

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE LETRAS

PRISCILA BORBA BORGES

**A RELAÇÃO ENTRE OS CINCO GRANDES FATORES DA PERSONALIDADE, O
SEXO E ASPECTOS LINGUÍSTICOS EM NARRATIVAS PESSOAIS ESCRITAS**

PORTO ALEGRE

2017

PRISCILA BORBA BORGES

**A RELAÇÃO ENTRE OS CINCO GRANDES FATORES DA PERSONALIDADE, O
SEXO E ASPECTOS LINGUÍSTICOS EM NARRATIVAS PESSOAIS ESCRITAS**

Monografia apresentada como requisito parcial para a obtenção do título de Bacharel em Letras – Tradutor Português e Inglês – pelo curso de Letras da Universidade Federal do Rio Grande do Sul.

Orientadora: Prof.^a Dr.^a Ana Beatriz Arêas da Luz
Fontes

PORTO ALEGRE

2017

PRISCILA BORBA BORGES

**A RELAÇÃO ENTRE OS CINCO GRANDES FATORES DA PERSONALIDADE, O
SEXO E ASPECTOS LINGUÍSTICOS EM NARRATIVAS PESSOAIS ESCRITAS**

Monografia apresentada como requisito parcial para a obtenção do título de Bacharel em Letras – Tradutor Português e Inglês – pelo curso de Letras da Universidade Federal do Rio Grande do Sul.

Aprovada em Porto Alegre, _____ de _____ de 2018.

BANCA EXAMINADORA:

Prof.^a Dr.^a Ana Beatriz Arêas da Luz Fontes

Sérgio Duarte Júnior

Prof.^a Dr.^a Maity Simone Guerreiro Siqueira

To J. P.

AGRADECIMENTOS

Em primeiro lugar, agradeço à minha orientadora, Prof.^a Dr.^a Ana Beatriz A. L. Fontes por ter aceitado orientar um projeto tão “fora da curva” e por ter sido minha estrela-guia nesse universo da pesquisa empírica. Em segundo lugar, agradeço a todos os que se dispuseram a fazer parte do estudo, dedicando seu tempo e seu esforço mental para ajudar uma colega. Também agradeço aos meus amigos, que me ofereceram auxílio tanto pra fazer o trabalho quanto para esquecer dele quando era necessário. Ainda, agradeço às minhas colegas do Laboratório de Processamento de Linguagem Bilíngue pelas valiosas trocas de ideias. Por fim, agradeço muito à minha irmã e à minha mãe, que me ajudaram direta e indiretamente nesta jornada e em todas as outras.

RESUMO

A linguagem pode ser um veículo de informações factuais sobre o mundo, mas também pode revelar informações subjetivas sobre quem a utiliza. Neste estudo correlacional, exploratório e quantitativo buscou-se relacionar aspectos lingüísticos de narrativas pessoais escritas, o sexo e fatores e facetas da personalidade dos participantes, de acordo com o modelo dos Cinco Grandes Fatores. Para isso, 85 indivíduos brasileiros escreveram sobre seus planos futuros e responderam a Bateria Fatorial de Personalidade (BFP) (NUNES; HUTZ; NUNES, 2010). Os textos foram processados pelo programa de análise textual *Linguistic Inquiry and Word Count* (LIWC), cuja dimensão Processos Linguísticos (com 25 categorias) foi selecionada para este trabalho. Os resultados do LIWC e as pontuações de cada faceta e fator da BFP foram então correlacionados utilizando o teste de correlação de Pearson no programa de análise estatística SPSS. Além disso, as médias das pontuações da BFP e das categorias do LIWC foram comparadas entre os sexos. Os resultados demonstraram correlações significativas entre 10 categorias linguísticas e quatro fatores da personalidade (Neuroticismo, Extroversão, Realização e Abertura), com valores de r variando entre .215 e -.323. Além disso, foram obtidas correlações significativas entre 15 variáveis do LIWC e 10 facetas da BFP (Depressão, Passividade, Instabilidade Emocional, Dinamismo, Interações Sociais, Amabilidade, Competência, Empenho, Liberalismo e Busca por Novidades), com valor de r variando entre .216 e .304. Ainda, foram encontradas diferenças significativas e marginalmente significativas nas médias das categorias linguísticas, com as mulheres apresentando maior uso das categorias *presente* ($p = .049$), *verbo* ($p = .003$) e *dicionário* ($p = .086$), e o masculino com médias maiores em *preposição* ($p = .089$), *palavras maiores de seis letras* ($p = .041$), *você* ($p = .006$), *ele/ela* ($p = .013$) e *artigo* ($p = .008$). Em relação às diferenças de personalidade entre os sexos, elas foram significativas no nível das facetas, mas não no dos fatores. Assim, as médias foram significativa ou marginalmente maiores para o sexo feminino nas facetas Empenho ($p = .084$), Vulnerabilidade ($p = .093$), e Instabilidade Emocional ($p = .004$); já o sexo masculino obteve médias marginalmente superiores na faceta Abertura a idéias ($p = .067$). De modo geral, os achados vão ao encontro do reportado na literatura, porém há algumas discrepâncias que podem ser o reflexo do perfil dos participantes da amostra bem como do dicionário em português utilizado no programa de análise textual LIWC.

Palavras-chave: Cinco Grandes Fatores. Diferenças entre sexos. Análise textual. LIWC.

ABSTRACT

Language can be a vehicle for factual information about the world, but it can also reveal subjective information about the person who uses it. In this exploratory, quantitative and correlational study, we investigated the relationship between linguistic aspects in written self-narratives, sex and personality traits according to the Five-Factor Model. In order to do that, 85 Brazilian individuals wrote about their plans for the future and answered the Bateria Fatorial de Personalidade (BFP) (NUNES; HUTZ; NUNES, 2010). The texts were then processed by the text analysis software Linguistic Inquiry and Word Count (PENNEBAKER; FRANCIS; BOOTH, 2001), whose Linguistic Processes dimension (with 25 categories) was selected for analysis. The results from LIWC and the scores from the BFP were then correlated with a Pearson correlation test. Furthermore, the means of the BFP and the LIWC categories were compared between the sexes using an independent t-test. Results showed significant correlations between 10 linguistic categories and four personality factors (Neuroticism, Extraversion, Conscientiousness and Openness), with r values ranging from .215 and -.323. In addition, there were significant correlations between 15 linguistic categories and 10 facets from the BFP (Depression, Passivity, Emotional Instability, Dynamism, Social Interactions, Tenderness, Competence, Diligence, Liberalism and Novelty Seeking), with r values ranging from .216 and .304. Furthermore, there were significant and approaching significant differences between the sexes in the LIWC categories, with females showing higher use of present tense ($p = .049$), verbs ($p = .003$) and words recognized by the LIWC dictionary ($p = .084$), and males using more prepositions ($p = .089$), second person singular ($p = .006$), third person singular ($p = .013$), articles ($p = .008$) and words bigger than six letters ($p = .041$). As for the differences in personality between the sexes, they were significant on the facet level, but not on the factor level. Thus, females showed higher means in the facet Emotional Instability ($p = .004$), Vulnerability ($p = .093$) and Diligence ($p = .084$), whereas males exhibited higher means in Openness to ideas ($p = .067$). As a whole, the results support what had been reported in past studies, but there are some discrepancies, which could be due to the sample's profile as well as to the Brazilian Portuguese dictionary used by the text analysis program.

Keywords: Big Five factors. Sex differences. Text analysis. LIWC.

LISTA DE ILUSTRAÇÕES

Figura 1 – Nuvem de palavras mais utilizadas por mulheres	27
Figura 2 – Nuvem de palavras mais utilizadas por homens	28
Figura 3 – Pontuações dos fatores da BFP	50

LISTA DE TABELAS

Tabela 1 – Dimensões e categorias do LIWC 2007	37
Tabela 2 – Correlações entre fatores da BFP e categorias linguísticas do LIWC	46
Tabela 3 – Médias masculinas e femininas nos fatores da BFP	49
Tabela 4 – Médias masculinas e femininas nas facetas da BFP	48
Tabela 5 – Correlações entre facetas da BFP e categorias do LIWC	52
Tabela 6 – Médias masculinas e femininas nas categorias linguísticas do LIWC	53

SUMÁRIO

1	INTRODUÇÃO	11
2	REFERENCIAL TEÓRICO	14
2.1	Personalidade	15
2.1.1	Teoria dos traços e os Cinco Grandes Fatores	15
2.1.2	Os Cinco Grandes Fatores no Brasil: classificação e mensuração	19
2.1.3	Linguagem e personalidade	24
2.2	Sexo	27
2.2.1	Personalidade e sexo	27
2.2.2	Linguagem e sexo	29
2.3	Análise textual	31
2.3.1	Desenvolvimento de ferramentas de análise textual automatizadas	31
2.3.2	<i>Linguistic Inquiry and Word Count – LIWC</i>	32
2.3.3	<i>Brazilian LIWC 2007 Dictionary</i>	33
3	MÉTODO	34
3.1	Participantes	34
3.2	Instrumentos	34
3.2.1	Bateria Fatorial de Personalidade	34
3.2.2	Tarefa de produção textual	36
3.2.3	<i>Linguistic Inquiry and Word Count – LIWC</i>	36
3.3	Procedimentos	42
3.4	Análise de dados	44
4	RESULTADOS	46
5	DISCUSSÃO	53
6	CONCLUSÃO	62
	REFERÊNCIAS	64

1 INTRODUÇÃO

Este trabalho originou-se a partir do interesse em se identificar de que maneira diferenças individuais (personalidade) e coletivas (sexo) podem moldar o comportamento verbal das pessoas, especificamente no que se refere a narrativas pessoais escritas. A ideia de que estados psicológicos diversos são refletidos na linguagem não é nova; pelo menos desde o surgimento da psicanálise, as manifestações verbais já eram vistas como indicadoras de diversas movimentações psíquicas (PENNEBAKER; MEHL; NIEDERHOFFER, 2003). É o caso, por exemplo, da noção de parapraxias – lapsos da língua e troca ou esquecimento de nomes –, estipulada por Sigmund Freud (1901) em *Sobre a psicopatologia da vida cotidiana* (Rio de Janeiro: Imago, 2006). Depois disso, uma série de trabalhos tem observado como a linguagem pode refletir estados psicológicos e até mesmo traços da personalidade.

Em pesquisas destinadas a medir os traços da personalidade, o modelo mais comumente utilizado tem sido o dos Cinco Grandes Fatores (CGF), que estabelece Neuroticismo, Realização, Abertura, Extroversão e Socialização como os níveis mais amplos de caracterização dos indivíduos. Assim, trabalhos que relacionam a personalidade à linguagem têm observado, por exemplo, que quem possui pontuações muito altas no fator Neuroticismo tende a utilizar muito mais pronomes de primeira pessoa do singular, o que reflete um autocentrismo exacerbado e pode estar ligado a sintomas de depressão e ansiedade (TAUSCZIK; PENNEBAKER, 2010). Outros padrões têm sido observados nos demais fatores da personalidade, tais como menor uso de palavras por pessoas com altas pontuações em Realização e maior uso de primeira pessoa do singular por quem é alto em Socialização (HOLTGRAVES, 2014).

Além disso, a noção de que a linguagem pode variar de acordo com o sexo da pessoa tem sido postulada há várias décadas, tendo como uma de suas proponentes iniciais a professora Robin Lakoff, que em 1975 escreveu sobre diferenças no modo de falar de homens e mulheres sob um ponto de vista político-social. Depois dela, diversas pesquisas têm demonstrado, por exemplo, que homens utilizam a linguagem mais como veículo de informações, enquanto mulheres a veem como um fim em si mesma e a apreciam pelo seu papel no estabelecimento e manutenção de relações sociais (TAUSCZIK; PENNEBAKER, 2010).

Sabe-se, portanto, que é possível extrair informações que vão muito além do conteúdo explícito de uma produção verbal, mas saber qual é a melhor forma de analisá-la não é tarefa fácil. A maioria dos pesquisadores de narrativas, por exemplo, acredita que a linguagem é, por

definição, contextual (PENNEBAKER et al., 2003). Desse modo, as frases e mesmo textos inteiros só teriam significado dentro de um contexto, que, por sua vez, só poderia ser determinado por um ser humano. Por outro lado, existem aqueles que acreditam que as palavras podem ser contadas e analisadas estatisticamente, independentemente do contexto. Essa última visão tem-se tornado cada vez mais popular nas últimas décadas e está baseada na ideia de que as palavras podem conter informações psicológicas que ultrapassam seu significado literal (TAUSCZIK; PENNEBAKER, 2010).

Dentre as ferramentas emergentes desse último tipo de perspectiva, destaca-se o *Linguistic Inquiry and Word Count* (LIWC), desenvolvido por Pennebaker, Francis e Booth em 2001. O *software* categoriza palavras em mais de 60 categorias, que incluem dimensões linguísticas tradicionais (pronomes, artigos, preposições etc.), processos psicológicos (emoções negativas e positivas, processos cognitivos) e dimensões tradicionais de conteúdo (profissão, família, morte, sexo). A partir do desenvolvimento de programas de análise textual automatizada, tais como o LIWC, os estudos sobre a linguagem como marcador de diferenças individuais e coletivas têm crescido de modo considerável nas últimas décadas e têm permitindo que aspectos historicamente ignorados, tais como os vocábulos gramaticais (preposições, pronomes, artigos etc.), sejam também incluídos.

No entanto, essa linha de pesquisa tem recebido pouca atenção no Brasil devido a pelo menos dois fatores. Em primeiro lugar, antes de 2007 não havia instrumentos que medissem a personalidade de modo abrangente conforme o modelo dos Cinco Grandes Fatores, que é o mais utilizado nas pesquisas no exterior; por exemplo, o Inventário de Personalidade NEO PI Revisado (NEO-PI-R), originalmente desenvolvido por Costa e McCrae (1992), só foi disponibilizado em 2007 pela pesquisadora Carmen E. Flores-Mendoza (FLORES-MENDOZA, 2007). Não é de se surpreender, portanto, que, entre os anos de 2001 e 2010, apenas 23 artigos tenham sido publicados no Brasil com base nesse modelo teórico (SILVA; NAKANO, 2011). Em segundo lugar, o LIWC, um dos programas de análise textual automatizada mais utilizados nas pesquisas internacionais, só começou a processar textos em português a partir do desenvolvimento do *Brazilian Portuguese 2007 LIWC Dictionary* (BALAGE FILHO; ALUÍSIO; PARDO, 2013), tendo sido utilizado em pouquíssimas publicações até o momento. Desse modo, apesar de haver alguns esforços recentes na área da Ciência da Computação para extrair a personalidade ou o sexo a partir de textos utilizando ferramentas de análise lexical (PAIM; CAMATI; ENEMBRECK, 2016; MACHADO, 2016; PONTE JUNIOR; GUEDES; BEZERRA, 2016), esses trabalhos são escassos e não resolvem a questão teórica relacionada à correlação entre aspectos linguísticos e a personalidade/sexo.

Conseqüentemente, procurou-se preencher essa lacuna na literatura no estudo presente, que teve como objetivo fazer uma análise exploratória das relações entre personalidade, sexo e aspectos linguísticos em textos de indivíduos brasileiros.

Para alcançar esse objetivo, foram traçados alguns procedimentos metodológicos. Em primeiro lugar, o gênero textual de narrativas pessoais (*self-narratives*) foi escolhido, tendo sido pedido aos participantes para discorrerem sobre seus planos futuros. Essa decisão foi motivada pelos resultados de estudos anteriores que investigaram a relação entre a personalidade e a linguagem, que apresentaram efeitos estatísticos maiores com esse tipo de produção verbal. De acordo com Hirsh e Peterson (2009), isso se deve ao fato de o formato combinar a escrita livre com a escrita estruturada, proporcionando um padrão entre os participantes, enquanto que os fluxos de consciência, por exemplo, que são um dos tipos textuais mais comuns nas pesquisas do gênero, apresentam um formato livre demais, que não confere padronização. Além disso, as *self-narratives* também são reveladoras das características pessoais dos indivíduos na medida em que as histórias produzidas estão intimamente relacionadas a aspectos importantes da identidade e da personalidade de cada um (MCLEAN; PASUPATHI; PALS, 2007). Em segundo lugar, foi escolhido o programa de análise textual automatizada *Linguistic Inquiry and Word Count* por ser um dos mais completos de sua categoria e por ser acessível a pesquisadores fora da área das Ciências da Computação. Os textos foram então analisados de acordo com uma dimensão específica do LIWC, qual seja a de processos linguísticos, que inclui, entre outros, pronomes, artigos e preposições, por se tratar de uma dimensão menos explorada na literatura e que tem maior relevância para o curso de Letras. Em terceiro lugar, foi escolhida a Bateria Fatorial de Personalidade (NUNES; HUTZ; NUNES, 2010) como instrumento de medição da personalidade dos participantes, por ele ser amplamente utilizado nas pesquisas dessa área no Brasil e por ter sido produzido tendo em mente as idiossincrasias da cultura brasileira. Ainda, foi considerado o sexo dos participantes, ignorando a questão das identidades de gênero, o que também foi uma escolha metodológica de todos os estudos similares encontrados (ver, por exemplo, a metanálise de NEWMAN; GROOM; HANDELMAN; PENNEBAKER, 2008).

Desse modo, os dados provenientes do LIWC, da BFP e de uma ficha de informações dos participantes foram analisados no SPSS (SPSS Inc., 2009) por meio de testes de correlações bicaudais de Pearson e de Spearman, dependendo se a variável estava normalmente distribuída ou não, e por meio de testes-t de amostras independentes ou testes não paramétricos de amostras independentes, seguindo o mesmo critério. Esses testes foram

aplicados para encontrar possíveis correlações entre as variáveis linguísticas do LIWC, os fatores de personalidade da BFP e o sexo dos participantes, bem como para identificar diferenças nas médias dos grupos masculino e feminino no que diz respeito ao uso linguístico e perfis de personalidade.

Como se trata de um estudo exploratório, hipóteses específicas não foram formuladas. No entanto, seguindo a linha de estudos anteriores (HIRSH; PETERSON, 2009; PENNEBAKER; KING, 1999), esperou-se que diferenças individuais na personalidade se manifestariam na escolha de palavras das narrativas pessoais dos participantes. Similarmente, com base em achados como os de Weisberg, DeYoung e Hirsh (2011) e Mehl e Pennebaker (2003), esperou-se encontrar diferenças relacionadas ao sexo dos participantes, tanto em relação ao uso de palavras quanto à prevalência de alguns traços de personalidade. Por fim, esperou-se que, a partir dos achados desta investigação, fosse fornecido suporte empírico para práticas de mineração de personalidade e sexo a partir de textos escritos, possibilitando o desenvolvimento de produtos educativos e comerciais customizados, ferramentas para análise e ensino de texto e incentivo práticas terapêuticas baseadas na escrita, entre outros. Em última instância, objetivou-se lançar luz sobre aspectos da linguagem e da personalidade humanas, contribuindo para a incipiente área de pesquisa baseada em análise textual automática no Brasil.

Na primeira parte deste trabalho, apresenta-se o referencial teórico, que está dividido em três seções: a primeira discute a teoria dos traços de personalidade e a relação entre personalidade e linguagem; a segunda analisa as diferenças entre os sexos na linguagem e na personalidade; e a terceira é dedicada à análise textual computadorizada. Na segunda parte do trabalho, reporta-se o método utilizado, com caracterização dos participantes, dos instrumentos e dos procedimentos adotados. Por fim, relatam-se os resultados e discutem-se suas implicações. A seguir, então, serão discutidos os aspectos teóricos que fundamentam o trabalho.

2 REFERENCIAL TEÓRICO

Este estudo teve como base a teoria do traço de personalidade, com foco especial no modelo dos Cinco Grandes Fatores (CGF), que possui uma longa história de desenvolvimento, sendo difícil atribuir-lhe um único pesquisador-pai. Além disso, este trabalho buscou fundamentos na literatura relacionada a diferenças de personalidade e comportamento linguístico entre os sexos. Por fim, foi considerada a história do

desenvolvimento de ferramentas de análise textual, com foco no programa *Linguistic Inquiry and Word Count*.

2.1 Personalidade

O conceito de personalidade não é simples de se definir: dependendo de que linha teórica se trabalhe, a resposta pode variar consideravelmente. Algumas podem focar na estrutura da personalidade, descrevendo seus aspectos mais estáveis, já outras podem focar nos processos, nos elementos dinâmicos da personalidade; ainda, algumas podem focar em aspectos internos do indivíduo, enquanto outras podem dar mais ênfase ao ambiente externo. De Freud (1923) a Gray (GRAY; MCNAUGHTON, 2003), passando por Jung (1933), Rogers (1961), Mischel (MISCHEL; SHODA, 1995), Eysenck (1990), entre outros, cada pensador postulou sua própria concepção de personalidade e forneceu seus próprios parâmetros de como melhor desenvolvê-la. Hoje, já existem abordagens que fazem sentido dessa miríade teórica de modo integrativo. É o caso de McAdams e Pals (2006), que definem o conceito de personalidade como (a) a variação única de um indivíduo dentro do plano evolutivo geral da natureza humana, que se manifesta em padrões de (b) traços individuais, (c) características adaptativas e (d) narrativas de vida, as quais, por sua vez, estão inseridas (e) na cultura e no contexto social. Neste trabalho, o foco recairá apenas sobre o segundo componente dessa concepção de personalidade, qual seja o dos traços individuais devido a suas aplicações práticas e riqueza empírica.

2.1.1 Teoria dos traços e o desenvolvimento do modelo dos Cinco Grandes Fatores

Os traços têm sido vistos como dimensões abrangentes de diferenças entre indivíduos, que se manifestam de modo consistente e contínuo no comportamento, pensamento e sentimento das pessoas em diversas situações e em períodos diferentes da vida (MCADAMS; PALS, 2006). Uma questão fundamental após essa primeira definição seria como esses traços poderiam ser classificados ou de que modo se poderia formar uma taxonomia dos atributos da personalidade. De acordo com John e Srivastava (1995), um ponto inicial para tal classificação seria a linguagem natural. Esse ponto de vista tem como base a *hipótese léxica*, que foi inicialmente concebida por Sir Francis Galton em 1884. De acordo com John, Angleitner e Ostendorf (1988, p. 174), essa hipótese estabelece que “As diferenças mais salientes e socialmente relevantes nas vidas das pessoas serão codificadas em algum momento

na linguagem” e que, “quanto mais importante essa diferença, mais provável será que ela seja expressa em uma única palavra” (tradução nossa)¹. Logo, o vocabulário contido em um dicionário de língua natural poderia fornecer um conjunto necessário, ainda que não suficiente, de atributos que os falantes da língua consideram úteis para as suas interações cotidianas (GOLDBERG, 1981).

Apesar de o escritor e cientista inglês Francis Galton ter sido o primeiro a procurar no dicionário por termos relacionados à personalidade, gerando uma lista de 1000 descritores, os primeiros esforços sistemáticos para organizar a linguagem da personalidade tiveram início com Klages (1926) e Baumgarten (1933). Klages sugeriu que uma análise minuciosa da linguagem poderia ser valiosa para a psicologia da personalidade. Baumgarten, por sua vez, aceitou a sugestão e reuniu adjetivos e substantivos que considerou descritores da personalidade. Com um total de 1629 palavras, a lista de Baumgarten influenciou Allport e Odbert (1936), que foram cruciais para estabelecer a fundamentação empírica da pesquisa taxonômica dos traços (JOHN et al., 1988).

Allport e Odbert (1936) examinaram um dicionário completo da língua inglesa e extraíram quase 18.000 termos, entre adjetivos, substantivos e partículas, que julgaram ser úteis para distinguir o comportamento dos indivíduos. Para organizar essa imensa quantidade de termos, com a ajuda de três juízes independentes, eles criaram quatro categorias: traços, com um total de 4.500 termos, que foram definidos como “[...] tendências determinantes pessoais e generalizadas - modos estáveis e consistentes de adaptação de um indivíduo ao seu ambiente²” (ALLPORT; ODBERT, 1936, p. 26, tradução nossa); estados temporários; julgamentos de cunho avaliativo relacionados à conduta e a reputação da pessoa; e características físicas, habilidades e termos de pouca relevância para a personalidade. Apesar de fornecerem uma base inicial para a formação do léxico da personalidade, as classificações de Allport e Odbert não serviram como uma taxonomia que tivesse aplicações práticas, pois não apresentaram um parâmetro sistemático para a distinção, ordenação e nomeação das diferenças individuais de comportamento (JOHN; SRIVASTAVA, 1995).

¹ “Those individual differences that are most salient and socially relevant in people’s lives will eventually become encoded into their language; the more important such a difference, the more likely is it to become expressed as a single word.” (ANGLEITNER; OSTENDORF, 1988, p. 174).

² “They [traits of personality] designate generalized and personalized determining tendencies – consistent and stable modes of an individual’s adjustment to his environment.” (ALLPORT; ODBERT, 1936, p. 26).

Logo, em 1943, tendo em mente a criação dessa taxonomia, Cattell baseou-se na lista de 4.500 termos relacionados a traços estáveis de Allport e Odbert (1936) e os reduziu a apenas 35 variáveis a partir de processos de agrupamento semânticos e empíricos. Após uma série de análises fatoriais oblíquas, Cattell (1943) concluiu que havia identificado 12 fatores de personalidade. Mais tarde, no entanto, Fiske (1949), ao tentar replicar os achados de Cattell, encontrou uma solução de apenas cinco fatores utilizando o método de rotação ortogonal. O mesmo aconteceu quando Tupes e Christal (1961 apud JOHN; SRIVASTAVA, 1995) concluíram que cinco fatores eram suficientes para explicar os dados por eles observados. Norman (1963 apud JOHN; SRIVASTAVA, 1995), Borgatta (1964 apud JOHN; SRIVASTAVA, 1995) e Smith (1967 apud JOHN; SRIVASTAVA, 1995) também encontraram indícios para apenas cinco fatores.

Apesar das evidências levantadas por esses trabalhos, a importância desses cinco fatores ficou relativamente escondida durante os anos 1960 e 1970, quando o debate “pessoa-situação” exerceu forte influência na área. Como um dos protagonistas nesse debate, Mischel (1968 apud MCADAMS; PALS, 2006) propôs que o comportamento é mais contingente ou dependente da situação do que a teoria do traço sugere, e que a pontuação nas escalas dos traços não corresponde ao que os indivíduos fazem em situações específicas da vida real. Apesar dessas críticas, nos anos 1980 e 1990 o modelo ressurgiu mais forte do que nunca. Um dos pesquisadores que mais contribuiu para esse resgate dos CGF foi Goldberg (1981), que, tomando como base a lista de Norman (1967), encontrou cinco fatores utilizando diferentes métodos de extração fatorial.

Além desses esforços provenientes da tradição léxica, outros pesquisadores, que trabalhavam com inventários de personalidade, também encontraram cinco fatores correspondentes ao *Big Five*. Os mais proeminentes desses pesquisadores foram Costa e McCrae, que lançaram o primeiro inventário de personalidade baseado nos CGF, o NEO PI (MCCRAE; COSTA, 1985), e que mostraram que era possível extrair cinco grandes fatores a partir de outros inventários renomados, tais como o California Q-Set (MCCRAE; COSTA; BUSCH, 1986). Essa outra tradição, baseada em inventários de personalidade, contribuiu para o desenvolvimento dos CGF principalmente do ponto de vista teórico já que a abordagem lexical limitava-se à análise de traços da personalidade representados na língua comum, sem levar em conta os conceitos pensados por teóricos da personalidade. Portanto, a correspondência dos achados resultantes de procedimentos empíricos com aqueles resultantes de instrumentos criados a partir de pressupostos teóricos fortaleceu o modelo dos CGF, em

parte remediando o problema da falta de fundamentação teórica (COSTA; MCCRAE, 1992; NUNES, 2000).

Os trabalhos de Costa e McCrae (1985, 1992) geraram a nomenclatura mais comumente utilizada hoje em dia para se referir aos CGF, qual seja: *Extraversion*, *Agreeableness*, *Conscientiousness*, *Neuroticism*, e *Openness to Experience*. Essas são dimensões hierárquicas básicas, que não esgotam as possibilidades de descrição da personalidade, podendo ainda ser divididas em subfatores ou facetas. Todos os indivíduos são classificados dentro do espectro de cada fator, correspondendo a diferentes pontuações em cada dimensão, sem gerar um único tipo específico.

Atualmente, o modelo dos Cinco Grandes Fatores encontra-se relativamente consolidado (MCADAMS; PALS, 2006) devido a uma série de razões. Em primeiro lugar, existem evidências para uma quase universalidade dos traços: estudos realizados em locais com culturas e línguas diferentes das que geraram os CGF originalmente têm produzido soluções fatoriais similares (CHURCH, 2000; HUTZ et al., 1998), ainda que em algumas culturas tenham sido encontrados números diferentes de fatores (LEE; ASHTON, 2004; DE RAAD; PEABODY, 2005), como é o caso dos *Big Three* encontrados por De Raad e Peabody (2005) na Holanda e dos seis fatores encontrados por Cheung, Leung, Zhang, Sun, Gan, Song e Xie (2001) no Havai. Em segundo lugar, diversos estudos demonstraram que a pontuação dos traços de personalidade implica diferenças em comportamentos de forma estatisticamente significativa, especialmente quando são vistos a partir de várias situações (MOSKOWITZ, 1990 apud MCADAMS; PALS, 2006). Além disso, os traços estão associados a diversas medidas relacionadas à qualidade de vida dos indivíduos, tais como desempenho no trabalho (BARRICK; MOUNT, 1991 apud MCADAMS; PALS, 2006), qualidade de relações sociais (ASENDORF; WILPERS, 1998 apud MCADAMS; PALS, 2006), bem-estar psicológico (MAGNUS; DIENER; FUJITA; PAVOT, 1993) e longevidade (FRIEDMAN; TUCKER; TOMLINSON-KEASEY; SCHWARTZ; WINGARD; CRIQUI, 1993). Em terceiro lugar, estudos longitudinais evidenciaram uma estabilidade em longo prazo dos traços de personalidade (COSTA; MCCRAE, 1994; ROBERTS; DELVECCHIO, 2000). Esses achados, por sua vez, estão relacionados à quarta razão pela qual a teoria dos traços ressurgiu com tanta força, qual seja a carga hereditária, que corresponde a cerca de 50% dos traços de personalidade, de acordo com estudos feitos com gêmeos (BOUCHARD; LYKKEN; MCGUE; SEGAL; TELLEGEN, 1990). Em quinto lugar, pesquisas na área da neurociência começaram a exibir relações entre alguns traços e funções cerebrais. Por exemplo, no estudo de Adelstein, Shehzad, Mennes, DeYoung, Zuo, Kelly, Margulies, Bloomfield, Gray,

Castellanos e Milham (2011), ao se analisar a conectividade funcional em repouso do cérebro dos participantes, foi encontrado que os fatores Neuroticismo e Extroversão previam conectividade a áreas associadas à regulação emocional, à autoavaliação, à motivação e à recompensa (córtex pré-frontal dorsomedial e regiões paralímbicas laterais, respectivamente), o que corresponde às características desses traços no modelo *Big Five*. Além disso, DeYoung, Peterson e Higgins (2005) encontraram que o fator Abertura está relacionado à inteligência cristalizada e fluída, bem como ao funcionamento do córtex pré-frontal dorsolateral, uma área conhecida pelo seu envolvimento em processos cognitivos, como o da memória de trabalho. Ademais, foram encontradas correlações entre volume cerebral e quatro dos cinco fatores, estando Extroversão associado ao volume do córtex orbitofrontal medial; Neuroticismo, ao volume de regiões cerebrais associadas a afeto negativo, ameaça e punição; Socialização associado ao volume em regiões que processam informações sobre intenções e estados mentais de outros indivíduos; e Realização associado ao volume do córtex pré-frontal lateral, uma região envolvida no planejamento e controle voluntário do comportamento (DEYOUNG; HIRSH; SHANE; PAPADEMETRIS; RAJEEVAN; GRAY, 2010). Em geral, apesar de incipientes, as pesquisas demonstram resultados promissores para a área da neurociência dos traços.

Diante dessa variedade de evidências, o modelo dos Cinco Grandes Fatores hoje se apresenta como uma alternativa sólida para a avaliação da personalidade de diversas populações. Por exemplo, a população brasileira, apesar de estar imbuída em uma cultura diferente da americana em muitos aspectos, apresentou um padrão de personalidade bastante similar ao dos ianques. Na seção seguinte, serão dados mais detalhes sobre a trajetória dos CGF no Brasil.

2.1.2 Os Cinco Grandes Fatores no Brasil: classificação e mensuração

Em 1998, Hutz, Nunes, Silveira, Serra, Anton e Wiczorek buscaram desenvolver marcadores para a avaliação da personalidade no modelo dos Cinco Grandes Fatores para uso no Brasil, para que fossem funcionalmente equivalentes ao inventário de Goldberg (1992). No seu estudo, 976 estudantes de diversas universidades da região sul do Brasil se autodescreveram, respondendo a um instrumento previamente desenvolvido, contendo 96 adjetivos derivados de um estudo prévio sobre descritores de traços mais frequentes em português. Após realizarem análises fatoriais com diferentes métodos de extração, eles

encontraram que uma solução de cinco fatores era apropriada e que esses fatores correspondiam aos descritos na literatura (HUTZ et al., 1998).

A partir do estudo de Hutz e colegas (1998), bem como de informações do DSM-IV (American Psychiatric Association, 1994 apud NUNES et al., 2010) e do *International Personality Item Pool* (<http://ipip.ori.org> apud NUNES et al., 2010), foram desenvolvidos os primeiros itens de escalas que avaliavam individualmente cada um dos cinco grandes fatores, a saber: Escala Fatorial de Neuroticismo (HUTZ; NUNES, 2001), Escala Fatorial de Extroversão (NUNES; HUTZ, 2007a), Escala Fatorial de Socialização (NUNES; HUTZ, 2007b), Escala Fatorial de Realização (PACICO; HUTZ, [s.d.]) e Escala Fatorial de Abertura (VASCONCELLOS; HUTZ, 2008). Posteriormente, por meio de análises semânticas e psicométricas das escalas individuais, foram selecionados 167 itens (28 para Socialização – Alfa de Cronbach de 0,85; 24 para Neuroticismo – Alfa de Cronbach de 0,89; 29 para Extroversão – Alfa de Cronbach de 0,89; 44 para Abertura – Alfa de Cronbach de 0,88; e 42 para Realização – Alfa de Cronbach de 0,88) para uma versão preliminar do que ficou conhecido como a Bateria Fatorial de Personalidade (BFP). Em decorrência de novas informações provenientes de estudos posteriores, foram eliminados alguns itens, de modo que a versão final da BFP contou com 126 itens no total (Nunes, Hutz e Nunes, 2010).

Desse modo, os fatores foram denominados pela BFP (NUNES et al., 2010) como (I) *Extroversão*, (II) *Socialização*, (III) *Realização*, (IV) *Neuroticismo* e (V) *Abertura*. O fator I, Extroversão, refere-se à quantidade e à intensidade das interações interpessoais, bem como à necessidade de estimulação e à capacidade de sentir emoções positivas. Indivíduos com pontuações altas nesse fator são considerados sociáveis, ativos, falantes, otimistas e afetuosos; por sua vez, os que têm baixa pontuação em Extroversão costumam ser reservados, comedidos, independentes e quietos. (COSTA; WIDIGER, 1993 apud NUNES, 2000).

De acordo com a Bateria Fatorial de Personalidade (NUNES et al., 2010), as facetas que compõem o fator Extroversão são E1 – Comunicação, E2 – Altivez, E3 – Dinamismo e E4 – Interações Sociais. A faceta Comunicação diz respeito ao quão comunicativas e expansivas as pessoas acreditam que são. Altas pontuações indicam indivíduos com facilidade em falar em público e em conhecer pessoas novas, além de apontarem para uma tendência do indivíduo a falar mais de si mesmo e a externar suas opiniões publicamente. O oposto é verdade para aqueles que apresentam baixa pontuação nessa faceta: falam pouco de si e não gostam de se expressar em público. A faceta Altivez indica indivíduos com percepções grandiosas sobre seu valor e suas capacidades. Além disso, ela está relacionada a uma maior propensão a falar de si e a uma maior necessidade de atenção alheia. Quem possui baixa

pontuação nessa faceta costuma ser mais humilde, precisar menos de atenção e ter dificuldade em reconhecer suas qualidades. A terceira faceta do fator Extroversão é Dinamismo. Escores mais altos indicam um indivíduo que toma iniciativa, que coloca suas ideias em prática e que está envolvido em diversas atividades ao mesmo tempo. Ao contrário, quem possui baixa pontuação nessa faceta tende a desempenhar uma tarefa por vez, sem precisar realizar várias atividades ao mesmo tempo para se sentir bem, podendo ainda demorar mais para colocar suas ideias em prática. A última faceta, Interações Sociais, descreve indivíduos que buscam situações que exigem interação social, que gostam de atividades em grupo, que se esforçam para manter contato com amigos e conhecidos e que se envolvem facilmente com as pessoas. Já quem possui um escore baixo nessa faceta prefere estar sozinho ou em pequenos grupos, demoram mais para se envolver com pessoas e não têm necessidade de relações sociais intensas (NUNES et al., 2010).

Por sua vez, o fator Socialização também consiste em uma dimensão interpessoal, referindo-se ao quão compassivos ou antagônicos os indivíduos são em suas interações. Aqueles com alta pontuação em Socialização costumam ser francos, bondosos, altruístas, prestativos, ingênuos e afáveis. Por outro lado, os indivíduos baixos em Socialização tendem a ser cínicos, não cooperativos, manipuladores e intransigentes (COSTA; WIDIGER, 1993 apud NUNES, 2000).

Socialização subdivide-se em três facetas (NUNES et al., 2010): S1 – Amabilidade, S2 – Pró-sociabilidade e S3 – Confiança nas pessoas. A primeira delas está relacionada ao quão empática, compreensiva, educada e atenciosa a pessoa é; quem tem uma pontuação alta nesse fator tende a se preocupar com o bem-estar alheio e a demonstrar apreço pelos outros. Pessoas baixas em Amabilidade são mais autocentradas, insensíveis e até hostis. Já a faceta Pró-sociabilidade indica sujeitos que evitam situações arriscadas, que não gostam de transgredir as regras e que não induzem os outros a fazerem coisas que não queiram. Por outro lado, quem possui baixa pontuação nessa faceta tende a envolver a si e aos outros em situações de perigo, não tem consideração por regras e pode ser manipulador e hostil em suas interações interpessoais. A faceta Confiança nas pessoas tem a ver com o quanto o indivíduo confia nos outros e o quanto imagina que podem prejudicá-lo. Altas pontuações indicam pessoas que creem que os outros são honestos e bem-intencionados, enquanto que baixas pontuações descrevem comportamentos céticos, onde está implícito que os outros podem ser desonestos e perigosos (COSTA; MCCRAE, 1992 apud NUNES et al., 2010).

O terceiro fator, Realização, está relacionado ao nível de organização, controle, persistência e motivação na conquista de objetivos. Pessoas com pontuações altas nesse fator

são organizadas, obstinadas, trabalhadoras, decididas, confiáveis, pontuais, escrupulosas e ambiciosas; ao contrário, aqueles que têm baixas pontuações em Realização não possuem objetivos claros, são menos confiáveis, e são mais preguiçosas, descuidadas e hedonistas (COSTA; WIDIGER, 1993 apud NUNES, 2010).

Realização está subdividido em três facetas (NUNES et al., 2010): R1 – Competência, R2 – Ponderação/Prudência e R3 – Empenho. A primeira delas diz respeito ao quão disposto a fazer sacrifícios e ao quão ativo é um indivíduo na busca pelos seus objetivos. Um escore alto nessa faceta reflete uma pessoa que gosta de atividades desafiadoras, que pode fazer várias coisas ao mesmo tempo e que sabe com clareza aquilo que quer para sua vida. Baixas pontuações indicam alguém que desiste fácil, que não gosta de atividades complexas e desafiadoras e que não possui objetivos bem definidos. A segunda faceta, Ponderação/Prudência, está relacionada ao cuidado que uma pessoa tem em manifestar suas opiniões e ao quanto ela analisa as consequências de suas ações. Altos escores descrevem indivíduos que refletem mais e procuram controlar seus impulsos, agindo de maneira ponderada. Ao contrário, escores baixos relacionam-se à impulsividade, no sentido de falta de planejamento e organização no falar e no agir. Finalmente, a terceira faceta, Empenho/Prudência, descreve altos níveis de detalhismo e exigência em relação à qualidade dos trabalhos realizados. As pessoas que apresentam uma pontuação elevada nessa faceta costumam ser perfeccionistas e muito dedicadas a atividades acadêmicas e profissionais. Já quem possui escores baixos é mais descuidado, relaxado e menos persistente em relação às tarefas que se propõe a realizar.

Por sua vez, o fator Neuroticismo tem a ver com o nível habitual de ajustamento emocional e de instabilidade. Altas pontuações em Neuroticismo indicam tendências a sofrimentos psicológicos, ansiedade, depressão, vulnerabilidade, autocrítica e impulsividade, além de poderem refletir a presença de ideias irrealis e uma baixa tolerância a frustrações (COSTA; WIDIGER, 1993 apud NUNES, 2000). O contrário é verdadeiro para baixas pontuações nesse fator.

Neuroticismo divide-se em quatro facetas (NUNES et al., 2010): N1 – Vulnerabilidade, N2 – Instabilidade Emocional, N3 – Passividade/Falta de energia e N4 – Depressão. A Vulnerabilidade descreve o quão frágil emocionalmente as pessoas se sentem. Uma pontuação muito alta indica baixa autoestima, insegurança, dependência e dificuldade em tomar decisões. A segunda faceta está relacionada ao quanto as pessoas se descrevem como nervosas, irritáveis e com grandes variações no humor. Quem possui uma pontuação muito alta nela tende a não conseguir controlar seus sentimentos, ter baixa tolerância à

frustração e agir de modo impulsivo. Já quem possui uma pontuação baixa nessa faceta tende a ter o humor mais constante, conseguir lidar bem com sentimentos negativos e controlar seus impulsos. Já a faceta Passividade/Falta de energia descreve, no caso de uma pontuação alta, indivíduos que têm dificuldade em iniciar tarefas, que procrastinam, que são indecisos e que abandonam atividades consideradas longas ou difíceis, precisando de estímulo de outras pessoas para alcançar seus objetivos. Já uma baixa pontuação reflete indivíduos proativos, que conseguem concluir as tarefas que se propõem a fazer e que são decididos e motivados. Por fim, pontuações altas na faceta Depressão indicam pessoas com expectativas negativas para o futuro, que não possuem objetivos claros e que não se consideram capazes de lidar com as dificuldades do cotidiano. Quem possui baixas pontuações nessa faceta, por outro lado, acredita ser capaz de lidar com os problemas e é esperançoso em relação ao futuro; em níveis extremos, pode indicar dificuldade em reconhecer problemas reais (NUNES et al., 2010).

Finalmente, uma alta pontuação em Abertura reflete um indivíduo que gosta de explorar coisas novas, é curioso, criativo, aberto a novas ideias e capaz de sentir uma ampla variedade de emoções. Já uma pontuação baixa nesse fator indica alguém mais convencional em suas ideias e atitudes, alguém conservador e mesmo dogmático, com crenças pouco flexíveis. Além disso, baixa Abertura corresponde a pessoas que respondem fracamente a emoções (COSTA; WIDIGER, 2002 apud NUNES et al., 2010).

As facetas de Abertura são (NUNES et al., 2010): A1 – Abertura a ideias, A2 – Liberalismo e A3 – Busca por novidades. A primeira descreve o nível de interesse por conceitos abstratos, discussões filosóficas, atividades artísticas e pensamento imaginativo ou fantasioso. Baixos escores refletem indivíduos que são pouco curiosas, conservadoras e rígidas em seus gostos artísticos e conceitos, sendo o oposto verdade para os escores altos. Já segunda faceta, Liberalismo, tem a ver com abertura a valores morais e relativização de regras sociais. Escores muito baixos refletem pessoas dogmáticas em suas crenças, enquanto escores muito altos refletem posturas relativistas. Por fim, a Busca por novidades descreve o quanto o indivíduo procura viver novas experiências. Quem apresenta uma alta pontuação nessa faceta relata não gostar de rotinas ou de tarefas repetitivas e se entendiam facilmente na ausência de eventos novos. Ao contrário, quem tem uma pontuação baixa não gosta de quebras na rotina e não apresenta interesse em fazer coisas novas.

É importante ressaltar que, por se tratar de um teste autoavaliativo (*self-report*), as pontuações nas facetas e nos fatores refletem o que a pessoa acredita ser o caso a respeito de si mesma, estando sujeitas a uma série de limitações típicas desse tipo de instrumento. Por exemplo, as pessoas podem ser incapazes de se enxergar de forma realista, apresentando uma

visão mais positiva ou negativa de si mesmas, ou podem não admitir comportamentos negativos por serem indesejáveis socialmente (MCDONALD, 2008). Entretanto, a grande maioria (mais de 95%) dos trabalhos publicados em revistas de psicologia da personalidade utilizam *self-reports*, ainda que por vezes incluam outras medidas, tais como a observação direta do comportamento dos participantes (MCDONALD, 2008). Além disso, diversos pesquisadores acreditam que ninguém tem mais acesso às informações sobre si e ninguém está mais motivado a falar sobre si do que a própria pessoa (MCDONALD, 2008). Finalmente, o instrumento específico utilizado neste trabalho, qual seja a Bateria Fatorial de Personalidade, apresenta índices de confiabilidade, que mede a consistência interna dos fatores das escalas, e de validade do construto, que avalia o quanto o instrumento está medindo o que se propõe a medir, que são considerados adequados (coeficiente Alfa de Cronbach dos fatores variando entre 0,74 e 0,89), tendo sido validado e aprovado pelo Sistema de Avaliação de Testes Psicológicos (SATEPSI) do Conselho Federal de Psicologia.

2.1.3 Linguagem e personalidade

Há algumas décadas tem-se postulado que a linguagem não é só um veículo dos pensamentos mas também da personalidade dos indivíduos (SANFORD, 1942). No entanto, a análise sistemática das maneiras com as quais o uso de palavras e a personalidade se relacionam são relativamente recentes (HIRSH; PETERSON, 2009). Um dos obstáculos para esse tipo de investigação consistia na identificação das unidades linguísticas que deveriam ser o foco e dos métodos mais eficazes de análise do comportamento verbal. Os vocábulos gramaticais, por exemplo, não eram considerados significativos o suficiente, tendo sido praticamente ignorados por séculos nas pesquisas do gênero. Além disso, só há algumas décadas é que se desenvolveram ferramentas automatizadas para a análise textual, tais como os programas *General Inquirer* (STONE et al., 1966) e *Linguistic Inquiry and Word Count* (PENNEBAKER; FRANCIS; BOOTH, 2001), que conferiram possibilidades novas de investigação. Por exemplo, a partir do surgimento desses programas computadorizados, os pesquisadores puderam considerar aspectos que antes não acreditavam ser relevantes, como é o caso dos vocábulos gramaticais.

Esses vocábulos não se referem a objetos ou noções fixas no mundo, isto é, eles têm sentido apenas no contexto em que se encontram, servindo principalmente para estabelecer relações entre as palavras (PENNEBAKER, 2011). Além disso, eles são curtos e possuem alta carga social. Por exemplo, a frase “Ele ficou lá até ela terminar aquilo” é impossível de ser

compreendida fora de um contexto social; um bilhete com essa frase deixado a alguém que não soubesse quem “ele”, “ela”, “lá” e “aquilo” significam seria inútil. De fato, até em nível cerebral essa classe de palavras está relacionada a funções sociais e atividades emocionais: a área de Broca, localizada no lobo frontal esquerdo, é essencial para seu processamento e é crucial para a fluência verbal e a calidez na fala, o que pode ser evidenciado nas afasias provocadas por danos majoritários a essa área. Por outro lado, os vocábulos lexicais, tais como substantivos, verbos e adjetivos, dependem mais da área de Wernicke, localizada no lobo temporal esquerdo (na maioria das pessoas); desse modo, pacientes com afasias causadas por danos na área de Wernicke apresentam fala fluida, mas com sérios déficits de sentido (PENNEBAKER, 2011).

Apesar de os vocábulos gramaticais serem os mais profícuos em se tratando de possíveis relações psicológicas/sociais, os vocábulos lexicais, como verbos e substantivos, também são importantes na medida em que indicam as prioridades e os objetivos das pessoas. Por exemplo, em uma conversa entre amigos, alguns tendem a falar mais de futebol enquanto outros preferem focar sobre algum conhecido: o conteúdo da conversa revela o que é importante para os falantes, incluindo seus objetivos, valores e, num nível mais amplo, a sua personalidade (PENNEBAKER, 2011, p. 84). Além disso, mesmo em situações idênticas, duas pessoas sempre irão focar em aspectos diferentes. Por exemplo, ao se depararem com a mesma praça, dois indivíduos com personalidades distintas poderiam tirar fotos de coisas muito distintas: um introvertido poderia focar nos detalhes dos paralelepípedos no chão, por exemplo, enquanto que um extrovertido poderia pedir que alguém tirasse a sua foto ao lado de um amigo ou de um monumento local. Assim, ao descreverem a experiência de estar naquela praça, cada um utilizaria vocábulos lexicais que apontariam para noções díspares.

Nas últimas décadas, considerando ambos vocábulos lexicais e gramaticais, têm-se investigado mais sobre a relação do uso das palavras com aspectos da personalidade, cujo modelo estrutural mais empregado tem sido o dos Cinco Grandes Fatores (HOLTGRAVES, 2014). Os achados não são tão robustos como se gostaria, mas alguns padrões estão emergindo. Por exemplo, o fator Extroversão tem sido associado ao uso de pronomes pessoais de segunda pessoa (YARKONI, 2010), palavras mais curtas (MEHL; GOSLING; PENNEBAKER, 2006) e palavras relacionadas a processos sociais (OBERLANDER; GILL, 2006). Já Socialização tem sido associado a uma linguagem mais polida, com maior uso de palavras com conotação positiva, autorreferências (provavelmente devido à utilização de expressões como “eu acho”, que indicam incerteza e polidez, e não à autorreferência ruminante presente na linguagem de indivíduos com pontuações altas em Neuroticismo)

(HOLTGRAVES, 2014), palavras que indicam processos sociais (MEHL et al., 2006), e menor uso de negações (NOWSON, 2006) e palavrões (HOLTGRAVES, 2011). Por sua vez, o fator Neuroticismo está associado a maiores índices de primeira pessoa do singular (PENNEBAKER; KING, 1999), emoções negativas (PENNEBAKER; KING, 1999) e palavrões (YARKONI, 2010). Já o fator Abertura deve ser dividido em suas facetas: as que descrevem abertura a ideias têm sido associadas a um estilo mais formal e informativo e menos narrativo de linguagem, com menos uso de pronomes, e mais uso de artigos, preposições e palavras longas; já as que descrevem abertura a sentimentos e estética está relacionada a maior uso de pronomes e emoções positivas (HOLTGRAVES, 2014). Finalmente, o fator Realização está caracterizado por uma linguagem mais cordial, com menor uso de palavrões e palavras de conotação negativa (MEHL et al., 2006). É válido observar que esses resultados provêm, em sua grande maioria, de estudos com falantes de inglês.

Além disso, as relações encontradas entre os indicadores da personalidade e aspectos linguísticos dependem também do tipo de produção verbal analisada. Por exemplo, os estudos que se basearam em tarefas de produção textual de narrativas pessoais, tais como os de Fast e Funder (2008) e Hirsh e Peterson (2009), encontraram correlações maiores com a personalidade do que os que se basearam em tarefas de produção de fluxos de consciência ou ensaio acadêmico, como o de Pennebaker e King (1999). Os primeiros foram mais vantajosos para a análise da personalidade porque, de acordo com Hirsh e Peterson (2009), as narrativas acerca da trajetória de vida de um indivíduo, incluindo descrições subjetivas do passado e futuro, são cruciais para a individualidade e identidade do sujeito, o que leva a uma maior relação entre os traços de personalidade e as características linguísticas desse tipo de produção textual.

Em resumo, a personalidade é um conceito ubíquo, que pode ser descrito por meio de traços mais ou menos estáveis de comportamento, os quais, por sua vez, podem se manifestar na linguagem que os indivíduos utilizam no dia-a-dia. Assim, a linguagem não é só um veículo de informações sobre o mundo externo ao indivíduo, podendo também transmitir informações sobre seu mundo interno. Na próxima seção, será apresentada uma maneira pela qual aspectos linguísticos podem dar indícios de mais um componente desse mundo interno.

2.2 Sexo

Assim como a personalidade, o sexo também pode ser um fator que influencia diversas vertentes do comportamento humano. Em inglês, por exemplo, a expressão “*Boys will be boys*” aponta para o fato de homens (ou meninos) muitas vezes agirem como agem em razão do seu sexo biológico. Similarmente, a canção *Garotos II*, cantada por Leoni e muito popular até hoje no Brasil, aponta para um padrão de comportamento tipicamente masculino nas relações românticas. Para além dessas manifestações na cultura popular, estudos científicos têm demonstrado a influência do sexo em diversos âmbitos da vida, incluindo o da produção linguística.

2.2.1 Linguagem e sexo

A influência do sexo sobre a forma e o conteúdo das manifestações verbais das pessoas já está bem estabelecida, com diversos estudos demonstrando diferenças tanto em estilo quanto em conteúdo da produção verbal (TAUSCZIK; PENNEBAKER, 2010). Por exemplo, no site do World Well-Being Project (WWBP), que mede o bem-estar psicológico baseado na análise linguística de mídias sociais, existem nuvens de palavras associadas a cada um dos sexos, desenvolvidas a partir de análises automáticas de publicações em redes sociais. A figura a seguir descreve as palavras mais utilizadas por mulheres (WORD..., [s.d.]):

Figura 1 - Nuvem de palavras mais utilizadas por mulheres



Fonte: World Well-Being Project (WORD..., [s.d.]).

palavras que denotam emoções e verbos no presente. Além disso, Pennebaker e King (1999) identificaram maior uso de primeira pessoa do singular por pessoas do sexo feminino.

Apesar de esses padrões de diferenças linguísticas entre os sexos se manifestarem em diversos contextos e culturas diferentes, é importante pensar que pode haver variações com base no tipo de amostra em questão. Por exemplo, um estudo constituído em sua maioria por pessoas jovens terá um perfil de uso da língua bem diferente de outro cujos participantes fossem idosos. Similarmente, um estudo com estudantes de Letras apresentaria um perfil de uso da língua diferente de outro com estudantes de Engenharia, tanto devido à maior familiaridade com a escrita que se pressupõe em um curso de Letras quanto pela diferença de personalidade entre os estudantes de cada grupo (NUNES et al., 2010)

2.2.2 Personalidade e sexo

Falar sobre diferenças intrínsecas entre homens e mulheres, como é o caso dos traços de personalidade, é tarefa espinhosa por pelo menos duas razões. Em primeiro lugar, o argumento de que homens e mulheres são fundamentalmente diferentes foi usado no passado para privar as mulheres de direitos e de liberdade, seja por meio da falta de representação política, seja pela imposição de barreiras no mercado de trabalho e na educação formal. Em segundo lugar, existe muito debate em relação às causas dessas diferenças, com alguns grupos privilegiando a influência do papel social (EAGLY, 1987) e outros dando mais ênfase a aspectos biológicos pautados na psicologia evolutiva (BUSS, 1995). Por isso, é importante enfatizar que o fato de haver diferenças não implica argumentos morais sobre a superioridade de um ou outro grupo, sendo, ao contrário, algo que deve ser considerado para poder trazer avanços à sociedade como um todo. Por exemplo, é crucial para a prevenção e o prognóstico de doenças que se reconheça que mulheres são mais suscetíveis a desenvolver, depressão, hepatite, enxaqueca, esclerose múltipla, entre outros, e que homens são mais propensos a ter infartos do miocárdio (WHITLEY; LINDSEY, 2009). Algo similar pode ser dito em relação a possíveis diferenças de personalidade na medida em que, caso se queira mudar o que por ventura for considerado inadequado, é necessário ter uma ideia clara sobre o quadro presente. Já no que tange à personalidade, é pertinente notar que metanálises de estudos feitos em culturas diferentes têm demonstrado que a causa mais provável das diferenças entre os sexos é uma combinação de fatores biológicos e sociais; isso porque o padrão de diferenças de personalidade não é o mesmo em todas as culturas do mundo, o que enfraquece a hipótese evolutiva, e porque que as maiores diferenças de personalidade entre os sexos são encontradas

em países mais desenvolvidos e igualitários, o que contradiz a abordagem baseada em papéis sociais (SCHMITT; REALO; VORACEK; ALLIK, 2008).

De qualquer forma, independentemente de sua origem, diferenças entre homens e mulheres existem: desde aspectos mais básicos no nível cromossômico, até diferenças em estrutura corporal, distribuição hormonal, comportamento neuroquímico e personalidade (NGUN; GHARAMAMI; SÁNCHEZ; BOCKLANDT; VILAIN, 2011; SCHMITT; LONG; MCPHEARSON; O'BRIAN; REMMERT; SHAH, 2016). Em relação aos traços de personalidade especificamente, as diferenças reportadas têm sido pequenas a moderadas, considerando pesquisas com homens e mulheres de diferentes culturas, etnias e classes sociais (SCHMITT et al., 2016; WEISBERG; DEYOUNG; HIRSH, 2011). Dentro do modelo dos Cinco Grandes Fatores, por exemplo, essas análises demonstram níveis maiores de Neuroticismo ($d = -0,40$), Socialização ($d = -0,15$), Realização ($d = -0,12$) e Extroversão ($d = -0,10$) para o sexo feminino, com o fator Abertura não apresentando diferenças significativas consistentes (SCHMITT; REALO; VORACEK; ALIK, 2008). Em relação aos subfatores ou facetas do modelo dos CGF, em sua metanálise transcultural, Costa, Terracciano e McCrae (2001) encontraram que, de modo geral, as mulheres tinham pontuações mais altas em: facetas que indicavam vulnerabilidade, insegurança e ansiedade, no fator Neuroticismo; todas as facetas do fator Socialização, tais como modéstia, cooperação, altruísmo, simpatia e confiança; facetas que indicavam calidez, gregariedade e emoções positivas, no fator Extroversão; facetas relacionadas a comportamento diligente, no fator Realização; e em facetas relacionadas a abertura a sentimentos e estética, no fator Abertura. Já as pontuações dos homens foram relatadas como maiores em facetas relacionadas a abertura a idéias, no fator Abertura, e a assertividade e busca de sensações fortes, no fator Extroversão (COSTA et al., 2001).

Em suma, existem diferenças na linguagem e na personalidade de homens e mulheres de maneira geral. Essas diferenças, no entanto, são modestas e não significam que todos os membros do grupo se comportam da mesma forma; ao contrário, é normal haver mais variação entre indivíduos do que entre grupos. De qualquer modo, é relevante saber quais padrões se revelam por trás das atividades humanas, e isso inclui as atividades linguísticas, para as quais diversos métodos de análise têm sido desenvolvidos. A seguir, serão expostos alguns desses métodos.

2.3 Análise textual

A análise do comportamento verbal dos indivíduos é feita diariamente por cada um de nós de modo intuitivo, porém costumamos dar mais importância ao conteúdo e aos elementos prosódicos das falas, tais como tom, volume e ritmo do discurso, do que ao estilo empregado pelo interlocutor. Por outro lado, quando temos um texto escrito à nossa frente, tendemos a privilegiar sua informação semântica, ignorando aspectos formais ou estilísticos. Para analisar tanto transcrições de fala quanto textos originais escritos sob o aspecto formal, as ferramentas de análise textual automatizadas podem ser uma ótima solução. Isso porque elas são capazes de captar detalhes que passam despercebidos para os humanos, tais como a frequência de preposições, artigos e pronomes utilizados pela pessoa. Essas palavras aparentemente insignificantes, podem, na verdade, fornecer uma série de informações acerca de estados psicológicos e até biológicos de quem as usa.

2.3.1 Desenvolvimento de ferramentas de análise textual automatizadas

As ferramentas de análise textual têm adquirido novos recursos e têm ampliado suas possibilidades de aplicação à medida que a tecnologia avança; os primeiros esforços para análise psicológica de fenômenos linguísticos eram muito mais rudimentares e dependentes de humanos do que os atuais, que funcionam de modo inteiramente automático. Um dos primeiros deles foi desenvolvido por Gottschalk e seus colegas (GOTTSCHALK; GLESER, 1969 apud TAUSCZIK; PENNEBAKER, 2010) e tinha como objetivo analisar o conteúdo para identificar temáticas freudianas nas transcrições dos fluxos de consciência de pacientes. Após as dividirem em frases gramaticais, juízes humanos buscavam nas transcrições indicadores de processos psicológicos, tais como ansiedade e hostilidade. Esse método foi usado posteriormente para o diagnóstico de disfunções cognitivas, alcoolismo, dano cerebral e transtornos mentais.

O primeiro programa *computadorizado* de análise psicológica de texto foi desenvolvido em 1966 por Philip Stone e seus colegas (STONE et al., 1966 apud TAUSCZIK; PENNEBAKER, 2010). O *software*, chamado *General Inquirer*, baseia-se em uma série de algoritmos desenvolvidos pelo autor e recebe *input* da tradição psicanalítica e da teoria de necessidades de McClelland (1961). Ele pode ser usado para estudar praticamente qualquer tópico a partir da criação de um dicionário produzido pelo usuário. Outros programas foram desenvolvidos depois dele tendo o objetivo de investigar outros aspectos

psicológicos, tais como tom emocional e o nível de abstração, como é o caso do TAS/C (MERGENTHALER, 1996 apud PENNEBAKER; MEHL; NIEDERHOFFER, 2003) e o tom verbal em discursos políticos, como é o caso do DICTION (HART, 2001 apud PENNEBAKER et al., 2003). Todos foram úteis para avançar as pesquisas na área de análise textual com foco em aspectos psicológicos.

De modo destacado, no entanto, é preciso apontar para o desenvolvimento do programa *Linguistic Inquiry and Word Count*. Em 1986, Pennebaker e Beal decidiram investigar quais aspectos da escrita sobre experiências de vida negativas poderiam prever benefícios para a saúde física de participantes (PENNEBAKER; BEAL, 1986). Para isso, eles analisaram textos que tratavam sobre traumas utilizando o julgamento de juízes humanos, que liam e classificavam as produções em diversas dimensões psicológicas e linguísticas. No entanto, esse sistema não era funcional, pois os juízes não chegavam a um acordo e o processo era muito lento e custoso (PENNEBAKER; TAUSCZIK, 2010). Em busca de um método que fosse mais eficiente, Pennebaker e seus colegas começaram a desenvolver um programa de análise textual computadorizado para contar as palavras e classificá-las em categorias psicológicas, cognitivas e linguísticas. O resultado é o *Linguistic Inquiry and Word Count* ou LIWC (pronuncia-se como o nome “Luke”), cuja primeira versão foi divulgada por Pennebaker, Francis e Booth em 2001.

2.3.2 *Linguistic Inquiry and Word Count* – LIWC

O LIWC possui duas características principais – seu processador e seus dicionários. A primeira é o programa em si, que pode analisar, categorizar e identificar com cores diversos formatos de arquivos textuais; a segunda corresponde aos dicionários, que podem ser baixados facilmente a partir do programa e que incluem línguas como russo, alemão, português e espanhol. Esses dicionários, desenvolvidos por terceiros (como é o caso no português) ou pelos próprios criadores do LIWC (como é o caso do espanhol), contêm uma lista de palavras pré-classificadas que servem como base para a identificação e classificação das palavras do texto-base. Cada palavra pertence a uma ou mais categorias. Por exemplo, “amei” é classificada em oito categorias diferentes, incluindo verbo, tempo verbal passado e palavras com conotações sexuais. No final da análise, os resultados são apresentados de acordo com a porcentagem de palavras de cada categoria presente no texto-base. Algumas dessas categorias são bastante transparentes, como é o caso dos artigos e pronomes; outras são mais subjetivas, tal como as categorias *emoção positiva* ou *emoção negativa*. No caso das

últimas, as palavras foram primeiro coletadas de dicionários, glossários e questionários, e depois foram classificadas por três juízes de modo independente ao longo de anos.

É importante destacar que o LIWC, assim como todas as ferramentas de análise textual automática, é um sistema probabilístico e, como tal, é suscetível a erros de classificação quando a interpretação é dependente do contexto (PENNEBAKER et al., 2003). Por exemplo, uma palavra como “louco” é classificada como negativa, mas na frase “Eu tô louco de faceiro” ela deveria ser classificada como positiva. Logo, essa incapacidade de levar em consideração o contexto implica a impossibilidade de captar ironias e outras figuras de linguagem que são amplamente utilizadas no dia-a-dia. De qualquer forma, ele é um ótimo auxílio para quem busca informações sobre os autores no estilo do texto e na frequência de uso das palavras, por exemplo.

2.3.3 *Brazilian Portuguese LIWC 2007 Dictionary*

O *Brazilian Portuguese LIWC 2007 Dictionary* (BALAGE FILHO et al., 2013) é o dicionário em português criado para o programa *Linguistic Inquiry and Word Count*. Ele foi desenvolvido a partir da tradução da versão de 2007 do dicionário LIWC original em inglês, e as classificações das palavras nas categorias linguísticas e psicológicas foram feitas de modo automático. A versão brasileira do léxico contém mais de 127 mil palavras, catalogadas em 64 categorias linguísticas e psicológicas diferentes. Seu conteúdo está disponível gratuitamente para download no endereço *web* do PortLex (PORT..., [s.d.]).

A disponibilização desse dicionário em português para o LIWC vem contribuindo para as pesquisas com base em análise textual no Brasil, pois o programa é um dos mais completos de sua categoria. Estudos como o de Machado et al. (2015), Paim, Camati e Enembreck (2016), Jaques e Nunes (2012), Ponte Junior, Guedes e Bezerra (2016) e Wainberg (2016) têm utilizado o LIWC para as mais diversas aplicações, incluindo a mineração de traços de personalidade, a construção de ambientes de aprendizagem online personalizados, a detecção de falsidade no discurso e a inferência do sexo de usuários de redes sociais. Mais detalhes acerca do dicionário e do programa em geral estão disponíveis na seção 3.2 Instrumentos.

De modo geral, portanto, pode-se pensar que ferramentas de análise textual automatizadas podem contribuir para a investigação do comportamento verbal das pessoas, as quais podem apresentar diferentes padrões de utilização da linguagem a depender de fatores como sexo e personalidade, entre outros. Contudo, sejam por limitações teóricas ou metodológicas, as pesquisas nessa área são relativamente incipientes. Com o objetivo de

contribuir para a consolidação desse ramo pouco explorado é que este estudo se apresenta. A seguir, então, serão explicitados os detalhes metodológicos e serão discutidos os resultados.

3 MÉTODO

3.1 Participantes

Realizaram as tarefas 85 indivíduos, 37 homens e 48 mulheres, brasileiros, cujas idades variaram entre 18 e 59 anos ($M = 27,3$, $DP = 8,6$). A amostra foi composta em sua maioria por estudantes universitários de cursos de humanas (90%), sendo que 72,2% era proveniente do curso de Letras da UFRGS. Sessenta e quatro por cento dos participantes possuía ensino superior incompleto, tendo os demais níveis maiores de escolaridade (superior completo, pós-graduação, mestrado ou doutorado). Ainda, a amostra incluiu sujeitos de etnias diferentes, incluindo brancos, negros e pardos.

3.2 Instrumentos

3.2.1 Bateria Fatorial de Personalidade

A Bateria Fatorial de Personalidade (BFP) é um instrumento psicológico construído para a avaliação da personalidade a partir do modelo Cinco Grandes Fatores (CGF), que inclui as seguintes dimensões e facetas, conforme descrito na seção 2.1.2: Extroversão (E1 – Comunicação; E2 – Altivez; E3 – Dinamismo; E4 – Interações Sociais); Socialização (S1 – Amabilidade; S2 – Pró-sociabilidade; S3 – Confiança nas pessoas); Realização (R1 – Competência; R2 – Ponderação / Prudência; R3 – Empenho / Comprometimento); Neuroticismo (N1 – Vulnerabilidade; N2 – Instabilidade emocional; N3 – Passividade / Falta de Energia; N4 – Depressão); e Abertura (A1 – Abertura a ideias; A2 – Liberalismo; A3 – Busca por novidades).

O fator Extroversão diz respeito à quantidade e à intensidade das interações interpessoais preferidas, nível de atividade, necessidade de estimulação e capacidade de alegrar-se. Socialização, apesar de também representar uma dimensão interpessoal, avalia o interesse pelo bem-estar dos outros, a confiança nas pessoas e a adesão a regras ou a normas sociais. Por sua vez, o fator Realização faz referência ao grau de organização, persistência, controle e motivação para alcançar objetivos. Já o fator Neuroticismo refere-se ao nível de

ajustamento emocional e instabilidade, envolvendo especificamente a vulnerabilidade à opinião dos outros, instabilidade de humor/emocional, presença de sintomas depressivos e comportamentos passivos ou falta de energia para agir em situações importantes. Por último, o fator denominado Abertura para experiências reflete comportamentos exploratórios e o reconhecimento da importância de vivenciar novas experiências, medindo o interesse por novas idéias e a busca por novidades (NUNES; HUTZ; NUNES, 2010).

A BFP foi desenvolvida em 2010 por Carlos Henrique S.S. Nunes, Claudio Simon Hutz e Mariana Farias Oliveira Nunes, levando em conta a linguagem, os valores culturais, as diversidades regionais e as especificidades dos quadros clínicos no Brasil (NUNES; HUTZ; NUNES, 2010). Ela é composta por 126 itens, que são respondidos em uma escala tipo likert de sete pontos, a qual indica o nível de identificação das pessoas com as características neles descritas; por exemplo, “Sou divertido” ou “Acredito que as pessoas têm boas intenções”. A BFP foi elaborada a partir de uma seleção dos itens com as melhores propriedades psicométricas das escalas individuais para a avaliação dos CGF, a saber: a Escala Fatorial de Neuroticismo (HUTZ; NUNES, 2001), a Escala Fatorial de Extroversão (NUNES; HUTZ, 2007a), a Escala Fatorial de Socialização (NUNES; HUTZ, 2007b), a Escala Fatorial de Abertura (VASCONCELLOS; HUTZ, 2008) e a Escala Fatorial de Realização (PACICO; HUTZ, [s.d.]).

As evidências de validade da BFP foram investigadas a partir de uma base de dados de 18 estudos independentes, com uma amostra total de 6.599 pessoas, sendo a maioria estudantes universitários ou de ensino médio, provenientes de quatro regiões do Brasil (com exceção do Norte). Para verificar as dimensões da BFP, foram extraídas soluções fatoriais utilizando a rotação *varimax* com 4, 5 e 6 fatores. A solução mais adequada foi encontrada considerando a consistência interna dos fatores e a sua pertinência em relação ao sistema utilizado (NUNES et al., 2010). As análises indicaram que a melhor solução consistia em cinco fatores, cujos coeficientes Alfa de Cronbach (de 0 a 1; quanto mais perto de 1, melhor) foram de 0,74, para Abertura; 0,83, para Realização; 0,84, para Extroversão; 0,85, para Socialização; e 0,89, para Neuroticismo. Também foram realizadas análises fatoriais exploratórias para a identificação das facetas de cada fator, dessa vez utilizando o método de rotação *direct oblimin*. Os resultados indicaram quatro facetas para os fatores Neuroticismo e Extroversão, e três facetas para os fatores Realização, Socialização e Abertura. A consistência interna das facetas, considerando o Alfa de Cronbach, variou entre 0,57 (A3 - Busca por Novidades) e 0,85 (S1 – Amabilidade), com a maior parte dos valores sendo considerada boa

ou muito boa. O instrumento foi validado e aprovado pelo SATEPSI, do Conselho Federal de Psicologia.

3.2.2 Tarefa de Produção Textual

Os participantes (em grupos de dois a doze indivíduos por sessão) realizaram uma tarefa de produção escrita em um formulário do Google Forms, onde deveriam responder à seguinte questão: “Como você gostaria que a sua vida fosse e que tipo de pessoa você gostaria de ser em 5 anos?” Esse tema foi escolhido por ir ao encontro do que foi feito por Hirsh e Peterson (2009), quando alunos da graduação da Universidade de Toronto produziram narrativas pessoais (*self-narratives*) discorrendo sobre seus planos futuros; o estudo também tinha como objetivo relacionar indicadores de personalidade do modelo *Big Five* e categorias do LIWC. Uma pergunta muito semelhante também é apresentada na página do programa de escrita pessoal (*self-authoring*) desenvolvido por Peterson, a saber: o que você espera conquistar em sua vida e que tipo de pessoa você quer ser (FUTURE..., 2017).

Não foi estipulado um número mínimo de palavras, já que queríamos incluir o número de palavras total nas correlações com os indicadores de personalidade. Em vez disso, apenas propusemos um mínimo de 30 minutos e um máximo de uma hora para que os participantes tentassem realizar a tarefa. Também não foi feito nenhum tipo de restrição quanto à linguagem, quer seja em termos de escolhas lexicais, registro (formal ou informal), estrutura ou gênero do texto (narrativo, dissertativo etc.). O objetivo era analisar a linguagem na sua forma mais espontânea, procurando minimizar a rigidez da linguagem acadêmica. Tal abordagem segue uma tradição metodológica para esse tipo de investigação, nas quais o objetivo é relacionar características linguísticas com traços de personalidade (HIRSH; PETERSON, 2009; PENNEBAKER; KING, 1999). Todos os textos foram escritos em português, a língua nativa de todos os participantes.

Os participantes foram informados de que suas tarefas seriam processadas por um programa de computador e de que seriam identificadas apenas por um número, garantindo seu anonimato durante a fase de análise de dados.

3.2.3 *Linguistic Inquiry and Word Count* – LIWC

O LIWC (PENNEBAKER; FRANCIS; BOOTH, 2001) é uma ferramenta de análise textual computadorizada que avalia palavras individualmente, tendo como base um dicionário

interno pré-definido pelo usuário. O *software* classifica as palavras em dezenas de dimensões linguísticas e psicológicas, que incluem processos sociais, cognitivos e afetivos. O módulo de análise textual identifica e categoriza as palavras e depois calcula a porcentagem total de palavras que corresponde às categorias pré-definidas no dicionário. A versão utilizada neste trabalho foi baseada no *Brazilian Portuguese LIWC 2007 Dictionary*, que possui 64 categorias, divididas em quatro dimensões, a saber: (i) processos linguísticos, que inclui artigos, preposições, advérbios, negações, pronomes, entre outras; (ii) processos psicológicos, que está subdividida em processos sociais, biológicos, cognitivos, perceptivos e relatividade; (iii) preocupações sociais, que inclui categorias relacionadas a trabalho, casa, religião, entre outras; (iv) categorias da fala, que inclui expletivos, expressões de concordância e de não fluência. Cada palavra pode ser classificada em várias categorias diferentes. Por exemplo, a palavra “aonde” é classificada como *adverb* (advérbio), *function* (vocábulo gramatical) e *relativ* (relatividade). O dicionário brasileiro foi baseado no dicionário original americano. Este, por sua vez, teve como critério de seleção as palavras que eram mais comuns na língua inglesa (PENNEBAKER; FRANCIS; BOOTH, 2001). As categorias e dimensões do LIWC, adaptadas para o português de acordo com a versão 2007 do programa, estão explicitadas na Tabela 1.

Tabela 1 Dimensões e categorias do LIWC 2007 (adaptado de PENNEBAKER et al., 2007)

DIMENSÕES	CATEGORIAS	EXEMPLOS
PROCESSOS LINGUÍSTICOS		
	Categorias métricas:	
	Palavras totais	
	Palavras por frase	
	Palavras > 6 letras	
	Dicionário	
	Categorias linguísticas	
	Vocábulo gramatical	
	Pronome	Vós, o, a, ela
	Pronome pessoal	Ele, nós
	Eu	Eu, mim
	Nós	Nós, nosso
	Você/tu	Vosso, teu, você
	Ele/ela	Ele, ela, dele
	Eles/elas	Elas, delas
	Pronome impessoal	Isso, aquilo
	Artigo	A, o, um, uma
	Verbo	Buscar, falamos
	Verbo auxiliar	Ser, sido, haveria
	Pretérito	Falei, cantávamos, falara
	Presente	Percorro, contém
	Futuro	Deveria, poderes, podiam
	Advérbio	Muito,

PROCESSOS PSICOLÓGICOS

Preposição	acolá,estupidamente
Conjunção	Da, para, acima
Negação	Contudo, mas, e Nada, jamais, nem, nenhum
Quantificador	Quantia, poucos, alguns, casal
Número	Quatro, dúzia, primeiro
Palavrão	Bosta, censurar, contar
Processos sociais	Gerir, aconselhar, declarei
Família	Enteada, maninha, marido
Amigos	Amigo, parceiro, peguete
Humanos	Embaracei, governasse, misantropo
Processos afetivos	Distraísse, amizade, divirtam
Emoção positiva	Calmamente, aceitar, carinhoso
Emoção negativa	Cansada, enfermo, lesou
Ansiedade	Desassossego, espantou, fóbico
Raiva	Castigo, ridículo, sordidamente
Tristeza	Sofrer, soluçei, culpa
Processos cognitivos	Soluções, executar, separei
<i>Insight</i>	Clarificar, conheci, determinamos
Causa	Contaminaram, gastei, furtaram
Discordância	Enganada, pesara, podíamos
Incerteza	Enigma, possivelmente, talvez
Certeza	Precisamente, provado, absolutamente
Inibição	Acanhado, adiei, admoestação
Inclusão	Incluí, aberto, abrange
Exclusão	Eliminada, contra, correto
Processos perceptivos	Acalmar, acelerou, soaram
Visão	Aceso, acompanhei, encoberto
Audição	Escutei, esguelando, falassem
Sensação	Felpa, flácido, gelaram
Processos biológicos	Gengiva, gestante, imobilizar
Corpo	Inalada, latejar, magro
Saúde	Malhar, medicar, nauseado
Sexo	Nudez, pegador, preservativo
Ingestão	Provar, ingerir, derrubar
Relatividade	Abaixo, posterior, importunara
Movimento	Abordou, entregue, fôramos
Espaço	Interior, mapear, nacionalidade
Tempo	Interrompendo, inverno,

PREOCUPAÇÕES PESSOAIS		pare
	Trabalho	Labutar, liderança, montou
	Conquista	Investimos, constituiu, proveitoso
	Lazer	Jardim, patinei, adestrado
	Casa	Acamado, hospedar, querido
	Dinheiro	Acumulada, libertar, prover
	Religião	Adivinha, Jesus, padre
	Morte	Túmulo, exaurido, execução
CATEGORIAS DA FALA		
	Concordância	Aham, awn, concordo
	Não fluência	Hm, aiai, ué
	Expletivos	Sei, blá, tipo

Fonte: Elaborada pela autora.

Ao observar o conteúdo dos exemplos da Tabela 1, é possível perceber que a classificação das palavras muitas vezes não é transparente. De acordo com o PortLex, o portal online que abriga o dicionário em português do LIWC, “A criação do Brazilian Portuguese LIWC 2007 Dictionary foi realizada via tradução, usando vários dicionários bilíngues Português-Inglês” (LIWC..., [s.d.]). A página *web* também aponta que “as conjugações foram inseridas automaticamente usando o dicionário Unitex-PB do NILC [Núcleo Interinstitucional de Linguística Computacional] e as categorias do dicionário foram levantadas automaticamente” (LIWC..., [s.d.]). Ainda, é enfatizado que não foi feita a revisão do trabalho manual de tradução. Esses fatores podem explicar as classificações contraintuitivas, como é o caso do verbo “contar” na categoria *palavrão*, ou do substantivo “invulgar” na categoria *quantificador*.

De fato, o dicionário apresenta uma série de limitações. Por exemplo, verbos como “irei”, “falará” e “planejaremos” não são classificados como futuro, porque a lista de palavras que pertence a essa categoria no inglês, que serviu de base para a lista em português, é muito reduzida. Isso se deve ao fato de a língua inglesa não apresentar marcação de tempo verbal futuro nos seus verbos, o que não é o caso em português. Assim, temos a palavra “poderia” (e suas variações, tais como “poderes” e “podiam”) como pertencente à categoria futuro porque a expressão “*could’ve*” original está classificada nessa categoria em inglês, mas esse resultado não faz sentido em português. Além disso, por ter sido traduzida do dicionário em inglês, uma língua onde a marcação dos sujeitos gramaticais não é dada pelos verbos, nenhum verbo em português tinha classificações referentes aos sujeitos gramaticais; ou seja, a palavra “amei”, por exemplo, não está na categoria *eu*, muito embora seja uma referência à primeira pessoa do

singular. Finalmente, nem todas as palavras foram classificadas em todas as categorias que poderiam encaixar-se e há uma inconsistência nas classificações, como é o caso do advérbio “velozmente”, que está classificado somente na categoria *tempo*, e a palavra “rapidamente”, que está classificada como *advérbio*, *vocábulo gramatical*, *emoção positiva*, *percepção*, *tempo*, entre outras.

Apesar de suas limitações, o dicionário em português do LIWC tem sido empregado para classificação de diversos tipos de texto no Brasil com resultados satisfatórios (FERREIRA; FERNANDES, 2017; MACHADO, 2016; PONTE JÚNIOR et al., 2016). Assim, para este trabalho, foram selecionadas as categorias pertencentes à primeira dimensão do LIWC, qual seja a de processos linguísticos; ainda, não foi considerada a categoria *futuro*, por essa não corresponder à estrutura da língua portuguesa. Desse modo, foram consideradas: *número de palavras total*, *palavras por frase*, *palavras maiores de seis letras*, *dicionário*, *vocábulo gramatical*, *pronomes*, *pronomes pessoais*, *eu*, *nós*, *ocê/tu*, *ele/ela*, *eles/elas*, *pronome impessoal*, *artigo*, *verbo*, *verbo auxiliar*, *pretérito*, *presente*, *advérbio*, *preposição*, *negação*, *conjunção*, *quantificador*, *número* e *palavrão*, somando ao todo 25 categorias. Os processos linguísticos foram escolhidos por serem objeto de estudo pertinente e flagrante deste curso de Letras e por apresentarem os itens menos óbvios no que tange a relações com a personalidade. Por exemplo, espera-se que alguém mais ambicioso fale mais sobre trabalho ou estudos, ou que alguém mais depressivo utilize mais referências à morte ou a emoções negativas, mas a relação entre uso de pronomes, artigos ou preposições não é transparente. Além disso, as relações entre as categorias psicológicas do LIWC e a personalidade foram mais exploradas na literatura (HIRSH; PETERSON, 2009), fazendo com que o estudo das variáveis linguísticas em relação aos indicadores de personalidade consista em uma lacuna maior a ser preenchida.

A primeira categoria a ser inserida nesta investigação, isto é, a de *número de palavras*, é bastante autoexplicativa: representa o número de palavras total utilizado pelos participantes em seus textos. Vale frisar que as palavras compostas só são contabilizadas se estiverem hifenizadas, pois o dicionário só trabalha com unigramas. Por exemplo, “mal-estar” é classificado como uma palavra, e “no entanto”, que só aparece separadamente no dicionário, é classificado como duas palavras diferentes. Já a segunda categoria escolhida refere-se ao número de palavras por frase, sendo estas delimitadas pelo espaço entre qualquer ponto final ou pontos de interrogação/exclamação. Por sua vez, a terceira categoria, *palavras no dicionário*, inclui todas as palavras que o dicionário interno do LIWC pode identificar, lembrando que o *Brazilian Portuguese LIWC 2007 Dictionary* contém 127 mil tokens, o que

não é muito comparado ao que existe em dicionários online como o Houaiss, com 230 mil verbetes (HOUAISS..., [s.d.]). A quarta categoria, palavras maiores de seis letras, a última da subcategoria *métrica*, contabiliza as palavras que possuem mais de seis letras, observados os critérios para identificação de palavras descritos acima (unigramas).

Já nas categorias linguísticas propriamente ditas, a *vocábulo gramatical* refere-se às chamadas *function words*, tais como conjunções (“mas”, “e”), verbos auxiliares (“ser”, “ter”), advérbios (“aqui”, “não”), pronomes (“eu”, “dela”) e artigos (“a”, “um”). A próxima categoria, *pronome*, engloba todos os pronomes - pessoais (“eu”, “mim”), possessivos (“nosso”, “dela”), demonstrativos (“estas”, “aquele”, indefinidos (“alguém”, “quaisquer”), interrogativos (“quais”, “como”) e relativos (“que”, “cujo”). Ressalta-se que o pronome interrogativo “por que” bem como outras locuções não são reconhecidos pelo dicionário. Já a categoria *pronome pessoal* remete aos pronomes pessoais oblíquos (“te”, “lhe”) e retos (“você”, “ele”). As próximas categorias, *eu*, *nós*, *você/tu*, *ele/ela*, *eles/elas* abarcam exclusivamente as referências explícitas que são feitas a esses pronomes (tanto na forma possessiva quanto na oblíqua quanto na reta) nos textos; ou seja, verbos conjugados, como “amaram” e “vendeu”, não são considerados na classificação. Já a categoria *pronome impessoal* engloba os pronomes relativos e indefinidos (“cujo”, “ninguém”, “alguém”, “outro”). A categoria *artigo*, por sua vez, inclui “a(s)”, “o(s)” e “um(a)”, “uns”, “umas”. Além disso, ela inclui a palavra “muito” e outras a ela relacionadas, cujo correspondente em inglês, “*alot*” (correspondente a expressão “*a lot of*”), encontra-se classificado na categoria *article* pois pode denotar a presença de um substantivo subsequente, que é uma das principais funções dos artigos em termos de análise textual.

Por sua vez, a categoria *verbo* inclui todos os tipos de verbos, incluindo regulares (“correr”), auxiliares (“ter”, “poder”), pronominais (“acalmado-se”) e verbos em forma nominal (infinitivo, gerúndio e particípio). É importante destacar que nem todos os verbos do dicionário obtiveram essa classificação; por exemplo, o verbo “fazer” recebeu a classificação, mas o verbo “desfazer”, não. Já a categoria *verbo auxiliar*, além dos principais “ser” e “ter”, inclui os verbos auxiliares modais “dever” e “poder”. A categoria *pretérito* diz respeito aos verbos que estão conjugados no pretérito perfeito (“fiz”), imperfeito (“fazíamos”) e mais-que-perfeito (“fizera”). Similarmente, categoria *presente* refere-se aos verbos conjugados no tempo presente simples (“apoias”) e também os que estão no gerúndio (“faltando”).

Já a categoria *advérbio* engloba todos os tipos de advérbios: de modo (“comumente”), de tempo (“agora”), de lugar (“ali”), de intensidade (“bastante”), de dúvida (“quicá”), de negação (“jamais”) e de afirmação (“certamente”). Por sua vez, a categoria *preposição* inclui

palavras como “acima”, “durante”, “da”, “conforme”. Em *negação* estão inseridos vocábulos como “não”, “nunca”, “absolutamente”, “negativo” e “tampouco”. Em *conjunção* estão presentes vocábulos como “conquanto”, “conforme”, “durante” e “embora”. Observa-se que as divisões entre conjunções, preposições e advérbios não são rígidas, pois alguns termos são classificados nas três categorias ao mesmo tempo, como é o caso de “como”. A categoria *quantificador* abarca expressões que indicam, como é de se esperar, noções de quantidade; por exemplo, “todo”, “seleção”, “sobra”, “singular”, “suficiente”, entre outras. Já em *número* temos expressões mais estritamente numéricas, tais como “dois”, “primeiro”, “zero” e “zilhão”.

Finalmente, a categoria *palavrão* inclui uma série de palavras de baixo calão, consideradas ofensivas ou vulgares, tais como “merda”, “cuzão”, “vaca”, “comer” e até expressões como “puxa” e “putz”. Ressalta-se que o *software* não consegue ler o contexto, motivo pelo qual palavras com sentido duplo, tais como “comer”, podem ser um problema. Também é importante apontar que essa foi a categoria com o maior número de palavras esdrúxulas no que diz respeito à classificação, pois muitas que não consideramos palavrão em português estavam assim classificadas; por exemplo: “contar”, “aberta”, “comum”, “absolver” e “amaldiçoar”. Mais uma vez, são evidenciadas as limitações do levantamento automático.

Todas as categorias, com exceção de *número de palavras* e *palavras por frase*, são dadas em porcentagem de palavras totais e, talvez com a exceção de *artigo*, nenhuma categoria é exaustiva em relação às palavras que contém. Ainda, vale frisar que, embora apresentem alguns elementos anômalos, as categorias contêm, em sua grande maioria, palavras classificadas adequadamente.

3.3 Procedimentos

O primeiro passo para a realização deste projeto foi encaminhar o pedido de autorização para o Comitê de Ética em Pesquisa da UFRGS (CEP/UFRGS). Após a aprovação pelo CEP/UFRGS, o segundo passo foi passar para um formulário do Google Forms as assertivas da Bateria Fatorial de Personalidade para facilitar a computação dos dados. A seguir, foi criada a tarefa de produção textual em outro formulário do Google Forms, que continha instruções para a tarefa bem como a pergunta que serviu de guia para o texto a ser produzido, qual seja “Como você gostaria que a sua vida fosse e que tipo de pessoa você gostaria de ser em 5 anos?”. O próximo passo foi entrar em contato com os possíveis participantes presencial (nas salas de aula da Faculdade de Letras) e virtualmente (em grupos

do Facebook da UFRGS). A seguir, os participantes selecionaram de uma lista com horários pré-definidos aquele que lhe era mais conveniente para realizar as tarefas. No dia e horário escolhidos, os participantes, em grupos de dois a 12 por vez, compareceram a um laboratório de informática do prédio da Faculdade de Letras da UFRGS e realizaram ambas as tarefas consecutivamente em um computador individual.

Em cada sessão, os participantes leram o Termo de Consentimento Livre e Esclarecido (TCLE), e assinaram quando assentiram em participar (não houve ocorrência do contrário). Depois disso, as tarefas foram explicadas a cada participante individualmente, sendo lidas as instruções das tarefas e esclarecidas quaisquer dúvidas remanescentes. Os participantes iniciaram respondendo a Bateria Fatorial de Personalidade no Google Forms, observando as instruções que constavam no manual do instrumento, a saber:

Por favor, leia atentamente cada uma das sentenças e pense o quanto você se identifica com elas. Para cada item, você tem a opção de marcar "1" a "7". Se você acha que a sentença absolutamente não o descreve adequadamente, marque "1". Se você acha que a frase o descreve muito bem, marque o "7". Se você considerar que a frase o descreve "mais ou menos", marque "4". Considere que quanto mais você acha que a frase é apropriada para descrevê-lo, maior deve ser o valor a ser marcado na escala (respostas 5, 6 e 7); quanto menos você identificar-se com a descrição feita, menor será o valor a ser registrado na escala (respostas 1, 2 e 3). Note que todos os valores da escala podem ser marcados.

Os participantes ficaram em torno de 20 minutos nessa primeira tarefa.

Após finalizarem a BFP, os participantes sinalizaram quando estavam por iniciar a próxima tarefa, a de produção textual, e eram informados do horário em que os 30 minutos mínimos necessário para tentativa de realização da tarefa já teriam passado. Por exemplo, se o participante sinalizasse que estava iniciando a tarefa de produção textual às 10h30, ele era informado que deveria tentar escrever até as 11h. A tarefa de produção textual era de formato e tamanho livres, mas uma pergunta-guia foi estipulada, a saber: “Como você gostaria que a sua vida fosse e que tipo de pessoa você gostaria de ser em 5 anos?”. Os participantes então utilizaram um formulário do Google Forms para responder a pergunta, observando as seguintes instruções:

Escreva livremente por pelo menos 30 minutos (no máximo 1 hora) sobre o tema abaixo. Não se preocupe com aspectos gramaticais ou com a estrutura textual. O formato e o tamanho do texto são livres, sendo apenas estipulado um período de 30 minutos para a tentativa da realização da tarefa. Durante a tarefa, por favor, mantenha

o foco (feche outras abas do navegador) e fique em um ambiente silencioso. Você pode escrever seu texto direto no espaço fornecido neste formulário (ele se expandirá automaticamente) ou, se preferir, pode escrever em um programa de edição de texto e depois copiar e colar o conteúdo para este formulário.

Como pode ser inferido pelas instruções, não foram impostas estruturas pré-definidas nem sequer foi determinado o gênero textual, estando os participantes livres para se expressarem do modo mais natural possível dentro das circunstâncias. Similarmente, buscando não causar desconforto aos participantes, foi permitido que terminassem antes de decorridos os 30 minutos estipulados, o que fez com que cerca de vinte por cento deles declarasse ter terminado a tarefa antes desse tempo mínimo. Ainda, aproximadamente vinte por cento dos participantes escolheu continuar escrevendo após o tempo mínimo ter passado, com alguns tendo escrito por até uma hora.

Além das sessões presenciais, alguns participantes (23) realizaram as tarefas em casa, tendo recebido, assinado e entregue o Termo de Consentimento Livre e Esclarecido por e-mail. Também foram instruídos e tiveram suas dúvidas esclarecidas por meio virtual.

De modo geral, os participantes relataram ter gostado das tarefas, demonstrando curiosidade com relação aos resultados do inventário de personalidade e desejo de receber posteriormente os textos que produziram. Os textos foram processados apenas pelo programa LIWC, não tendo sido analisados pelas pesquisadoras. O resultado da BFP e os textos produzidos pelos participantes serão entregues a eles por e-mail.

3.4 Análises de dados

Este projeto adotou um desenho exploratório, quantitativo, de cunho correlacional, e investigou a relação entre os indicadores de personalidade, de acordo com a Bateria Fatorial de Personalidade, as 25 categorias linguísticas processadas pelo *Linguistic Inquiry and Word Count*, e o sexo dos participantes.

Após o final do processo de coleta de dados, os resultados provenientes do inventário de personalidade e da tarefa de produção textual, que estavam no Google Forms, foram transformados em planilhas do Excel. A pontuação da BFP foi calculada manualmente e com a ajuda das funções disponíveis no Excel, com base nas instruções presentes na seção “Aplicação e levantamento da Bateria Fatorial de Personalidade” do manual do instrumento (NUNES; HUTZ; NUNES, 2010). Este descreve dois tipos de procedimentos para a apuração do escore bruto das facetas: um para aquelas que não possuem itens negativos e outro para as

que apresentam itens negativos. No primeiro caso, fazia-se a soma das respostas dadas aos itens da faceta e dividia-se o resultado pelo número de itens respondidos. Assim, o resultado final da faceta sem itens negativos era dado pela fórmula:

$$EB_{\text{faceta}} = \text{SOMA}_{\text{faceta}} / \text{NTIR}$$

Em que: EB_{faceta} = escore bruto na faceta, $\text{SOMA}_{\text{faceta}}$ = resultado da soma das respostas da faceta e NTIR = número total de itens respondidos. Já o segundo procedimento, para facetas com itens negativos, envolvia um processo de cinco etapas: 1) somar as respostas dadas aos itens positivos ($EB_{\text{positivos}}$); 2) somar as respostas aos itens negativos ($EB_{\text{negativos}}$); 3) multiplicar o número de itens negativos respondidos por 8 (INR); 4) calcular o número total de itens respondidos da faceta (NTIR); 5) aplicar a seguinte fórmula para obter o escore bruto final (EB): $EB = EB_{\text{positivos}} + (\text{INR} - EB_{\text{negativos}}) / \text{NTIR}$.

Por sua vez, os fatores eram calculados a partir do escore bruto de cada faceta, somando-se os resultados e dividindo pelo número de facetas do fator. Por exemplo, para calcular o escore bruto de Socialização, somavam-se os escores das facetas S1 - Amabilidade, S2 - Pró-sociabilidade e S3 - Confiança nas pessoas, e dividia-se por três.

Já as respostas da tarefa de produção textual passaram por um processo de “limpeza” seletiva segundo o *Operator's Manual* do LIWC 2015 (PENNEBAKER; BOOTH; BOYD; FRANCIS, 2015), tendo sido unidos alguns termos para que o programa conseguisse classificá-los adequadamente (por exemplo, “sei lá” quando era um expletivo era transformado em “seilá”, tudo junto, de acordo com a classificação do dicionário do LIWC em português). Depois disso, o dicionário em português foi escolhido da base de dados do programa, e os textos foram processados por ele em um só arquivo .xlsx, com as respostas de cada participante correspondendo a uma linha da planilha. Em seguida, utilizando o módulo “Análise textual”, o *Linguistic Inquiry and Word* gerou, para cada um dos textos produzidos, porcentagens de utilização das 64 categorias linguísticas e psicológicas disponíveis no dicionário brasileiro. Dessas categorias, foram selecionadas as 22 que pertenciam à dimensão *Processos linguísticos*, além das variáveis métricas *número de palavras*, *palavras dicionarizadas* e *palavras maiores de seis letras*, somando ao todo 25 variáveis. A seguir, criou-se uma planilha com os resultados de cada faceta/fator da BFP e das 25 variáveis do LIWC de cada participante. Essa planilha foi então aberta no programa SPSS (SPSS INC., 2009) onde foram realizados diversos testes estatísticos.

Primeiramente, realizou-se uma análise descritiva de todas as variáveis, gerando as médias e as curvas de distribuição de cada uma. Foi então observado que a média do número de palavras nos textos era de 491 palavras (DP = 299) e que a amostra não estava normalmente distribuída. Como nem todos os participantes escreveram pela mesma quantidade de tempo, excluímos os textos cujo número de palavras correspondesse a dois desvios padrões acima ou abaixo da média (5), o que normalizou a variável (M = 442, DP = 225) e resultou em um $n = 85$ (48 do sexo feminino), cinco a menos do que o número original. Em seguida, identificamos que outras variáveis não estavam normalmente distribuídas, a saber: as categorias *palavras por frase*, *pronome pessoal*, *nós*, *negação* e *quantificador* do programa LIWC. Para essas variáveis foi aplicada uma transformação logarítmica para se obter normalidade. Ainda, observou-se que as facetas N3 - Passividade, R2 - Ponderação e S2 - Pró-sociabilidade apresentavam assimetria negativa (as pontuações da amostra estavam mais altas que o normal) e que as facetas N4 - Depressão e E2 - Altivez apresentavam assimetria positiva (as pontuações da amostra estavam mais baixas que o normal). Também para essas variáveis foram aplicadas transformações logarítmicas, porém três delas (N3, R2 e S2) permaneceram anormais. Assim, testes não paramétricos foram utilizados com essas facetas.

O resultado do processamento das variáveis linguísticas do LIWC bem como a pontuação dos cinco fatores e das facetas fornecida pela BFP foram comparados em suas médias com base no sexo dos participantes por meio do teste-T de amostras independentes. A pontuação dos fatores da BFP e os resultados do LIWC foram correlacionados por meio do teste de correlação de Pearson, com exceção das facetas assimétricas, que foram testadas com o teste de correlação de Spearman.

4 RESULTADOS

Como pode ser observado na Tabela 2, foram constatadas correlações significativas entre 10 das 25 variáveis linguísticas do LIWC com pelo menos um dos cinco grandes fatores da Bateria Fatorial de Personalidade. As correlações variaram de $r = .217$ ($p = .046$) a $r = .323$ ($p = .003$).

Tabela 2 Correlações (r) entre fatores da BFP e categorias linguísticas do LIWC

Categorias do LIWC	Extroversão	Realização	Neuroticismo	Abertura
palavras > 6 letras	-.221*			
pronomes pessoais			.217*	

eu		.244*	
dicionário	.288**		
palavras por frase			.237*
verbo auxiliar		-.224*	
passado	-.253*		
preposição	.287**		
palavrões	.224*	.215*	
número			-.322**

Fonte: Elaborada pela autora

* $p < 0,05$ ** $p < 0,01$

Mais especificamente, o fator Extroversão apresentou uma correlação negativa marginal ($p < .064$) com a categoria *palavrão* e uma correlação negativa significativa com *palavras maiores de seis letras* ($p = .042$); ou seja, quanto maior a pontuação de um indivíduo no fator Extroversão, menor o uso de palavras maiores de seis letras e menor o uso de palavrões. Já o fator Realização obteve uma correlação positiva com palavras presentes no dicionário do LIWC ($p = .008$) e com preposições ($p = .008$). Isso quer dizer que, quanto mais altas as pontuações em Realização, mais chances a pessoa tinha de fazer uso de preposições e de palavras que eram reconhecidas pelo dicionário do programa de análise textual, cujo critério de seleção era a frequência de uso das palavras. Realização também apresentou uma correlação negativa com o uso do pretérito ($p = .020$); ou seja, quanto mais a pessoa pontuou no fator Realização, menos ela empregou esse tempo verbal. Quanto ao fator Neuroticismo, foram observadas correlações negativas com o uso de verbos auxiliares ($p = .039$) e correlações positivas com as categorias *eu* ($p = .024$), *pronome pessoal* ($p = .046$) e *palavrão* ($p = .048$); isto é, indivíduos com pontuações mais altas em Neuroticismo utilizavam menos verbos auxiliares e mais pronomes pessoais e da primeira pessoa e palavrões. Finalmente, o fator Abertura apresentou uma correlação positiva com a categoria *palavras por frase* ($p = .028$) e foi negativamente relacionado à categoria *número* ($p = .003$); ou seja, quem obteve pontuações altas em Abertura escreveu frases mais longas e fez menos referência a números em seus textos. Pode-se observar que o fator que mais apresentou correlações com as variáveis linguísticas do LIWC foi Neuroticismo, com quatro correlações significativas, enquanto que o fator Socialização não obteve nenhuma correlação. Já a categoria *palavrão* do LIWC foi a variável que mais apresentou correlações (duas) com os fatores da Bateria Fatorial de Personalidade.

Para aprofundar mais o entendimento sobre as relações entre as categorias linguísticas do LIWC e os indicadores de personalidade, as facetas da BFP também foram consideradas, e os testes correlacionais com as variáveis do LIWC demonstraram uma série de resultados

significativos (Tabela 5). As correlações positivas, em que, quanto maior a pontuação na faceta, maior o uso da categoria lingüística, foram as seguintes: 1) categoria *dicionário* – facetas Competência e Empenho ($r = .243$ e $r = .304$, $p = .025$ e $p = .005$, respectivamente); ou seja, quem obteve pontuações altas nas facetas Competência e/ou Empenho teve mais chance de usar palavras que eram reconhecidas pelo dicionário do LIWC; 2) *preposição* – Empenho ($r = .268$, $p = .013$); isto é, pontuações altas na faceta Empenho previam maior uso de preposições; 3) *pronome pessoal* – Busca por Novidades ($r = .227$, $p = .037$): quanto mais altas as pontuações em Busca por Novidades, mais chance de utilizar pronomes pessoais; 4) *nós* – Liberalismo ($r = .922$, $p = .026$, $n = 5$): pontuações altas nessa faceta estavam relacionadas a maior uso de pronome da primeira pessoa do plural, mas apenas 5 sujeitos obtiveram pontuações válidas após a transformação logarítmica da variável, sendo a amostra reduzida consideravelmente; 5) *eu* – Passividade e Depressão ($r = .240$, $r = .216$, $p = .027$ e $p = .047$, respectivamente); isto é, quem obteve altas pontuações em Passividade e/ou Depressão usou mais a primeira pessoa do singular, lembrando que a faceta Passividade foi correlacionada por meio do teste de correlação não paramétrico de Spearman, pois não estava normalmente distribuída; 6) *número de palavras* – Instabilidade Emocional ($r = .238$, $p = .028$); ou seja, pessoas altas em Instabilidade Emocional tendiam a escrever textos mais longos; 7) *palavrão* – Depressão ($r = .239$, $p = .028$): indivíduos com altas pontuações em Depressão utilizaram mais palavras da categoria *palavrão*; 8) *pronome impessoal* – Amabilidade ($r = .223$, $p = .040$): quem pontuou mais nessa faceta teve tendência a usar mais pronomes impessoais; 9) *verbo auxiliar* – Dinamismo ($r = .231$, $p = .033$): quanto maior a pontuação em Dinamismo, maior a chance de usar verbos auxiliares; e 10) *conjunção* – Interações Sociais ($r = .240$, $p = .040$); ou seja, quem obteve altas pontuações nessa faceta teve maior probabilidade de utilizar conjunções. Também foram encontradas relações negativas, ou seja, aquelas em que, quanto mais alta a pontuação na faceta, menor o uso das categorias lingüísticas, a saber: 11) entre a categoria *ele/ela* e a faceta Empenho ($r = -.218$, $p = .045$); isto é, quanto maior a pontuação em Empenho, menor a chance de fazer referências à terceira pessoa do singular; 12) entre *pretérito* e Competência ($r = -.302$, $p = .005$): quem pontuou mais em Competência usou menos o tempo pretérito; 13) entre *número*, Liberalismo e Busca por Novidades ($r = -.234$, $r = -.243$, $p = .031$ e $p = .025$, respectivamente): indivíduos com pontuações mais altas em Liberalismo e/ou Busca por Novidades fizeram menos referências a números; 14) entre *número de palavras* e Busca por Novidades ($r = -.230$, $p = .034$); ou seja, quanto maior a pontuação nessa faceta, maior a chance de escrever um texto mais curto; 15) entre *palavrão* e Dinamismo ($r = -.245$, $p = .024$); isto é, pessoas com altas

pontuações nessa faceta utilizaram menos palavras; e 16) entre *palavras maiores de seis letras* e Interações Sociais ($r = -.276, p = .011$): quanto maior a pontuação nessa faceta, menos a pessoa utilizou palavras maiores de seis letras. Como pode ser visto, metade dessas correlações também foram observadas em nível de fator (1, 2, 5, 7, 12, 13, 15, 16) e metade limitaram-se às facetas (3, 4, 6, 8, 9, 10, 11, 14), o que mostra a importância de se considerar a personalidade nesse nível de resolução mais elevado.

Tabela 3 – Médias masculinas e femininas nos fatores da BFP

Fator	Sexo	N	Média	Desvio padrão
Abertura	F	48	4,86	,666
	M	37	5,06	,729
Realização	F	48	4,74	,739
	M	37	4,67	,908
Socialização	F	48	5,04	,627
	M	37	4,91	,729
Extroversão	F	48	3,85	1,005
	M	37	3,72	,971
Neuroticismo	F	48	4,18	,870
	M	37	3,88	1,113

Fonte: Elaborada pela autora

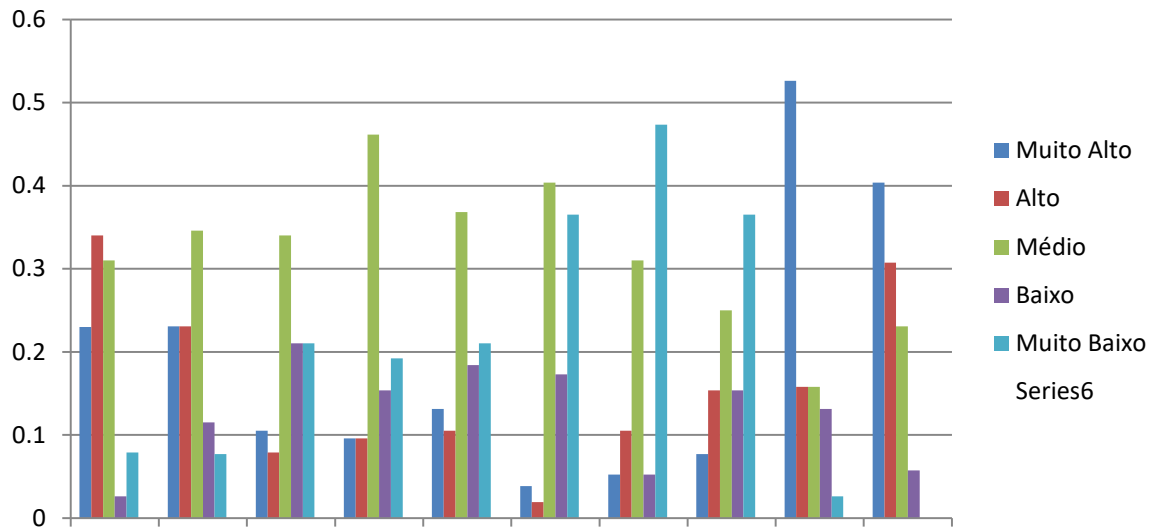
Tabela 4 – Médias masculinas e femininas nas facetas da BFP

Faceta	Sexo	N	Média	Desvio Padrão
Vulnerabilidade	F	48	4,43	,202
	M	37	3,91	,235
Instabilidade Emocional	F	48	4,54	,178
	M	37	3,65	,247
Passividade	F	48	4,55	,173
	M	37	4,73	,218
Depressão	F	48	3,21	,145
	M	37	3,22	,203

Comunicação	F	48	3,80	,216
	M	37	3,67	,223
Ativez	F	48	3,56	,144
	M	37	3,45	,180
Dinamismo	F	48	4,13	,164
	M	37	3,96	,217
Interações Sociais	F	48	3,90	,164
	M	37	3,80	,202
Amabilidade	F	48	5,49	,111
	M	37	5,42	,110
Pró-sociabilidade	F	48	5,26	,132
	M	37	4,87	,176
Confiança nas pessoas	F	48	4,37	,150
	M	37	4,45	,163
Competência	F	48	4,44	,126
	M	37	4,48	,180
Ponderação	F	48	5,04	,183
	M	37	5,21	,194
Empenho	F	48	4,74	,157
	M	37	4,31	,192
Abertura a idéias	F	48	5,06	,143
	M	37	5,45	,152
Liberalismo	F	48	5,34	,117
	M	37	5,48	,125
Busca por novidades	F	48	4,19	,153
	M	37	4,26	,215

Fonte: Elaborada pela autora

Figura 3 – Pontuações dos fatores da BFP



Fonte: Elaborada pela autora

Como pode ser observado na Figura 3, as pontuações da amostra geral demonstraram níveis predominantes de pontuações muito altas em Neuroticismo para ambos os sexos, níveis altos de Abertura para o sexo masculino e níveis muito baixos de Extroversão para ambos os sexos. Já os fatores Socialização e Realização demonstraram maior prevalência de pontuações médias para ambos os sexos. Apesar de apresentarem valores brutos diferentes, o teste-T de amostras independentes, utilizado para comparar as médias de cada grupo, não verificou diferenças estatisticamente significativas no nível dos cinco grandes fatores (Tabela 3); ou seja, na amostra em questão, constituída em sua maioria por estudantes do curso de Letras brasileiros com média de idade de 27,3 anos (DP = 8,5), os indicadores de personalidade segundo a BFP de homens e mulheres não se diferenciou de modo significativo no nível dos fatores. Entretanto, houve diferença significativa ($t(83) = 2,9, p = .004$) na faceta N2 – Instabilidade Emocional, com médias superiores do sexo feminino (ver Tabela 4 para os valores das médias em cada faceta). Além disso, foram observadas diferenças marginalmente significativas ($p = .093, p = .067, p = .084$, respectivamente) nas facetas N1 – Vulnerabilidade ($t(83) = 1,69, p = .093$), com médias superiores para o sexo feminino; A1 – Abertura a idéias ($t(83) = - 1,85, p = .067$), com médias maiores para o sexo masculino; e R3 – Empenho ($t(83) = 1,7, p = .084$), com médias superiores para o sexo feminino.

Quanto às diferenças no uso da língua (Tabela 6), tais como analisado pelo programa LIWC, foram encontrados valores significativos no teste-T de amostras independentes em relação ao sexo dos participantes. Os homens em geral apresentaram médias superiores nas

categorias *palavras maiores de seis letras* ($t(83) = -2,7, p = .041$); *você/tu* ($t(83) = -2,8, p = .006$); *ele/ela* ($t(83) = -2,5, p = .013$); *preposição* ($t(83) = -1,7, p = .089$) e *artigo* ($t(83) = -2,7, p = .008$). Já as mulheres tiveram médias maiores no uso de verbos ($t(83) = 3,1, p = .003$), tempo verbal presente ($t(83) = -1,7, p = .089$) e palavras reconhecidas pelo dicionário do LIWC ($t(83) = 1,7, p = .086$).

Tabela 5 – Correlações entre facetas da BFP e categorias do LIWC

	Depressão	Instabilidade Emocional	Passividade***	Dinamismo	Interações Sociais	Amabilidade	Competência	Empenho	Liberalismo	Busca por Novidades
Dicionário							.243*	.304**		
Ele/ela								-.218		
Verbo auxiliar				.231*			.282**			
Pretérito							-.302**			
Preposição								.268*		
Número									-.234*	-.243*
Número de palavras		.238*								-.230*
Pronome pessoal										.227*
Nós									.922*	
Eu	.216*		.240*							
Palavrão	.239*			-.245*						
Palavras > 6 letras					-.276*					
Pronome impessoal						.223				
Conjunção					.223*					

Fonte: Elaborada pela autora

* $p < .05$; ** = $p < .01$; *** = Rô de Spearman

Tabela 6 - Médias masculinas e femininas nas categorias linguísticas do LIWC

LIWC	Média [M]	Desvio Padrão [M]	Média [F]	Desvio Padrão [F]
Preposição	17,18	3,23	16,10	2,54
Palavras > 6 letras	24,58	4,85	22,63	3,84
Dicionário	84,74	3,28	86,04	3,50
Você	5,46	1,45	4,62	1,26
Ele/ela	5,10	1,59	4,33	1,18
Artigo	7,51	1,50	6,55	1,70
Verbo	14,78	3,20	16,90	3,05
Presente	9,24	2,65	10,37	2,51

Fonte: Elaborada pela autora

5 DISCUSSÃO

Este estudo teve como objetivo explorar as possíveis relações existentes entre indicadores da Bateria Fatorial de Personalidade, que segue o modelo dos Cinco Grandes Fatores (CGF) (COSTA; MCCRAE, 1992), o sexo dos participantes e aspectos linguísticos de autonarrativas. Foram observadas relações significativas entre aspectos linguísticos, tais como medidos pelo programa LIWC com o dicionário português, e os cinco grandes fatores de personalidade medidos pela Bateria Fatorial de Personalidade (NUNES; HUTZ; NUNES, 2010). Este trabalho, até onde sabemos, é o primeiro a utilizar essa metodologia, baseando-se em autonarrativas e considerando aspectos linguísticos dos textos por meio do *Linguistic Inquiry and Word Count* para relacioná-los aos CGF. Outros estudos no exterior já haviam utilizado o LIWC para analisar outros tipos de materiais, tais como transcrições de fala (MEHL; GOSLING; PENNEBAKER, 2006), fluxos de consciência e ensaios acadêmicos (LEE; KIM; SEO; CHUNG, 2007; PENNEBAKER; KING, 1999). Além disso, Hirsh e Peterson (2009) utilizaram autonarrativas (*self-narratives*) e o LIWC em seu estudo, mas só consideraram as categorias psicológicas do programa, ignorando as linguísticas.

Parte dos achados deste estudo vão ao encontro do que foi reportado em estudos passados, em sua grande maioria com sujeitos falantes de inglês. Por exemplo, a correlação negativa do fator Extroversão com palavras maiores de seis letras já havia sido reportada por

Mehl, Gosling e Pennebaker (2006), quando analisaram transcrições da fala de estudantes americanos. De acordo com Tausczik e Pennebaker (2010), esse achado pode ser explicado ao se entender que altas pontuações em Extroversão indicam indivíduos com menos inibição, e que essa falta de inibição também se manifesta na produção linguística da pessoa, o que pode levar a uma linguagem menos complexa, com palavras mais curtas. Além disso, palavras mais curtas são geralmente mais comuns, pelo menos na língua inglesa, e caracterizam um discurso mais “amigável” na medida em que são mais fáceis de compreender. Essa facilidade em ser compreendido pode ser pensada como uma das prioridades de indivíduos com altas pontuações em Extroversão, por esses valorizarem muito as relações interpessoais. De fato, à inspeção mais aprofundada, a faceta Interações Sociais, que descreve uma tendência a buscar mais contato com outras pessoas e a manter boas relações, foi a única desse fator que apresentou correlações negativas significativas com a variável *palavras maiores de seis letras*. Por outro lado, a correlação negativa da categoria *palavrão* com o fator Extroversão parece estar mais atrelada a características descritas pela faceta Dinamismo, já que ela foi a única com a qual houve correlações significativas. De acordo com a BFP, essa faceta corresponde ao nível de atividade e iniciativa do indivíduo; assim, quem toma menos iniciativa e tem um nível de atividade menor usa mais palavrões, uma relação cujas razões últimas permanecem incógnitas. De qualquer modo, a relação negativa entre o uso de palavrões e o fator Extroversão já havia sido reportada por Yee, Harris, Jabon, e Bailenson (2011) e quiçá possa ser entendida como um reflexo do comportamento pró-social embutido em pontuações altas nesse fator como um todo. Ainda, é importante lembrar que as correlações encontradas no estudo presente foram baseadas na classificação do dicionário em português do LIWC, o qual, como já apontado em seção anterior, contém palavras anômalas especialmente na categoria *palavrão*. Assim, deve-se ter reserva ao tirar quaisquer conclusões sobre esse achado.

Outro resultado congruente com o reportado na literatura tem a ver com o maior uso de primeira pessoa por pessoas altas em Neuroticismo. Por exemplo, Pennebaker e King (1999) já haviam encontrado essa relação em seu estudo utilizando fluxos de consciência e ensaios acadêmicos, e Yarkoni (2010) reproduziu esse achado em seu estudo de *blogs*. De acordo com o manual da Bateria Fatorial de Personalidade (NUNES et al., 2010), o fator Neuroticismo está relacionado a níveis de depressão, ansiedade, vulnerabilidade e instabilidade emocional. Quando analisado em nível de faceta, o uso maior de primeira pessoa está associado de modo significativo apenas a Depressão e Passividade, não sendo as relações com Vulnerabilidade e Instabilidade Emocional estatisticamente significativas. A faceta

Depressão indica o quanto um indivíduo indica estar desesperançoso em relação ao futuro, o quão vazia ele acha que sua vida é, e o quão difíceis as tarefas cotidianas lhe parecem. Já a faceta Passividade está relacionada a problemas em iniciar e terminar tarefas e a dificuldades em tomar decisões. A razão pela qual o fator Neuroticismo e algumas de suas facetas relacionam-se a um maior uso de primeira pessoa do singular tem a ver com o maior nível de foco em si mesmo ou de autocentrismo (*self-involvement*), que é comum em sintomas depressivos e ansiosos, os quais caracterizam indivíduos com altas pontuações nesse fator (TAUSCZIK; PENNEBAKER, 2010; PENNEBAKER; MEHL; NIEDERHOFFER, 2003). Similarmente, a relação positiva entre Neuroticismo e *palavrão* já havia sido reportada na literatura (YARKONI, 2010). Essa relação está ligada à faceta Depressão, a única a apresentar correlações positivas com a variável do LIWC e cujas características estão descritas acima. Os motivos para essa relação entre Neuroticismo e uso de palavras têm sido associados a maiores níveis de irritabilidade, impaciência e hostilidade, na literatura (JAY, 2009), porém aqui parecem estar ligados mais à falta de esperança, sentido e objetivos claros na vida, que são típicas em indivíduos com altos escores em Depressão.

Já a correlação negativa encontrada entre Neuroticismo e verbos auxiliares não havia sido descrita anteriormente, consistindo em um achado sobre cuja explicação podemos apenas especular. De acordo com Langacker (2008), os verbos auxiliares possuem caráter esquemático, não apresentando conteúdo lexical. Ademais, de acordo com Tausczik e Pennebaker (2010), essas partículas sinalizam linguagem informal e uso de voz passiva. Assim, pode-se pensar que os textos de indivíduos com pontuações mais altas em Neuroticismo contiveram menos palavras indicando relações esquemáticas, com menor uso de construções formais e voz passiva. É válido notar que, apesar de significativa em nível de fator, essa correlação não foi significativa em nível de faceta, havendo apenas relações marginais.

Quanto às relações encontradas entre variáveis linguísticas do LIWC e o fator Realização, duas encontram-se descritas na literatura. A primeira diz respeito ao uso de preposições e descreve uma relação positiva com Realização; ou seja, quanto maior a pontuação em Realização, maior o uso de preposições, lembrando que Realização descreve o nível de ambição, organização, meticulosidade, persistência e ponderação do indivíduo (NUNES; HUTZ, 2002). Já em 2007, Mairesse, Walker, Mehl e Moore descreviam uma relação positiva entre esse fator e a categoria *preposição* do LIWC, um achado replicado por Yarkoni (2010). Entretanto, as razões para tal correlação não são evidentes. Na concepção da gramática cognitiva (LANGACKER, 2008), as preposições indicam uma relação não

processual entre duas coisas e podem indicar localização ou instrumentalização; por exemplo: “Ele nasceu *em* Agosto” ou “Ele fez o furo *com* uma furadeira”. Ao se inspecionar mais de perto, percebe-se que a faceta Empenho, a qual descreve o quão detalhista, exigente e meticulosa a pessoa é em relação às tarefas que faz, foi a única que apresentou relações significativas com a variável do LIWC. Considerando essas características, é plausível conceber que o uso de preposições pode estar relacionado a um maior nível de detalhe e concretude na descrição dos planos futuros, que era o tema da tarefa de produção textual que os participantes fizeram para este estudo. De fato, Tausczik e Pennebaker (2010) afirmam que o uso de preposições, entre outros fatores, é indicativo de concretude e detalhismo no discurso. Por exemplo, é possível dizer apenas “Ela saiu”, ou pode-se dizer “Ela saiu do quarto”, o que já confere uma concretude maior à imagem evocada no leitor. A outra correlação encontrada também foi positiva e tem a ver com a variável *dicionário*, que corresponde às palavras presentes no dicionário do LIWC. À examinação mais minuciosa, percebe-se que tanto a faceta Empenho quanto a faceta Competência foram relacionadas com essa variável. A última descreve um comportamento ativo na busca de objetivos, com disposição a fazer sacrifícios para alcançar os resultados esperados. Para entendermos essa correlação, é necessário lembrar que o dicionário foi desenvolvido com base nas palavras mais comuns da língua inglesa. Desse modo, a linguagem utilizada por indivíduos com altas pontuações no fator Realização, especificamente em relação às facetas Competência e Empenho, tende a conter menos palavras raras, ou uma linguagem mais “pé-no-chão”, o que tem certa consistência com as características do fator e das facetas correlacionadas. A relação entre Realização e a categoria *dicionário* também foi encontrada por Mairesse e colegas (2007).

Além desses resultados, que já haviam sido reportados na literatura, verificamos uma correlação negativa com o tempo verbal pretérito e o fator Realização, que não foi relatada em pesquisas passadas. Desta vez, a faceta Competência foi a única a obter correlações significativas em nível inferior aos fatores. Apesar de ser uma relação inédita, não é difícil encontrar suas possíveis raízes. Por exemplo, pode-se pensar que, devido ao fato de Realização estar ligado a maiores níveis de ambição, em especial no que diz respeito à faceta Competência, que descreve a clareza de objetivos do indivíduo e a sua capacidade de fazer sacrifícios para futuro, os textos escritos por indivíduos com altas pontuações nesse fator conteriam menos referências ao passado e mais ao futuro. De fato, Mairesse e colegas (2007) descrevem que existe uma correlação positiva entre Realização e o uso de tempo verbal futuro. Cabe lembrar que, neste trabalho, a variável *futuro* não foi considerada, por conter

muitas divergências em relação à estrutura linguística da categoria original, em inglês, e a traduzida, em português.

Finalmente, o fator Abertura apresenta duas correlações já descritas na literatura. A primeira diz respeito à categoria *palavras por frase*, que apresentou uma relação positiva com esse fator, mas não apresentou relações significativas com suas facetas. Assim, quanto mais alta a pontuação em Abertura, maiores eram as frases escritas pelos participantes. A mesma relação em nível de fator encontra-se reportada em Mairesse e colegas (2007). Para auxiliar na interpretação do resultado, é pertinente recordar que Abertura descreve o nível de liberalismo, interesse por ideias abstratas, curiosidade, imaginação, criatividade e pensamentos não convencionais das pessoas (NUNES et al., 2010), o que talvez contribua para sentenças mais longas devido a essa riqueza associativa mental. Já a segunda correlação diz respeito ao uso de números, sendo negativamente relacionada à Abertura e a duas de suas facetas, quais sejam Liberalismo e Busca por Novidades. Assim, indivíduos com pontuações mais altas nesse fator e nessas facetas utilizaram menos referências a números em seus textos. A relação com o fator já havia sido descrita por Yarkoni (2010). Essa correlação talvez pudesse ser lida à luz do fato de números serem elementos que conferem especificidade e concretude ao discurso, o que não condiz com a descrição do fator Abertura, que, em suas pontuações mais altas, descreve indivíduos inclinados a abstrações. Entretanto, as relações em nível de faceta demonstram que os aspectos mais determinantes para se utilizar menos números estão ligados às facetas Liberalismo, que aponta para relativismo moral e abertura a valores, e Busca por Novidades, que descreve baixa tolerância a atividades rotineiras ou repetitivas e ânsia por vivenciar eventos novos. Ou seja, o menor uso de números parece estar mais relacionado a essas características do que a propensões a abstrações em si.

Além disso, foram encontradas relações em nível de facetas que não apareceram no nível dos fatores. Por exemplo, a faceta Busca por Novidades teve relação negativa com o número total de palavras escritas pelos participantes ($p = .034$) e relação positiva ($p = .037$) com o uso de pronomes pessoais, mas o fator Abertura, da qual ela faz parte, não apresentou essas relações. Da mesma forma, a faceta Competência apresentou relação positiva com o uso de verbos auxiliares ($p = .009$), o que não apareceu no fator Realização, do qual ela faz parte. Ainda, a faceta Liberalismo, que faz parte de Abertura, teve uma relação positiva com a variável *nós* ($p = .026$), mas essa relação não foi transferida para o nível do fator. Em relação à faceta Instabilidade Emocional, foi encontrada uma relação positiva ($p = .028$) entre sua pontuação e o número total de palavras escritas pelos participantes, sem que essa relação fosse visível no nível do fator Neuroticismo, ao qual a faceta pertence. Já Dinamismo obteve uma

correlação positiva com o uso de verbos auxiliares ($p = .033$), sem que isso se refletisse no fator no qual ela se insere, qual seja Extroversão. Ainda, a faceta Interações Sociais apresentou relações positivas com a categoria *palavras maiores de seis letras* ($p = .011$) e *conjunção* ($p = .040$). Finalmente, foi observada correlação positiva entre *pronome impessoal* ($p = .040$) e Amabilidade, uma faceta do fator Socialização, que foi o único a não apresentar nenhuma relação com as variáveis do LIWC. Devido ao fato de os estudos passados terem analisado apenas as relações entre aspectos linguísticos e os Cinco Grandes Fatores, sem conter detalhes sobre as relações entre facetas, não sabemos se relações semelhantes foram encontradas ou não. De qualquer forma, caso positivo, elas também não foram transferidas para o nível dos fatores, pois nenhuma relação foi observada na literatura com as variáveis linguísticas e os fatores às quais as facetas que apresentaram correlações pertencem. Por exemplo, a relação com o uso de verbos auxiliares, que apareceu na faceta Competência, do fator Realização, não foi constatada com o fator Realização em pesquisas anteriores, e assim sucessivamente com as facetas supramencionadas.

É importante salientar que, apesar de a maioria dos achados corroborar o que havia sido descrito em estudos anteriores, as correlações variam dependendo de uma série de fatores, incluindo o tipo de teste de personalidade aplicado, a ferramenta utilizada para extração dos aspectos linguísticos, o perfil dos participantes e o tipo de texto analisado. Desse modo, é possível encontrar na literatura uma variedade de correlações que não foram reportadas no presente trabalho. Algumas delas possuem raízes que são difíceis de identificar; por exemplo, a relação negativa entre o uso de artigos e o fator Extroversão, encontrada por Pennebaker e King (1999) e Mairesse e colegas (2007); ou a relação negativa entre o uso de artigos e o uso de palavras e o fator Socialização, encontrada por Mehl e colegas (2006); ou ainda a relação negativa entre o uso de artigos e a relação positiva entre negações e tempo verbal presente e o fator Neuroticismo, encontradas por Mairesse e colegas (2007); ou, por fim, a relação negativa com o tempo verbal pretérito e a relação positiva com o uso de artigos e preposições e o fator Abertura, encontradas por Yarkoni (2010) e Nowson (2006). Para nenhuma dessas relações foram encontrados motivos aparentes que justificassem o seu não aparecimento no estudo presente. Já outras podem ser explicadas pelas limitações do dicionário em português do LIWC; por exemplo, a relação entre pronomes de primeira pessoa do singular e o fator Extroversão, que apareceu em Holtgraves (2011), pode não ter sido identificada no presente trabalho porque o dicionário não é capaz de identificar as menções implícitas aos sujeitos que são feitas pela conjugação dos verbos em português, limitando-se à identificação de pronomes que estão explícitos no texto. O mesmo pode ser dito em relação

aos achados que relacionam a primeira pessoa do singular ao fator Socialização (MEHL et al., 2006; YARKONI, 2010) e aos que relacionam a primeira pessoa do plural de forma negativa ao fator Abertura (PENNEBAKER; KING, 1999; YARKONI, 2010). Ainda, alguns estudos encontraram relações positivas entre o uso de palavras maiores de seis letras e os fatores Socialização (MEHL et al., 2006; NOWSON, 2006) e Abertura (MEHL et al., 2006; PENNEBAKER; KING, 1999), o que não foi observado no trabalho presente e que poderia ser explicado pelas diferenças formais das línguas inglesa e portuguesa, já que a última apresenta muito mais palavras longas. Assim, apesar de, em geral, as palavras mais longas serem mais complexas e indicarem palavras menos comuns, por as palavras em português serem mais longas do que as em inglês, que em sua maioria são monossilábicas, esse limiar de tamanho talvez tivesse que ser maior no dicionário em português para se poder encontrar o mesmo efeito.

Por fim, é pertinente apontar que, nos estudos citados, foi encontrada pouquíssima ou nenhuma explicação para as correlações apresentadas, o que evidencia o caráter relativamente hermético das categorias linguísticas estritas, como preposições, conjunções e artigos, em suas relações com os traços da personalidade; por outro lado, as categorias psicológicas, tais como palavras com conotações positivas ou palavras que se relacionam à família ou a amigos, são muito mais transparentes e são descritas de modo mais aprofundado na literatura. Por isso, recomenda-se cautela na consideração das explicações fornecidas acima, pois são necessárias mais investigações a respeito da natureza dessas correlações antes de se poder afirmar algo categoricamente.

Em relação aos indicadores de personalidade da BFP em homens e mulheres, foram observadas diferenças significativas nas médias de apenas uma faceta, com todos os fatores apresentando médias semelhantes (Tabela 5). A faceta em questão, Instabilidade Emocional, avalia o quanto as pessoas se descrevem como nervosas, irritáveis e com variações de humor (NUNES et al., 2010) e teve em média uma pontuação mais alta para o sexo feminino (ver Tabela 6 para as médias de todas as facetas). Além disso, houve diferenças marginalmente significativas nas facetas Vulnerabilidade, com médias superiores para o sexo feminino, Abertura a idéias, com médias maiores para o sexo masculino, e Empenho, com médias maiores para o sexo feminino. A primeira delas reflete uma menor autoestima, maior insegurança, dependência e fragilidade emocional (NUNES et al., 2010), no caso desse estudo, por parte das mulheres; a segunda faceta, Abertura a idéias, diz respeito aos níveis de curiosidade, imaginação, fantasia e propensões filosóficas e artísticas (NUNES et al., 2010), que nesse estudo foram maiores para os homens; já a faceta Empenho está relacionada a

níveis de detalhismo, perfeccionismo, dedicação e exigência com trabalhos realizados, que foram maiores para as mulheres neste estudo. Ademais, apesar de não significativas estatisticamente, as médias do sexo feminino foram superiores nos fatores Neuroticismo, Extroversão, Socialização e Realização (Tabela 5), o que vai ao encontro do que foi levantado por análises transculturais, onde mulheres apresentaram maiores pontuações médias exatamente nesses fatores (SCHMIT; VORACEK; REALO; ALLIK, 2008; WEISBERG; DEYOUNG; HIRSH, 2011), só que com valores significativos estatisticamente. Já os homens apresentaram médias maiores, ainda que não significativas, em Abertura. Em relação a esse fator, as diferenças entre os sexos não têm sido consistentes na literatura, sabendo-se apenas que os homens, em geral, apresentam médias maiores na faceta Abertura a idéias, o que também foi o caso neste estudo, que demonstrou uma diferença marginal ($t(83) = 1,8, p = .067$) entre homens e mulheres nessa faceta. De qualquer forma, os estudos destinados a encontrar diferenças entre os sexos nos indicadores de personalidade segundo o modelo dos CGF têm reportado efeitos pequenos a moderados, sendo as diferenças mais robustas presentes nos níveis das facetas, o que é problemático porque as divisões das facetas variam de acordo com o instrumento utilizado.

Em comparação às médias dos fatores na amostra brasileira, reportadas em Nunes, Hutz e Nunes (2010), as deste estudo foram maiores nos fatores Neuroticismo e Abertura e menores nos fatores Realização, Socialização e Extroversão, tanto para homens quanto para mulheres. Para fazer sentido desses achados, é importante considerar que a amostra em questão consistiu quase exclusivamente (90%) em estudantes de cursos de humanas, pois, como mostra a revisão sistemática de Vedel (2016), o curso universitário também reflete perfis específicos de traços da personalidade. No caso das humanidades, os alunos tendem a possuir níveis mais altos de Abertura, Neuroticismo e Socialização bem como níveis mais baixos de Realização e Extroversão. Com exceção do resultado referente ao fator Socialização, o que consiste em uma questão a ser investigada em pesquisas futuras, todos os demais refletem esse padrão. Por fim, é essencial ter em mente que as diferenças entre grupos existem, mas que há grandes variações intragrupais, o que impede que qualquer pessoa seja pré-julgada com base exclusiva no seu pertencimento a um grupo determinado; ou seja, muitas pessoas não se encaixam no padrão de personalidade descrito para o seu sexo, curso acadêmico ou qualquer outro tipo de grupo.

Em relação às diferenças entre os sexos no que se refere às categorias linguísticas do LIWC, foram encontradas maiores porcentagens de uso pelos homens nas categorias *você*, *ele/ela*, *preposição*, *artigo* e *palavras maiores de seis letras*. As três últimas categorias

apresentaram maior uso por indivíduos do sexo masculino na metanálise de mais de 14 mil textos escritos por americanos e outros falantes de inglês (NEWMAN, GROOM, HANDELMAN E PENNEBAKER, 2008). Além disso, essas categorias (*preposição, artigo e palavras maiores de seis letras*) são marcadores de linguagem complexa, também tipicamente associada ao sexo masculino (TAUSCZIK; PENNEBAKER, 2010). Além disso, o uso de artigos e palavras longas está associado a discursos mais formais e com menos envolvimento social, algo também relacionado aos homens (PENNEBAKER et al., 2003) Já o uso de segunda e terceira pessoa tem sido associado a mulheres (NEWMAN et al., 2008), o oposto do que se verificou neste trabalho. Considerando que o dicionário capta apenas referências explícitas aos sujeitos do discurso, é possível que este achado reflita um maior uso *explícito* desses pronomes por parte do sexo masculino, mas não um uso maior no total. De qualquer forma, no estudo presente, por razões que não são claras, os homens utilizaram mais a segunda e a terceira pessoa do singular, demonstrando maior interação aparente com o interlocutor e maior inclusão explícita de terceiros nas suas narrativas pessoais.

Por sua vez, as mulheres empregaram mais verbos, tempo verbal presente e palavras reconhecidas pelo dicionário LIWC. O maior uso de tempo verbal presente pelo sexo feminino já havia sido reportado na literatura (NEWMAN et al., 2008) e tem sido considerado um indicador de maior interação social (PENNEBAKER ET al., 2003). Os outros dois achados não foram diretamente relatados na literatura; entretanto, considerando que homens tendem a usar palavras mais longas (que, em inglês, língua do dicionário original, tendem a ser mais incomuns), faz sentido que as mulheres apresentem textos com maior proporção de palavras captadas pelo LIWC, devido ao fato de seu critério de seleção ser basicamente a frequência de uso das palavras (PENNEBAKER et al., 2001). Por sua vez, o uso de verbos não foi relacionado ao sexo masculino, mas também não apareceu significativamente maior no sexo feminino em estudos anteriores. Para entender melhor esse resultado, é válido lembrar que na categoria *verbos* estão incluídos os verbos modais (que não possuem categoria própria), e que foram relacionados ao sexo feminino na literatura (XIA, 2013).

Já em relação ao tipo de texto escrito pelos participantes, qual seja o das narrativas pessoais, percebeu-se que ele foi suficiente, ainda que talvez não ideal, para a detecção de diferenças em uso linguístico com base na personalidade e no sexo dos participantes. Não existe na literatura um estudo que tenha usado esse tipo de texto e que tenha considerado as variáveis linguísticas do LIWC para se fazer uma comparação; entretanto, os resultados descritos neste estudo apontam para a adequação das narrativas pessoais (*self-narratives*) para a detecção de diferenças no uso da linguagem com base no sexo e nos perfis de personalidade,

pois puderam reproduzir achados anteriores e apresentaram correlações significativas. Além disso, constatou-se que os vocábulos gramaticais, tais como pronomes, preposições e verbos auxiliares, podem, de fato, conter informações de cunho psicológico e social, corroborando as observações feitas por diversos pesquisadores, com destaque para James Pennebaker (2011).

Por fim, cabe apontar que este estudo conteve limitações. Em primeiro lugar, a amostra da população foi constituída, em sua maioria, de pessoas provenientes do mesmo curso universitário, o que, em se tratando de estudos envolvendo a personalidade, pode ser um problema, já que a escolha do curso acadêmico tem sido relacionada a perfis de personalidade específicos (VEDEL, 2016). Em segundo lugar, a ferramenta que serviu de base para a análise linguística, qual seja o dicionário do programa LIWC, apresenta limitações relacionadas a não correspondência de aspectos estruturais do inglês, língua do dicionário original, com o português, tais como o fato de os verbos conjugados não serem identificados em suas referências aos sujeitos ou em suas referências ao tempo verbal futuro. Por fim, foram realizados múltiplos testes correlacionais sem que fossem feitas correções do nível de significância estatística, podendo, assim, haver maior risco de erro Tipo I, onde a hipótese nula é erroneamente rejeitada. Por outro lado, uma correção desse tipo levaria a maior risco de erro Tipo II, onde a hipótese nula é erroneamente aceita.

6 CONCLUSÃO

Existem evidências de que o comportamento verbal de um indivíduo pode variar em função dos traços de personalidade (de acordo com o modelo dos Cinco Grandes Fatores) e de seu sexo (HIRSH; PETERSON, 2009; NEWMAN; GROOM; HANDELMAN; PENNEBAKER, 2008; MAIRESSE; WALKER; MEHL; MOORE, 2007; PENNEBAKER; MEHL; NIEDERHOFFER, 2003). Um veículo pelo qual esse comportamento se manifesta é o texto escrito, e um tipo de texto que pode fornecer *insight* sobre aspectos psicológicos do indivíduo é o da narrativa pessoal. Ferramentas de análise textual automatizadas têm auxiliado no processamento desses textos, possibilitando uma abordagem baseada na contagem de palavras, a qual confere uma perspectiva complementar à feita por juízes humanos.

De modo geral, este trabalho contribuiu para a aproximação da análise linguística com os estudos sobre a personalidade e as diferenças entre os sexos. Longe de apresentar resultados conclusivos, a pesquisa aponta diversos pontos para investigação futura. Um deles diz respeito às diferenças de pontuações da Bateria Fatorial de personalidade entre os sexos, pois, devido às características da amostra em questão, constituída principalmente por

estudantes de Letras, pode ser inapropriado estender os achados à população geral. Outro tema que deve ser melhor investigado consiste no uso de elementos linguísticos por diferentes perfis de personalidade, já que as descobertas deste estudo, por serem baseadas nas classificações do dicionário em português do LIWC, contêm limitações. Assim sendo, elas devem ser replicadas futuramente utilizando ferramentas mais precisas. Por fim, devem ser examinadas mais a fundo as diferenças no comportamento verbal de homens e mulheres, replicando os achados deste estudo e fornecendo explicações mais sólidas para seus resultados.

REFERÊNCIAS

ADELSTEIN, J. S.; SHEHZAD, Z.; MENNES, M; DEYOUNG, C. G.; ZUO, X. N.; KELLY, C.; MARGULIES, D. S.; BLOOMFIELD, A.; GRAY, J. R.; CASTELLANOS, F. X.; MILHAM, M. P. Personality is reflected in the brain's intrinsic functional architecture. **PLoS ONE**, v. 6, e27633, 2011.

ALLPORT, G. W.; ODBERT, H. S. Trait-names: a psycho-lexical study. **Psychological Monographs**, v. 47, n. 1, p. i-171, 1936.

AMBADY, N.; KOO, J; LEE, F.; ROSENTHAL, R. More than words: linguistic and nonlinguistic politeness in two cultures. **Journal of personality and social psychology**, v. 70, p. 996-1011, 1996.

BALAGE FILHO, P. P.; ALUÍSIO, S. M.; PARDO, T. A. S. An evaluation of the Brazilian Portuguese LIWC dictionary for sentiment analysis. In: SIMPÓSIO BRASILEIRO DE TECNOLOGIA DA INFORMAÇÃO E LINGUAGEM HUMANA, 9., 2013, Fortaleza. **Anais...** Fortaleza: s.n., 2013. p. 215-219.

BAUMGARTEN, F. Die Charktereigenschaften. In: **Beitraege zur Charakter Und Persoenlichkeitsforschung**. Bern: A. Francke, 1933.

BOUCHARD, T. J. JR.; LYKKEN, D. T.; MCGUE, M.; SEGAL, N. L.; TELLEGEN, A. Sources of human psychological differences: the Minnesota Study of Twins Reared Apart. **Science**, New York, v. 250, n. 4978, p. 223-228, Oct., 1990.

BROWN, R.; GILMAN, A. Politeness theory in Shakespeare's four major tragedies. **Language in society**, v. 18, p. 159-212, 1989.

BUSS, D. M. Psychological sex differences: Origins through sexual selection. **The American Psychologist**, v. 50, n. 3, p. 164-168, 1995.

CATTELL, R. B. The description of personality: Basic traits resolved into clusters. **Journal of Abnormal and Social Psychology**, v, 38, p. 476-506, 1943.

CHEUNG, F. M.; LEUNG, K, ZHANG, J. X.; SUN, H. F.; GAN, Y. Q.; W. Z. SONG; XIE, D. Indigenous Chinese Personality Constructs. **Journal of Cross-Cultural Psychology**, v. 32, n. 4, p. 407-433, Jul., 2001.

CHURCH, A. T. Culture and personality: Toward an integrated cultural trait psychology. **Journal of Personality**. v. 68, p. 651–703, 2000.

COSTA, P. T. JR.; MCCRAE, R. R. **NEO-PI-R professional manual**: Revised NEO Personality Inventory (NEO-PI-R) and NEO-Five Factor Inventory (NEO-FFI). Odessa: Psychological Assessment Resources, 1992.

COSTA, P. T. JR.; MCCRAE, R. R. **The NEO Personality Inventory manual**. Odessa: Psychological Assessment Resources, 1985.

COSTA, P. T. JR.; MCCRAE, R. R. Set like plaster? Evidence for the stability of adult personality. In: HEATHERTON, T. F.; WEINBERGER, J. F. (Ed.). **Can personality change?** Washington: American Psychology Association, 1994. p. 21-40.

COSTA, P. T. JR.; TERRACCIANO, A.; MCCRAE, R. R. Gender differences in personality traits across cultures: robust and surprising findings. **Journal of personality and social psychology**, v. 81, n. 2, p. 322-331, ago. 2001.

DE RAAD, B.; PERUGINI, M.; HREBICKOVA, M.; SZAROTA, P. Lingua franca of personality: taxonomies and structures based on the psycholexical approach. **Journal of cross- cultural psychology**, v. 29, p.212-232, 1998.

DE RAAD, B.; PEABODY, D. Cross-culturally recurrent personality factors: analyses of three factors. **European Journal of Personality**, v. 19, n. 6, p. 451-474, Sep., 2005.

DEYOUNG, C. G.; HIRSH, D. M.; SHANE, M. S.; PAPADEMETRIS, X.; RAJEEVAN, N.; GRAY, J. R. Testing predictions from personality neuroscience. Brain structure and the big five. **Psychological Science**, v. 21, n. 6, p. 820-828, Jun., 2010.

DEYOUNG, C. G.; PETERSON, J. B.; HIGGINS, D. M. Sources of openness/intellect: cognitive and neuropsychological correlates of the fifth factor of personality. **Journal of Personality**, v. 73, n. 4, p. 825-858, Aug., 2005.

DIGMAN, J. M. Personality structure: Emergence of the five-factor model. **Annual Review of Psychology**, v. 41, p. 417- 440, 1990.

EAGLY, A. H. **Sex differences in social behavior**: A social-role interpretation. Hillsdale: Erlbaum, 1987.

EYSENCK, H. J. Genetic and environmental contributions to individual differences: the three major dimensions of personality. **Journal of Personality**, v. 58, n. 1, p. 245-261, Mar., 1990.

FAST, L. A.; FUNDER, D. C. Personality as manifest in word use: correlations with self-report, acquaintance report, and behavior. **Journal of personality and social psychology**, v. 94, n. 2, p. 334, 2008.

FAUL, F.; ERDFELDER, E.; BUCHNER, A.; LANG, A.-G. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. **Behavior Research Methods**, v. 41, n. 4, p. 1149-1160, 2009.

FERREIRA, T.; FERNANDES, M. Detecção de traços de personalidade em textos para apoiar a formação de grupos para colaboração. In: Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE), 28., 2017, Recife. **Anais...** Recife: UFP, 2017, p. 1627.

FIGUEIRÓ, M. T.; MARTINS, L. N. R.; LEITE, L. R. C., ZILLOTTO, J. M.; BACHA, M. M. Traços de personalidade de estudantes de Psicologia. **Psicólogo inFormação**, v. 14, n. 14, p. 13-28, 2010.

FLORES-MENDOZA, C. E. **Inventário de personalidade NEO-Revisado**. Manual técnico. São Paulo: Vetor Editora, 2007.

FOLTZ, P.W. Special issue: quantitative approaches to semantic knowledge representation. **Discourse Process**, v. 25, p. 127-363, 1998.

FREUD, S. **Sobre a psicopatologia da vida cotidiana**. Rio de Janeiro: Imago, 2006. 311 p. (Sigmund Freud, v. 6).

FURNHAM, A. Language and personality. In: GILES, H; ROBINSON, W. P. **Handbook of language and social psychology**. New York: Wiley, 1990. p. 73-95.

FUTURE AUTHORING PROGRAM. Self-authoring, [S.l.], 2017. Disponível em: <<https://selfauthoring.com/future-authoring.html>>. Acesso em: 19 dez. 2017.

GALTON, F. Measurement of character. **Fortnightly Review**, v. 36, p. 179-185, 1884.

GOLDBERG, L. R. From ace to zombie: some explorations in language of personality. In: SPIELBERGER, C. D.; BUTCHER, J. N (Ed.). **Advances in personality assessment**. v. 1. Hillsdale: Erlbaum, 1982. p. 203-234.

GOLDBERG, L. R. Language and individual differences: the search for universals in personality lexicons. In: WHEELER, L. (Ed.). **Review of personality and social psychology**, v. 2. Beverly Hills: Sage, 1981. p. 141-165.

GOLDBERG, L. R. The development of markers for the big-five factor structure. **Psychological Assessment**, v. 4, n. 1, p. 26-42, 1992.

GOTTSCHALK, L. A.; GLEISER, G. C. **The measurement of psychological states through the content analysis of verbal behavior**. Berkeley: University of California Press, 1969.

GOTTSCHALK, L. A.; GLEISER, G. C.; DANIELS, R.; BLOCK, S. The speech patterns of schizophrenic patients: a method of assessing relative degree of personal disorganization and social alienation. **Journal of nervous and mental disease**, v. 127, p. 153-166, 1958.

GRAY, J. A.; MCNAUGHTON, N. **The Neuropsychology of Anxiety**. 2. ed. Oxford: Oxford University Press, 2003.

HIRSH, J. B.; PETERSON, J. B. Personality and language use in self-narratives. **Journal of research in personality**, n. 43, p. 524-527, 2009.

HOLTGRAVES, T. Text messaging, personality, and the social context. **Journal of research in personality**, v. 45, n. 1, p. 92-99, 2011.

HOLTGRAVES, T. (Ed.). **The Oxford handbook of language and social psychology**. Oxford: Oxford Library of Psychology, 2014.

HOUAISS. Houaiss online, [S.l.]: [s.d.]. Disponível em: <<https://houaiss.uol.com.br/pub/apps/www/v33/html/index.php>>. Acesso em: 17 dez. 2017.

HUTZ, C. S.; NUNES, C. H. S. S. **Escala Fatorial de Neuroticismo**. São Paulo: Casa do Psicólogo, 2001

HUTZ, C. S.; NUNES, C. H. S. S.; SILVEIRA, A. D.; SERRA, J.; ANTON, M.; WIECZOREK, L. S. O desenvolvimento de marcadores para a avaliação da personalidade no modelo dos cinco grandes fatores. **Psicologia: Reflexão e Crítica**, v. 11, n. 2, p. 395-409, 1998.

JAY, T. B. The utility and ubiquity of taboo words. **Perspectives on Psychological Science**, v. 4, p. 153-161, 2009.

JAQUES, P. A.; NUNES, M. A. S. N. Ambientes Inteligentes de Aprendizagem que inferem, expressam e possuem emoções e personalidade. In: JORNADA DE ATUALIZAÇÃO EM INFORMÁTICA NA EDUCAÇÃO, 1., 2012, Rio de Janeiro. **Anais...** Rio de Janeiro: Universidade Federal do Rio de Janeiro, 2012. p. 30-81.

JOHN, O. P.; ANGLEITNER, A.; OSTENDORF, F. The lexical approach to personality: A historical review of trait taxonomic research. **European Journal of Personality**, v. 2, n. 3, p. 171-203, 1988.

JOHN, O. P.; NAUMANN, L. P.; SOTO, C. J. Paradigm shift to the integrative big five trait taxonomy. In: JOHN, O. P.; ROBINS, R. W.; PERVIN, L. A. (Ed.). **Handbook of personality: theory and research**. New York: Guilford Press, 2008. p. 114-158.

JOHN, O. P.; SRIVASTAVA, S. The Big Five trait taxonomy: history, measurement and theoretical perspectives. In: PERVIN, L. A.; JOHN, O. P. (Ed.). **Handbook of personality: theory and research**. 2. ed. New York: Guilford Press, 1999. p. 102-138.

JUNG, C. G. **Modern Man in Search of a Soul**. London: Kegan Paul, 1933.

KLAGES, L. **The science of character**. London: Allen & Unwin, 1932.

LAKOFF, R. T. **Language and woman's place**. New York: Harper & Row, 1975.

LANGACKER, R. W. **Cognitive grammar: A basic introduction**. Oxford: Oxford University Press, 2008.

LEE, C. H.; KIM, K.; SEO, Y. S.; CHUNG, C. K. The relations between personality and language use. **The Journal of General Psychology**, v. 134, n. 4, p. 405-413, 2007.

LIWC. PortLex, [S.l.], [s.d.]. Disponível em:

<<http://143.107.183.175:21380/portlex/index.php/pt/projetos/liwc>>. Acesso em: 17 dez. 2017.

MACHADO, A. A. A. **Inferência de personalidade a partir de textos de rede social utilizando um léxico afetivo em português brasileiro**. 2016. Dissertação (Mestrado em Ciência da Computação)– Universidade Federal de Sergipe, São Cristóvão, 2016.

MACHADO, A.; LONGUI, M.; NUNES, M. A. S. N.; PARDO, T. Personalitatem lexicon: um léxico em português brasileiro para mineração de traços de personalidade em textos. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 26., Maceió. **Anais...** Maceió: Universidade Federal de Alagoas, 2015. p. 1122-1126.

MAGNUS, K.; DIENER, E.; FUJITA, F.; PAVOT, W. Extraversion and neuroticism as predictors of objective life events: A longitudinal analysis. **Journal of Personality and Social Psychology**, v. 65, n. 5, p. 1046-1053, 1993.

MAIRESSE, F.; WALKER, M. A.; MEHL, M. R.; MOORE, R. K. Using linguistic cues for the automatic recognition of personality in conversation and text. **Journal of artificial intelligence research**, v. 30, p. 457-500, 2007.

MCADAMS, D. P. The psychology of life stories. **Review of General Psychology**, v. 5, p. 100–122, 2001.

MCADAMS, D. P.; PALS, J. A new big five: fundamental principles for an integrative science of personality. **The American Psychologist**, v. 61, n. 3, p. 204-217, Apr., 2006.

MCCRAE, R. R.; COSTA, P. T.; BUSCH, C. M. Evaluating comprehensiveness in personality systems: The California Q-Set and the five-factor model. **Journal of Personality**, v. 54, n. 2, p. 430-446, 1986.

MCCRAE, R. R.; JOHN, O. P. An introduction to the five-factor model and its applications. **Journal of Personality**, v. 60, n. 2, p. 175-215, 1992.

MCDONALD, J. D. Measuring Personality Constructs: The Advantages and Disadvantages of Self-Reports, Informant Reports and Behavioural Assessments. **Enquire**, v. 1, n. 1, p. 75-94, 2008.

MCLEAN, K. C.; PASUPATHI, M.; PALS, J. L. Selves creating stories creating selves: A process model of self-development. **Personality and Social Psychology Review**, v. 11, p. 262, 2007.

MCCLELLAND, D. C. **The achieving society**. Princeton: Van Nostrand, 1961.

MEHL, M. R.; GOSLING, S. D.; PENNEBAKER, J. W. Personality in its natural habitat: manifestations and implicit folk theories of personality in daily life. **Journal of personality and social psychology**, v. 90, n. 5, p. 862, 2006.

MEHL, M. R.; PENNEBAKER, J. W. The sounds of social life: a psychometric analysis of students' daily social environments and natural conversations. **Journal of personality and social psychology**, v. 84, n. 4, p. 857-870, 2003.

MISCHEL, W.; SHODA, Y. A Cognitive-Affective System Theory of Personality: Reconceptualizing Situations, Dispositions, Dynamics, and Invariance in Personality Structure. **Psychological Review**, v. 102, n. 2, p. 246-268, 1995.

MORAND, D. A. Language and power: an empirical analysis of linguistic strategies used in superior-subordinate communication. **Journal of organizational behavior**, v. 21, p. 235-248, 2000.

NEWMAN, M. L.; BERRY, D. S.; RICHARDS, J. M. Lying words: predicting deception from linguistic styles. **Personality and social psychology bulletin**, v. 29, n. 5, p. 665-675, maio 2003.

NEWMAN, M. L.; GROOM, C. J.; HANDELMAN, L. D.; PENNEBAKER, J. W. Gender differences in language use: An analysis of 14,000 text samples. **Discourse Processes**, v. 45, n. 3, p. 211-236, 2008.

NEWMAN, M. L.; GROOM, C. J.; HANDELMAN, L. D.; PENNEBAKER, J. W. Gender differences in language use: An analysis of 14,000 text samples. **Discourse Processes**, v. 45, n. 3, p. 211-236, 2008.

NGUN, T.; GHAMRANI, N.; SÁNCHEZ, F. J.; BOCKLANDT, S.; VILAIN, E. The genetics of sex differences in brain and behavior. **Frontiers in neuroendocrinology**, v. 32, n. 2, p. 227-246, 2011.

NORMAN, W. T. **2,800 personality trait descriptors**: Normative operating characteristics for a university population. Ann Arbor: University of Michigan, 1967.

NOWSON, S. **The Language of Weblogs**: A study of genre and individual differences. 2006. Tese (Doutorado em Engenharia)– University of Edinburgh, Edinburgh, 2006.

NUNES, C. H. S.S. **A construção de um instrumento de medida para o fator neuroticismo/estabilidade emocional dentro do modelo de personalidade dos cinco grandes fatores**.2000. Dissertação (Mestrado em Psicologia)– Instituto de Psicologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2000.

NUNES, C. H. S. S.; HUTZ, C. S. Construção e validação da escala fatorial de Socialização no modelo dos Cinco Grandes Fatores de Personalidade. **Psicologia: Reflexão e Crítica**, v. 20, n. 1, 2007.

NUNES, C. H. S. S.; HUTZ, C. S. **Escala Fatorial de Extroversão**: Manual técnico.São Paulo, SP: Casa do Psicólogo, 2007a.

NUNES, C. H. S. S.; HUTZ, C. S. **Escala Fatorial de Socialização**: Manual técnico.São Paulo, SP: Casa do Psicólogo, 2007b

NUNES, C. H. S.; HUTZ, C. S.; NUNES, M. O. **Bateria fatorial de personalidade**: manual técnico. Casa do Psicólogo: São Paulo, 2010.

NUNES, M. A. S. N.; TELES, F. R.; DE SOUZA, J. G. Inferindo personalidade via tweets. **GEINTEC-Gestão, Inovação e Tecnologias**, v. 3, n. 3, p. 45-57, 2013.

OBERLANDER, J; GILL, A. J. Language with character: A stratified corpus comparison of individual differences in e-mail communication. **Discourse Processes**, v. 42, n. 3, p. 239-270, 2006.

OXMAN, T. E.; ROSENBERG, S. D.; TUCKER, G. J. The language of paranoia. **American Journal of Psychiatry**, v. 139, p. 275-282, 1982.

PACICO, J. C.; HUTZ, C. S. Desenvolvimento e evidências de validade da escala fatorial de realização-EFR. No prelo.

PAIM, A. M.; CAMATI, R. S.; ENEMBRECK, F. Inferência de personalidade a partir de textos em português utilizando léxico linguístico e aprendizagem de máquina. In: ENCONTRO NACIONAL DE INTELIGÊNCIA ARTIFICIAL E COMPUTACIONAL, 13., 2016, Recife. **Anais...** Recife: s.n., 2016. p. 481-492.

PENNEBAKER, J. W. **The secret life of pronouns**. New York: Bloomsbury, 2011.

PENNEBAKER, J. W.; FRANCIS, M.E.; BOOTH, R. J. **Linguistic inquiry and word count: LIWC 2001**. Mahway: Lawrence Erlbaum Associates, 2001.

PENNEBAKER, J. W.; KING, L. A. Linguistic styles: language use as an individual difference. **Journal of personality and social psychology**, v. 77, n. 6, p. 1296-1312, 1999.

PENNEBAKER, J. W.; MEHL, M. R.; NIEDERHOFFER, K. Psychological aspects of natural language use: Our words, our selves. **Annual Review of Psychology**, v. 54, p. 547-577, 2003.

PENNEBAKER, J. W.; STONE, L. D. Words of wisdom: language use over the life span. **Journal of personality and social psychology**, v. 85, n. 2, p. 291-301, ago. 2003.

PONTE JUNIOR, L. A.; GUEDES, G. P.; BEZERRA, E. Inferindo o sexo de usuários de redes sociais utilizando o LIWC em português do Brasil. In: CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO, 36., 2016, Rio de Janeiro. **Anais...** Rio de Janeiro: UFRJ, 2016. p. 217-222.

POPPING, R. **Computer-assisted text analysis**. London: Sage, 2000.

RICOEUR, P. **Interpretation theory: discourse and the surplus of meaning**. Fort Worth: Texas Christian University Press, 1976.

ROBERTS, B. W.; DELVECCHIO, W.F. The rank-order consistency of personality traits from childhood to old age: a quantitative review of longitudinal studies. **Psychology Bulletin**, v. 126, n. 1, p. 3-25, Jan. 2000.

ROGERS, C. **On becoming a person**. Boston: Houghton Mifflin, 1961.

SANFORD, F. H. Speech and personality. **Psychological Bulletin**, v. 39, p. 811–845, 1942.

SCHERWITZ, L.; CANICK, J. Self-reference and coronary heart disease risk. In: HOUSTON, K.; SNYDER, C. R. **Type A Behavior Pattern: Theory, and Intervention**. New York: Wiley, 1988. p. 146-167.

SCHIFFRIN, D. **Approaches to discourse**. Cambridge, MA: Blackwell, 1994.

SCHMITT, D. P.; LONG, A. E.; MCPHEARSON, A.; O'BRIEN, K.; REMMERT, B.; SHAH, S. H. Personality and gender differences in global perspective. **International Journal of Psychology**, Mar. 2016.

SCHMITT, D. P.; REALO, A.; VORACEK, M.; ALLIK, J. P. Why can't a man be more like a woman? Sex differences in Big Five personality traits across 55 cultures. **Journal of personality and social psychology**, v. 94, n. 1, p. 168, 2008.

SMITH, C.P. **Motivation and personality: handbook of thematic content analysis**. Cambridge, MA: Cambridge University Press, 1992.

SILVA, I. B.; NAKANO, T. C. Modelo dos cinco grandes fatores da personalidade: análise de pesquisas. **Avaliação psicológica**, v. 10, n. 1, p. 51-62, 2011.

SPSS INC. Released 2009. **PASW Statistics for Windows**, Version 18.0. Chicago: SPSS Inc, 2009.

STONE, P. J.; DUNPHY, D. C.; SMITH, M. S.; OGILVIE, D. M. **The general inquirer: A computer approach to content analysis**. Cambridge: MIT Press, 1966.

TAUSCZIK, Y. R.; PENNEBAKER, J. W. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. **Journal of Language and Social Psychology** v. 29, n. 1, p. 24-54, 2010.

VASCONCELLOS, S. J. L.; HUTZ, C. S. Construção e validação de uma escala de abertura à experiência. **Avaliação Psicológica**, Porto Alegre, v. 7, n. 2, ago. 2008.

VEDEL, A. Big Five personality group differences across academic majors: A systematic review. **Personality and individual differences**, v. 92, p. 1-10, 2016.

WAINBERG, J. A. Lula, Dilma e FHC: a mentira e a verdade na comunicação. **Contemporânea**, v. 14, n. 2, p. 235-252, maio-ago., 2016.

WEINTRAUB, W. **Verbal behavior in everyday life**. New York: Springer, 1989.

WEISBERG, Y. J.; DEYOUNG, C. G.; HIRSH, J. B. Gender differences in personality across the ten aspects of the Big Five. **Frontiers in psychology**, v. 2, p. 178, 2011.

WEST, M. D. **Theory, method and practice in computer content analysis**. New York: Ablex, 2001.

WHITLEY, H.; LINDSEY, W. Sex-based differences in drug activity. **American family physician**, v. 80, n. 11, p. 1254-1258, 2009.

WORD CLOUDS. World well-being project, [S.l.], [s.d.]. Disponível em: <http://www.wwbp.org/gender_wc.html>. Acesso em: 17 dez. 2017.

XIA, X. Gender differences in using language. **Theory and Practice in Language Studies**, v. 3, n. 8, p. 1485, 2013.

YARKONI, Tal. Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. **Journal of research in personality**, v. 44, n. 3, p. 363-373, 2010.

YEE, N.; HARRIS, H.; JABON, M.; BAIENSON, J. N. The expression of personality in virtual worlds. **Social Psychological and Personality Science**, v. 2, p. 5-12, 2011.