISTRAIDOR( $D, (A_i, V)$ )
13:   if  $V = \text{null}$  then
14:     return 0
15:   else
16:     return  $\{x \in D : \text{valor}(x, A) \neq V\}$ 

```

Fonte: Adaptado de Ferreira e Paraboni (2017)

O algoritmo Incremental recebe como entrada três parâmetros: o conjunto com o contexto, o objeto a ser referenciado e a lista da ordem de preferência. Para início do algoritmo o conjunto L , que consistirá na descrição do objeto-alvo final, é inicializado com vazio, e D , que é a lista de distraidores, é inicializada com o conjunto resultante da extração do objeto-alvo do conjunto do contexto.

Após estas inicializações, o algoritmo itera de acordo com a lista de preferências dos atributos seguindo os seguintes passos. Na linha 5 é atribuído à V o valor da propriedade de t relativo ao atributo iterado e é verificado na linha 6 se esta propriedade elimina algum distraidor por meio da função auxiliar *eliminaDistraidor*, a qual retorna os objetos do conjunto D que possuem valor diferente para o mesmo atributo iterado. Caso haja eliminação, na linha 7 é realizada a inclusão da propriedade verificada em L e na linha 8 é realizada a exclusão dos distraidores retornados da função *eliminaDistraidor*. Seguindo, tem-se as duas condições de parada do algoritmo, na linha 9 quando não há mais distraidores ou quando todos os atributos da lista de preferência são iterados.

Seja o objeto-alvo $r = p1$ e os distraidores $D = \{p2, p3\}$ do contexto da figura 1, e a ordem preferencial de atributos $P = \{mf, race, emotion\}$. O algoritmo Incremental inicia com uma lista L vazia, e itera de acordo com a lista de preferências P na ordem estabelecida. Na primeira iteração, o valor do atributo *mf* (*female*) de r é comparado com o valor dos distraidores ($p2$ e $p3$), e como o valor do mesmo atributo de $p2$ é diferente (*male*), o atributo *mf* é inserido em L . Consequentemente, o distraidor $p2$ é removido de D .

Uma vez que o conjunto D ainda não se encontra vazio, o algoritmo considera o próximo atributo em P , ou seja, *race*, incluindo a propriedade [*race, caucasian*] em L e eliminando o objeto $p3$ que tem o valor diferente de $p1$ ($p3 = [race, black]$). Como todos os distraidores foram excluídos, o algoritmo encerra retornando $L = \{[mf, female], [race, caucasian]\}$, que poderia ser realizada como “A mulher caucasiana”.

Em Dale e Reiter (1995) sugere-se que este algoritmo esteja próximo do comportamento humano, pois humanos tendem a produzir expressões de referência utilizando atributos de acordo com a sua preferência. No entanto, esta ordem de preferência não é estabelecida pelo algoritmo, e tende a ser dependente do domínio.

3.1.4 Algoritmo Baseado em Grafos

Em Krahmer, Erk e Verleg (2003) é descrita uma abordagem do problema de geração de expressões de referência que utiliza uma estrutura do tipo grafo rotulado e direcionado para representar os objetos de um contexto. Nesta abordagem, o contexto é representado por um grafo cujos vértices são os objetos, e cujas arestas são as propriedades – atômicas (laços) e relacionais (arestas de conexão entre vértices) – de cada objeto. O objetivo do algoritmo é encontrar um subgrafo único que represente a descrição de referência de um objeto-alvo no contexto representado.

Para selecionar os atributos é utilizada uma função de custo que permite adequar a geração de expressões de referência a situações diferentes. Por exemplo, é possível atribuir a cada aresta um custo de valor 1 para produzir o menor subgrafo que descreve o objeto-alvo, ou atribuir custos de acordo com uma lista de atributos preferidos (e.g. com valores menores para as propriedades preferidas) simulando assim a implementação do algoritmo Incremental (conforme seção anterior), dentre muitas outras possibilidades.

Para demonstrar o funcionamento desta abordagem, em Krahmer, Erk e Verleg (2003) foi proposto um algoritmo que recebe como entrada um objeto-alvo r e um grafo $G = \{V, A\}$, onde V é o conjunto de vértices representando os objetos no contexto ($V = D$) e A é o conjunto de arestas que representam as propriedades (tanto atômicas como relacionais) dos objetos. O objetivo do algoritmo é retornar um subgrafo H com menor custo em G , tal que represente a descrição de r de forma única, ou seja, tal que não haja nenhum subgrafo idêntico a H em outra região de G .

O algoritmo considera inicialmente um grafo H (que representa a descrição a ser produzida) contendo apenas o vértice que representa o objeto-alvo. Este grafo é incrementado recursivamente com o acréscimo de arestas para vértices adjacentes ou laços com base no seu custo. Para cada iteração, verifica-se se os custos de H são maiores do que os custos do melhor subgrafo encontrado até esse momento, e faz-se um teste de isomorfismo no grafo verificando-se se H é um subgrafo único. Neste ponto, apenas alternativas com custos menores do que esta melhor solução encontrada precisam ser testadas. Posteriormente, é retirado do conjunto de distraidores o objeto excluído pelo atributo analisado. A condição de parada do algoritmo é atingida quando este conjunto estiver vazio ou os custos de H excederem os custos do subgrafo encontrado até o momento.

No final, o algoritmo retorna o subgrafo único com menor custo, se existir, ou o subgrafo vazio.

3.2 A Variação Humana em GER

Esta seção descreve estudos que, partindo da premissa que a tarefa de GER deve ser capaz de replicar o conteúdo referencial semântico produzido por humanos, utilizam técnicas para analisar o comportamento referencial de indivíduos e procuram gerar as descrições com base nas informações obtidas a partir destes.

3.2.1 Extensões do algoritmo Incremental e Full Brevity com variação humana

Em Bohnet (2008) é avaliado o desempenho de extensões do Algoritmo Incremental e Full Brevity no corpus TUNA (BELZ; GATT, 2007) que utilizam informações sobre locutores individuais.

A variação humana utilizando o Algoritmo Incremental se deu pela utilização de uma lista de preferências ordenada pela frequência dos atributos por participante.

Já o algoritmo IS-FP, uma extensão do algoritmo Full Brevity apresentada no estudo, utilizou a técnica de aprendizado do *Vizinho Mais Próximo* (KNUTH, 1998) para selecionar a expressão mais curta obtida da comparação entre as expressões que são mais semelhantes às expressões de referência do participante, e as de um outro participante que tenha expressões semelhantes às do primeiro (a semelhança é processada a partir do coeficiente Dice entre elas). Como resultado, o algoritmo retorna a expressão mais breve e, caso haja empate, é retornada a expressão cujos atributos sejam mais frequentes.

Nos resultados em Bohnet (2008) observou-se que o uso de listas de preferências individuais supera o uso de uma única ordem de preferência computada a partir de todos os dados de treinamento, ou seja, o algoritmo Incremental que levava em consideração a variação dos participantes superou o algoritmo Incremental sem a variação, assim como o algoritmo IS-FP superou o algoritmo de *baseline* que não utilizava a variação individual do participante.

3.2.2 Uso de exemplos de descrições recentes

O trabalho em Fabbrizio, Stent e Bangalore (2008b) abordou as questões de seleção de conteúdo e realização textual de GER considerando a variação na preferência dos participantes nestas tarefas. Como o enfoque do presente trabalho está na seleção de conteúdo, a discussão a seguir é limitada a esta tarefa, que como na seção anterior, também fez uso dos algoritmos Full Brevity e Incremental para gerar descrições a partir do corpus TUNA.

A variação humana em Fabbrizio, Stent e Bangalore (2008b) foi implementada com o uso do Algoritmo Incremental ordenando a lista de preferência de atributos P de acordo com a preferência individual de cada participante. Para a abordagem do algoritmo Full Brevity, a variação humana foi implementada por meio de dois critérios: selecionar uma expressão de referência com os atributos mais frequentes do participante, ou selecionar os atributos mais usados recentemente pelo participante.

No resultado em Fabbrizio, Stent e Bangalore (2008b) concluiu-se que, o Algoritmo Incremental obteve com variação humana melhor desempenho em relação ao mesmo algoritmo sem esta variação, e o algoritmo Full Brevity que utilizou o critério das frequências dos participantes obteve melhor desempenho, inclusive em relação ao algoritmo utilizando atributos recém utilizados.

3.2.3 Variação Humana nos corpus GRE3D3 e GRE3D7

O trabalho em Viethen e Dale (2010) apresenta experimentos com o uso de árvore de decisão para prever padrões no processo de seleção de conteúdo. O estudo foi baseado nos corpus de cenas visuais GRE3D3 e GRE3D7 descritos na seção 2.2.

O experimento utilizou indução de árvores de decisão implementada pelo algoritmo C4.5 (QUINLAN, 2014) para treinar e testar os conjuntos de expressões de referência destes corpus. Foram utilizadas as combinações de atributos (tipo, cor, tamanho, etc.) escolhidos nas expressões produzidas nos corpus para classificar padrões de conteúdo.

Para modelar o experimento, identificou-se quais atributos foram utilizados nas expressões de referência dos corpus GRE3D3 e GRE3D7, chegando-se a uma lista de 9 atributos (4 que descrevem o objeto-alvo, 4 que descrevem o objeto-relacionado e 1 que descreve o relacionamento entre os dois objetos). Cada descrição contida nos corpus foi

caracterizada em termos de um padrão de conteúdo definido pela presença ou ausência de cada um destes 9 atributos. Encontrou-se 37 padrões de conteúdo diferentes nos dois corpúsculos. O corpúsculo GRE3D3 abrangeu 31 destes e o GRE3D7 27, sendo que 21 ocorreram em ambos.

Estas combinações foram utilizadas como classes das respectivas árvores de decisão para classificar a combinação adequada de atributos do objeto-alvo e do objeto-relacionado para compor a expressão de referência, treinadas com e sem poda. A variação humana neste estudo foi implementada por meio da inclusão da identificação do participante como um atributo, juntando-se este dado aos conjuntos das combinações de atributos.

Os resultados em Viethen e Dale (2010) mostram que o classificador que tem acesso aos identificadores do participante supera os que não têm. Esta abordagem é estendida em Viethen (2011), no qual faz-se o experimento de treinar classificadores individuais para cada atributo, com o intuito de prever se cada atributo deve ou não compor a descrição de saída. A avaliação dos resultados desta última abordagem demonstrou melhores resultados que a primeira abordagem (relativa aos padrões de conteúdo completo).

3.2.4 Variação Humana usando algoritmos baseados em grafos

Em Viethen, Mitchell e Krahmer (2013) é avaliado o uso de informações relacionais da abordagem de grafos posposta em Krahmer, Erk e Verleg (2003). Além disso, o estudo apresenta uma extensão do algoritmo baseado em grafos, o LogestFirst, que objetiva encontrar a descrição mais longa – seguindo a observação de que os seres humanos tendem a superespecificar (ENGELHARDT; BAILEY; FERREIRA, 2006) – e com menor custo.

O algoritmo LogestFirst teve os pesos das arestas do grafo treinados por meio de funções de custo definidas em diferentes granularidades para produzir a melhor correspondência com os dados humanos. Para implementar a variação humana, ao invés de utilizar a identificação do participante, o experimento agrupou falantes com comportamento semelhante e definiu funções de custo separadas para cada grupo. Foi usada a técnica de clusterização k-means (JAIN, 2010) para agrupar os participantes em dois grupos: um com participantes que preferiram descrições longas, muitas vezes com atributos relacionais, e outro com participantes que preferiram descrições mais curtas e não-relacionais. Para cada um dos grupos, foram treinadas funções de custo de ordens de preferências.

Como resultado do experimento em Viethen, Mitchell e Krahmer (2013) verificou-se que o algoritmo LongestFirst reproduziu corretamente as relações usadas por pessoas que utilizaram atributos relacionais.

3.2.5 Variação Humana em múltiplos domínios

Assim como em Viethen (2011), o estudo em FERREIRA e PARABONI (2017) aborda o uso de classificadores de atributos individuais para decidir se cada atributo deve ou não compor uma expressão de referência, no entanto, o estudo combina a saída dos classificadores para construir expressões de referência mais completas e reais.

O trabalho em FERREIRA e PARABONI (2017) discute a questão da variação humana em GER por meio de modelos que consideraram o comportamento referencial de locutores individuais, tanto no que diz respeito à escolha de atributo como em nível de especificação, para produzir descrições que se assemelhem às produzidas por cada indivíduo.

O experimento realizado avaliou duas abordagens: uma abordagem baseada no Algoritmo Incremental para gerar descrições relacionais, nomeada de Algoritmo Incremental Estendido (EIA), e uma abordagem utilizando modelos de Aprendizado de Máquina, na qual foi considerado duas implementações alternativas: uma abordagem padrão baseada em Árvores de Decisão (DT) (BREIMAN et al., 1984), como em Viethen e Dale (2010); e uma segunda versão usando Máquinas de Vetores de Suporte com *kernel Gaussiano* (SVM) (CORTES; VAPNIK, 1995), como em Ferreira e Paraboni (2014). Cada uma dessas abordagens foi analisada em versões com (+) e sem (-) informações do locutor.

Na abordagem EIA original, o algoritmo faz uso de uma única ordem da lista de preferências P para cada tipo de objeto no domínio. Na abordagem EIA+ a variação humana foi implementada modificando-se a maneira pela qual a ordem da lista de preferências é computada, de forma semelhante a vista em Fabbrizio, Stent e Bangalore (2008b), ou seja, baseada na frequência dos atributos por locutor.

Nas abordagens de aprendizado de máquina originais (SVM- e DT-) fez-se o uso de atributos de contexto que modelam objetos de domínio e suas similaridades. Para levar em consideração a questão da variação humana, este modelo foi ampliado com duas características: aquelas que modelam informações pessoais sobre os locutores e aquelas que modelam seu comportamento referencial. Considerou-se também de forma breve três

soluções intermediárias: um modelo que considera apenas o uso de identificadores de locutores, um modelo que usa informações de idade e gênero, e um modelo que usa as listas de atributos preferidos do indivíduo.

Como conjuntos de dados para treinamento e testes, seis corpúscos foram considerados: TUNA Furniture, TUNA People, GRE3D3, GRE3D7, Stars e Stars2 .

Para avaliar os modelos propostos, utilizou-se o coeficiente Dice e a Acurácia, e comparou-se algoritmos que não consideravam a variação humana, como os que consideravam. Mais especificamente, comparou-se EIA- com EIA +, SVM- com SVM + e DT- com DT +. Além disso comparou-se a abordagem AIE com ambos os modelos de aprendizado de máquina (SVM e DT).

Para validação dos algoritmos utilizados no experimento, em todos os casos foi utilizada a validação cruzada de 6 *folds* para os corpúscos Stars e TUNA, e a de 10 *folds* para GRE3D3, GRE3D7, e Stars2. A escolha do número de *folds* para os corpúscos foi feita de modo que o número de expressões de referência de cada locutor por *fold* fosse mantido equilibrado.

No caso especial das variações SVM, os melhores valores para os parâmetros de custo (C) e os parâmetros γ do *kernel Gaussiano* foram determinados executando-se o método *Grid Search*. Para cada iteração da tarefa de validação cruzada, um *fold* f foi reservado para o teste, e os outros foram usados em uma tarefa de validação cruzada secundária em que os classificadores foram treinados usando diferentes combinações de C (1, 10, 100 e 1000) e valores de γ (1, 0,1, 0,001 e 0,0001). Os parâmetros que produziram maior acurácia foram selecionados, e então aplicados à tarefa de validação cruzada primária para testar o *fold* f . Este procedimento foi repetido para cada iteração da validação primária.

Como resultados do estudo, notou-se que algoritmos baseados em métodos de aprendizado de máquina (SVM e DT) superaram a abordagem Incremental (EIA) em ambas as versões (com e sem informação do locutor). Dentre os modelos de aprendizado de máquina, a estratégia de melhor desempenho foi a SVM+. Em relação ao uso de informações relacionadas ao locutor, constatou-se que os modelos que utilizavam essa informação se saíram melhor, principalmente no caso dos modelos de aprendizado de máquina. Já em relação ao teste das três outras alternativas de abordagem, o uso do conjunto completo de características dependentes do participante superou as alternativas que consideram subconjuntos destas.

3.2.6 Considerações

Nesta seção apresentamos uma visão geral de estudos que levam em consideração a variação humana na tarefa de GER. O quadro 1 resume os trabalhos apresentados neste capítulo.

Quadro 1 – Resumo dos trabalhos apresentados que consideram a variação humana para a tarefa de GER

Estudo	Modelo utilizado	Tipo de conhecimento para a variação	Córpus
Bohnet (2008)	AI e IS-FP (extensão do <i>Full Brevitty</i>)	Frequências dos atributos e das descrições do mesmo locutor.	TUNA
Fabbrizio, Stent e Bangalore (2008a)	AI e <i>Full Brevitty</i>	Frequência dos atributos e atributos recentes.	TUNA
Viethen e Dale (2010)	Árvore de Decisão C4.5 (QUINLAN, 2014)	Identificador do locutor.	GRE3D e GRE7D
Viethen, Mitchell e Krahmer (2013)	LongestFirst (extensão da abordagem de grafos)	Frequência dos atributos por grupo de locutores.	GRE3D3
FERREIRA e PARABONI (2017)	Algoritmo Incremental Estendido; Árvore de Decisão e SVM	Frequência dos atributos e identificador do locutor	TUNA; GRE3D; GRE7D; <i>Stars</i> e <i>Stars2</i>

Fonte: Danielle Sampaio Monteiro, 2018

Como observação da revisão destes estudos, em todas as situações foi concluído que o uso da variação humana gera expressões referenciais semelhantes às produzidas por humanos. Notou-se também que as abordagens baseadas em métodos de aprendizado de máquina superaram as outras abordagens. Com destaque, os estudos em Viethen e Dale (2010), Viethen (2011) e FERREIRA e PARABONI (2017) se mostraram os mais relacionados a presente proposta. Neles verificou-se qual o estilo geral de expressão a produzir usando variação humana; o uso da variação humana para decidir quais atributos selecionar; e finalmente a abordagem sendo incorporada a um algoritmo de GER completo para geração de expressões completas com base na variação humana.

Em comum, todas estas abordagens exigem um cópua de GER anotado com um grande conjunto de exemplos de cada locutor de interesse. Uma alternativa a métodos deste tipo será proposta no capítulo 6, no qual são anotados também os traços de personalidade de cada locutor.

4 Construção do corpus *b5-ref*

Em um projeto de pesquisa de maior escopo sobre o uso de modelos de personalidade humana na interpretação e geração de língua natural, foi construído um corpus paralelo de textos e inventários de personalidade, denominado *b5* (RAMOS et al., 2018). Um dos componentes deste corpus é a base *b5-ref* de expressões de referência. Estas expressões foram coletadas em colaboração com outros alunos deste grupo de pesquisa, e anotada pela autora desta proposta e seu orientador. Estas atividades serão descritas a seguir.

O corpus *b5-ref* é uma coleção de expressões de referência que oferece suporte a estudos para a tarefa computacional de GER com base em um conjunto de traços de personalidade. O objetivo deste corpus é modelar as decisões de seleção de conteúdo das expressões de referência produzidas por indivíduos com diferentes traços de personalidades (PARABONI; MONTEIRO; LAN, 2017).

A coleta de dados foi realizada por meio de um experimento presencial no qual o participante deveria identificar entidades por meio de estímulos visuais. A tarefa dos participantes do experimento consistia em descrever de forma única a entidade destacada com a borda vermelha, de tal maneira que fosse possível uma outra pessoa discriminar essa entidade em relação às outras da mesma cena somente com o auxílio desta descrição. Para cada imagem, os participantes foram solicitados a identificar unicamente um alvo específico dentro do contexto visual, completando uma frase no formulário ‘A pessoa destacada em vermelho é o/a ...’.

O domínio de estímulo visual do *b5-ref* é baseado nas imagens da base de dados Face Place (RIGHI; PEISSIG; TARR, 2012). Esta base consiste em fotografias de jovens anotadas com informações de sexo, etnia, emoções, dentre outras. Um exemplo de imagem de estímulo utilizada no domínio *b5-ref* é ilustrado na figura 7.

O domínio do experimento contou com 12 contextos visuais com 6 imagens cada um. As situações de referência consideradas envolviam contextos com diferentes graus de ambiguidade (isto é, exigindo descrições mais curtas ou mais longas) e propriedades físicas e afetivas com diferentes graus de saliência (por exemplo, uma cena contendo apenas uma pessoa sorridente ou uma única mulher). Além das 12 cenas de objetivo geral, foram intercaladas mais 12 cenas com *fillers* extraídos da base de dados Greebles (GAUTHIER;

Figura 7 – Exemplo de cena de estímulo visual para o *cópus b5-ref*.

Fonte: Adaptada de Ramos et al. (2017)

TARR, 1997)¹, que serviram para evitar que os participantes criassem um padrão de descrição.

O inventário de personalidade considerado foi o IGFP-5 (ANDRADE, 2008) , que foi validado para o português brasileiro e é baseado no modelo dos Cinco Grandes Fatores. O conjunto de inventários de personalidade dos participantes foi coletado por meio de um aplicativo de autoavaliação para a rede social Facebook. Para os participantes que não eram usuários da rede social, foi disponibilizada uma versão off-line que deveria ser respondida presencialmente junto com a segunda etapa do experimento.

Todos os dados do *cópus b5-ref* foram coletados com consentimento explícito dos participantes, antes da realização do experimento, e mediante aprovação do Comitê de Ética em Pesquisa em Seres Humanos - CEP da Escola de Artes, Ciências e Humanidades (EACH) da Universidade de São Paulo (USP). O experimento contou com a participação de 152 pessoas falantes do português brasileiro, sendo 86 do sexo feminino e 56 do sexo masculino, com faixa etária média era de 25,8 anos. Os participantes produziram 1822 descrições.

O processamento dessas expressões produzidas se deu com a correção ortográfica das mesmas e a anotação dos dados por meio de um esquema de 27 atributos, apresentados no apêndice A, que representaram semanticamente os dados extraídos das expressões de referência produzidas pelos participantes. Na tabela 1 são expostos alguns dos atributos mais frequentes nas descrições coletadas.

¹ O uso de imagens da base Greebles é uma cortesia de Michael J. Tarr, Center for the Neural Basis of Cognition and Department of Psychology, Carnegie Mellon University.

Tabela 1 – Atributos mais frequentes.

Atributo	Valores Possíveis	Ocorrências
gender	{male,female}	1707
race	{asian,black,caucasian}	794
smile	{yes,no}	784
isyoung	{yes}	705
hair.colour	{dark,blonde}	633
hair.length	{short,long }	434
emotion	{positive,negative,neutral}	266
eye.colour	{light,dark}	191
ponytail	{yes,no}	174
eyebrows	{other}	156

Fonte: Adaptado de Paraboni, Monteiro e Lan (2017)

A escolha destes atributos foi motivada pela frequência com que ocorreram nas descrições. Foram incluídas informações relacionadas às características físicas e emocionais, e excluídos do esquema, embora ainda permaneçam na cadeia de caracteres originais, informações infrequentes, como quantificadores (“a única”, “muito”, “pouco”, etc.), comparativos (“maior/menor que”, etc.), referências a outras entidades (“próximo ao moço asiático”, etc.) e informações que fugiam ao propósito de identificação (“a moça com aparência de pessoa enjoada”).

Além da exclusão destas informações, houve a combinação de alguns atributos afins para evitar uma grande quantidade de atributos esparsos. Por exemplo, todas as referências a pelos faciais, como bigode, barba e afins, foram representadas somente por um atributo *facial.hair*, que recebia os valores *yes* ou *no* indicando se havia alguma referência a pelos faciais ou não. E também houve o tratamento de informações que não poderiam ser mensuradas de forma absoluta, como formato de rosto ou tamanho de nariz e sobrancelha, por exemplo. Ou seja, as informações que dependiam e podiam mudar de acordo com o ponto de vista que cada pessoa tem ao identificar a característica, foram modeladas para receber o valor *other* em atributos relacionados quando mencionadas.

Com a estrutura dos atributos normalizada, primeiramente representou-se todas as imagens de cada um dos 12 contextos do domínio com os 27 atributos. Assim como TUNA e outros corpú de GER, a notação utilizada foi a linguagem XML. A atribuição dos valores às entidades deu-se considerando as informações fornecidas na própria base Face Place, como a etnia (asiática, caucasiana ou negra) e emoções (alegre, triste, etc.). Quando não havia especificação de valor para algum atributo na base, o valor atribuído foi o que estava presente na maioria das descrições para aquela entidade.

Finalmente, as descrições coletadas foram representadas em 152 arquivos XML, um para cada participante. Cada arquivo é estruturado com o identificador do participante, a expressão original produzida em formato de *string* (somente com os ajustes de ortografia), e a representação desta por meio dos atributos (e seus valores) escolhidos pelo participante.

Sobre a coleção de dados em si, observamos que os objetos alvo neste domínio são, em princípio, identificáveis por meio de descrições curtas (de uma a três propriedades cada). Na prática, no entanto, as descrições são ligeiramente mais longas, com uma média de quatro propriedades cada. Isto significa que a superespecificação referencial desempenha um papel significativo neste domínio.

O corpus *b5-ref* encontra-se completo e pronto para uso, já tendo inclusive sido empregado, em uma versão preliminar, em um estudo sobre superespecificação referencial baseada em personalidade, apresentado em Paraboni, Monteiro e Lan (2017).

5 Experimento piloto

O foco principal dessa pesquisa é analisar a influência da utilização de dados traços de personalidade de um locutor alvo de interesse na tarefa de GER de seleção de conteúdo para geração de expressões de referência que leva em consideração. Neste capítulo descrevemos um experimento simples que incorpora informação de personalidade ao algoritmo Incremental (DALE; REITER, 1995) e seus resultados.

5.1 O experimento

O objetivo do experimento foi a implementação de uma versão do algoritmo Incremental na qual a ordem das listas de preferência é determinada pela personalidade, o qual terá seus resultados comparados às descrições geradas por uma implementação da versão tradicional do algoritmo Incremental. Espera-se que o coeficiente Dice médio resultante da comparação entre as expressões originais disponibilizadas pelo *cópus b5-ref* e das geradas pela versão que leva em conta a personalidade, aqui chamada AIP (Algoritmo Incremental por Personalidade), seja significativamente maior, demonstrando que a versão proposta tem o desempenho mais próximo ao humano do que a versão tradicional do algoritmo Incremental.

O algoritmo proposto é uma implementação padrão do algoritmo Incremental, porém recebendo como entrada listas de preferências com base na informação de personalidade computadas com base na informação de personalidade disponibilizada no *cópus b5-ref*. Na versão tradicional do algoritmo Incremental, os atributos são ordenados em uma lista única de preferência de acordo com suas frequências em todo o conjunto de dados, enquanto que no AIP as listas de preferência de atributos são ordenadas pela frequência dos atributos no conjunto de dados por cada um dos dois graus em cada fator de personalidade, ou seja, o algoritmo é executado com 10 listas possíveis. Por exemplo, há uma lista que representa a frequência dos atributos selecionados pelo grupo de indivíduos com alto grau de Extroversão, outra para os indivíduos com baixo grau de Extroversão, e assim por diante.

Diante do fato de que as pontuações de personalidade são valores escalares, convencionou-se que um indivíduo possui grau alto ou baixo de personalidade quando sua

pontuação está situada no quartil, respectivamente, superior ou inferior do conjunto de dados. Indivíduos de personalidade intermediária foram desconsiderados neste experimento, totalizando 1810 descrições analisadas.

5.2 Resultados

Os algoritmos foram executados recebendo como entrada todo o conjunto de dados *b5-ref*, e no caso do algoritmo Incremental, também as cinco dimensões de personalidade do locutor-alvo que produziu aquela descrição, computando-se o coeficiente de Dice de cada uma em relação à descrição original. A tabela 2 sumariza o coeficiente de Dice médio obtido pelos dois algoritmos avaliados.

Tabela 2 – Resultado Dice médio do experimento piloto

	Baseline AIP	
Dice médio	0,55	0,51

Fonte: Danielle Monteiro, 2018

Como forma de ilustrar o comportamento do algoritmo proposto, a tabela 3 apresenta os resultados de Dice médios para cada uma das dez dimensões de personalidade consideradas por este algoritmo.

Tabela 3 – Resultado Dice médio e desvio padrão (DP) por polaridade de cada dimensão

Dimensão	Dice	DP
Extroversão-alta	0,55	0,26
Conscienciosidade-alta	00,55	0,26
Amabilidade-alta	0,55	0,26
Neuroticismo-alta	0,55	0,26
Abertura a experiências-alta	0,55	0,26
Extroversão-baixa	0,54	0,25
Conscienciosidade-baixa	0,27	0,23
Amabilidade-baixa	0,55	0,25
Neuroticismo-baixa	0,54	0,25
Abertura a experiências-baixa	0,51	0,24

Fonte: Danielle Monteiro, 2018

Neste caso, observa-se que em relação ao uso das listas de preferência com os fatores de personalidade com pontuação alta obtivemos os mesmos valores de coeficiente Dice que o obtido no algoritmo Incremental original, já com os resultados dos fatores de personalidade com pontuação baixa, os valores, mesmo que bem próximos, foram um pouco inferiores ao obtido no algoritmo Incremental original.

5.3 *Discussão*

Este capítulo apresentou um experimento para servir de comparativo do uso do algoritmo Incremental com o uso de listas de preferência por personalidade. Como resultado, observamos que o simples uso de informação de personalidade na ordenação de listas de preferências pode não ser suficiente para a geração de expressões de referência com base no algoritmo Incremental. Com base nestas observações, no próximo capítulo será discutida uma alternativa que combina aprendizado de máquina com dados de personalidade.

6 Modelo computacional proposto

Este capítulo apresenta uma proposta de modelo, chamado *PersonalityGER*, que utiliza classificadores e informações de personalidade para a geração de expressões de referências.

Para o *PersonalityGER*, considera-se um conjunto A com atributos que podem descrever entidades de um determinado domínio D , e o conjunto de classificadores binários $C = \{c^1, c^2, \dots, c^A\}$, os quais decidem se determinado atributo deve ou não compor a expressão de referência. Para a tarefa de gerar uma expressão de referência para um objeto alvo t , onde $t \in D$, c^i é treinado para prever se o atributo $a^i \in A$ deve compor a expressão de referência de t . Estes classificadores decidem selecionar ou não cada atributo com base em informações do contexto e levam em consideração uma personalidade-alvo fornecida como entrada, conforme será detalhado a seguir.

6.1 Classificadores

O modelo proposto faz uso de uma série de classificadores binários individuais treinados a partir de um corpus de expressões de referência rotuladas com os cinco fatores de personalidade do modelo CGF. No presente trabalho, o corpus considerado é o corpus *b5-ref* descrito no capítulo 4.

Cada classificador indica se o atributo correspondente deve ou não ser selecionado para compor a expressão de referência. Por exemplo, o classificador *race* indica se o atributo *race* do domínio *b5-ref* deve ou não ser selecionado.

Considerando que diversos atributos do domínio *b5-ref* são muito infrequentes para fins de aprendizagem, o modelo proposto implementa apenas nove classificadores que correspondem aos nove atributos mais frequentes do corpus, conforme a tabela 1 no capítulo 4, com exceção do atributo *gender*, o qual é utilizado em quase todas as expressões, e portanto o uso de *gender* não depende da personalidade do participante.

6.2 Características de aprendizagem

No modelo *PersonalityGER*, um vetor de características é a entrada fornecida a um classificador para decidir se um determinado atributo (a classe binária a ser aprendida) deve ser selecionado ou não. Para decidir se seleciona um atributo para compor uma expressão L , há dois tipos de características: características do domínio, que são referentes ao contexto ou cena na qual o objeto-alvo se encontra, e características de personalidade do autor da expressão.

As características de domínio são extraídas da representação do contexto no qual cada expressão pertence. Estas características incluem o poder discriminatório de cada atributo possível e as frequências de uso de cada atributo observadas em uma porção de dados de treinamento conforme será discutido no capítulo 7. Para cada contexto, o vetor de características possui uma lista de valores de poderes discriminatórios de cada atributo do objeto-alvo, por exemplo, $LD = \{isYoung: 0, race: 2, \dots\}$, isso significa que para o alvo de uma imagem I , o uso do atributo *isYoung* não eliminaria nenhum distraidor, o atributo *race* eliminaria dois objetos, etc. Outra lista é adicionada ao vetor de características, a lista de frequência dos atributos, por exemplo, $LF = \{isYoung: 170, race: 155, smile: 22, \dots\}$, isso significa que o atributo *isYoung* aparece 170 vezes nos dados de treinamento, etc.

Além das características relacionadas ao poder discriminatório e à frequência de uso de cada atributo no contexto, cada classificador inclui cinco características representando os valores escalares de cada uma das dimensões de personalidade CGF do locutor que produz a expressão de referência L .

6.3 Combinação das predições dos classificadores

Após os classificadores de cada atributo terem sido treinados, a decisão de quais atributos devem compor a expressão de referência tem como base as predições. Um algoritmo foi implementado para combinar estas predições e gerar a expressão de referência de acordo com o contexto em questão, como demonstrado no algoritmo 2.

O algoritmo recebe como entrada o objeto-alvo t a ser referido relacionado a um determinado contexto, uma lista P com todas as predições dos classificadores relacionada ao contexto referido e uma lista de atributos A que são os correspondentes aos classificadores

Algoritmo 2 Algoritmo Combinação das previsões

```

1: function GETDESCRICAO( $t, Pred, A$ )
2:    $L \leftarrow \{\}$ 
3:   for  $A_i \in A$  do
4:      $V \leftarrow valor(t, A_i)$ 
5:     if  $Pred[A_i] == 1$  then
6:        $L \leftarrow L \cup (A_i, V)$ 
7:   return  $L$ 

```

 Fonte: Danielle Sampaio Monteiro, 2018

implementados. A saída do algoritmo é o conjunto L que representa uma descrição do objeto-alvo.

Para início do algoritmo o conjunto L , que consistirá na descrição do objeto-alvo final, é inicializado com vazio. Após esta inicialização, o algoritmo itera de acordo com a lista de atributos dos classificadores seguindo os seguintes passos.

Na linha 4 é atribuído à V o valor da propriedade de t relativo ao atributo iterado e é verificado na linha 5, se o classificador relacionado ao atributo iterado prediz a adição ou não do mesmo. Se o classificador prediz que o atributo deve compor a expressão de referência, na linha 6 é realizada a inclusão da propriedade verificada em L . Na linha 7 é retornada a lista L como a expressão montada até todos os atributos da lista A serem iterados.

Para exemplificar o desempenho do algoritmo, considere o contexto da figura 8. Seja a cena do contexto t com os objetos $\{p1, p2, p3\}$ e P conjunto de atributos $A = \{ gender, race, smile, isyoung, hair.colour, hair.length, emotion, eyebrows, ponytail, eye.colour \}$,

Figura 8 – Contexto com três entidades $p1$, $p2$ e $p3$

Fonte: Imagem construída a partir de exemplos da base Face Place (RIGHI; PEISSIG; TARR, 2012)

- $p1 = \{ [gender, female], [race, caucasian], [smile, no], [isyoung, yes], [hair.colour, blonde], [hair.length, short], [emotion, neutral], [eyebrows, other], [ponytail, no], [eye.colour, dark] \}$

- $p2 = \{ [gender, male] , [race, caucasian], [smile, yes], [isyoung, yes], [hair.colour, blonde], [hair.length, short], [emotion, positive], [eyebrows, other], [ponytail, no], [eye.colour, light] \}$
- $p3 = \{ [gender, female] , [race, black], [smile, yes], [isyoung, yes], [hair.colour, black], [hair.length, short], [emotion, positive], [eyebrows, other], [ponytail, no], [eye.colour, dark] \}$

Neste exemplo, considera-se $p2$ o objeto-alvo e P um possível conjunto exemplo de predições feitas pelos classificadores dos atributos de A definido por:

$$P = \{ [gender, 1] , [race, 1], [smile, 1], [isyoung, 0], [hair.colour, 0], [hair.length, 0], [emotion, 0], [eyebrows, 0], [ponytail, 0], [eye.colour, 0] \}$$

O algoritmo de combinação de predições inicia com uma lista L vazia, e itera de acordo com a lista de atributos A na ordem estabelecida. Na primeira iteração, o valor *male* é atribuído, e o valor do atributo predito pelo classificador do atributo *gender* é verificado e a propriedade $[gender, male]$ é inserido em L . Na próxima iteração o algoritmo considera o próximo atributo em P , ou seja, *race*, incluindo a propriedade $[race, caucasian]$ em L , e assim por diante, até percorrer toda a lista P . Após o algoritmo iterar por todos os atributos de P , o algoritmo encerra retornando $L = \{[gender, male], [race, caucasian], [smile, yes]\}$, que poderia ser realizada como “O homem caucasiano que está sorrindo”.

7 Experimentos para avaliação do modelo *PersonalityGER*

Este capítulo descreve uma série de experimentos de avaliação do modelo *PersonalityGER* proposto no capítulo anterior. A avaliação faz uso do corpus *b5-ref* de expressões de referência rotuladas com informação de personalidade apresentado no capítulo 4, e objetiva comparar o modelo *PersonalityGER* a três modelos de *baseline* descritos a seguir.

Os experimentos implementados neste trabalho, incluindo os algoritmos que serviram de *baseline*, as variações propostas como modelos computacionais de GER, os leitores de corpus e as métricas de avaliação foram implementados na linguagem Python e os que utilizam técnicas de aprendizado de máquina utilizou-se a biblioteca *Scikit-learn* (PEDREGOSA et al., 2011), mais especificamente, os classificadores foram implementados utilizando Árvore de Decisão C4.5 (QUINLAN, 2014). Todas as implementações estão disponíveis no repositório *GitHub*¹. Além disso, o corpus *b5-ref* foi utilizado como conjunto de dados para os testes.

7.1 Modelos considerados

Nos experimentos feitos para explorar a influência da personalidade na tarefa de GER, além do modelo *PersonalityGER* descrito no capítulo 5, foram considerados três sistemas de *baseline*: uma implementação tradicional do algoritmo Incremental (DALE; REITER, 1995), que será utilizado para ilustrar o desempenho de abordagens tradicionais não baseadas em aprendizado de máquina; uma versão do modelo proposto, chamada *ContextGER*, cujo objetivo é ilustrar o desempenho do modelo proposto sem uso de personalidade; e uma versão que inclui um identificador de cada locutor do corpus ao invés de incluir as informações de personalidade do mesmo, denominada *ParticipantGER*, cujo objetivo é ilustrar o desempenho do modelo utilizando somente a identificação dos locutores.

O *ContextGer* utiliza os classificadores para selecionar os atributos da expressão resultante omitindo as cinco dimensões de personalidade de cada participante. O *ParticipantGER* objetiva representar uma aproximação ao modelo proposto em FERREIRA e PARABONI (2017) para prever a seleção de atributos individualizadas por locutor sem,

¹ <https://github.com/danisamon/ger-personality>

no entanto, utilizar conjuntos de exemplos de descrições de cada locutor individualmente como naquele caso.

7.2 Hipóteses

Para a avaliação do desempenho dos modelos propostos, dividimos a hipótese principal da dissertação em três hipóteses. Nos três casos, a medida de desempenho a ser considerada é o coeficiente Dice médio obtido pelo modelo em questão na tarefa de geração de um conjunto de expressões de teste. As hipóteses a serem verificadas são as seguintes:

- **H1:** o uso de uma abordagem de GER baseada em classificadores proporciona superiores ao algoritmo Incremental.

Esta hipótese será testada comparando-se os coeficientes Dice obtidos pelo modelo e pelo *baseline* Incremental. Espera-se que *ContextGER* apresente, em média, resultado superior.

- **H2:** o uso de informação de personalidade proporciona resultados superiores ao mesmo modelo sem acesso a esta informação.

Esta hipótese será testada comparando-se os coeficientes Dice obtidos pelo modelo *PersonalityGER* e pelo *baseline ContextGER*. Espera-se que *PersonalityGER* apresente, em média, resultado superior.

- **H3:** o uso de informação de personalidade proporciona resultados aproximados ao mesmo modelo com a informação individual de cada participante.

Esta hipótese será testada comparando-se os coeficientes Dice obtidos pelo modelo *PersonalityGER* e pelo *baseline ParticipantGER*. Espera-se que *PersonalityGER* apresente, em média, resultado aproximado.

7.3 Dados

Para fins de treinamento e teste dos modelos avaliados, neste trabalho utilizou-se o corpus *b5-ref*, o qual contou com 152 participantes, e totaliza uma amostragem de 1822

expressões de referência, as quais foram representadas por seus valores semânticos de acordo com a seleção de 27 atributos, conforme descrito no capítulo 4.

O *cópus* foi dividido em dois conjuntos: um para treinamento, com 1582 instâncias, e outro para teste, com 240 instâncias. A divisão se deu de forma balanceada atribuindo ao conjunto de teste 20 instâncias para cada contexto (cena), onde cerca de duas descrições de cada participante foram reservadas para o mesmo.

7.4 *Procedimento*

Para o *baseline* do algoritmo Incremental, a lista de preferência utilizada foi computada de acordo com a frequência com que os atributos foram escolhidos por todos os participantes. Para os modelos baseados em classificação, foi feito um balanceamento utilizando SMOTE, técnica para sobreamostragem para criar exemplos artificiais por interpolação (CHAWLA et al., 2002), utilizando cinco vizinhos como parâmetro, pois em testes preliminares constatou-se o forte desbalanceamento de classes, e com o balanceamento verificou-se resultados superiores na tarefa de GER.

Para testar as hipóteses utilizou-se o conjunto de teste com as 240 instâncias calculando-se o coeficiente Dice médio e Acurácia, e comparando cada expressão humana (obtida do *cópus*) com a mesma expressão gerada de forma automática pelos modelos computacionais implementados no experimento. Para fins de complementação da avaliação, testes estatísticos foram realizados, o teste de Wilcoxon foi utilizado sobre os coeficientes Dice, e o teste Chi-Quadrado foi realizado sobre a Acurácia.

8 Resultados

A avaliação do modelo proposto e dos modelos implementados para comparação foi realizada com base no *córpus b5-ref* obedecendo a mesma ordenação e divisão dos dados para todos os experimentos, conforme descrito na Seção 7.3. Neste capítulo são apresentados os resultados da avaliação dos modelos computacionais de GER implementados na presente pesquisa.

8.1 Resultados da tarefa de classificação

Para validação dos classificadores individuais usados pelos modelos *PersonalityGER*, *ContextGER* e *ParticipantGER*, a metodologia de validação cruzada foi utilizada sobre os dados de treinamento utilizando 10 *folds*. Os resultados das medidas de Precisão, Revocação e medida F dos classificadores estão na tabela 4.

Tabela 4 – Resultado dos classificadores na fase de treinamento

Atributo	<i>PersonalityGER</i>			<i>ContextGER</i>			<i>ParticipantGER</i>		
	Prec.	Revoc.	F	Pre.	Revoc.	F	Prec.	Revoc.	F
isYoung	0,50	0,66	0,58	0,56	0,50	0,52	0,34	0,60	0,48
race	0,63	0,62	0,65	0,78	0,81	0,79	0,40	0,58	0,49
emotion	0,70	0,83	0,73	0,68	0,57	0,62	0,55	0,82	0,64
smile	0,58	0,65	0,60	0,70	0,65	0,69	0,38	0,64	0,50
eyebrows	0,74	0,89	0,79	0,62	0,68	0,59	0,61	0,85	0,69
hair.colour	0,65	0,72	0,66	0,70	0,59	0,65	0,47	0,69	0,57
hair.length	0,66	0,75	0,67	0,69	0,79	0,72	0,52	0,78	0,59
ponytail	0,89	0,91	0,89	0,86	0,98	0,87	0,81	0,89	0,83
eye.colour	0,78	0,89	0,81	0,68	0,76	0,72	0,60	0,82	0,67
Média	0,68	0,77	0,71	0,70	0,70	0,69	0,52	0,74	0,61

Fonte: Danielle Sampaio Monteiro, 2018

Com base nas métricas, principalmente a da medida F, visualiza-se que o modelo computacional de GER que incorpora as informações de personalidade obteve resultados superiores quase sempre, o que sugere em princípio que o modelo proposto neste trabalho obtém desempenho superior.

8.2 Resultados na tarefa de GER

Esta seção apresenta os resultados da tarefa de geração das expressões do conjunto de teste pelo modelo *PersonalityGER* e pelos modelos de *baseline*: algoritmo Incremental, *ContextGer* e *ParticipantGER*.

Na tabela 5 são apresentados os coeficientes de Dice médio para cada uma das 12 cenas do domínio do corpus *b5-ref* gerados por cada um dos modelos avaliados para as três hipóteses citadas na seção 7.2.

Tabela 5 – Coeficiente de Dice médio e Desvio Padrão (DP) dos modelos computacionais de GER.

Cena	<i>PersonalityGER</i>		<i>ContextGer</i>		<i>ParticipantGER</i>		<i>Incremental</i>	
	Dice	DP	Dice	DP	Dice	DP	Dice	DP
1	0,77	0,17	0,65	0,10	0,78	0,16	0,62	0,33
2	0,70	0,23	0,51	0,18	0,56	0,29	0,56	0,29
3	0,66	0,24	0,56	0,17	0,62	0,17	0,53	0,16
4	0,65	0,22	0,56	0,18	0,58	0,18	0,56	0,13
5	0,66	0,16	0,53	0,14	0,54	0,22	0,40	0,12
6	0,67	0,21	0,62	0,20	0,63	0,23	0,34	0,26
7	0,67	0,23	0,55	0,16	0,62	0,15	0,39	0,15
8	0,73	0,22	0,60	0,16	0,70	0,25	0,79	0,21
9	0,74	0,17	0,49	0,19	0,67	0,21	0,71	0,27
10	0,65	0,23	0,54	0,17	0,61	0,24	0,57	0,32
11	0,63	0,19	0,65	0,14	0,62	0,17	0,71	0,30
12	0,68	0,18	0,57	0,16	0,59	0,23	0,43	0,19
Média	0,69	0,20	0,57	0,17	0,63	0,21	0,55	0,27

Fonte: Danielle Sampaio Monteiro, 2018

Tabela 6 – Resultados das métricas Precisão, Revocação e Medida F dos classificadores dos modelos implementados

Atributo	<i>PersonalityGER</i>			<i>ContextGER</i>			<i>ParticipantGER</i>		
	Prec.	Revoc.	F	Pre.	Revoc.	F	Prec.	Revoc.	F
isYoung	0,83	0,77	0,76	0,53	0,36	0,42	0,82	0,78	0,73
race	0,85	0,80	0,83	0,81	0,84	0,81	0,88	0,86	0,83
emotion	0,76	0,43	0,32	0,74	0,46	0,34	0,78	0,23	0,23
smile	0,67	0,67	0,70	0,74	0,67	0,71	0,69	0,68	0,62
eyebrows	0,90	0,40	0,33	0,57	0,53	0,12	0,79	0,27	0,11
hair.colour	0,72	0,54	0,59	0,78	0,67	0,67	0,72	0,48	0,68
hair.length	0,73	0,46	0,50	0,59	0,70	0,46	0,68	0,45	0,44
ponytail	0,87	0,50	0,56	0,76	1	0,43	0,85	0,27	0,25
eye.colour	0,85	0,57	0,50	0,72	0,85	0,26	0,69	0,48	0,27
Média	0,80	0,57	0,57	0,69	0,68	0,47	0,77	0,50	0,46

Fonte: Danielle Sampaio Monteiro, 2018

Os resultados na tabela 5 mostram que os modelos que utilizaram aprendizado de máquina superaram os resultados do algoritmo Incremental. Visualiza-se também que o modelo que incorpora informações relacionadas à personalidade geralmente superou aqueles que não levam.

Na tabela 6 é demonstrado o desempenho dos modelos implementados que utilizaram aprendizado de máquina. Nota-se de que o modelo que agregou a informação de personalidade obteve resultados geralmente superiores em comparação aos outros modelos.

8.3 Discussão

A partir dos resultados apresentados na seção anterior, as hipóteses H1, H2 e H3 propostas no capítulo 7 foram verificadas por meio do teste de Wilcoxon sobre o coeficiente dos pares de modelos avaliados.

Com relação à hipótese H1 (algoritmo Incremental x *ContextGER*), a tabela 5 sugere que o modelo *ContextGER* obteve valores superiores do coeficiente Dice médio comparados aos obtidos pelo algoritmo Incremental. A comparação entre *ContextGER* e algoritmo Incremental não possui diferença significativa entre seus coeficientes de Dice: $W=1854/p=0,3576$, entretanto, o modelo que utiliza aprendizado de máquina obteve grau de acurácia significativamente superior ao algoritmo Incremental, conforme teste Chi-Quadrado: $\chi^2 = 31,2916/p<0,00001$, constatando-se que o modelo *ContextGER* gerou um maior número de expressões de referência exatamente iguais a descrições produzidas pelos participantes do corpú do que o algoritmo Incremental.

Com relação à hipótese H2 (*ContextGER* x *PersonalityGER*), a tabela 5 sugere que o modelo *PersonalityGER* obteve valores superiores do coeficiente Dice médio comparados aos obtidos pelo modelo *ContextGER*. A diferença entre seus coeficientes de Dice é significativa: $W=13618/p<0,0001$ e teste Chi-Quadrado: $\chi^2 = 25,7204/p<0,00001$, comprovando a hipótese que o modelo *PersonalityGER* possui um desempenho superior em relação ao modelo que não utiliza informações sobre a personalidade dos locutores.

Com relação à hipótese H3 (*PersonalityGER* x *ParticipantGER*), a tabela 5 sugere que o modelo *PersonalityGER* obteve valores superiores do coeficiente Dice médio comparados aos obtidos pelo modelo *ParticipantGER*. A diferença entre seus coeficientes de Dice é significativa: ($W=6104/p<0,0001$) e teste Chi-Quadrado: $\chi^2 = 4,0897/p=0,043145$,

constatando-se que ambos os modelos erram/acertam praticamente o mesmo número de expressões idênticas as produzidas pelos locutores. Em vista desse resultado, constata-se que não somente o modelo *PersonalityGER* apresenta resultados aproximados aos apresentados pelo modelo que leva em consideração a identificação do locutor, como também superou esse modelo.

Estes resultados corroboram com os identificados no trabalho Paraboni, Monteiro e Lan (2017), embora a o presente trabalho foque especificamente na tarefa de seleção de atributos, e não na questão de superespecificação referencial, além de que o conjunto de teste empregado nestes experimentos é diferente do que foi utilizado no artigo e, portanto, a comparação não é totalmente direta muito embora os resultados tenham sido muito próximos.

Além dessa avaliação, foi realizada uma análise por meio de comparação entre as expressões geradas por cada um dos pares de modelos avaliados para as hipóteses (algoritmo Incremental - *ContextGER*, *ContextGER* - *PersonalityGER* e *ParticipantGER* - *PersonalityGER*) para verificar em quais casos os modelos com desempenho superior perderam e identificar as possíveis causas. A partir dessa análise não foram identificados padrões e diferenças significativas.

8.4 Considerações

A partir dos resultados apresentados neste capítulo, a hipótese H1 e H2 foram verificadas e confirmadas estatisticamente conforme os testes de Wilcoxon e Chi-quadrado. Já a hipótese H3 não foi confirmada, porém o resultado obtido foi superior à expectativa, pois *PersonalityGER* obteve resultados superiores ao *ParticipantGER*.

Estes resultados comprovaram a hipótese geral do projeto de que modelos computacionais de GER que utilizam informações sobre personalidade podem gerar expressões de referência semelhantes às produzidas por outros locutores com personalidade semelhante, e superam os modelos tradicionais de GER que não utilizam a informação de personalidade.

8.5 Teste complementar com atributos afetivos

No trabalho Paraboni, Monteiro e Lan (2017), a análise da influência dos dados de personalidade teve um dos focos voltados para a superespecificação, isto é, nos efeitos da personalidade sobre os tipos de informação que um locutor pode adicionar além do que é estritamente necessário para a desambiguação. Os resultados da superespecificação referencial sugeriram que atributos de teor mais afetivo como *smile* e *emotion* poderiam ser mais suscetíveis a diferenças de personalidade; assim, foi realizado um teste com base em um esquema de anotação alternativo do *cópus b5-ref*.

Para analisar se os atributos afetivos podem fornecer correlações úteis com os traços de personalidade, foi necessário considerar a polaridade também. Isso significa que a seleção de conteúdo precisou considerar o uso de propriedades, ou seja, pares atributo-valor como *emoção: positiva* ou *emoção: negativa* e não apenas os atributos.

Neste teste, os atributos *smile* e *emotion* foram desmembrados em *smile.yes*, *smile.no*, *emotion.negative*, *emotion.positive* e *emotion.neutral* e os modelos propostos foram avaliados. Os resultados deste experimento, entretanto, não apresentaram resultados superiores aos dos modelos aqui discutidos, e por isso foram desconsiderados. Este comportamento é explicado por uma característica inerente do domínio *b5-ref*, no qual um objeto-alvo nunca possui mais de um valor para o mesmo atributo. Assim, por exemplo, um indivíduo rotulado como *emotion-neutral* jamais pode assumir a propriedade *emotion-negative* e, portanto, o desmembramento por valor não tem efeito prático perceptível.

9 Conclusão

Esta monografia apresentou um estudo em nível de mestrado com foco no problema computacional de seleção de conteúdo de GER baseado em traços de personalidade, considerando a lacuna existente na área, que consiste na falta de equilíbrio entre diminuir o custo computacional de personalizar a tarefa para cada indivíduo, assemelhando-se assim às expressões geradas por sistemas e a linguagem humana, e a generalização dessa produção.

Os resultados comprovam a hipótese inicial de que algoritmos de GER que levam em consideração os traços de personalidades podem gerar expressões de referência mais semelhantes às produzidas por seres humanos do que algoritmos que não levam esta questão em consideração.

Ao propor um modelo inédito de GER baseado em personalidade, este estudo avança o conhecimento sobre o problema computacional de GER e sobre o uso de informação de personalidade em sistemas de GLN e, de modo mais específico, contribui para o avanço da pesquisa em GLN desenvolvida para o português brasileiro.

A pesquisa teve como principais contribuições o corpus *b5-ref*, integrante do corpus *b5* (RAMOS et al., 2018); o modelo *PersonalityGER*; e a validação da principal hipótese por meio da comparação de modelos *baseline* com o corpus *b5-ref*. Além da publicação em Ramos et al. (2018), um estudo prévio da implementação do modelo proposto foi publicado em Paraboni, Monteiro e Lan (2017).

Vale salientar que esse trabalho trata apenas da etapa de seleção de conteúdo, não abordando a tarefa de realização textual de um sistema GER. Além disso, as expressões geradas foram avaliadas somente na forma intrínseca utilizando a base de dados do corpus de referência *b5-ref*, a qual está em português e utiliza o modelo de personalidade CGF.

Como trabalhos futuros, consideramos identificar qual o papel de cada dimensão de personalidade analisando se o modelo é superior para alguns tipos de personalidade e se sim, identificar quais. Adicionalmente, planeja-se analisar a criação de um modelo padrão do algoritmo proposto para contextos menos restritos.

Referências¹

- AIRES, J. P. B. M. *Automatic Generation of Sports News*. Dissertação (Mestrado) — FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO. Citado na página 16.
- ALLPORT, F. H.; ALLPORT, G. W. Personality traits: Their classification and measurement. *The Journal of Abnormal Psychology and Social Psychology*, American Psychological Association, v. 16, n. 1, p. 6, 1921. Citado na página 29.
- ALLPORT, G. W.; ODBERT, H. S. Trait-names: A psycho-lexical study. *Psychological monographs*, Psychological Review Company, v. 47, n. 1, p. i, 1936. Citado na página 29.
- ANDERSON, A. H. et al. The hcr map task corpus. *Language and speech*, SAGE Publications Sage UK: London, England, v. 34, n. 4, p. 351–366, 1991. Citado na página 22.
- ANDRADE, J. M. de. *Evidências de validade do inventário dos cinco grandes fatores de personalidade para o Brasil*. Tese (Doutorado) — Universidade de Brasília, 2008. Citado 4 vezes nas páginas 28, 29, 31 e 45.
- BELZ, A.; GATT, A. The attribute selection for gre challenge: Overview and evaluation results. *Proceedings of UCNLG+ MT: Language Generation and Machine Translation*, Copenhagen, p. 75–83, 2007. Citado na página 38.
- BOHNET, B. The fingerprint of human referring expressions and their surface realization with graph transducers. In: *Fifth International Natural Language Generation Conference*. [S.l.]: Association for Computational Linguistics, 2008. p. 207–210. Citado 3 vezes nas páginas 17, 38 e 43.
- BREIMAN, L. et al. *Classification and regression trees*. [S.l.]: CRC press, 1984. Citado na página 41.
- CATTELL, R. The description of personality: basic traits resolved into clusters. *The Journal of Abnormal and Social Psychology*, American Psychological Association, v. 38, n. 4, p. 476, 1943. Citado na página 29.
- CHAWLA, N. V. et al. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, v. 16, p. 321–357, 2002. Citado na página 57.
- CORTES, C.; VAPNIK, V. Support-vector networks. *Machine learning*, Springer, v. 20, n. 3, p. 273–297, 1995. Citado na página 41.
- COSTA, P. T.; MACCRAE, R. R. *Revised NEO personality inventory (NEO PI-R) and NEO five-factor inventory (NEO-FFI): Professional manual*. [S.l.]: Psychological Assessment Resources, Incorporated, 1992. Citado na página 30.
- 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 30.

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

- DALE, R. Cooking up referring expressions. In: *27th annual meeting on Association for Computational Linguistics*. [S.l.]: Association for Computational Linguistics, 1989. p. 68–75. Citado 3 vezes nas páginas 27, 33 e 34.
- DALE, R. The return of the chatbots. *Natural Language Engineering*, Cambridge University Press, v. 22, n. 5, p. 811–817, 2016. Citado na página 16.
- DALE, R.; HADDOCK, N. Generating referring expressions involving relations. In: *fifth conference on European chapter of the Association for Computational Linguistics*. [S.l.]: Association for Computational Linguistics, 1991. p. 161–166. Citado na página 34.
- DALE, R.; REITER, E. Computational interpretations of the gricean maxims in the generation of referring expressions. *Cognitive science*, Wiley Online Library, v. 19, n. 2, p. 233–263, 1995. Citado 7 vezes nas páginas 22, 27, 34, 35, 36, 48 e 55.
- DEEMTER, K. v. et al. Generation of referring expressions: Assessing the incremental algorithm. *Cognitive science*, Wiley Online Library, v. 36, n. 5, p. 799–836, 2012. Citado na página 23.
- 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 28.
- ENGELHARDT, P. E.; BAILEY, K. G.; FERREIRA, F. Do speakers and listeners observe the gricean maxim of quantity? *Journal of Memory and Language*, Elsevier, v. 54, n. 4, p. 554–573, 2006. Citado 2 vezes nas páginas 22 e 40.
- EUGENIO, B. D. et al. The agreement process: An empirical investigation of human–human computer-mediated collaborative dialogs. *International Journal of Human-Computer Studies*, Elsevier, v. 53, n. 6, p. 1017–1076, 2000. Citado na página 22.
- EUGENIO, B. di et al. PatientNarr: Towards generating patient-centric summaries of hospital stays. In: *8th International Natural Language Generation Conference (INLG-2014)*. Philadelphia, Pennsylvania, USA: Association for Computational Linguistics, 2014. p. 6–10. Citado na página 16.
- FABBRIZIO, G. D.; STENT, A. J.; BANGALORE, S. Referring expression generation using speaker-based attribute selection and trainable realization (attr). In: *Fifth International Natural Language Generation Conference*. [S.l.]: Association for Computational Linguistics, 2008. p. 211–214. Citado na página 43.
- FABBRIZIO, G. D.; STENT, A. J.; BANGALORE, S. Trainable speaker-based referring expression generation. In: *Twelfth Conference on Computational Natural Language Learning*. [S.l.]: Association for Computational Linguistics, 2008. p. 151–158. Citado 2 vezes nas páginas 39 e 41.
- FERREIRA, T. C.; PARABONI, I. Referring expression generation: taking speakers’ preferences into account. In: SPRINGER. *International Conference on Text, Speech, and Dialogue*. [S.l.], 2014. p. 539–546. Citado na página 41.
- FERREIRA, T. C.; PARABONI, I. Generating natural language descriptions using speaker-dependent information. *Natural Language Engineering*, Cambridge University press, p. 1–22, 2017. Citado na página 35.

- FERREIRA, T. C.; PARABONI, I. Generating natural language descriptions using speaker-dependent information. *Natural Language Engineering*, Cambridge University Press, p. 1–22, 2017. Citado 3 vezes nas páginas 41, 43 e 55.
- GATT, A.; KRAHMER, E. Survey of the state of the art in natural language generation: Core tasks, applications and evaluation. *arXiv preprint arXiv:1703.09902*, 2017. Citado na página 16.
- GATT, A.; SLUIS, I. V. D.; DEEMTER, K. V. Evaluating algorithms for the generation of referring expressions using a balanced corpus. In: *Eleventh European Workshop on Natural Language Generation*. [S.l.]: Association for Computational Linguistics, 2007. p. 49–56. Citado 2 vezes nas páginas 22 e 23.
- GAUTHIER, I.; TARR, M. J. Becoming a greeble expert: Exploring mechanisms for face recognition. *Vision Research*, v. 37, n. 12, p. 1673–1682, 1997. Citado na página 45.
- 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 na página 17.
- GRICE, H. P. Logic and conversation. 1975, p. 41–58, 1975. Citado 2 vezes nas páginas 21 e 33.
- HUTZ, C. S. et al. O desenvolvimento de marcadores para a avaliação da personalidade no modelo dos cinco grandes fatores. *Psicologia: reflexão e crítica*. Porto Alegre. Vol. 11, n. 2 (1998), p. 395–409, 1998. Citado 2 vezes nas páginas 29 e 30.
- IACOVELLI, D.; GALINDO, M. R.; PARABONI, I. Lausanne: A framework for collaborative online nlp experiments. In: SPRINGER. *International Conference on Computational Processing of the Portuguese Language*. [S.l.], 2014. p. 280–285. Citado na página 27.
- JAIN, A. K. Data clustering: 50 years beyond k-means. *Pattern recognition letters*, Elsevier, v. 31, n. 8, p. 651–666, 2010. Citado na página 40.
- JAQUES, P. A.; NUNES, M. A. S. Ambientes inteligentes de aprendizagem que inferem, expressam e possuem emoções e personalidade. *Jornada de Atualização em Informática na Educação*, v. 1, n. 1, p. 30–81, 2013. Citado na página 29.
- JOHN, O. P.; DONAHUE, E. M.; KENTLE, R. L. *The big five inventory—versions 4a and 54*. [S.l.]: Berkeley, CA: University of California, Berkeley, Institute of Personality and Social Research, 1991. Citado na página 31.
- JOHN, O. P.; SRIVASTAVA, S. The big five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research*, Guilford, v. 2, n. 1999, p. 102–138, 1999. Citado 2 vezes nas páginas 29 e 30.
- JORDAN, P. et al. TBI-Doc: Generating patient & clinician reports from brain imaging data. In: *8th International Natural Language Generation Conference (INLG)-2014*. Philadelphia, Pennsylvania, USA: Association for Computational Linguistics, 2014. p. 143–146. Citado na página 16.
- KNUTH, D. E. *The art of computer programming: sorting and searching*. [S.l.]: Pearson Education, 1998. v. 3. Citado na página 38.

KRAHMER, E.; DEEMTER, K. V. Computational generation of referring expressions: A survey. *Computational Linguistics*, MIT Press, v. 38, n. 1, p. 173–218, 2012. Citado 4 vezes nas páginas 20, 22, 28 e 35.

KRAHMER, E.; ERK, S. V.; VERLEG, A. Graph-based generation of referring expressions. *Computational Linguistics*, MIT Press, v. 29, n. 1, p. 53–72, 2003. Citado 2 vezes nas páginas 37 e 40.

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 na página 31.

MAIRESSE, F.; WALKER, M. A. A personality-based framework for utterance generation in dialogue applications. In: *AAAI Spring Symposium: Emotion, Personality, and Social Behavior*. [S.l.: s.n.], 2008. p. 80–87. Citado na página 32.

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 na página 32.

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 na página 32.

MCCRAE, R. R.; COSTA, P. T. The structure of interpersonal traits: Wiggins’s circumplex and the five-factor model. *Journal of personality and social psychology*, American Psychological Association, v. 56, n. 4, p. 586, 1989. Citado na página 30.

OLIVEIRA, H. G. Automatic generation of poetry inspired by twitter trends. In: SPRINGER. *International Joint Conference on Knowledge Discovery, Knowledge Engineering, and Knowledge Management*. [S.l.], 2015. p. 13–27. Citado na página 16.

PARABONI, I.; GALINDO, M. R.; IACOVELLI, D. Stars2: a corpus of object descriptions in a visual domain. *Language Resources and Evaluation*, Springer, p. 1–24, 2016. Citado 3 vezes nas páginas 22, 26 e 27.

PARABONI, I.; MONTEIRO, D. S.; LAN, A. G. J. Personality-dependent referring expression generation. In: *Text, Speech and Dialogue (TSD-2017) Lecture Notes in Artificial Intelligence vol. 10415*. Prague, Czech Republic: Springer-Verlag, 2017. p. 20–28. Citado 6 vezes nas páginas 44, 46, 47, 61, 62 e 63.

PARABONI, I. et al. Generating underspecified descriptions of landmark objects. In: SPRINGER. *International Conference on Text, Speech, and Dialogue*. [S.l.], 2014. p. 76–83. Citado 3 vezes nas páginas 22, 25 e 26.

PEDREGOSA, F. et al. Scikit-learn: Machine learning in python. *Journal of machine learning research*, v. 12, n. Oct, p. 2825–2830, 2011. Citado na página 55.

PERVIN, L. A.; JOHN, O. P. *Personalidade: teoria e pesquisa*. [S.l.]: Artmed Editora, 2009. Citado na página 28.

QUINLAN, J. R. *C4. 5: programs for machine learning*. [S.l.]: Elsevier, 2014. Citado 3 vezes nas páginas 39, 43 e 55.

- RAMOS, R. et al. Building a corpus for personality-dependent natural language understanding and generation. In: *11th International Conference on Language Resources and Evaluation (LREC-2018)(to appear)*, Miyasaki, Japan, ELRA. [S.l.: s.n.], 2018. Citado 2 vezes nas páginas 44 e 63.
- RAMOS, R. M. S. et al. *O corpus b5 de textos e inventários de personalidade (v. 1.0.)*. São Paulo, 2017. Citado 3 vezes nas páginas 31, 45 e 71.
- REITER, E. Natural language generation. *The handbook of computational linguistics and natural language processing*, Wiley Online Library, p. 574–598, 2010. Citado na página 16.
- REITER, E.; DALE, R.; FENG, Z. *Building natural language generation systems*. [S.l.]: MIT Press, 2000. v. 33. Citado 3 vezes nas páginas 16, 20 e 32.
- REITER, E. et al. Choosing words in computer-generated weather forecasts. *Artificial Intelligence*, v. 167, p. 137–169, 2005. Citado na página 16.
- RIGHI, G.; PEISSIG, J. J.; TARR, M. J. Recognizing disguised faces. *Visual Cognition*, v. 20, n. 2, p. 143–169, 2012. Citado 3 vezes nas páginas 20, 44 e 53.
- TUPES, E. C.; CHRISTAL, R. E. *Recurrent personality factors based on trait ratings*. [S.l.], 1961. Citado na página 29.
- VIETHEN, H. A. E. *The generation of natural descriptions: corpus-based investigations of referring expressions in visual domains*. [S.l.]: Macquarie University, 2011. Citado 3 vezes nas páginas 40, 41 e 43.
- VIETHEN, J.; DALE, R. The use of spatial relations in referring expression generation. In: *Fifth International Natural Language Generation Conference*. [S.l.]: Association for Computational Linguistics, 2008. p. 59–67. Citado 3 vezes nas páginas 22, 23 e 24.
- VIETHEN, J.; DALE, R. Speaker-dependent variation in content selection for referring expression generation. In: MELBOURNE, VIC. *Proceedings of the 8th Australasian Language Technology Workshop*. [S.l.], 2010. p. 81–89. Citado 5 vezes nas páginas 17, 39, 40, 41 e 43.
- VIETHEN, J.; DALE, R. Gre3d7: A corpus of distinguishing descriptions for objects in visual scenes. In: *UCNLG+ Eval: Language generation and evaluation workshop*. [S.l.]: Association for Computational Linguistics, 2011. p. 12–22. Citado 3 vezes nas páginas 22, 24 e 25.
- VIETHEN, J.; DALE, R. et al. Referring expression generation: what can we learn from human data. In: *Proceedings of the Pre-Cogsci Workshop on Production of Referring Expressions: Bridging the Gap between Computational and Empirical Approaches to Reference*. [S.l.: s.n.], 2009. v. 29. Citado na página 21.
- VIETHEN, J.; MITCHELL, M.; KRAHMER, E. Graphs and spatial relations in the generation of referring expressions. In: *Proceedings of the 14th European Workshop on Natural Language Generation*. [S.l.: s.n.], 2013. p. 72–81. Citado 3 vezes nas páginas 40, 41 e 43.
- ZHANG, X.; LAPATA, M. Chinese poetry generation with recurrent neural networks. In: *2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, 2014. p. 670–680. Citado na página 16.

Apêndice A – Base de conhecimento para o corpus *b5-ref*

Atributos que representam semanticamente os dados extraídos das expressões de referência produzidas pelos participantes no experimento.

Quadro 2 – Quadro da base de conhecimento.

Atributo	Valor	Descrição
<i>mf</i>	{ <i>female, male</i> }	gênero explícito (homem, rapaz) ou implícito (japonesa, loira, o/a)
<i>isyoung</i>	{ <i>yes</i> }	jovem, rapaz, moça, garota
<i>race</i>	{ <i>asian, black, caucasian</i> }	<i>asian</i> : oriental, asiático, japonês, chinês; <i>black</i> : raça negra, afrodescendente (sem menção à pele) como em ‘mulher negra’; <i>caucasian</i> : branco, ocidental, homem branco. Convenções: ‘alemão’ é anotado como cor de cabelo, não como raça) mas sem menção à pele; ‘mestiço’ é anotado como <i>race-asian</i> e <i>skin-dark</i>
<i>skin</i>	{ <i>fair, dark</i> }	menção explícita à cor da pele: clara, escura, negra - sempre com menção à pele. Convenção: ‘mestiço’ é anotado como <i>race-asian</i> e <i>skin-dark</i>
<i>emotion</i>	{ <i>positive, negative, neutral</i> }	emoção positiva / negativa / neutra: semblante / expressão alegre, feliz, extrovertida (sem menção explícita ao sorriso). Ou triste, cansado, impaciente, rancoroso, com raiva, chateado, entediado, emburrado, fechado; ou normal, cara de paisagem, ‘pokerface’, calmo, tranquilo, sem fazer bico. Olhar distante.
<i>smile</i>	{ <i>yes, no</i> }	com ou sem sorriso (sério)
<i>mouth</i>	{ <i>shut, open</i> }	boca fechada, aberta; mostrando ou não mostrando os dentes
<i>facial.hair</i>	{ <i>yes</i> }	toda referência à barba, barbicha, cavanhaque, pelos ou penugem no queixo ou face (independentemente de cor)
<i>shape</i>	{ <i>other</i> }	referência ao formato do rosto (arredondado, fino) mas não bochecha. Obesidade só é marcada quando menciona face, caso contrário é <i>face-others</i> .
<i>spots</i>	{ <i>yes</i> }	toda referência a espinhas/marcas/sinais/sardas no rosto, testa ou pescoço
<i>lips</i>	{ <i>other</i> }	toda referência a lábios, exceto quando cerrado/aberto, que é marcado como <i>mouth-shut</i> ou <i>mouth-open</i> .
<i>nose</i>	{ <i>other</i> }	toda referência a nariz (tamanho, formato, aparência etc.)

<i>eyebrows</i>	{ <i>other</i> }	toda referência a sobrancelhas (grossa, fina, formato, curvatura, cor)
<i>face</i>	{ <i>other</i> }	qualquer outro aspecto da face (e.g., magro, obeso, com sobrepeso) ou pele (macia - mas não quando referente a cor), queixo ou bochechas (e.g., grandes), covinhas tamanho da boca (grande, pequena); pescoço.
<i>hair.colour</i>	{ <i>dark, blonde</i> }	<i>dark</i> : cabelo escuro / preto / castanho / castanho escuro; <i>blonde</i> : cabelo claro / loiro / castanho / castanho claro. Convenção: ‘moreno’ geralmente é anotado como <i>hair.colour-dark</i> , a menos que haja referência explícita à pele morena. No entanto, para as imagens 9 e 11 (mostrando pessoas negras), ‘moreno’ foi entendido como <i>race-black</i> .
<i>hair.length</i>	{ <i>short, long</i> }	cabelo curto / raspado / na altura do pescoço ou ombros / do queixo para cima, ou cabelo longo / comprido
<i>hair.style</i>	{ <i>straight, curly</i> }	liso ou crespo / ondulado / encaracolado
<i>unkempt</i>	{ <i>yes</i> }	cabelo desarrumado, despenteado, assanhado
<i>ponytail</i>	{ <i>yes, no</i> }	rabo-de-cavalo, cabelo preso
<i>fringe</i>	{ <i>yes</i> }	toda referência à franja, cabelos cobrindo a face, caindo sobre a testa, mecha de cabelo na testa, cabelos presos
<i>hair</i>	{ <i>other</i> }	qualquer outro aspecto especial do cabelo (grande, volumoso, espesso)
<i>narrow.eyed</i>	{ <i>yes, no</i> }	para orientais: olhos fechados, quase fechados, estreitos, profundos, ‘como os de um japonês’, olhos orientais / asiáticos. Para outras raças, ‘olhos fechados’ é marcado como <i>eyes-other</i> .
<i>eye.colour</i>	{ <i>light, dark</i> }	olhos claros / azuis / verdes ou escuros / pretos
<i>glasses</i>	{ <i>yes, no</i> }	com ou sem óculos
<i>eyes</i>	{ <i>other</i> }	qualquer outro aspecto especial dos olhos (e.g., pequenos, grandes, estrábicos, com olheiras, maquiados com lápis); olhar para a frente;
<i>ears</i>	{ <i>other</i> }	toda referência a orelhas (e.g., pequenas, grandes, visíveis ou não – mas não quando falando do comprimento do cabelo, anotado em <i>hair.length</i>)

<i>earring</i>	$\{yes, no\}$	com ou sem brincos, com ou sem furo na orelha (mas orelhas furadas são marcadas como <i>ears-other</i>)
----------------	---------------	--

Fonte: Ramos et al. (2017)