

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO
PUC-SP**

MARIA CAROLINA ZUPPARDO

**DIMENSÕES DE VARIAÇÃO EM MANUAIS AERONÁUTICOS: um
estudo baseado na análise multidimensional**

**MESTRADO EM LINGUÍSTICA APLICADA E ESTUDOS DA
LINGUAGEM**

**SÃO PAULO
2014**

MARIA CAROLINA ZUPPARDO

**DIMENSÕES DE VARIAÇÃO EM MANUAIS AERONÁUTICOS: UM ESTUDO
BASEADO NA ANÁLISE MULTIDIMENSIONAL**

Dissertação apresentada à Banca Examinadora da Pontifícia Universidade Católica de São Paulo, como exigência parcial para obtenção do título de MESTRE em Linguística Aplicada e Estudos da Linguagem, sob a orientação do Prof. Dr. Antonio Paulo Berber Sardinha.

**PUC - SP
2014**

BANCA EXAMINADORA

Aos meus pais que sempre apoiaram as minhas escolhas.

AGRADECIMENTOS

Agradeço primeiramente ao meu orientador, Prof. Dr. Tony Berber Sardinha. Obrigada, Tony, por incentivar a minha paixão pela pesquisa ligada à estatística, acreditar em mim, abrir as portas do mundo da academia, e me incentivar a buscar cada vez mais novos horizontes, tanto na pesquisa acadêmica quanto na minha vida profissional.

Agradeço também aos membros do Grupo de Estudos de Linguística de *Corpus* (GELC). Obrigada pelos questionamentos por vezes difíceis, mas que me possibilitaram crescer como pesquisadora, e pelas sugestões e ajuda sempre bem-vindas. Agradeço ao Etelvo Ramos Filho e a Renata Condi de Souza, amigos mais do que especiais, que contribuíram para o meu desenvolvimento como pesquisadora e como ser humano. Obrigada pelas conversas e especialmente por suas contribuições valiosas!

Agradeço a Prof. Dra. Randi Reppen pelos ensinamentos acerca da pesquisa em AMD ao longo de uma semana de congresso, por se dispor a me ajudar com a análise e pela validação dos meus dados.

Agradeço a amiga e colega de profissão Katherine Asega pelo incentivo e pelos ensinamentos sobre inglês para aviação. Aos pilotos e mecânicos que tanto me ajudaram na coleta de dados, agradeço pela oportunidade de conhecer e entender seu trabalho, em especial, ao Enio Luis dos Santos pelas explicações e esclarecimentos.

Agradeço também ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela bolsa integral concedida e ao Programa de Estudos Pós-Graduados em Linguística Aplicada e Estudos da Linguagem (LAEL), sua coordenação, professores e funcionárias pelo apoio durante essa jornada. À escola de inglês para aviação que me concedeu a oportunidade de aprender e trabalhar com essa área da língua inglesa.

Por fim, agradeço à minha família. Aos meus pais pelo incentivo e apoio. Ao Rodrigo, pela paciência, ajuda, explicações e informações, e por estar ao meu lado nessa jornada.

RESUMO

O trabalho teve como objetivos principais identificar as dimensões de variação próprias da linguagem dos manuais aeronáuticos, identificar onde se situam os manuais de manutenção de aeronaves nas dimensões de variação do inglês e verificar se há variação estatística significativa entre os fabricantes de aeronaves, os modelos de aeronaves e os tipos de manuais aeronáuticos em relação às dimensões de variação dos manuais aeronáuticos.

A pesquisa foi embasada nos pressupostos teóricos da Linguística de *Corpus*, uma área de pesquisa da Linguística Aplicada que se baseia na visão probabilística da linguagem e cujo estudo de uma língua ou variedade linguística envolve o uso de ferramentas computacionais e interativas aplicadas à grandes coletâneas de textos em formato eletrônico (*corpora*).

Mais especificamente, o estudo se baseia na Análise Multidimensional (AMD), uma metodologia baseada em *corpus* para o estudo das relações entre características linguísticas e registros em grandes quantidades de textos, por meio de procedimentos estatísticos a fim de identificar dimensões de variação.

O *corpus* empregado na pesquisa é composto de manuais operacionais e de manutenção de aeronaves comerciais e executivas, contendo 10.009.040 palavras distribuídas em 154 textos.

A análise da localização do *corpus* nas dimensões de variação do inglês definidas por Biber em 1988 indicou que manuais aeronáuticos são essencialmente informativos e não narrativos, contendo referências explícitas, expressão não explícita de persuasão e informações abstratas. Além disso, a análise multidimensional revelou três dimensões próprias dos manuais aeronáuticos que foram denominadas: (1) Ampla descrição de sistemas; (2) Discurso de tomada de decisões *versus* Discurso de procedimentos; e (3) Discurso de solução de problemas *versus* Descrição de partes específicas.

A pesquisa pretende ter fornecido uma contribuição original para a área de inglês para aviação e para a área de pesquisas baseadas em *corpus* ao identificar as dimensões próprias dos manuais aeronáuticos e suas características linguísticas mais salientes. O trabalho ainda apresenta e discute as limitações da pesquisa e possíveis pesquisas para o futuro, além de possíveis sugestões de aplicações pedagógicas dos resultados da pesquisa.

Palavras-chave: Manuais Aeronáuticos; Linguística de *Corpus*; Análise Multidimensional.

ABSTRACT

This study set out to identify the dimensions of variation that characterize aircraft manuals, apply the 1988 dimensions for English to aircraft manuals, and verify if there is statistically significant variation among aircraft manufacturers and models as well as types of manuals in terms of the dimensions of variation of aircraft manuals.

The main theoretical underpinning for the research is provided by Corpus Linguistics, an area of Applied Linguistics based on the view of language as a probability system whose studies of language or language varieties involve the use of computational and interactive tools applied to large collections of texts held in electronic format (corpora).

More specifically, the study presented here is informed by Multidimensional Analysis, a corpus-based approach for the study of associations between linguistic features and registers in large amounts of text through the use of statistical procedures in order to identify dimensions of variation.

The corpus compiled for the study is comprised of operational and maintenance aircraft manuals of commercial and corporate aircraft, containing 10,009,040 words distributed across 154 texts.

On the whole, the findings suggest that, in terms of placement of the *corpus* in the dimensions of variation for English defined by Biber in 1988, aircraft manuals are highly informational, non-narrative, with explicit references, abstract information, and covert persuasion. In addition, the multidimensional analysis revealed three dimensions of variation for aircraft manuals, namely: (1) Broad system descriptions; (2) Decision-making discourse *versus* Procedural discourse; e (3) Problem-solving discourse *versus* Specific parts descriptions.

The study presented here may have made an original contribution to the English for Aviation areas in addition to the existing body of corpus-based research in that it determined the dimensions of variation of aircraft manuals and their salient linguistic features. The study also presents and discusses possible limitations and further research, as well suggestions for pedagogical applications of the findings.

Keywords: Aircraft Manuals; Corpus Linguistics; Multidimensional Analysis.

Lista de Quadros

Quadro 1	Exemplo de texto anotado pelo <i>Biber Tagger</i>	57
Quadro 2	Exemplo de expressão numérica para cálculo do escore de fator.....	79
Quadro 3	Trecho de texto – Dimensão 1 – Mapeamento CAM/Biber.....	89
Quadro 4	Trecho de texto – Dimensão 2 – Mapeamento CAM/Biber.....	90
Quadro 5	Trecho de texto – Dimensão 3 – Mapeamento CAM/Biber.....	92
Quadro 6	Trecho de texto – Dimensão 4 – Mapeamento CAM/Biber.....	94
Quadro 7	Trecho de texto – Dimensão 5 – Mapeamento CAM/Biber.....	95
Quadro 8	CAM – Trecho de texto do Fator 1 – Polo positivo.....	107
Quadro 9	CAM – Trecho de texto do Fator 1 – Polo negativo.....	108
Quadro 10	CAM – Trecho de texto do Fator 2 – Polo positivo.....	112
Quadro 11	CAM – Trecho de texto do Fator 2 – Polo negativo.....	113
Quadro 12	CAM – Trecho de texto do Fator 3 – Polo negativo.....	116
Quadro 13	CAM – Trecho de texto do Fator 3 – Polo positivo.....	117

Lista de Gráficos

Gráfico 1	SPSS 20 – Análise Fatorial – <i>Scree Plot</i>	73
Gráfico 2	Localização do corpus CAM na Dimensão 1 – Mapeamento CAM/Biber.....	88
Gráfico 3	Localização do corpus CAM na Dimensão 2 – Mapeamento CAM/Biber.....	90
Gráfico 4	Localização do corpus CAM na Dimensão 3 – Mapeamento CAM/Biber.....	91
Gráfico 5	Localização do corpus CAM na Dimensão 4 – Mapeamento CAM/Biber.....	93
Gráfico 6	Localização do corpus CAM na Dimensão 5 – Mapeamento CAM/Biber.....	95
Gráfico 7	CAM – <i>Scree Plot</i>	98
Gráfico 8	CAM – Distribuição dos registros no Fator 1.....	109
Gráfico 9	CAM – Distribuição dos registros no Fator 2.....	114
Gráfico 10	CAM – Distribuição dos registros no Fator 3.....	118

Lista de Tabelas

Tabela 1	Distribuição dos registros dos <i>corpora Brown</i> e LOB (cf. SOUZA, 2012).....	29
Tabela 2	<i>Corpus</i> utilizado na AMD realizada por Biber em 1988 (adaptada de SOUZA, 2012).....	40
Tabela 3	Nomenclatura das dimensões do inglês em 1988 e 2009 (adaptada de Berber Sardinha, 2013).....	43
Tabela 4	Lista de fabricantes, aeronaves e manuais que perfazem o <i>corpus</i> CAM.....	47
Tabela 5	Tipos de Manuais do <i>corpus</i> CAM.....	49
Tabela 6	Manuais subdivididos e não subdivididos em capítulos....	51
Tabela 7	Limites de corte.....	51
Tabela 8	Composição do <i>corpus</i> CAM.....	54
Tabela 9	SPSS 20 – Análise Fatorial – Teste KMO e Bartlett.....	71
Tabela 10	SPSS 20 – Análise Fatorial – Comunalidades (recorte)....	71
Tabela 11	SPSS 20 – Análise Fatorial – Variância Total Explicada (recorte).....	72
Tabela 12	SPSS 20 – Análise Fatorial – Variância Total Explicada com rotação (recorte).....	76
Tabela 13	SPSS 20 – Análise Fatorial – Matriz de padrões (recorte).....	76
Tabela 14	SPSS 20 – ANOVA – Exemplo de tabela.....	83
Tabela 15	Dimensões de 1988 consideradas no mapeamento do CAM.....	85
Tabela 16	Variáveis linguísticas dos Fatores 1 a 5 (cf. Biber, 1988).....	86
Tabela 17	CAM – Teste KMO e teste de esfericidade de Bartlett.....	97
Tabela 18	CAM – Variância Total Explicada.....	99
Tabela 19	CAM – Matriz de Correlação entre Fatores.....	100
Tabela 20	CAM – Matriz de padrões.....	100
Tabela 21	Dimensões do <i>corpus</i> CAM.....	104
Tabela 22	CAM – Variáveis do Fator 1.....	105
Tabela 23	CAM – Variáveis do Fator 2.....	110
Tabela 24	CAM – Variáveis do Fator 3.....	115
Tabela 25	ANOVA – Fabricantes de Aeronaves.....	120
Tabela 26	ANOVA – Modelos de Aeronaves.....	120
Tabela 27	ANOVA – Tipos de Manuais.....	121

Lista de Figuras

Figura 1	Formulário preenchido para realização de amostragem aleatória.....	53
Figura 2	Números dos textos escolhidos aleatoriamente para inclusão no <i>corpus</i>	54
Figura 3	<i>Biber Tagger</i>	57
Figura 4	<i>TagCount</i>	58
Figura 5	SPSS 20 – Estatísticas Descritivas – Dimensões de 1988..	66
Figura 6	SPSS 20 – Redução de Dimensão – Análise Fatorial.....	67
Figura 7	SPSS 20 – Análise Fatorial – Inserção de variáveis.....	68
Figura 8	SPSS 20 – Análise Fatorial – Descritivos.....	69
Figura 9	SPSS 20 – Análise Fatorial – Extração.....	69
Figura 10	SPSS 20 – Análise Fatorial – Rotação.....	75
Figura 11	SPSS 20 – Análise Fatorial – Opções.....	75
Figura 12	SPSS 20 – Estatísticas Descritivas – Padronização.....	78
Figura 13	SPSS 20 – ANOVA.....	82
Figura 14	SPSS 20 – ANOVA – Exemplo de cálculo.....	82

Lista de Abreviaturas e Siglas

.pdf	Extensão indicativa de texto em PDF
.txt	Extensão indicativa de texto sem formatação
AMD	Análise Multidimensional
AMM	<i>Aircraft Maintenance Manual</i>
ANAC	Agência Nacional de Aviação Civil
ANOVA	<i>Analysis of Variance</i>
ATA	<i>Air Transport Association</i>
ATR	<i>Avions de Transport Regional</i>
BNC	<i>British National Corpus</i>
CAM	<i>Corpus of Aircraft Manuals</i>
	Conselho Nacional de Desenvolvimento Científico e Tecnológico
CNPq	
COCA	<i>Corpus of Contemporary American English</i>
COMET	Corpus Multilíngue para Ensino e Tradução
CorTec	Corpus Técnico-Científico
FAA	<i>Federal Aviation Administration</i>
FAPESP	Fundação de Amparo à Pesquisa do Estado de São Paulo <i>Flight Crew Operating Manual ou Flight Crew Operations Manual</i>
FCOM	
FIM	<i>Fault Isolation Manual</i>
GELC	Grupo de Estudos em Linguística de Corpus
ICAO	<i>International Civil Aviation Organization</i>
ICLE	<i>International Corpus of Learner English</i>
KMO	Teste Kaiser-Meyer-Olkin
LAEL	Linguística Aplicada e Estudos da Linguagem
LC	Linguística de Corpus
LLC	<i>London-Lund Corpus of Spoken English</i>
LOB	<i>Lancaster/Oslo/Bergen Corpus</i>
MEL	<i>Minimum Equipment List</i>
MMEL	<i>Master Minimum Equipment List</i>
MPP	<i>Maintenance Practices and Procedures</i>
MRO	<i>Maintenance/Repair Organizations</i>
NASA	<i>National Aeronautics and Space Administration</i>
NASA ASRS	<i>NASA Aviation Safety Reporting System</i>
OM	<i>Operational Manual</i>
P&P	<i>Practices and Procedures</i>
PDF	<i>Portable Document Format</i>
POH	<i>Pilot's Operating Handbook</i>
PROMAX	Proporção máxima de rotação
PUCSP	Pontifícia Universidade Católica de São Paulo
QRH	<i>Quick Reference Handbook</i>

SAS	<i>Statistical Analysis System</i>
SDS	<i>Systems Description Section</i>
SEE	<i>Survey of Spoken English</i>
SEU	<i>Survey of English Usage</i>
SGML	<i>Standard Generalized Markup Language</i>
SPSS	<i>Statistical Package for the Social Sciences</i>
TSM	<i>Troubleshooting Manual</i>
TTM	<i>Technical Training Manual</i>
UK	<i>United Kingdom</i>
VARIMAX	<i>Variação máxima de rotação</i>
XML	<i>Extensible Markup Language</i>

SUMÁRIO

Introdução	15
1 Fundamentação Teórica	26
1.1 Linguística de <i>Corpus</i>	26
1.2 Análise Multidimensional.....	36
2 Metodologia	47
2.1 Desenho e coleta do <i>Corpus</i>	47
2.2 Procedimentos de Análise.....	56
2.2.1 Anotação e Etiquetagem.....	56
2.2.2.1 SPSS e análise qualitativa: Mapeamento do CAM nas dimensões de 1988....	63
2.2.2.2 SPSS e análise qualitativa: Identificação das dimensões do CAM.....	66
2.2.2.2.1 Redução de dimensão: extração inicial.....	67
2.2.2.2.2 Redução de dimensão: extração rotacionada.....	74
2.2.2.2.3 Cálculo dos escores dos fatores.....	77
2.2.2.2.4 Análise qualitativa das dimensões.....	79
2.2.2.2.5 Análise de variância – ANOVA.....	81
3 Apresentação e Discussão dos Resultados	85
3.1 Mapeamento do <i>Corpus</i> CAM nas dimensões da língua inglesa (Biber, 1988).....	85
3.2 As dimensões do <i>Corpus</i> CAM.....	96
3.2.1 Dimensão 1	105
3.2.2 Dimensão 2	110
3.2.3 Dimensão 3	114
3.3 Variáveis independentes no <i>Corpus</i> CAM (ANOVAS).....	119
Considerações Finais	122
Bibliografia	125
Anexos	
Anexo A – Lista de etiquetas.....	132
Anexo B – Descrições das Etiquetas do <i>Biber Tagger</i>	137
Anexo C – CAM – Tabela de Variância Total Explicada.....	143
Anexo D – CAM – Estatísticas Descritivas.....	147
Anexo E – Gráficos de distribuição.....	150

Introdução

O desenvolvimento da aviação comercial foi impulsionado pelas demandas criadas pelas guerras mundiais. No período anterior à primeira guerra mundial, nos primórdios da aviação, o francês era tido como a língua principal devido a grande contribuição para a terminologia dessa nova área especializada. No entanto, com a produção de aeronaves sendo feita também em outros países, principalmente na Alemanha e nos Estados Unidos, surgem obras plurilíngues, entre elas dicionários, glossários e enciclopédias.

O ano de 1927 marca o início da aviação comercial no Brasil, quando uma empresa alemã e outra francesa foram autorizadas a realizar cabotagem¹ no território nacional e a primeira empresa aérea comercial brasileira foi criada, a VARIG (BOCORNY, 2011). Nessa época, pilotos e controladores eram obrigados a se comunicar em diversos idiomas, os mais comuns sendo inglês, francês, espanhol e alemão. O desenvolvimento da aviação comercial após a primeira guerra mundial e, especialmente, durante e após a segunda guerra mundial evidenciou a necessidade de uma língua em comum para a comunicação entre pilotos e controladores de voo. Uma vez que os Estados Unidos eram considerados a maior potência mundial no pós-guerra, foi o inglês que de fato se tornou a língua franca da aviação mundial. A consolidação do inglês como língua oficial da aviação ocorreu em 1951, quando a recém criada *International Civil Aviation Organization* (ICAO) recomendou o uso da língua inglesa entre pilotos e controladores que falassem línguas diferentes (BOCORNY, 2011).

Em termos linguísticos, nos primórdios da aviação, o foco estava na compilação e tradução da terminologia dedicada a essa área, com a publicação de obras terminológicas em francês e glossários plurilíngues. Com os avanços da aviação

¹ A cabotagem aérea é o transporte, por uma empresa estrangeira, de cargas ou passageiros entre dois pontos do território de um mesmo país e é vetada pelo atual marco regulatório brasileiro. Por exemplo, uma empresa estrangeira como a American Airlines não está autorizada a fazer um voo exclusivamente entre cidades brasileiras (<http://www2.senado.gov.br/bdsf/item/id/167>).

comercial e a ocorrência de incidentes e acidentes² em que umas das causas foi a questão da comunicação oral entre pilotos e controladores de voo, o foco dos estudos de inglês para aviação passou a ser a comunicação oral. Entretanto, o uso da língua inglesa na aviação permeia diversas esferas, tais como: a comunicação oral entre pilotos e controladores de voo; a comunicação oral e escrita entre pilotos e mecânicos; a comunicação escrita dos manuais, fichas de inspeção e documentações oficiais das aeronaves, sendo que aeronaves do mesmo modelo mas com matrículas diferentes podem ter alterações em seus manuais; e as diretrizes de aeronavegabilidade fornecidas pelas agências regulatórias da aviação, entre outras. É importante ressaltar que o foco da presente pesquisa é apenas uma dessas esferas: a comunicação escrita dos manuais aeronáuticos.

Apesar da comunicação oral entre pilotos e controladores de voo constituir apenas uma das esferas em que o inglês é utilizado na aviação, é a mais evidente uma vez que as consequências de uma quebra nessa comunicação, seja por carência de conhecimento da língua ou não, podem levar a acidentes fatais. Como resultado, a esmagadora maioria dos estudos a respeito do impacto da falta de proficiência nessa língua tem como foco a comunicação oral entre esses profissionais. Além disso, a exigência de proficiência na língua inglesa falada é aplicada pela ICAO apenas a pilotos e controladores de voo:

o único objeto dos requisitos de proficiência linguística da ICAO é a comunicação da radiofonia aeronáutica, uma subcategoria especializada da linguagem da aviação que corresponde à parcela limitada de usos de linguagem de somente dois profissionais – controladores de tráfego aéreo e tripulação de voo.³ (ICAO, 2010, p.32).

No entanto, conforme descrito acima, o inglês é usado em outras esferas da aviação e outros profissionais, como os técnicos de manutenção de aeronaves, dependem do conhecimento nessa língua para realizarem seu trabalho com

² De acordo com o documento Doc 9156 Annex 13 da ICAO (Disponível em: http://www.cad.gov.rs/docs/udesi/an13_cons.pdf) e o Manual do Comando da Aeronáutica MCA 3-6 2011 (Disponível em: <http://www.cenipa.aer.mil.br/cenipa/index.php/legislacao/category/7-mca-manual-do-comando-da-aeronautica>), são considerados acidentes ocorrências em que haja intenção de voo e uma das seguintes situações ocorra: lesões graves a pessoas ou morte; danos graves à aeronave; aeronave desaparecida. Um incidente é qualquer ocorrência relacionada à operação de uma aeronave, em que haja intenção de voo e que não possa ser classificada como acidente, que afete ou possa afetar a segurança da operação.

³ *“The sole object of ICAO language proficiency requirements is aeronautical radiotelephony communications, a specialized subcategory of aviation language corresponding to a limited portion of the language uses of only two aviation professions – controllers and flight crews.”*

segurança. De acordo com a FAA (2005), técnicos de manutenção de aeronaves passam de 25% a 40% do tempo de serviço buscando, usando ou documentando informações escritas. Essas informações escritas devem estar em língua inglesa, em conformidade com as normas da ICAO. Em sua rotina de trabalho, esses profissionais são responsáveis por consolidar informações relativas aos procedimentos de manutenção provenientes de fontes diversas, tais como: diferentes tipos de manuais da aeronave, fichas de inspeção, diário de bordo, relatórios de falhas de sistema, e diretrizes de aeronavegabilidade. Esse grande volume de informações que precisam ser integradas para que o trabalho possa ser realizado transforma o uso da documentação escrita em um enorme desafio para o técnico de manutenção. Nesse caso, a carência de conhecimento específico em inglês pode ter sérias consequências, uma vez que problemas com documentação escrita podem levar a erros, retrabalho, atrasos, e até mesmo incidentes e/ou acidentes (FAA, 2012).

Uma análise recente dos relatórios de incidentes/acidentes do banco de dados NASA ASRS⁴ realizada pela FAA (2012) mostrou que, de 2001 a 2011, aproximadamente 64% (cerca de 9.000 dos 14.267 relatórios submetidos) dos incidentes relatados estavam associados a problemas com documentação técnica e/ou procedimentos. Em 2001, a FAA (FAA, 2005) havia constatado que o uso inadequado de documentações de manutenção foi o fator que mais contribuiu para a ocorrência de diversos incidentes e acidentes de manutenção de aeronaves, com um total de treze acidentes, entre 1995 e 2003. As principais causas de incidentes de manutenção apontadas pelo Comitê de Segurança de Voo do Reino Unido (*UK Flight Safety Committee*) em 2004 são: (1) a falha em seguir as instruções ou informações técnicas, (2) o uso de procedimentos não autorizados e não previstos nas informações técnicas e (3) a aceitação por supervisores dessas duas práticas.

A documentação técnica usada por técnicos de aeronaves, conforme supracitado, compreende uma variedade de materiais escritos, porém o contato maior desses profissionais é com manuais operacionais e de manutenção. Cada aeronave tem o seu próprio conjunto de manuais fornecido pelo fabricante. A quantidade e a

⁴ *Aviation Safety Reporting System* (ASRS) é um programa da FAA que, por intermédio da NASA, recebe, processa e analisa dados de relatórios de incidentes e/ou acidentes, preenchidos confidencial e voluntariamente por profissionais da aviação. Esses relatórios não podem ser usados para propósitos de punição dos profissionais e/ou empresas envolvidos. Disponível em: <http://asrs.arc.nasa.gov/index.html>.

variedade de manuais depende do fabricante, do modelo da aeronave e da customização da aeronave exigida pelo cliente, tanto para aeronaves executivas quanto aeronaves comerciais.

Técnicos de manutenção de aeronaves podem trabalhar para empresas aéreas, empresas de taxi aéreo, para empresas privadas, e/ou em centros de manutenção, conhecidos como *Maintenance Repair & Overhaul* (MRO). Em todos os casos, esses profissionais estão em contato com aeronaves de modelos diferentes e de fabricantes diferentes. Portanto, eles precisam lidar com um conjunto de manuais diferente para cada aeronave atendida, o que pode dificultar o seu trabalho, uma vez que manuais de diferentes aeronaves podem ter características linguísticas diferentes.

A partir desses dados pode-se perceber o quão relevante é a investigação das características linguísticas mais salientes dos manuais de manutenção aeronáutica, como a realizada nesta pesquisa. Na verdade, estudos sobre essas características podem ser úteis a esse campo de atividade e à sociedade, já que podem contribuir no sentido de reduzir incidentes no momento da manutenção desse meio de transporte de passageiros e cargas, especificamente, causados por desconhecimento em relação à linguagem adotada nesses documentos oficiais das empresas, os manuais operacionais e de manutenção. Os diversos profissionais que trabalham na manutenção de aeronaves estão em constante contato com esses manuais, entretanto, pouco se sabe sobre a constituição linguística desses textos.

Atualmente, o maior contingente de aeronaves comerciais em atividade no mundo é proveniente de poucos fabricantes, sendo os principais: *Airbus* (França), *Boeing* (Estados Unidos) e *Embraer* (Brasil). Fabricantes como a *ATR* (França) e a *Bombardier* (Canadá) têm menor participação no mercado de aeronaves comerciais de grande porte.

A manutenção dessas aeronaves no mundo é feita, em sua maior parte, por centros de manutenção terceirizados, denominados *Maintenance/Repair Organizations* (MROs), nos quais a presença de falantes não nativos de inglês é maioria, e por técnicos em países cuja língua oficial não é o inglês (DRURY, MA, 2003 e 2005). Apesar das documentações escritas necessárias para a realização do trabalho do técnico de manutenção de aeronaves estar em inglês, conforme mencionado, a ICAO não prevê critérios universais de proficiência para esses

profissionais, uma vez que o foco do ensino de inglês está concentrado em pilotos e controladores de tráfego aéreo. Além disso, não há pesquisas abrangentes a respeito do conjunto de características gramaticais e lexicais mais salientes do inglês para aviação que técnicos de manutenção de aeronaves necessitam para realizar a manutenção aeronáutica de forma eficaz. Os poucos estudos existentes têm como foco os erros de comunicação e fatores humanos em geral que resultaram em incidentes de manutenção (DRURY, MA, 2003 e 2005; FAA, 2005). Esta pesquisa pretende preencher em parte esta lacuna, ao suprir a análise linguística abrangente e profunda dos manuais de manutenção aeronáutica, na medida em que os resultados dessa análise possam vir a ser usados como subsídio para melhorar a formação desses profissionais, muito embora esse não seja um objetivo da presente pesquisa.

Assim como na maior parte do mundo, no Brasil também não há requisitos estabelecidos em nível nacional para a proficiência em língua inglesa dos técnicos de manutenção de aeronaves, pois as provas para obtenção de licenças profissionais ministradas pela Agência Nacional de Aviação Civil (ANAC) são realizadas em português. A proficiência nessa língua é a única exigida (ANAC, 2001). Desse modo, por não haver exigências formais de proficiência, os centros de manutenção de aviação executiva e comercial brasileiros, em sua maioria, exigem o conhecimento em língua inglesa como parte dos requisitos para a contratação desses profissionais.

Drury e Ma (2003) afirmam que apesar de o trabalho dos mecânicos de aeronaves ser essencial para a segurança de voo e ser baseado na leitura de manuais em inglês, o conhecimento de língua inglesa por esses profissionais não é garantido, o que favorece a ocorrência de acidentes. O trabalho do mecânico depende da leitura de manuais relativos à aeronave trabalhada e devido à carência de conhecimento de inglês, essa leitura é, muitas vezes, baseada nas figuras, sendo o texto relegado à última instância, o que pode comprometer a segurança, tanto do técnico quanto da aeronave. Espera-se que a identificação das características linguísticas mais salientes dos manuais aeronáuticos realizada nesta pesquisa possa contribuir para o melhor entendimento desses textos, pois os resultados da análise poderão ser utilizados na formação dos técnicos de manutenção de aeronaves, favorecendo a segurança de voo e desses profissionais.

Em geral, o ensino de inglês para aviação no Brasil é realizado por um grupo restrito de professores particulares e institutos de idiomas especializados nessa área, os quais, devido à carência de materiais didáticos voltados especificamente para essa área, desenvolvem seus próprios materiais de ensino baseados no ensino de língua inglesa para fins específicos, sem respaldo acadêmico ou científico (SARMENTO, 2008). O livro *English for Aircraft Maintenance* (SHAWCROSS, 1992) é o único conhecido para o ensino de técnicos de manutenção de aeronaves; no entanto, não é muito utilizado no Brasil, pois não está disponível em larga escala.

Nesse sentido, esta pesquisa visa a fornecer uma contribuição original para a área de inglês para aviação, visto que há uma escassez de trabalhos que lidem com a identificação das características linguísticas mais salientes de manuais de manutenção aeronáutica. Assim, um dos possíveis desdobramentos desta pesquisa é o uso dessas características para informar a elaboração de materiais didáticos para esse público, em especial, os técnicos de manutenção de aeronaves, haja vista que não há materiais didáticos disponíveis no mercado que sejam informados por pesquisa acadêmica. A pesquisa pode colaborar não somente com autores de materiais a serem publicados, mas também com professores particulares e de institutos de idiomas, os quais desenvolvem seus próprios materiais e/ou apostilas.

Esta pesquisa está situada na Linguística Aplicada, a qual, segundo Signorini, tem como objeto de pesquisa “o estudo de práticas específicas de uso da linguagem em contextos específicos” (1998, p.101). Além disso, está embasada nos pressupostos teórico-metodológicos da Linguística de *Corpus* (doravante LC), e, mais especificamente, da sua vertente conhecida por Análise Multidimensional (doravante AMD).

A LC é baseada na visão probabilística de linguagem, ou seja, a linguagem é vista como um sistema probabilístico em que certos traços são mais frequentes que outros (HALLIDAY, 1991, 1992 apud BERBER SARDINHA, 2000). Nesse caso, a linguagem está condicionada a conjuntos de escolhas linguísticas, ou seja: (1) ao sistema (escolhas disponíveis aos usuários da língua; (2) às funções (escolhas feitas pelos usuários); e (3) aos seus contextos. (FIRTH, 1957; BEAUGRANDE, 1991; SINCLAIR, 1991 apud SOUZA, 2012).

A LC tem por objetivo o estudo de uma língua ou variedade linguística. O ponto de partida é a observação de dados provenientes da linguagem em uso, extraídos por meio de ferramentas computacionais e interativas. Isto é, a LC, ao se dedicar à coleta criteriosa e exploração de conjuntos de dados linguísticos, denominados corpora, possibilita a pesquisa de uma língua ou variedade linguística em uso, empiricamente e com o uso de computador (BERBER SARDINHA, 2004b). De acordo com Stubbs (2007), os métodos de análise e os dados provenientes do estudo de corpora fornecem novas maneiras de se pesquisar a relação entre a linguagem e o seu uso.

Mais especificamente, o trabalho aqui proposto fundamenta-se na área de pesquisa baseada em *corpus* que se preocupa com a Análise Multidimensional de variação de registro (sigla AMD) (BIBER, 1988). A questão central desta área é a identificação de dimensões de variação textual entre registros. Por registro, entende-se “uma variedade linguística definida por aspectos situacionais, incluindo o propósito do falante, a relação entre falante e o ouvinte, e o contexto de produção”⁵ (BIBER, 2009, p.823). Na AMD, dimensões são parâmetros situacionais e funcionais dos registros analisados, sendo que “cada dimensão é composta por um grupo independente de características linguísticas coocorrentes, e os padrões de coocorrência podem ser interpretados em termos funcionais”⁶ (BIBER, 1988).

Os primeiros estudos com a abordagem da Análise Multidimensional desenvolvidos por Douglas Biber investigaram as associações entre registros orais e escritos da língua inglesa (BIBER, 1984, 1986 e 1988). A partir desses estudos, as dimensões de variação da língua inglesa foram identificadas (BIBER 1988), sendo elas: (1) produção com interação *versus* produção informacional; (2) preocupações narrativas *versus* não-narrativas; (3) referências explícitas *versus* referências dependentes do contexto; (4) expressão explícita de persuasão *versus* não-explícita; (5) informação abstrata *versus* não-abstrata; e (6) elaboração informacional online. O uso das dimensões de variação do inglês estabelecidas por Biber (1988) permite a pesquisadores entender novos registros ou fazer comparações entre registros especializados e os registros falados e escritos da língua inglesa. Além da aplicação

⁵ “a cover term for any language variety defined by its situational characteristics, including the speaker’s purpose, the relationship between speaker and hearer, and the production circumstances.”

⁶ “Each dimension comprises an independent group of co-occurring linguistic features, and each co-occurrence pattern can be interpreted in functional terms.”

das dimensões de variação do inglês à novos registros, a metodologia da Análise Multidimensional permite a identificação de dimensões de variação próprias de outros registros e de outras línguas.

Algumas pesquisas e trabalhos foram publicadas na área de inglês para aviação, tais como Sarmiento (2008) e Bocorny (2008). Ambas as pesquisas abordaram aspectos específicos da linguagem especializada da aviação, verbos modais e unidades nominais, respectivamente, a fim de gerar subsídios para a produção de material didático e aprimorar o ensino de inglês para aviação, ausentando-se da investigação de outros aspectos linguísticos presentes nos manuais aeronáuticos. Ademais, suas pesquisas foram baseadas em um *corpus* composto por manuais de somente uma aeronave. Almeida e Prado (2011) apresentam o processo de produção do conteúdo programático de um curso especializado em linguagem para técnicos de manutenção aeronáutica, guiado por palavras-chave e padrões lexicogramaticais mais frequentes em um *corpus* de manuais de manutenção de aeronaves. Entretanto, nenhuma dessas pesquisas enfocou o problema da variação linguística entre os diferentes registros textuais encontrados em uma área de especialidade como a aviação, conforme proposto aqui.

Desse modo, pode-se constatar que nossa proposta de estudo difere-se, ao buscar descrever, de forma completa, a composição linguística dos manuais de manutenção de aeronaves, analisando as dimensões de variação de suas características sintáticas e lexicais, e considerando onde esses manuais se situam nas dimensões de variação da língua inglesa estabelecidas por Biber (1988).

Ao mencionar o estudo de um uso específico da língua inglesa, em geral, logo vem à mente a área de Inglês para Fins Específicos (ESP, na sigla em Inglês), a qual se dedica à analisar as características linguísticas de registros especializados em inglês. De acordo com Biber *et al.* (2000), o maior objetivo dessa área é o desenvolvimento de materiais e atividades didáticas fiéis ao registros estudados para que os alunos possam entendê-los e reproduzi-los. No entanto, são poucos os estudos baseados em *corpus* realizados nessa área de pesquisa, ou seja, a grande maioria dos estudos focam um ou alguns poucos aspectos linguísticos e não contemplam uma grande seleção de textos ou a combinação de técnicas de análise quantitativas com interpretações qualitativas (BIBER *et al.*, 2000). Uma desvantagem

desse tipo de abordagem é que “estudos de poucas e seletas características linguísticas não fornecem descrições abrangentes dos registros e, por vezes, podem até mesmo deturpar a caracterização das relações entre os registros”⁷ (BIBER *et al.*, 2000, p.157). Já a análise multidimensional, conforme supracitado, fornece uma descrição mais abrangente e completa das características linguísticas dos registros sendo estudados, além de compará-los em termos linguísticos e funcionais. Dessa forma, optou-se por utilizar a Análise Multidimensional ao invés de uma investigação baseada nos princípios da área de Inglês para Fins Específicos.

Além do que foi mencionado, há uma motivação pessoal para a pesquisa. Ao longo de dois anos ministrando aulas para pilotos, comissários de voo e técnicos de manutenção de aeronaves, pode-se perceber alguns aspectos especiais sobre o ensino de inglês para aviação na formação de técnicos de manutenção de aeronaves que levaram à escolha de se desenvolver a pesquisa nessa área. Primeiramente, as agências reguladoras da aviação civil brasileira não possuem exigência de proficiência em língua inglesa para os mecânicos de aeronaves, apenas para pilotos. Além disso, o curso de formação desses profissionais assim como as provas para obtenção das licenças de trabalho são inteiramente em português, sendo que a única exigência educacional para ingresso no curso de formação é ensino médio completo ou em andamento.

No entanto, após formado e com as licenças de trabalho em mãos, o técnico de manutenção de aeronaves precisa enfrentar entrevistas de emprego em inglês, cujo ensino em sua formação foi inexistente⁸. As entrevistas em inglês ocorrem porque todos os manuais com os quais o técnico de manutenção aeronáutica irá trabalhar estão em língua inglesa e, de acordo com as regulamentações internacionais, não podem ser traduzidos. Em geral, é nesse momento, ou logo após o ingresso no mercado de trabalho, que o mecânico procura cursos para aprender inglês técnico e se depara com uma difícil realidade: há um número muito reduzido de escolas com

⁷ “*studies of a few, selected linguistic features do not provide comprehensive register descriptions, and at times they may even be misleading in their characterization of the relationships among registers.*”

⁸ A lista das matérias ministradas no curso de formação de mecânico de manutenção aeronáutica consta do Regulamento Brasileiro de Homologação Aeronáutica 065 (RBHA065), disponível em <http://www2.anac.gov.br/biblioteca/rbha/rbha065.pdf>. Há uma proposta para o Regulamento Brasileiro da Aviação Civil – RBAC 147 (“Centros de Instrução de Aviação Civil, para formação e qualificação de mecânicos de manutenção aeronáutica”) em processo de audiência pública, cujo anexo traz a inclusão de uma matéria de inglês técnico no curso básico de mecânicos (<http://www2.anac.gov.br/transparencia/pdf/15/anexo.pdf>).

cursos específicos de inglês para aviação e, dessas, poucas possuem cursos para técnicos de manutenção aeronáutica. Ademais, conforme mencionado acima, há apenas um livro didático específico para esse público, o qual não está disponível em larga escala no mercado brasileiro. Desse modo, os materiais de ensino para esses profissionais são elaborados por coordenadores ou professores dessas escolas de idiomas ou por professores particulares, sem qualquer embasamento teórico ou técnico, o que evidencia a necessidade de uma pesquisa abrangente nessa área que possa ser divulgada para esses profissionais de ensino.

A motivação pessoal para essa pesquisa encontra-se principalmente nesse ponto. Como professora em uma escola de inglês para aviação, essa pesquisadora ministrava aulas para os mecânicos com o material desenvolvido pela diretoria, o qual era baseado na tradução de termos técnicos considerados importantes pela diretoria e com os quais os mecânicos haviam relatado dificuldades. O incômodo com essa situação é devido ao fato de que apenas termos já conhecidos e/ou encontrados pelos mecânicos eram contemplados no material, o ensino era feito por meio de tradução de termos isolados e descontextualizados, e não havia preocupação de se estudar o registro como um todo, apenas poucas de suas características linguísticas.

Dentro desse quadro, o objetivo geral deste projeto é identificar as dimensões de variação de manuais de manutenção aeronáutica, por meio da Análise Multidimensional. Os objetivos específicos do projeto são: identificar onde se situam os manuais de manutenção de aeronaves nas dimensões de variação do inglês, de acordo com Biber (1988); identificar as dimensões de variação próprias da linguagem dos manuais de manutenção de aeronaves; e verificar se há diferença estatística significativa entre os manuais de diferentes fabricantes e modelos de aeronaves em relação às dimensões de variação próprias dos manuais aeronáuticos. Dado esse objetivo, as questões de pesquisa a serem investigadas no projeto são elencadas abaixo:

1. Como os manuais de manutenção de aeronaves podem ser mapeados em relação às dimensões de variação do inglês sugeridas por Biber (1988)?
2. Quais são as dimensões de variação próprias dos manuais de manutenção de aeronaves?

3. Em que proporção há diferença estatística significativa entre os manuais dos diferentes fabricantes e os manuais dos diferentes modelos de aeronaves em relação às dimensões de variação próprias dos manuais aeronáuticos?

A partir desses questionamentos, pretende-se contribuir para a área de inglês para aviação com a identificação das características linguísticas mais salientes dos manuais de manutenção aeronáutica, que poderão em futuras pesquisas, serem utilizadas para informar a produção de materiais didáticos específicos para o inglês para aviação e, assim, aprimorar a formação dos profissionais dessa área.

A dissertação está organizada da seguinte maneira. O capítulo a seguir discute a fundamentação teórica da pesquisa, apresentando os principais conceitos e trabalhos prévios nas áreas de Linguística de *Corpus* e Análise Multidimensional. O capítulo 2 apresenta em detalhes a metodologia empregada na pesquisa, incluindo a descrição do *corpus*, bem como a especificação dos procedimentos de coleta, processamento e análise dos dados. O capítulo 3 descreve e interpreta os resultados obtidos por meio da análise dos dados para cada uma das questões de pesquisa, além de apresentar uma discussão dos resultados, fazendo uma apreciação crítica da contribuição dos achados perante a pesquisa prévia apresentada no capítulo 2. Por fim, o capítulo 'Considerações Finais' faz um fechamento do estudo. A bibliografia e os anexos encerram a dissertação.

1 Fundamentação Teórica

Neste capítulo são apresentados os pressupostos teórico-metodológicos da Linguística de *Corpus* (LC), e, mais especificamente, da sua vertente conhecida por Análise Multidimensional (AMD), os quais norteiam a pesquisa.

1.1 Linguística de *Corpus*

Conforme descrito na Introdução, esta pesquisa tem como fundamentação teórica principal a Linguística de *Corpus* (LC), que pode ser definida, de acordo com Berber Sardinha, como a área que

ocupa-se da coleta e da exploração de corpora, ou conjuntos de dados linguísticos textuais coletados criteriosamente, com o propósito de servirem para a pesquisa de uma língua ou variedade linguística. Como tal, dedica-se à exploração da linguagem por meio de evidências empíricas, extraídas por computador (2004b, p.3).

Portanto, a LC enquadra-se dentro do conceito formado pela abordagem empirista com a visão da linguagem como sistema probabilístico. Nesse caso, utilizar a abordagem empirista significa priorizar os dados oriundos da observação da linguagem, em contrapartida a posição racionalista da linguagem.

Para Berber Sardinha, “a visão da linguagem como sistema probabilístico pressupõe que, embora muitos traços linguísticos sejam possíveis teoricamente, não ocorrem com a mesma frequência” (2004b, p.30). Isso significa que embora muitas combinações e características da estrutura da língua sejam possíveis, não ocorrem todas com a mesma frequência. E “o mais importante da diferença de frequências entre os traços é não serem aleatórias” (BERBER SARDINHA, 2004b, p.31). Se essas frequências fossem aleatórias não acrescentariam informações a respeito da estrutura, e o fato das estruturas ocorrerem com frequências diferentes não seria significativo. De acordo com Stubbs (2007), tanto a metodologia da LC quanto os dados resultantes dos estudos de *corpora* fornecem novas formas de se analisar as

relações entre o sistema da língua e a linguagem em uso. Quando o uso de um padrão se torna altamente recorrente em grandes quantidades de texto, ele consolida-se como parte do sistema. Assim, a “frequência no *corpus* é uma evidência observável da probabilidade do sistema”⁹ (HALLIDAY, 1991, 1992, 1993 *apud* STUBBS, 2007, p.130). Portanto, a descrição do que é comum e típico é um dos principais objetivos de estudo da LC.

Berber Sardinha (2004b) acrescenta que há uma correlação entre características linguísticas e contextos de ocorrência da linguagem, isto é, há uma variação sistemática e não-aleatória de grupos de traços linguísticos em relação a textos provenientes de situações comunicativas específicas. Portanto, de acordo com o autor, pode-se afirmar que a linguagem é padronizada (*patterned language*) e essa “padronização se evidencia pela recorrência, isto é, uma colocação, coligação ou estrutura que se repete significativamente mostra sinais de ser, na verdade, um *padrão* lexical ou léxico-gramatical” (2004b, p.31). Esses padrões apresentam regularidade (com frequências que podem ser comparadas em *corpora* diferentes) e variação sistemática (correlação com diversas variedades textuais, dialetais etc.) (BERBER SARDINHA, 2004b).

Segundo Biber, Conrad e Reppen (1998), a identificação desses padrões típicos da linguagem em uso é possível por meio da abordagem baseada em *corpus*, a qual permite a análise de grandes volumes de dados linguísticos e fatores contextuais ou traços linguísticos por meio do uso de computadores. As características essenciais da análise baseada em *corpus* (plural *corpora*), de acordo com os autores (1998), são:

- Empírica, com a análise dos padrões reais de uso em textos naturais;
- Utilização de uma coletânea grande e criteriosa (*principled*) de textos naturais, mais conhecida como um *corpus*, como base para análise;
- Uso extensivo de computadores para análise, com a utilização de técnicas tanto automáticas quanto interativas;
- Depende de técnicas de análise quantitativas e qualitativas.

⁹ “Frequency in the corpus is observable evidence of probability in the system.”

De acordo com Biber, Conrad e Reppen (1998), um *corpus* é uma coletânea grande e criteriosa (*principled*) de textos de linguagem natural. Isso significa que uma coleção de textos qualquer não perfaz um *corpus*. Por coletânea criteriosa entende-se que um desenho de *corpus* com critérios específicos de coleta foi elaborado e seguido, abrangendo questões de representatividade e equilíbrio. Nesse sentido, Hunston afirma que “um *corpus* por si só nunca é bom ou ruim, mas sim adequado ou não a um propósito específico. As decisões acerca do que um *corpus* deve conter são baseadas na finalidade do *corpus*, além do que está disponível”¹⁰ (2002, p.26).

Atualmente, parte-se do pressuposto que todos os textos que perfazem um *corpus* estão em formato eletrônico. No entanto, a coleta de *corpora* para análise linguística é anterior ao advento do computador. Os primeiros trabalhos com *corpora* manuais datam da Antiguidade e eram compostos por citações da Bíblia. Muitas pesquisas de descrição de linguagem foram realizadas por meio de *corpora* desde o início do século XX por pesquisadores como Thorndike e Boas. A grande diferença entre as pesquisas do início e meio do século XX e a época atual é que os *corpora* eram coletados, armazenados e analisados manualmente (BERBER SARDINHA, 2000).

A revolução tecnológica, em especial a partir da década de 1960, tornou possível estudos linguísticos com *corpora* de maiores proporções. Um marco da LC foi o lançamento do primeiro *corpus* eletrônico de linguagem escrita, o *Brown University Standard Corpus of Present-Day American English*, em 1964, contendo 1 milhão de palavras distribuídas em 500 textos de diversos registros escritos em inglês americano e publicados em 1961.

A grande contribuição do *corpus Brown* foi ter estabelecido um padrão para *corpora* eletrônicos ao sistematizar e documentar a coleta de textos e ter sido disponibilizado a outros pesquisadores (SOUZA, 2012). Sua criação serviu de inspiração para a criação do *corpus* Lancaster/Oslo/Bergen (LOB), um *corpus* espelho em inglês britânico. O *corpus* LOB foi coletado entre 1970 e 1978 em uma parceria entre pesquisadores da Universidade de Lancaster, Universidade de Oslo e do Centro de Computação para Humanidades Norueguês. Assim como o *corpus Brown*, O LOB

¹⁰ “a corpus is neither good nor bad in itself, but suited or not suited to a particular purpose. Decisions about what should go into a corpus are based on what the corpus is going to be used for, but also about what is available.”

possui aproximadamente um milhão de palavras distribuídas em 500 textos de linguagem escrita publicados em 1961, sendo que cada texto contém uma amostra de 2000 palavras. Ambos estão agrupados em 15 categorias de texto, conforme descrito abaixo (Tabela 1):

Categoria / tipo de texto	Número de textos - <i>Brown</i>	Número de textos - <i>LOB</i>
Imprensa: reportagem	44	44
Imprensa: editorial	27	27
Imprensa: resumos ¹¹	17	17
Religião	17	17
Habilidades, negócios e hobbies	36	38
Histórias populares	48	44
Belles lettres, biografias e ensaios	75	77
Miscelânea (documentos de governo, relatórios de fundações, relatórios de indústrias, catálogos de faculdades)	30	30
Escrita de aprendiz e escritos científicos	80	80
Ficção geral	29	29
Ficção de mistério e de detetive	24	24
Ficção científica	6	6
Ficção: aventura e velho oeste	29	29
Romance e história de amor	29	29
Humor	9	9

Tabela 1. Distribuição dos registros dos *corpora Brown* e *LOB* (SOUZA, 2012)

O *Survey of English Usage* (SEU) da *University College London* é outro *corpus* de grande importância para a LC e tido como referência para subsequentes *corpora*. Compilado manualmente a partir de 1953 por um time de pesquisadores liderados por Randolph Quirk, o *corpus* continha um milhão de palavras de linguagem escrita e oral quando encerrado. Seu desenho seguiu critérios específicos, sendo composto por um número fixo de textos (200) contendo 5000 palavras cada. O SEU tornou-se um *corpus* eletrônico em 1989, contendo não somente os textos originais, mas também textos de linguagem falada de um projeto irmão, o *Survey of Spoken English* (SEE) da *Lund University*. A parceria entre os dois projetos, SEU e SEE, resultou na criação do *London-Lund Corpus of Spoken English* (LLC). O LLC continha 87 textos em sua

¹¹ Pode indicar resumos ou resenhas.

versão original, porém 13 textos foram posteriormente adicionados, totalizando 100 textos de linguagem oral.

Embora todos os *corpora* mencionados acima sejam igualmente importantes para a LC em geral, os *corpora* LOB e LLC têm maior relevância para esta pesquisa. Douglas Biber (1985, 1988, 1989) baseou-se nesses dois *corpora* para determinar as dimensões de variação da língua inglesa falada e escrita, sendo sua teoria e metodologia a base desta pesquisa.

Nos dias de hoje, além das pesquisas baseadas nos *corpora* *Brown*, LOB e LLC, os *corpora* de inglês geral mais utilizados para pesquisa em LC são: o *British National Corpus* (BNC) de 1995, que contém 100 milhões de palavras e é composto por inglês britânico, falado e escrito; o *Bank of English* lançado em 1987, que contém 450 milhões de palavras e é composto por inglês britânico (contém um *subcorpus* com 56 milhões de palavras disponível para o ensino de língua inglesa); e o *Corpus of Contemporary American English* (COCA) lançado em 2008, que contém mais de 460 milhões de palavras e é composto por inglês americano, falado e escrito. Há diversos tipos de *corpora* e a classificação deles varia conforme o tamanho, propósito e maneira como forem compilados. Sarmiento (2009) apresenta uma lista dos tipos de *corpora* sugeridos por Sinclair (1 a 4) e Hunston (5 a 8):

1. *Corpus* Geral: contém diversos tipos de textos, sendo eles provenientes da linguagem falada, escrita ou ambas; produzidos em um determinado país ou em vários. Por ser um *corpus* de língua geral, é necessário ter a maior variedade de tipos de textos possível, com dimensões muito maiores do que um *corpus* específico. São muitas vezes chamados de *Corpus* de Referência por serem utilizados para contrastar estudos com *corpora* especializados. Um dos mais famosos é o *British National Corpus* (BNC).

2. *Corpus* Monitor: projetado para averiguar as mudanças recentes em uma língua. É alimentado com frequência (diária, mensal ou anualmente) e sempre com a mesma proporção de tipos de textos a fim de que se possam comparar cada período com o anterior. O *Corpus of Contemporary American English* (COCA) é um exemplo desse tipo, contendo atualmente 464.020.256 palavras.

3. *Corpus* Comparável: composto por dois ou mais *corpora* em línguas diferentes ou diferentes variedades de uma mesma língua, cuja compilação segue os mesmos critérios. Esses tipos de *corpora* são utilizados em sua maioria por tradutores e/ou aprendizes de uma língua para identificar suas diferenças e equivalências. Um exemplo é o CorTec do Projeto COMET da Universidade de São Paulo, que contém dezenove *corpora* técnicos comparáveis.
4. *Corpus* Paralelo: composto por dois ou mais *corpora* em línguas diferentes contendo textos originais e suas respectivas traduções ou textos produzidos ao mesmo tempo em duas ou mais línguas, como, por exemplo, as normas da União Europeia.
5. *Corpus* de Aprendiz: composto por textos ou redações produzidas por aprendizes de uma determinada língua. Esse tipo de *corpus* tem o objetivo de identificar os mecanismos de aquisição de segunda língua ou língua estrangeira e/ou contrastar aspectos em que os aprendizes diferem entre si e em relação a falantes nativos. É provável que o maior *corpus* de aprendiz seja o *International Corpus of Learner English* (ICLE) contendo produções escritas de aprendizes de língua inglesa falantes de 19 línguas nativas diferentes, incluindo o português, cujo *subcorpus* Br-ICLE está armazenado na PUCSP.
6. *Corpus* Pedagógico: composto pela linguagem que se expõe a um aprendiz. Pode conter livros didáticos e gravações.
7. *Corpus* Histórico ou Diacrônico: composto por textos provenientes de diferentes períodos de tempo. Seu propósito é identificar a evolução de determinados aspectos de uma língua através do tempo. Um exemplo é o *corpus* da revista TIME, que abrange o conteúdo publicado na revista de 1923, ano em que foi lançada, à 2006.
8. *Corpus* Especializado: composto por um tipo específico de texto, registro ou gênero textual, definido de acordo com a tipologia de *corpus*, cujo objetivo é representar um determinado tipo de texto ou linguagem. Geralmente pode ser compilado pelo próprio pesquisador. Apesar de não haver limites para o grau de especialização envolvido, devem-se seguir os parâmetros da tipologia textual escolhida. Esse é o tipo de *corpus* utilizado nesta pesquisa.

Segundo O’Keeffe (2010), um *corpus* especializado permite um equilíbrio entre as análises quantitativa e qualitativa, sendo possível encontrar sutilezas que não necessariamente apareceriam em um *corpus* de linguagem geral. Para a autora, em um *corpus* especializado, de dimensões menores, pode-se encontrar maior concentração de padrões da língua, padrões de uso localizados e especialização pragmática. Nesse caso, há maior concentração de cobertura do registro de linguagem sendo estudado e maior consciência do contexto de uso por parte do usuário do *corpus*. O’Keeffe lista as vantagens da utilização de *corpora* especializados como sendo:

Em primeiro lugar, porque eles são cuidadosamente orientados, consistem em dados que provavelmente representam o domínio mais fielmente que *corpora* que tenham como objetivo capturar tudo sobre a língua como um todo. Em segundo lugar, léxico e estruturas especializadas são prováveis de ocorrer com maior padronização e distribuição regular, mesmo com quantidades relativamente pequenas de dados. Em terceiro lugar, os objetivos pedagógicos em termos de como eles são utilizados e aplicados são susceptíveis de serem mais fáceis de definir e limitar (2007, p. 198).¹²

O’Keeffe (2007) afirma que, geralmente, exercemos maior cuidado ao estabelecer os critérios de construção do *corpus* para garantir sua representatividade. Segundo a autora, uma vez que os *corpora* são armazenados em computadores, é possível compilar quantidades maciças de textos para serem analisados por *software* específicos. No caso, não há um padrão pré-estabelecido para o tamanho ideal de um *corpus*. Ele está relacionado à velocidade e eficiência do software de acesso a esse *corpus*, à capacidade de armazenamento do computador, e, mais importante, ao objetivo da pesquisa em que será utilizado. Biber (1993a) discorre acerca da relevância da representatividade para o desenho de *corpus*. Para o autor, a representatividade está associada a extensão de variabilidade de uma determinada população presente na amostra de linguagem que perfaz o *corpus*. Nesse caso, o desenho do *corpus* deve contemplar essa variabilidade sob perspectivas linguísticas e situacionais a fim de determinar a representatividade da amostra de linguagem, ou

¹² “*Firstly, because they are carefully targeted, the data they consist of is likely to represent the target domain more faithfully than corpora which set out to capture everything about a language as a whole. Secondly, specialized lexis and structures are likely to occur with more regular patterning and distribution, even with relatively small amounts of data. Thirdly, the pedagogical goals in terms of how they are used and applied are likely to be easier to define and delimit.*”

seja, deve incluir a diversidade de tipos de texto e a distribuição linguística existentes numa língua (BIBER, 1993a). Além disso, de acordo com Biber e Jones, na LC,

pesquisas de uso de linguagem são, em geral, concebidas como estudos quantitativos com o objetivo de alcançar resultados generalizáveis correspondentes a um domínio de uso. O desenho de *corpus* é essencial para alcançar esse objetivo: o *corpus* deve ser representativo do domínio de uso para que análises subsequentes possam ser generalizadas acerca desse domínio¹³ (2009, p.1287).

Biber e Jones (2009) afirmam que tanto a composição, ou seja, as categorias de texto contempladas, quanto o tamanho em número de palavras ou textos são fatores determinantes da representatividade de um *corpus*.

Segundo Atkins (1992), a simples existência de um *corpus* não é suficiente para levantar dados linguísticos. Pesquisas baseadas em *corpus* requerem o uso de um grupo de ferramentas computacionais de análise quantitativa de dados, sendo que a análise qualitativa baseada nos resultados quantitativos obtidos é realizada posteriormente. Em geral, o processamento do *corpus* para análise inclui a anotação e/ou etiquetagem dos textos, análise lexical e análise estatística.

A anotação é o primeiro passo na análise de dados e, de acordo com Souza, consiste no “processo de adicionar marcações ao conjunto de texto que permitam facilitar a análise a partir da classificação de cada uma das palavras do texto de acordo com uma categoria, seja ela morfológica, sintática ou semântica” (2012, p.87). Assim, na anotação, informações referentes aos metadados e marcações textuais e linguísticas são inseridas no *corpus*.

Metadados são informações descritivas sobre o texto e sua fonte. McEnery e Hardie (2011) afirmam que metadados podem ser informações a respeito do autor, da publicação, falantes, entre outras. No caso de marcações textuais, informações estruturais do texto, tais como quebras de parágrafo, indicações de final e início de fala de falantes são anotadas no *corpus*. Esses tipos de marcações podem ser feitas manualmente ou por meio de ferramentas computacionais, em metalinguagens

¹³ “*corpus linguistic investigations of language use are usually designed as quantitative studies with the goal of generalizable findings representing some domain of use. The design of a corpus is a fundamentally important consideration for achieving this goal: the corpus must be representative of the target domain of use in order for subsequent analyses to be generalized to that domain.*”

padrão como *Standard Generalized Markup Language* (SGML) e *Extensible Markup Language* (XML), sendo XML a mais utilizada atualmente.

De acordo com Berber Sardinha (2004), há quatro tipos de anotação linguística: (1) morfossintática ou marcação de partes do discurso (*part of speech* – POS), (2) sintática (*parsing*), (3) semântica (*semantic*) e (4) discursiva (*discourse*). A etiquetagem ou marcação morfossintática, semântica e discursiva é realizada por programas denominados “etiquetadores” (*taggers*) e consiste em adicionar “um código a cada palavra do *corpus*, indicando a parte do discurso”¹⁴ (HUNSTON, 2002, p.18) a que ela pertence, por exemplo, a palavra *aircraft* receberia o código de substantivo. A marcação sintática é realizada por programas denominados *parsers*, os quais adicionam etiquetas contendo informações que identificam estruturas sintáticas, como por exemplo, sintagmas nominais e verbais.

Tanto os etiquetadores quanto os *parsers* podem ser baseados em regras ou serem probabilísticos. Etiquetadores e *parsers* baseados em regras inserem etiquetas com base em regras gramaticais, no caso de etiquetadores, e em concepções de língua, no caso de *parsers*, pré-estabelecidas no programa. Já etiquetadores e *parsers* probabilísticos “atribuem etiquetas com base na probabilidade estatística de uma determinada etiqueta ocorrer em um determinado contexto” (MEYER, 2002, p.88 *apud* SOUZA, 2012, p.89).

É importante ressaltar que nenhuma etiquetagem, seja ela manual ou automática, é 100% eficaz, ou seja, sempre há uma margem de erro. Segundo McEnery e Hardie (2011), é necessário considerar a consistência da anotação linguística. Anotações tanto manuais quanto automáticas são inconsistentes, pois uma das propriedades de toda análise linguística é que uma interpretação representa uma escolha entre várias possibilidades. Apesar disso, para os autores, a anotação linguística é válida pois as escolhas de um analista ou do software estão presentes de forma explícita no texto, deixando em evidência para os usuários dos dados as inconsistências. Desse modo, a replicabilidade da pesquisa é facilitada.

Nesta pesquisa, o etiquetador *Biber Tagger* foi utilizado para a anotação do *corpus*, conforme descrito na seção 2.2.1. O *Biber Tagger* opera em dois níveis: na

¹⁴ “[addition of] a code to each word in a corpus, indicating the part of speech”

"(1) identificação inicial da categoria gramatical de cada palavra, e (2) solução de ambiguidades nos casos em que o dicionário apresenta mais de uma possibilidade de categoria gramatical para uma palavra"¹⁵ (BIBER, 1988, p.214). A primeira versão do *tagger* (BIBER, 1988) fazia uso exclusivo de uma gama de algoritmos dependentes de contexto para resolver ambiguidades. No entanto, mais recentemente, a abordagem probabilística foi incorporada ao etiquetador (BIBER, 1993b). De acordo com Biber (1988), a margem de erro do *tagger* é de menos de 10%, ou seja, tem mais de 90% de eficácia. Além disso, o *Biber Tagger* pode ser considerado confiável pois é o *de facto tagger* usado em pesquisas de AMD.

A análise lexical do *corpus* é feita por meio de ferramentas computacionais que permitem a observação de padrões linguísticos recorrentes e a comparação entre *corpora*. Em geral, análises lexicais por meio da LC incluem, mas não estão limitadas a: linhas de concordância (ou KWIC – *key word in context*), listas de frequência de palavras, listas de colocados, listas de pacotes lexicais e listas de palavras chave. Para esse tipo de análise o *corpus* pode estar anotado ou não. Para que a análise seja realizada, essas informações são extraídas por meio de softwares específicos para esse fim, tais como *WordSmith Tools*¹⁶ e *AntConc*¹⁷. Ambos os programas contêm ferramentas semelhantes para a extração de dados do *corpus*; no entanto, o *WordSmith Tools* foi desenvolvido para uso por pesquisadores e o *AntConc* para o uso na sala de aula. Justamente por ter uma interface mais simples para o uso por professores e alunos, o *AntConc* tem sido usado em um número crescente de pesquisas.

Análises estatísticas também podem ser usadas na análise de dados provenientes de um *corpus* anotado. Para isso, pacotes estatísticos são usados, sendo o *Statistical Analysis System* (SAS) e o *Statistical Package for the Social Sciences* (SPSS) os mais comuns. Nesse tipo de análise, não são os textos que compõem o *corpus* que são inseridos na ferramenta, mas sim a contagem de frequência das etiquetas anotadas nos textos. Ferramentas computacionais denominadas *Counters* são usadas para contabilizar as etiquetas presentes no *corpus*

¹⁵ "(1) initial identification of the grammatical category of each word, and (2) in cases where the dictionary lists more than one possible grammatical category for a word, resolution of ambiguities."

¹⁶ O software pago *WordSmith Tools* foi desenvolvido por Mike Scott em 1996. Atualmente está na versão 6.0.

¹⁷ O software livre *AntConc* foi desenvolvido por Laurence Anthony em 2002. Atualmente está na versão 3.4.1.

e os resultados são apresentados em planilhas (no Microsoft Excel, por exemplo). Essas planilhas, por sua vez, são inseridas no pacote estatístico escolhido para serem processadas. A partir dos dados importados para o pacote estatístico, torna-se possível agrupar textos de acordo com padrões de co-ocorrência ou variação linguística. Para isso, estatísticas descritivas e análises fatoriais são empregadas.

Nesta pesquisa, a análise estatística figura como metodologia de análise quantitativa para a identificação de padrões de co-ocorrência e variação linguística nos manuais aeronáuticos. O uso do pacote estatístico SPSS é descrito nas subseções 2.2.2.1 e 2.2.2.2.

1.2 Análise Multidimensional

Tradicionalmente, o foco da Linguística de *Corpus* encontra-se no estudo dos padrões de linguagem. Porém, o estudo da comparação da linguagem de diferentes textos ou registros e da variação das características linguísticas que compõem esses textos ou registros são igualmente importantes (BIBER *et al.*, 2000), sendo este o foco da Análise Multidimensional (AMD).

A AMD, desenvolvida na obra de Douglas Biber (1988), é uma metodologia da LC para o estudo de variação entre variedades textuais situacionalmente definidas (registros), que engloba análises tanto quantitativa, realizada por meio de procedimentos estatísticos, quanto qualitativa, ao interpretar os conjuntos de variáveis resultantes da aplicação de soluções estatísticas. Berber Sardinha afirma que a AMD “foi criada por Douglas Biber com o objetivo de permitir uma descrição rica e complexa de *corpora* inteiros de textos por meio estatísticos bem como a extração precisa de características textuais em comum entre *corpora*” (2000, p.101).

Essa metodologia baseia-se no pressuposto de que padrões de co-ocorrência de traços linguísticos são marcadores de dimensões funcionais da linguagem e, por isso, destina-se a “(1) identificar os padrões de co-ocorrência salientes da linguagem

(...) e (2) comparar registros no espaço linguístico definido por tais padrões”¹⁸ (BIBER, 1988; 2009, p. 824). Inicialmente, a metodologia foi desenvolvida em 1985, sendo refinada e publicada posteriormente na obra de 1988 (BIBER, 1988).

Desse modo, a questão central é a identificação de dimensões de variação textual entre registros de uma língua ou variedade linguística. Por registro, entende-se “uma variedade linguística definida por aspectos situacionais, incluindo o propósito do falante, a relação entre falante e o ouvinte, e o contexto de produção”¹⁹ (BIBER, 2009, p. 823). Na AMD, dimensões são parâmetros situacionais e funcionais dos registros analisados, sendo que “cada dimensão é composta por um grupo independente de características linguísticas co-ocorrentes, e os padrões de co-ocorrência podem ser interpretados em termos funcionais”²⁰ (BIBER, 1988, p. 14).

Assim, cada manual utilizado pelos técnicos de manutenção aeronáutica compõe um registro diferente, pois é definido pelo seu contexto de uso. Por exemplo, em termos gerais, o manual denominado *Systems Description Section* (SDS) é usado para entender como funciona os sistemas da aeronave e o manual denominado *Minimum Equipment List* (MEL) é usado para reparar falhas e liberar uma aeronave para voo. É importante ressaltar que um registro não é uma unidade linguística, mas sim uma unidade situacional e/ou contextual. Portanto, dois registros podem ser iguais linguisticamente e continuar sendo dois registros. Por exemplo, os registros reunião de técnicos de manutenção e reunião de inspetores de manutenção em um hangar provavelmente são idênticos linguisticamente, mas são dois registros diferentes porque os participantes, os rituais, as intenções e a finalidade de cada registro são diferentes. As dimensões são o resultado do agrupamento desses registros de acordo com as características linguísticas, situacionais e funcionais identificadas, destacando um continuum entre suas semelhanças e diferenças.

¹⁸ “(1) identify the salient linguistic co-occurrence patterns in a language, in empirical/quantitative terms; and (2) compare registers in the linguistic space defined by those co-occurrence patterns.”

¹⁹ “a cover term for any language variety defined by its situational characteristics, including the speaker’s purpose, the relationship between speaker and hearer, and the production circumstances.”

²⁰ “Each dimension comprises an independent group of co-occurring linguistic features, and each co-occurrence pattern can be interpreted in functional terms.”

Segundo Biber (2009), essa variação é inerente à linguagem, uma vez que as escolhas lexicais, morfológicas, gramaticais e de pronúncia, entre outras, de cada falante refletem uma gama de fatores situacionais.

Diversos autores estudaram a distribuição de características linguísticas em diferentes situações e grupos sociais, ou mencionaram padrões de co-ocorrência como fatores de análise de variação linguística. Bernstein (1970; 1972) descreve códigos elaborados *versus* restrito; Ervin-Tripp (1972) e Irvine (1979) discorrem acerca da variação entre registros formais e informais; e Ervin-Tripp (1972) e Hymes (1974) definem as relações de co-ocorrência entre características linguísticas como ‘estilos de fala’, sendo cada estilo um registro marcado por características linguísticas co-ocorrentes.

De acordo com Ervin-Tripp (1972), as mudanças em co-ocorrências realizadas em diversos níveis de estrutura linguística de uma dada linguagem podem ser analisadas em termos de estilo de linguagem. A autora define estilo como sendo o “termo geralmente usado para denominar as mudanças co-ocorrentes em diversos níveis de estrutura linguística presentes em uma determinada linguagem” (1972, p.235).

Um dos estilos de fala definidos por Ervin-Tripp (1972) está relacionado à linguagem profissional. De acordo com a autora, as mudanças na linguagem ocupacional podem surgir de três fatores: (1) valores, conceitos e objetos inerentes a determinada atividade profissional podem ser conhecidos somente por seus especialistas; (2) escolhas léxico-gramaticais adequadas na fala ou na escrita podem fazer alusão à solidariedade ou identificação compartilhada por um grupo de trabalho; e (3) algumas características linguísticas podem surgir de situações de comunicação específicas, tais como mudanças na sintaxe em pedidos de restaurantes, por exemplo.

Ervin-Tripp (1972) afirma que todos os registros pertencentes a uma linguagem possuem características sintáticas e lexicais, sendo que os detalhes das “regras” de co-ocorrência dessas características precisam ser estudadas. Um estudo detalhado dessas co-ocorrências foi realizado por Biber (1988). O autor parte dos aspectos linguísticos para os aspectos funcionais da linguagem ao analisar, por meio da análise

multidimensional, os padrões de co-ocorrência de características gramaticais e lexicais de diversos registros da língua inglesa.

Apesar de o conceito de padrões de co-ocorrência, conforme supracitado, estar bem fundamentado na literatura da sociolinguística, Biber (1988, 1993b) afirma que há três diferenças essenciais entre os conceitos definidos anteriormente e o conceito de dimensão usado na AMD. Primeiramente, a maioria dos estudos anteriores avaliou a variação linguística baseando-se em um único parâmetro ou dimensão. A AMD parte do princípio que a descrição de variação linguística de uma determinada linguagem deve ser sempre multidimensional. Segundo, esses estudos tratam as dimensões de variação como dicotomias ao invés de variação contínua. Biber identifica as dimensões de variação da AMD como “parâmetros de variação contínua e quantificável, ou seja, como uma escala” (1988, p.22). Desse modo, estilos, registros, gêneros e tipos de texto são classificados de acordo com os graus de semelhança ou diferença em relação a cada dimensão ao invés de serem relacionados em termos de dois opostos. Isso é possível devido ao uso de técnicas de estatística e do conceito quantitativo de dimensão. Por fim, estudos anteriores foram baseados em análises funcionais para a identificação de características linguísticas distintas entre diferentes registros. Na AMD, a análise é feita com base na frequência das características linguísticas nos textos, para que, a partir disso, soluções estatísticas possam fornecer uma especificação quantitativa precisa dos padrões de co-ocorrência de um grupo de características linguísticas.

Conforme mencionado anteriormente, a análise dos resultados estatísticos provenientes da fatoração a fim de nomear as dimensões de variação é qualitativa. Desse modo, o pesquisador deve, a partir da leitura dos textos do *corpus*, analisar, não somente os padrões de co-ocorrência de grupos de características linguísticas, mas também as funções situacionais, comunicativas e funcionais que esses padrões representam.

O trabalho realizado por Douglas Biber (1988) foi pioneiro e possibilitou a identificação das dimensões de variação da língua inglesa, falada e escrita. Para isso, o autor utilizou uma seleção de textos de dois *corpora*, o LOB *Corpus*, contendo 15 registros de linguagem escrita, e o *London-Lund Corpus of Spoken English* (LLC), contendo seis registros de textos orais, além de dois tipos de cartas pessoais,

totalizando 23 registros orais e escritos a fim de representar a variedade de registros falados e escritos presentes na língua inglesa. A tabela abaixo apresenta a composição do *corpus* utilizado por Biber (1988) (Tabela 2).

Registro	Número de textos	Número aproximado de ocorrências
Linguagem escrita - LOB Corpus (1-15)		
1. Biografia (<i>biography</i>)	14	30.000
2. Cultura popular (<i>popular lore</i>)	14	30.000
3. Documento oficial (<i>official document</i>)	14	28.000
4. Imprensa – editorial (<i>editorial</i>)	27	54.000
5. Imprensa – reportagem (<i>press reportage</i>)	44	88.000
6. Imprensa – resumo (<i>press review</i>)	17	34.000
7. Ficção científica (<i>science fiction</i>)	6	12.000
8. Ficção de aventura (<i>adventure fiction</i>)	13	26.000
9. Ficção de mistério (<i>mystery fiction</i>)	13	26.000
10. Ficção geral (<i>general fiction</i>)	29	58.000
11. Ficção romântica (<i>romantic fiction</i>)	13	26.000
12. Hobbies (<i>hobbies</i>)	14	30.000
13. Humor (<i>humor</i>)	9	18.000
14. Prosa acadêmica (<i>academic prose</i>)	80	160.000
15. Religião (<i>religion</i>)	17	34.000
16. Carta pessoal (<i>personal letter</i>)	6	6.000
17. Carta profissional (<i>professional letter</i>)	10	10.000
Linguagem oral - LLC Corpus		
18. Conversa face a face (<i>face-to-face conversation</i>)	44	115.000
19. Conversa telefônica (<i>telephone conversation</i>)	27	32.000
20. Discurso espontâneo (<i>spontaneous speech</i>)	16	26.000
21. Discurso preparado (<i>prepared speech</i>)	14	31.000
22. Entrevista (<i>interview</i>)	22	48.000
23. Transmissões (<i>broadcasts</i>)	18	38.000
TOTAL	481	960.000

Tabela 2. *Corpus* utilizado na AMD realizada por Biber em 1988 (adaptada de SOUZA, 2012)

Uma análise da literatura foi realizada para identificar uma gama de características linguísticas possivelmente importantes e que pudessem estar relacionadas a funções comunicativas diversas. No total, foram selecionadas 67

características linguísticas, de natureza lexical e gramatical, divididas em 16 categorias, a saber: (1) marcadores de tempos e aspectos verbais (presente, passado, perfeitos); (2) advérbios de lugar e de tempo; (3) pronomes; (4) interrogações; (5) formas nominais; (6) passivas; (7) verbos de estado (*do* como verbo principal e *there* indicando existência); (8) características de subordinação (complementos, infinitivos, formas de particípio, pronomes relativos, orações subordinadas adjetivas e adverbiais); (9) frases preposicionadas, adjetivos (atributivos e predicativos) e advérbios (total); (10) especificidade lexical (razão forma/item e tamanho das palavras); (11) classes lexicais (conjunções, salvaguardas, pronomes demonstrativos etc.); (12) modais (de possibilidade, de necessidade e de predição); (13) classes especiais de verbo (*dicend*²¹, de cognição e de persuasão); (14) formas reduzidas (contrações, por exemplo); (15) coordenação; e (16) negação (sintética e analítica).

As características linguísticas foram utilizadas para anotar, manual e automaticamente, os textos que perfaziam o *corpus*. A anotação foi realizada por meio de etiquetagem morfossintática, processo pelo qual cada palavra do texto recebe uma etiqueta, isto é, um código contendo informações lexicais e/ou gramaticais a seu respeito. Após a conclusão da anotação do *corpus*, Biber deu prosseguimento à análise. A análise é dividida em duas partes, quantitativa e qualitativa. Na primeira, Biber (1988) utilizou um procedimento estatístico denominado Análise Fatorial. Nesse tipo de análise, utiliza-se a contagem de frequência das características linguísticas, sendo cada traço linguístico correspondente a uma variável original, para identificar “fatores”. Os fatores são combinações lineares das variáveis originais, e representam áreas que podem ser resumidas ou generalizadas. De acordo com Biber, “cada fator representa uma área de grande variação nos dados, um grupo de traços linguísticos que coocorrem com grande frequência”²² (BIBER, 1988, p.79). Isto é, na AMD, a frequência de cada característica linguística é contabilizada em cada um dos textos que perfazem o *corpus*. Com base nessas frequências e por meio de soluções estatísticas, as características linguísticas que tendem a co-ocorrer podem ser agrupadas de forma empírica. As características linguísticas podem ocorrer positiva e negativamente em cada fator e não há juízo de valor na identificação de características

²¹ Verbos *dicendi* são aqueles que antecedem uma pergunta ou uma declaração, tais como afirmar, dizer, exclamar, perguntar, responder, em português, ou *assert, declare, state, hint, report*, em inglês.

²² “each factor represents an area of high shared variance in the data, a grouping of linguistic features that co-occur with a high frequency.”

positivas e negativas. Tanto as características positivas quanto as negativas tendem a co-ocorrer nos textos, sendo que a presença de diversas características positivas indica a tendência de escassez de características negativas. Isso significa que um texto com escore positivo pode conter características negativas, porém o grupo de características positivas que co-ocorrem nesse texto são mais salientes. Portanto, é importante ressaltar que os fatores não representam dicotomias, e sim, um contínuo. De acordo com Berber Sardinha (2004a, p.6), “em cada dimensão, os textos podem se situar ao longo de uma escala que vai de ‘mais’ a ‘menos’ em relação a cada traço comunicativo”.

Na segunda parte da análise, os fatores são interpretados e nomeados, transformando-se em dimensões. Nessa etapa, a análise qualitativa é empregada para que as funções relacionadas a cada agrupamento de características linguísticas co-ocorrentes possam ser interpretadas (BIBER, 2009). De acordo com Biber,

O conteúdo linguístico de uma dimensão é composto por um grupo de características linguísticas (ex. nominalizações, sintagmas preposicionais, adjetivos atributivos) que co-ocorrem com frequência elevada nos textos. Baseando-se no pressuposto de que co-ocorrência reflete funções compartilhadas, esses padrões de co-ocorrência são interpretados em termos das funções situacionais, sociais e cognitivas mais amplamente compartilhadas pelas características linguísticas²³ (2009, p.824).

Assim, a interpretação dos fatores considera que funções comunicativas subjacentes são refletidas nos padrões de co-ocorrência linguística, ou seja, características linguísticas co-ocorrem porque têm funções comunicativas similares (BIBER, 2009). Além disso, deve-se levar em consideração por que conjuntos de características positivas e negativas ocorrem em oposição.

Originalmente, Biber identificou sete fatores. Durante o processo de interpretação, o sétimo fator foi descartado por não permitir uma interpretação completa. Sendo assim, seis fatores foram mantidos, resultando na subsequente identificação de seis dimensões, a saber:

²³ “*The linguistic content of a dimension comprises a group of linguistic features (e.g., nominalizations, prepositional phrases, attributive adjectives) that co-occur with a high frequency in texts. Based on the assumption that co-occurrence reflects shared function, these co-occurrence patterns are interpreted in terms of the situational, social, and cognitive functions most widely shared by the linguistic features.*”

- (1) produção com interação *versus* produção informacional;
- (2) preocupações narrativas *versus* não-narrativas;
- (3) referências explícitas *versus* referências dependentes do contexto;
- (4) expressão explícita de persuasão *versus* não-explícita;
- (5) informação abstrata *versus* não-abstrata; e
- (6) elaboração informacional *on-line*.^{24 e 25}

Posteriormente, em novos estudos, Biber (2009) descartou a sexta dimensão e revisou o nome das dimensões. Além disso, a polaridade da terceira dimensão foi revertida. Vale destacar que esses novos estudos e essa revisão de nomenclatura realizados em 2009 foram feitos com base no *corpus* e na fatoração da pesquisa de 1988. Observe a nomenclatura revisada a seguir (Tabela 4):

Dimensão	Biber 1988	Biber 2009
1	Produção com interação <i>versus</i> informacional	Produção com envolvimento <i>versus</i> informacional
2	Preocupações narrativas <i>versus</i> não-narrativas	Discurso narrativo <i>versus</i> não-narrativo
3	Referências explícitas <i>versus</i> dependentes do contexto	Referência dependente da situação <i>versus</i> elaborada
4	Persuasão explícita <i>versus</i> não-explícita	Argumentação explícita
5	Informação abstrata <i>versus</i> não-abstrata	Estilo abstrato <i>versus</i> não-abstrato
6	Elaboração informacional <i>on-line</i>	---

Tabela 3. Nomenclatura das dimensões do inglês em 1988 e 2009 (adaptada de BERBER SARDINHA, 2013)

A primeira dimensão, “produção com interação *versus* produção informacional”, compreende diferenças entre a linguagem falada e escrita. O polo positivo é marcado por conteúdo interacional, geral e com foco em posicionamento, tendo como principais registros conversas telefônicas e face a face. As características do polo positivo são:

²⁴ Termos traduzidos por Oliveira (1997, p.66).

²⁵ “(1) *involved versus informational production*, (2) *narrative versus non-narrative concerns*, (3) *explicit versus situation-dependent reference*, (4) *overt expression of persuasion*, (5) *abstract versus non-abstract information*, (6) *on-line information elaboration*” (BIBER, 1988p.122).

a presença de verbos (de cognição, verbos *do* e *be*, modais e verbos no presente), formas contraídas, pronomes (pessoal em primeira e segunda pessoas, *it*, demonstrativo, nominal, interrogativo *wh-*), advérbios (de qualidade, delimitador e amplificador), conjunções (subordinada causal e coordenada com conectivo), preposição desacompanhada, partícula do discurso e da omissão de *that*. O polo negativo é marcado por registros informacionais, como documentos oficiais e prosa acadêmica, ou seja, pela alta densidade informacional e escolha precisa de palavras, e contém as seguintes características linguísticas: presença de substantivos, preposições, adjetivos atributivos, palavras longas (*word length*) e relação entre tipos e ocorrências.

A segunda dimensão, “preocupações narrativas *versus* não-narrativas”, compreende a presença e ausência de características que marcam textos narrativos. Apenas um polo contém características estatisticamente relevantes, o positivo, a saber verbos no tempo passado e no aspecto perfeito, por verbos *dicendi*, orações com particípio presente, pronomes em terceira pessoa e negação sintética. A dimensão é marcada, por um lado, por todos os registros de ficção, o que reflete a ênfase em discursos narrativos, e, por outro, por registros com menos preocupações narrativas, como transmissões e documentos oficiais.

A terceira dimensão, “referências explícitas *versus* referências dependentes do contexto”, compreende, no polo positivo, registros que trazem referências explícitas e elaboradas para evitar ambiguidade, como documentos oficiais e cartas profissionais, e, no polo negativo, registros contendo referências a tempos e lugares que fazem parte de um saber compartilhado, tais como transmissões, conversas telefônicas e face a face. As características presentes no polo positivo são: uso de pronomes iniciados por *wh-* em orações relativas na posição de sujeito e de objeto com a presença ou não da preposição em posição frontal (*pied-piping*), conjunção coordenada com conector frasal e nominalização com substantivo no singular, que denotam maior explicitação. O polo negativo contém advérbios de tempo e lugar e outros advérbios.

A quarta dimensão, “expressão explícita de persuasão *versus* não-explícita”, assim como a dimensão 2, contém apenas características positivas, marcando um continuum entre a presença e ausência de características que marcam persuasão. As

características que perfazem o polo positivo e que, de acordo com Biber (1988, 2009), juntas funcionam para estruturar argumentos e apresentar o posicionamento do autor, são: verbos no infinitivo; verbos modais de predição, necessidade e possibilidade; verbos de persuasão (por exemplo, *agree*, *ask*, *insist*, *recommend*); advérbios; e conjunções subordinadas condicionais. Editoriais jornalísticos destacam-se no polo positivo por seu caráter opinativo.

A quinta dimensão, “informação abstrata *versus* não-abstrata”, também apresenta apenas características linguísticas positivas. São elas: voz passiva com e sem agente ou modificador pós-nominal e conjunções. O uso de construções de voz passiva aponta para textos mais técnicos e abstratos. Os registros mais marcados por essas características são prosa acadêmica e documentos oficiais, sendo conversas telefônicas e face a face os menos marcados.

Em seu trabalho, Biber (1988, 2009) mostra a inter-relação entre registros escritos e falados estabelecida nas dimensões de variação da língua inglesa identificadas. Essa descoberta de características comuns entre os diferentes registros do *corpus* corrobora a visão da Análise Multidimensional como metodologia para análise empírica da variação existente entre textos, falados e escritos.

Os estudos principais em AMD foram realizados com a finalidade de identificar dimensões de diversas línguas, como o inglês (BIBER, 1988; CROSSLEY e LOUWERSE, 2007; DE MÖNNINK, BROM e OOSTDIJK, 2003; LEE, 2000), o coreano (KIM e BIBER, 1994), o somali (BIBER e HARED, 1994), o espanhol (BIBER, DAVIES, JONES e TRACY-VENTURA, 2006; PARODI, 2007), entre outros. As dimensões de variação do português brasileiro foram identificadas por meio de estudo realizado por membros do GELC/CNPq²⁶ e financiado pela FAPESP e pelo CNPq (BERBER SARDINHA *et al.*, no prelo). O estudo foi feito com um *corpus* contendo 48 registros orais e escritos em português brasileiro e revelou seis dimensões de variação.

No Brasil, as pesquisas na área de Análise Multidimensional vêm crescendo, como podemos observar em dissertações e teses do Programa de Pós-Graduação em Linguística Aplicada e Estudos da Linguagem (LAEL) da PUC-SP, entre as quais se encontram: Oliveira (1997), que analisou a variação entre composições de alunos

²⁶ Grupo de Estudos de Linguística de *Corpus*, do LAEL/PUCSP. Uma das áreas de atuação do GELC/CNPq é a pesquisa em Análise Multidimensional (AMD).

de inglês e de falantes nativos; Conde (2002), que fez comparações entre composições de alunos de uma escola bilíngue e alunos de institutos de idiomas; Shergue (2003), que comparou comunicações em congressos e artigos acadêmicos da área médica; Kauffmann (2005), que identificou as dimensões de variação na escrita jornalística de um jornal brasileiro; Bértoli-Dutra (2010), que mapeou a variação entre letras de música popular anglo-americana; Souza (2012), que identificou as dimensões de variação nas matérias de capa da revista TIME; Pinto (2013), que identificou as dimensões de variação em um *corpus* diacrônico de filmes de Hollywood; e Delegá-Lúcio (2013), que identificou as dimensões de variação na escrita de redação de aprendizes de língua inglesa brasileiros.

Pretende-se que a Análise Multidimensional realizada nesta pesquisa contribua para os estudos de inglês para aviação ao identificar os traços linguísticos e a variação existentes entre os diferentes manuais de manutenção aeronáuticos utilizados por técnicos de manutenção de aeronaves no Brasil e no mundo.

A seguir, discutem-se as questões metodológicas que orientaram esta pesquisa.

2 Metodologia

Neste capítulo, o *corpus* de estudo, composto por manuais operacionais e de manutenção de aeronaves, e seu processo de coleta são apresentados. Em seguida, as metodologias utilizadas na análise do *corpus* são descritas.

2.1 Desenho e coleta do *Corpus*

O desenho e a coleta do *corpus* desta pesquisa seguiram critérios linguísticos específicos a fim de atingir os objetivos propostos. Conforme supracitado, o *corpus*, denominado *Corpus of Aircraft Manuals* (CAM), é um *corpus* especializado, pois contém apenas textos extraídos de manuais de aeronaves. O *corpus* coletado para a pesquisa tenta abranger a maior variedade possível de manuais, fabricantes e aeronaves. O *corpus* CAM é composto por 154 textos provenientes de sete tipos de manuais de 15 modelos de aeronaves, sendo 11 aeronaves comerciais e quatro executivas, de cinco dos principais fabricantes mundiais de aeronaves, a saber: *Airbus*, *ATR*, *Boeing*, *Cessna* e *Embraer*. Os manuais provenientes da *Airbus*, *Boeing*, *Cessna*, *Embraer* e *ATR* correspondem a 25,3%, 35,7%, 16,9%, 16,9% e 5,2% do *corpus*, respectivamente. Observe a seguir os tipos de manuais coletados para esta pesquisa agrupados de acordo com os fabricantes e modelos de aeronaves (Tabela 4):

Fabricante	Aeronave	Tipo de Manual
Airbus	A319-320-321 ²⁷	Práticas e Procedimentos
		<i>Troubleshooting</i>
		Treinamento (Operacional)
		Treinamento (Manutenção)
		<i>Checklist</i> (Manutenção)
	A330	Treinamento (Operacional)
		<i>Checklist</i> (Operacional)
<i>Checklist</i> (Manutenção)		

²⁷ A319-320-321 refere-se a todos os modelos de aeronave da família A320, ou seja, A319, A320 e A321. Como essas aeronaves apresentam apenas pequenas mudanças de configuração (ex: disposição dos assentos), os manuais são os mesmos para todas. Nesta pesquisa este grupo de aeronaves é referido como A320.

ATR	ATR-42	Treinamento (Operacional)
		<i>Checklist</i> (Operacional)
	ATR-45	Treinamento (Operacional)
		<i>Checklist</i> (Operacional)
ATR-72/600	Treinamento (Operacional)	
	<i>Checklist</i> (Operacional)	
ATR-75	Treinamento (Operacional)	
	<i>Checklist</i> (Operacional)	
Boeing	MD11	Práticas e Procedimentos
		<i>Checklist</i> (Manutenção)
	B737NG	Treinamento (Operacional)
		<i>Checklist</i> (Operacional)
		<i>Checklist</i> (Manutenção)
	B767	Treinamento (Operacional)
		<i>Checklist</i> (Operacional)
		<i>Checklist</i> (Manutenção)
	B777	Práticas e Procedimentos
		Descritivo
<i>Troubleshooting</i>		
Treinamento (Operacional)		
<i>Checklist</i> (Operacional)		
Cessna	CE-680 (Citation Sovereign)	Treinamento (Manutenção)
		<i>Checklist</i> (Manutenção)
		<i>Checklist</i> (Operacional)
	CE-650 (Citation III)	<i>Checklist</i> (Manutenção)
	CE-750 (Citation X)	Treinamento (Manutenção)
<i>Checklist</i> (Manutenção)		
Embraer	E190-195 ²⁸	Descritivo
		Práticas e Procedimentos
		<i>Checklist</i> (Manutenção)
	Phenom – 100	<i>Checklist</i> (Operacional)
		Treinamento (Operacional)
		<i>Checklist</i> (Manutenção)

Tabela 4. Lista de fabricantes, aeronaves e manuais que perfazem o *corpus* CAM

Foram selecionados sete tipos de manuais, os quais técnicos de manutenção de aeronaves utilizam em sua rotina de trabalho. Os manuais coletados são divididos

²⁸ *E190-195* refere-se a todos os modelos de aeronave da família *E190*, ou seja, *E190* e *E-195*. Como essas aeronaves apresentam apenas mudanças na configuração interna (ex: disposição dos assentos), os manuais são os mesmos para todas. Nesta pesquisa este grupo de aeronaves é referido como *E190*.

em manuais de manutenção e manuais operacionais, sendo que os manuais de manutenção abrangem cinco tipos e os operacionais, dois tipos.

Os manuais de manutenção são usados apenas por profissionais de manutenção de aeronaves e podem conter informações a respeito de sistemas e partes da aeronave (Descritivo e Treinamento), manutenção de rotina e corretiva (Práticas e Procedimentos), identificação e solução de defeitos e/ou falhas na aeronave (*Troubleshooting*), e investigação de problemas para liberação da aeronave para voo (*Checklist*).

Os manuais operacionais são destinados a pilotos; no entanto, também são utilizados por técnicos de manutenção de aeronaves, especialmente os responsáveis pelos sistemas eletrônicos da aeronave (aviônica). Os manuais operacionais coletados abrangem dois tipos de informação: (1) descrição e operação de sistemas da aeronave (Treinamento) e (2) procedimentos em situações de rotina e anormais (*Checklist*).

A composição do *corpus* mostra que há mais de um manual correspondente a cada tipo. Isso ocorre porque, apesar de serem do mesmo tipo, os nomes dos manuais variam de acordo com o fabricante e o modelo da aeronave. Para a composição do *corpus* foram consideradas as funções de cada manual para classificá-los em um dos sete tipos identificados. Essa informação é relevante, pois os textos foram nomeados e classificados a partir do modelo da aeronave e do nome de cada manual. Observe a seguir os tipos de manuais coletados para esta pesquisa e seus respectivos nomes (Tabela 5):

TIPOS DE MANUAIS		NOMES DOS MANUAIS
MANUTENÇÃO	Descritivo	AMM SDS
	Práticas e Procedimentos	AMM, AMM MPP, AMM P&P
	Troubleshooting	FIM, TSM
	Treinamento	TTM, TRAINING MANUAL
	Checklist	MMEL FAA
OPERACIONAL	Treinamento	FCOM, POH
	Checklist	QRH, QRH ANAC

Tabela 5. Tipos de Manuais do *corpus* CAM

O primeiro passo na coleta do *corpus* foi a extração dos textos dos manuais de CD-ROMs obtidos nos centros de manutenção de aeronaves de companhias aéreas

brasileiras. Os manuais que compõem o *corpus* foram selecionados a partir de listas de manuais aeronáuticos presentes em *websites* dos fabricantes dessas aeronaves e da disponibilidade de acesso a esses manuais. Os manuais são distribuídos às companhias aéreas e empresas privadas, no caso de aeronaves executivas, em formato de CD-ROM devido à enorme quantidade de informações contidas nesses textos.

Inicialmente foram coletados 520 textos, correspondendo a um total de 53.863.518 ocorrências (*tokens*) e 218.661 formas (*types*). No entanto, a quantidade de textos foi reduzida a fim de viabilizar a Análise Multidimensional. Conforme descrito no Capítulo Fundamentação Teórica, *corpora* utilizados para a identificação de dimensões de variação devem ser construídos com a finalidade de representar a maior variedade textual e de registros possível, sendo que o número de ocorrências não deve ser considerado como o critério principal para a coleta de textos.

Portanto, a fim de capturar a quantidade máxima de variação de textos no *corpus*, decidiu-se por reduzir criteriosamente o número de textos. Essa redução foi feita com base em critérios estatísticos.

A primeira decisão a ser tomada foi o que equivale a um texto. Essa decisão é de extrema relevância pois na AMD a unidade de análise é o texto. No *corpus* CAM, alguns manuais são subdivididos em capítulos (*ATA Chapters*²⁹) e outros não. Assim, foi necessário decidir se um texto equivale a um manual completo, seja ele subdividido ou não, a um manual completo apenas quando não subdividido, a cada uma das subdivisões ou capítulos, ou ao conjunto todo. Nesse caso, a decisão tomada foi: quando um manual for subdividido em capítulos, cada capítulo será considerado um texto; quando um manual não for subdividido em capítulos, o manual como um todo será considerado um texto. A tabela a seguir (Tabela 6) apresenta os manuais que são subdivididos em capítulos e os que não são:

²⁹ Cada capítulo de todo e qualquer manual de aeronave, independente de fabricante, corresponde a um sistema único da aeronave e é numerado e nomeado de acordo com um conjunto de diretrizes padrão, estabelecido pela *Air Transport Association* (ATA) e denominado *ATA Specification 2200* (iSpec 2200). A lista completa de capítulos pode ser adquirida em: <https://publications.airlines.org/>

Manuais Subdivididos em Capítulos 1 CAPÍTULO = 1 TEXTO	Manuais <u>não</u> subdivididos em Capítulos 1 MANUAL = 1 TEXTO
Descritivo Práticas e Procedimentos <i>Troubleshooting</i> Treinamento (Manutenção)	Treinamento (Operacional) <i>Checklist</i> (Operacional) <i>Checklist</i> (Manutenção)

Tabela 6. Manuais subdivididos e não subdivididos em capítulos

Uma vez determinado o que é um texto, decidiu-se por manter todos os manuais não subdivididos em capítulos e reduzir o número de textos dos manuais subdivididos em capítulos. Para isso, o primeiro critério estatístico aplicado foi estabelecer limites de corte (*cutoffs*) para os manuais subdivididos em capítulos. Esses limites foram estabelecidos com base no tamanho do texto em número de palavras (ocorrências). Desse modo, uma lista de palavras por frequência foi criada no programa *WordSmith Tools* para cada capítulo dos manuais subdivididos, ou seja, para cada texto. Em seguida, gráficos de distribuição foram criados para cada manual (Anexo E). Os limites de corte foram estabelecidos a partir da análise desses gráficos. De acordo com os limites de corte, os textos extremos (*outliers*), isto é, textos com número de ocorrências abaixo ou acima dos limites estabelecidos com base em gráficos de distribuição, foram eliminados. Por exemplo, o gráfico de distribuição do manual AMM da aeronave A320 mostrou textos com menos de 25.000 palavras e com mais de 400.000 palavras como extremos (*outliers*); portanto, foram mantidos apenas os textos contendo entre 25.001 e 399.999 palavras ($25.000 > x < 400.000$). Os limites estabelecidos por aeronave e por manual são (Tabela 7):

Aeronave	Manual	Limites de corte (nº de palavras)	Quantidade de textos eliminados
A320	AMM	$25.000 > x < 400.000$	20
	TSM	$10.000 > x < 150.000$	15
	TTM	$5.000 > x < 30.000$	5
MD11	AMM	$15.000 > x < 200.000$	10
B777	FIM	$15.000 > x < 200.000$	13
	AMM P&P	$15.000 > x < 200.000$	13
	AMM SDS	$2.500 > x < 70.000$	14
CE-680	Training Manual	$500 > x < 65.000$	2
CE-750	Training Manual	$500 > x < 45.000$	2

E190	AMM MPP	25.000 > x < 400.000	19
	AMM SDS	10.000 > x < 80.000	13

Tabela 7: Limites de corte

Os demais manuais não estão representados na tabela acima porque, conforme supracitado, não são subdivididos em capítulos. Todos os manuais não subdivididos coletados foram, a princípio, incluídos no *corpus*. Em relação aos manuais subdivididos, após a remoção dos extremos, a quantidade de textos restantes por aeronave e por manual continuava desequilibrada. Assim, o segundo passo foi estabelecer o número de textos por aeronave e por manual para a composição final do *corpus* de modo que a variedade de aeronaves e manuais fosse mantida. Observou-se que 11 era o menor número de textos de uma única aeronave, ou seja, o mínimo denominador comum dentre aeronaves era 11 textos. Assim, estabeleceu-se que cada manual subdividido de uma determinada aeronave teria 11 textos, representando 11 capítulos. Além disso, os textos compostos por manuais completos seriam provenientes de 11 aeronaves diferentes.

Como havia mais do que 11 textos por manual para a maioria das aeronaves representadas no *corpus*, um terceiro critério estatístico foi usado para selecionar os textos que fariam parte da composição final do *corpus*: a amostragem aleatória. A amostragem aleatória permite que um subconjunto de elementos (amostra) seja selecionado totalmente ao acaso, isto é, com imparcialidade, a partir de um conjunto maior de elementos (população). Essa amostragem foi escolhida porque, nesse tipo de procedimento, qualquer um dos elementos presentes na população possui igual probabilidade de ser selecionado (Biber, 1993a). Nesta pesquisa, os elementos são os textos que compõem o *corpus*. A amostragem aleatória pode ser realizada por meio de tabelas randômicas, fórmulas em programas como o Microsoft Excel ou ferramentas computacionais disponíveis na internet. Há diversos *websites* que disponibilizam ferramentas computacionais para realizar esse tipo de seleção de forma automática, os quais são úteis devido a sua disponibilidade e facilidade de uso. Portanto, um desses *websites* foi escolhido para a realização da amostragem aleatória. O *website* escolhido foi o *Research Randomizer*³⁰ por oferecer a ferramenta gratuitamente e possuir uma interface fácil de usar.

³⁰ O site *Research Randomizer* (<http://www.randomizer.org>) fornece uma ferramenta gratuita para a realização de amostragem aleatória.

Primeiramente, os textos do *corpus* foram agrupados de acordo com os manuais e aeronaves a que pertencem e um número foi atribuído a cada texto. Em seguida, os textos foram selecionados por meio da ferramenta computacional disponibilizada pelo *website*, também denominada *Research Randomizer*. O mesmo procedimento foi seguido para a seleção de todos os textos. Por exemplo, após a remoção dos extremos, restaram 36 textos do manual AMM P&P correspondente à aeronave B777. Então, cada um desses textos recebeu um número de 1 a 36. Em seguida, os campos do formulário disponível no *website* foram preenchidos com os dados dos textos. Um único conjunto de números foi gerado para cada manual contendo 11 números cada. O intervalo de numeração variou de acordo com cada manual, como no caso do exemplo acima, que foi de 1 a 36. Foi determinado que cada número no conjunto deveria ser único, ou seja, não haveria repetição de números selecionados. Além disso, os números selecionados foram classificados em ordem crescente (Figura 1).

To generate random numbers, enter your choices below (using integer values only):

How many sets of numbers do you want to generate? [Help](#)

How many numbers per set? [Help](#)

Number range (e.g., 1-50): From: To: [Help](#)

Do you wish each number in a set to remain unique? Yes No [Help](#)

Do you wish to sort the numbers that are generated? Yes: Least to Greatest No [Help](#)

How do you wish to view your random numbers? Place Markers On Place Markers Off [Help](#)

Site Overview

Randomize Now [▶](#)
Use the Randomizer form to instantly generate random numbers.

Quick Tutorial [▶](#)
See some examples of how Research Randomizer can be used for random sampling and random assignment.

Related Links [▶](#)
Visit links on random sampling, random assignment, and research methods.

About Research Randomizer [▶](#)
Learn more about Research Randomizer and read our User Policy.

Randomizer Box
Add this tool to your website and generate your own number sets.

Figura 1. Formulário preenchido para realização de amostragem aleatória

Após o preenchimento do formulário, deve-se clicar no botão *Randomize Now!* para que a amostragem aleatória seja realizada. O resultado traz uma sequência de 11 números selecionados aleatoriamente pela ferramenta. Esses números equivalem aos textos que deverão ser incluídos no *corpus*, como pode ser visto na Figura 2:



Research Randomizer Results

1 Set of 11 Unique Numbers Per Set

Range: From 1 to 36 -- Sorted from Least to Greatest

Job Status: Finished

Set #1:

1, 2, 3, 4, 8, 15, 20, 22, 25, 30, 36

Figura 2. Números dos textos escolhidos aleatoriamente para inclusão no *corpus*

Esse procedimento foi realizado para todos os manuais contendo mais de 11 textos restantes após o corte dos extremos e para os manuais completos quando houvesse mais do que 11 textos por tipo de manual.

A versão final do *corpus* CAM contém 10.009.040 ocorrências (*tokens*) e 235.177 formas (*types*) (Tabela 8):

	Tipo de Manual	Fabricante - Aeronave	Textos	Ocorrências
Manutenção	Descritivo	Boeing - B777	11	185.556
		Embraer - ERJ190	11	243.435
	Práticas e Procedimentos	Airbus - A320	11	1.662.071
		Boeing - B777	11	666.146
		Boeing - MD11	11	953.155
		Embraer - ERJ190	11	1.464.926
	Troubleshooting	Airbus - A320	11	691.026
		Boeing - B777	11	727.475
	Treinamento	Airbus - A320	11	162.562
		Cessna - Citation X	11	44.001
		Cessna – Sovereign	11	26.136
	Checklist	Airbus - A320 / A330	11	258.746
Boeing - B737NG / B767 / B777 / MD11				
Cessna - Citation III / X / Sovereign				
Embraer - ERJ190 / Phenom 100				
Operacional	Checklist	Airbus - A320 / A330	11	292.651
		ATR - 42 / 45 / 72-600 / 75		
		Boeing - B737NG / B767 / B777		
		Cessna - Citation Sovereign		
	Embraer – Phenom			
	Treinamento	Airbus - A320 / A330	11	2.631.154
ATR - 42 / 45 / 72-600 / 75				

	Boeing - B737NG / B767 / B777 / MD11		
	Embraer – Phenom		
	TOTAL	154	10.009.040

Tabela 8. Composição do *corpus* CAM

Assim, a composição final do *corpus* compreende quatro níveis de informação, a saber:

Nível 1 – Propósito do manual

Nível 2 – Tipo de manual

Nível 3 – Fabricantes de aeronaves

Nível 4 – Modelos de aeronaves

É importante ressaltar que todos os manuais coletados foram livremente doados por setores de manutenção de companhias aéreas brasileiras e correspondiam a versões anteriores à última versão em uso pelas companhias na época da coleta de dados, sendo usados única e exclusivamente para fins acadêmicos, não acarretando nenhuma quebra de direitos autorais.

O passo seguinte à coleta foi a limpeza dos textos. Os textos do *corpus* estavam em formato *.pdf* ou no formato proprietário do software do fabricante, sendo ilegíveis para os programas utilizados na análise linguística. Por isso, foi necessário fazer a conversão desses arquivos para arquivos de texto sem formatação (*.txt*). Em seguida, informações de cabeçalho e rodapé, informações de identificação (nomes da empresa aérea e do fabricante, modelo e matrícula da aeronave, versão do manual), além de símbolos e numerais foram retiradas do *corpus*, pois essas informações não configuram na análise de dados linguísticos. Essa limpeza foi feita semi-automaticamente por meio de um *script* personalizado, criado especialmente para esta pesquisa por Berber Sardinha, em código *Unix Shell* e processado pela ferramenta *Cygwin*³¹. Uma revisão manual da limpeza foi realizada para garantir a

³¹ *Cygwin* é uma aplicação que permite emular o sistema operacional *Unix* em um ambiente *Windows*. O conjunto de ferramentas disponíveis no *Cygwin* permite a realização de tarefas automáticas de pré-processamento de textos, tais como busca, substituição, eliminação de caracteres, entre outros. O programa é gratuito e está disponível para download na Internet em <http://www.cygwin.com>.

remoção de informações de identificação. Após a conclusão da limpeza dos textos, o *corpus* foi considerado pronto para ser processado.

2.2 Procedimentos de Análise

Nesta subseção as metodologias de análise do *corpus* são apresentadas por meio da descrição da etiquetagem morfossintática e semântica aplicada ao *corpus*, das ferramentas computacionais utilizadas, da aplicação dos passos da AMD e da análise qualitativa dos dados.

2.2.1 Anotação e Etiquetagem

O *corpus* de estudo foi anotado automaticamente por um etiquetador computacional, denominado *Biber Tagger*³² e desenvolvido por Douglas Biber, que insere etiquetas (códigos) morfossintáticas nos textos, conforme mencionado no capítulo Fundamentação Teórica. O procedimento para obtenção do arquivo etiquetado é simples, sendo composto por 4 passos: (1) seleccionar o tipo de texto (*exposition, fiction* ou *spoken*); (2) informar se há cabeçalhos no texto; (3) informar se deseja inserir comentários nos resultados; e (4) iniciar a etiquetagem (Figura 3).

³² O programa *Biber Tagger* está disponível no Banco de Dados do Grupo de Pesquisa em Linguística de Corpus (GELC)

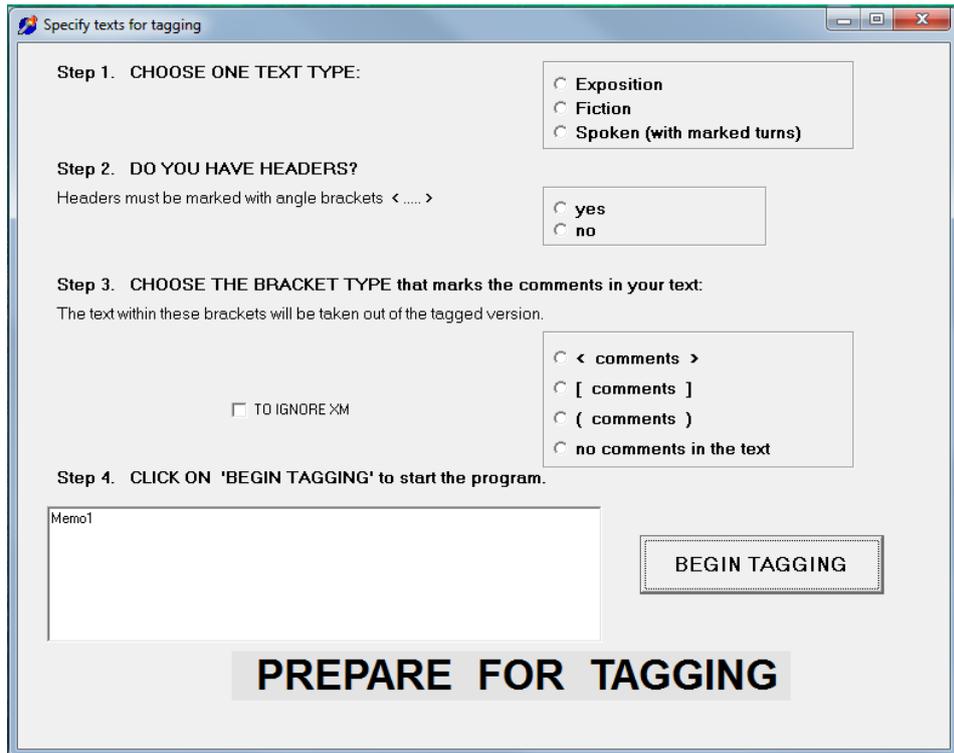


Figura 3. *Biber Tagger*

Os resultados são apresentados em um arquivo de texto sem formatação (.txt) contendo as anotações nos textos com siglas em inglês, conforme pode ser visto no exemplo³³ abaixo (Quadro 1).

```

Three ^cd++++=Three
cargo ^nn++++=cargo
compartments ^nns+nom+++=compartments
are ^vb+ber+vrb++=are
under ^in++++=under
the ^ati++++=the
cabin ^nn++++=cabin
floor ^nn++++=floor.
.^zz++++=EXTRAWORD

```

Quadro 1: Exemplo de texto anotado pelo *Biber Tagger*

A leitura do texto anotado é feita na vertical. Cada linha do arquivo contém uma unidade ortográfica do texto, localizada à esquerda (por exemplo: *Three*), seguida pelo símbolo ^ e os campos da etiqueta. O símbolo ^ marca o início da etiqueta. As etiquetas trazem a categorização morfossintática de cada unidade ortográfica do texto. A palavra *compartments*, por exemplo, é categorizada como *nns* (*plural common*

³³ A sentença exemplificada no quadro 1, “*Three cargo compartments are under the cabin floor.*” foi retirada de texto pertencente ao *corpus* CAM.

noun), substantivo plural comum, e *nom* (*nominalization*), nominalização. O símbolo + separa os campos da etiqueta e o símbolo = marca o início do texto em si, exceto para pontuação. Todas as marcas de pontuação (por exemplo: .) são codificadas como *EXTRAWORD* (“palavra extra”).

Subsequente ao processamento pelo *Biber Tagger*, os textos anotados e etiquetados foram inseridos na ferramenta computacional *TagCount*³⁴, também desenvolvida por Douglas Biber, para serem processados (Figura 4).

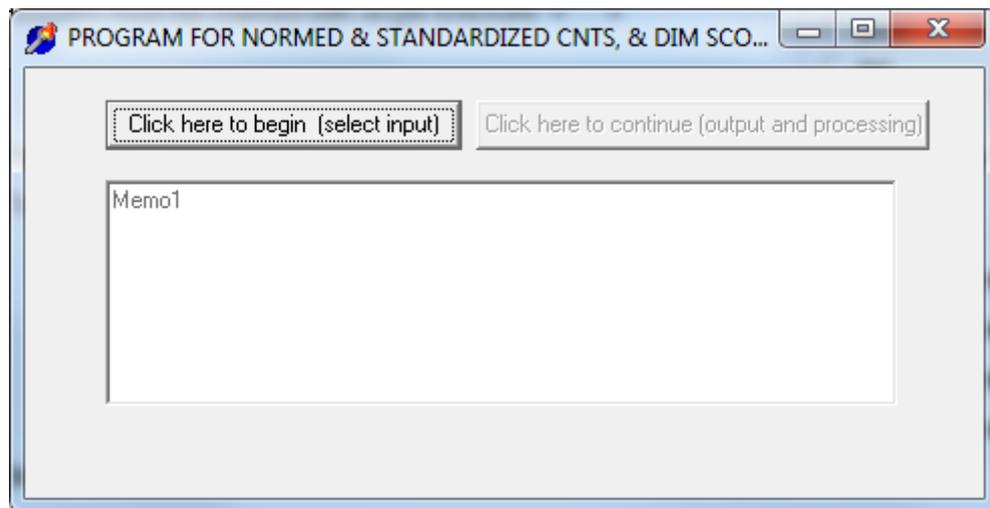


Figura 4: *TagCount*

O programa *TagCount* tem as seguintes funções: etiquetar as características lexicais do texto relacionadas a posicionamento e tipos de verbos e de substantivos, fazer a leitura das etiquetas inseridas pelo *Biber Tagger*, contabilizar as etiquetas como variáveis e estabelecer a frequência das etiquetas em cada texto do corpus, e agrupar as etiquetas em categorias morfossintáticas, semânticas e de marcação de posicionamento. Os resultados do processamento pelo *TagCount* são apresentados em um arquivo de texto sem formatação e foram transferidos para uma planilha em *Excel* para melhor visualização.

³⁴ O site do American National Corpus (<http://www.americannationalcorpus.org/SecondRelease/Biber-tags.txt>) traz uma descrição completa das etiquetas do *Biber TagCount*. Último acesso em: 13 nov. 2013.

No processamento do *corpus* CAM pelo *TagCount* foram identificadas 128³⁵ variáveis das 132³⁶ presentes no *tagset* do programa, o qual está disponível no Banco de Dados do GELC.

A contagem da frequência de características e categorias linguísticas resultantes do processamento pelo *TagCount* é realizada a fim de possibilitar a análise estatística que resultará no mapeamento dos textos do *corpus* nas dimensões do inglês estabelecidas por Biber (1988) e na identificação das dimensões próprias dos manuais aeronáuticos.

As características e categorias linguísticas identificadas pelo *TagCount* são características morfossintáticas, semânticas, ortográficas e de marcação de posicionamento. Dentre essas, estão as quatro classes lexicais de palavras, descritas por Biber *et al.* (1999 e 2002): adjetivos, advérbios, substantivos e verbos.

Em termos sintáticos, os adjetivos são classificados como atributivo (*attributive adjective*), o qual faz parte do sintagma nominal, ou predicativo (*predicative adjective*), o qual caracteriza o sintagma nominal embora não faça parte dele. Em termos de categoria semântica, os adjetivos podem ser classificados como descritores (*descriptors*, que abrangem descrições referentes a cor, tamanho, peso, idade e emoções) ou classificadores (*classifiers*, por exemplo: *very, general*).

Os advérbios são classificados em modificador (*modifier*), que modificam um adjetivo ou outro advérbio, e adverbial (*adverbial*), que exercem as funções típicas de um advérbio, ou seja, modificam adjetivos e verbos. As categorias semânticas dos advérbios são: advérbios de lugar (*place adverbs*), advérbio de tempo (*time adverbs*), advérbios de modo (*manner adverbs*), advérbios de grau (*degree adverbs*), advérbios delimitadores/aditivos (*additive/restrictive adverbs*), advérbios de posicionamento (*stance adverbs*) e advérbios de ligação (*linking adverbs*). Advérbios de posicionamento são subdivididos pelo programa em advérbio factivo (*factive adverb*) e advérbio não factivo (*non-factive adverb*). A categoria semântica de partículas do

³⁵ Disponíveis no Anexo A. Este número inclui as variáveis que contemplam as somas de todas as ocorrências de um tipo de categoria. Por exemplo: <alladj> inclui a soma de todas as categorias de adjetivos.

³⁶ Fonte: documento *biber_counter_vars* pertencente ao Banco de Dados do Grupo de Pesquisa em Linguística de Corpus (GELC). As descrições dessas variáveis estão disponíveis no Anexo B.

discurso é classificada junto aos advérbios por Biber (1988), por fazer o papel de um advérbio.

Os substantivos, de acordo com Biber *et al.* (1999 e 2002), podem ter classificações morfológicas, sintáticas e semânticas. Em termos morfológicos, substantivos podem receber sufixos; em termos sintáticos, substantivos estão presentes na posição principal de um sintagma nominal; e em termos semânticos, podem ser concretos ou abstratos. O programa possui dez classificações para os tipos de substantivos identificados: substantivo, nominalização, substantivo abstrato ou referente a processos, substantivo animado, substantivo cognitivo, substantivo concreto, substantivo referente a um grupo ou instituição, substantivo indicativo de lugar, substantivo indicativo de quantidade, e substantivo técnico/concreto.

A categoria de verbos contempla além de verbos lexicais, verbos auxiliares e verbos usados como verbo principal em orações, suas características morfológicas. Além disso, os verbos lexicais também são divididos em categorias semânticas, como por exemplo, verbos referentes a ações ou processos. As características morfológicas são classificadas de acordo com os tempos (presente, passado ou futuro), aspectos (por exemplo, contínuo ou perfeito), as modalidades (por exemplo, verbos modais), e as vozes (por exemplo: ativa ou passiva).

Além das quatro classes gramaticais, o programa também classifica as conjunções, que podem ser coordenadas ou subordinadas, e as preposições, que podem ser classificadas como preposições (classificação geral dos usos da preposição) ou preposições desacompanhadas (por exemplo: *the electrical power is on*).

Os pronomes são classificados de acordo com suas características morfossintáticas, a saber: pronomes pessoais (*personal pronouns*), reflexivos (*reflexive pronouns*), recíprocos (*reciprocal pronouns*) ou possessivos (*possessive pronouns*) em primeira, segunda e terceira pessoas; pronomes indefinidos (*indefinite pronouns*); pronomes demonstrativos (*demonstrative pronouns*); e usos do pronome iniciado por *WH-*, como pronome relativo (*relative pronoun*) ou interrogativo (*interrogative pronoun*). A omissão do pronome *that* também é classificada.

Além disso, há a classificação de construções complexas associadas à marcação de posicionamento, ou seja, como o falante expressa suas atitudes ou opiniões. Essas construções complexas são chamadas de orações complementares e podem ser controladas por um verbo ou adjetivo e conter orações com *that*, *to*, *ing* ou *ed*, ou pronome *wh-* (BIBER *et al.*, 1999 e 2002). De acordo com Souza, “a marcação de posicionamento é feita normalmente por meio do uso combinado com outras formas, tais como adjetivos e advérbios [...]. Essas outras categorias controlam as orações de posicionamento.” (2012, p.145). A soma das características também é fornecida pelo *Biber Tagger* (por exemplo, a soma de todos os verbos ou adjetivos).

As metodologias de análise quantitativa e qualitativa da AMD são descritas na subseção a seguir.

2.2.2 Análises quantitativa e qualitativa

De acordo com Berber Sardinha (2004a, p. 305), a aplicação da AMD obedece, resumidamente, aos seguintes passos:

1. Levantamento das características linguísticas relevantes para análise por meio de ampla consulta à literatura disponível.
2. Coleta ou adoção de um corpus de dados linguísticos representativo e compatível com as metas da análise.
3. Transformação das características linguísticas em variáveis quantificáveis.
4. Codificação dos dados baseada nas variáveis selecionadas, usando-se ferramentas computacionais para análise automática, semiautomática (interativa), ou manual.
5. Conferência manual da codificação feita por computador para se assegurar de sua exatidão.
6. Computação de frequência médias de cada variável.
7. Padronização das frequências (em geral por 1000 palavras), para permitir a comparação entre variedades (textos, registros ou corpora) de extensões diferentes.
8. Análise Fatorial inicial, a fim de se obter os pesos ('loadings') de cada variável em cada variedade.
9. Determinação do número de fatores, por meio da aplicação de técnicas como observação dos valores eigen ('eigenvalues') em um gráfico scree ('scree plot').
10. Análise Fatorial posterior, fazendo-se a rotação dos fatores.

11. Cálculo de escores de cada texto por fator pela padronização dos escores com base na média e no desvio padrão.
12. Cálculo de escores médios de cada variedade por fator.
13. Interpretação de cada fator e rotulação das dimensões.

Na análise multidimensional realizada originalmente por Biber (1988), os dados do *corpus* foram submetidos a todas as etapas de análise, quantitativas e qualitativas, a fim de se identificarem as dimensões de variação existentes no *corpus*. Na prática, há duas formas diferentes de se utilizar a metodologia de AMD: uma análise completa, como a original desenvolvida por Biber (1988), e outra partindo-se das dimensões de variação identificadas. Nessa segunda forma, as dimensões pré-existentes são o ponto de partida da análise. Isso é permitido devido ao caráter cumulativo dessa abordagem, ou seja, “a descrição de um *corpus* de uma certa variedade multidimensionalmente permite a comparação desta descrição com a descrição de outras variedades posteriormente” (BERBER SARDINHA, 2004a; 2000: 103). De acordo com Biber (2009), o uso das dimensões do inglês previamente estabelecidas permite a comparação de novos registros ou registros de linguagem especializada à grande variedade de registros orais e escritos da língua inglesa.

Nesta pesquisa, as duas formas de metodologia foram utilizadas. Primeiramente, o *corpus* de estudo foi mapeado nas dimensões da língua inglesa estabelecidas por Biber (1988) a fim de verificar onde os manuais aeronáuticos se situam em relação aos outros registros da língua inglesa. Nesse mapeamento, as mesmas variáveis observadas e destacadas na pesquisa de 1988 foram utilizadas. Em seguida, uma análise fatorial completa foi realizada para definir quais são as dimensões de variação próprias dos manuais aeronáuticos. Além disso, também foram feitas análises de correlação e variância (ANOVA) a fim de verificar se há diferenças estatisticamente significativas entre fabricantes e modelos de aeronaves em relação às dimensões de variação próprias dos manuais aeronáuticos. Análises de correlação estabelecem a força da relação existente entre duas variáveis e análises de variância comparam diferenças entre grupos independentes (GOMÉZ, 2013). Essas análises são explicitadas em detalhes na seção 2.2.2.2.5.

O uso do computador é necessário para a realização dos passos 2 ao 12. Os recursos computacionais utilizados nesta pesquisa são: *scripts* personalizados para limpeza dos textos, software de produtividade (*Microsoft Excel*) e pacote estatístico

(SPSS). Os resultados de cada passo precisam ser interpretados para que o próximo passo seja realizado. O passo 13 corresponde à análise qualitativa, em que as dimensões são nomeadas e os textos são analisados manualmente. Na próxima subseção, as duas formas de aplicação da metodologia AMD e seus respectivos passos são descritos em detalhes.

2.2.2.1 SPSS e análise qualitativa: Mapeamento do CAM nas dimensões de 1988

A análise quantitativa, etapas 6 a 12 dos passos da AMD descrito acima, foi realizada com o uso de um pacote estatístico denominado *Statistical Package for the Social Sciences* (SPSS). O SPSS é um pacote de *software* integrados para a realização de análises estatísticas. Primeiramente criado para o uso pelas ciências sociais, o pacote estatístico é hoje utilizado nas mais diversas áreas acadêmicas e profissionais, tais como: ciências médicas, tecnologia da informação, comunicação social, marketing, relações públicas, direito, economia, política e educação. O SPSS é um dos pacotes estatísticos mais usados em pesquisas, pois compreende uma grande variedade de tipos de análise estatística disponíveis em uma interface intuitiva e de fácil manuseio. Entre os principais tipos de análise estatística realizados pelo SPSS estão: a estatística descritiva ou univariada (por exemplo, tabulação, frequências, descritivas); a estatística bivariada (por exemplo, médias e desvio padrão, teste t, ANOVA, correlação, testes não paramétricos); a previsão de resultados numéricos (por exemplo, regressão linear); e a identificação de grupos (por exemplo, análise fatorial, análise de agrupamentos, análise discriminante linear). Nesta pesquisa, foram utilizadas as ferramentas de análise descritiva, estatística bivariada e identificação de grupos. Essas ferramentas são descritas a seguir.

Conforme descrito anteriormente, a primeira etapa de processamento estatístico foi o mapeamento dos manuais nas dimensões de variação da língua inglesa de 1988. Esse processamento foi feito a partir dos resultados da etiquetagem do *corpus* CAM pelos programas *Biber Tagger* e *TagCount*. O programa utiliza as mesmas variáveis e fórmulas da pesquisa de 1988 para calcular os escores de fator de cada texto do *corpus* em cada uma das cinco dimensões de 1988.

Os resultados do *TagCount* em formato de planilha do Excel, contendo a contagem de frequência das etiquetas morfossintáticas e semânticas produzidas pelo *Biber Tagger e TagCount*, foram importados para o pacote estatístico SPSS 20³⁷. A planilha de Excel continha as seguintes colunas de dados: (1) nome do arquivo (*filename*), (2) etiquetas e (3) dimensões de 1 a 5 da língua inglesa. Os dados passaram a ser chamados de variáveis a partir do momento em que foram inseridos no pacote estatístico SPSS versão 20.

A inserção dos dados no SPSS 20 é feita de modo semelhante à outros programas computacionais. Primeiro, clicou-se em “*file*”, depois em “*open*” e, por último, em “*data*”, onde a opção de tipo de arquivo Excel foi escolhida e a planilha selecionada. Após optar pela importação de dados do Excel, uma janela apareceu pedindo a confirmação de que a primeira linha da planilha corresponde ao nome das variáveis. As outras informações na janela foram mantidas. Ao clicar em “ok”, todos os dados foram importados para o SPSS 20.

Após a importação dos dados para o SPSS 20, foram adicionadas três variáveis independentes à planilha de dados. Essas variáveis foram adicionadas a fim de verificar se haveria variação estatisticamente significativa entre elas e as características linguísticas dos manuais aeronáuticos. Um resultado é estatisticamente significativo quando as diferenças encontradas são grandes o suficiente para não serem atribuídas ao acaso. Dessa forma, é necessário estabelecer um nível de significância para a análise. Em pesquisas linguísticas, o nível aceito é de 5% de erro ($p < 0,05$), ou seja, há menos de 5% de chance de que rejeitar a hipótese nula seja um erro (GOMÉZ, 2013). No caso desta pesquisa, resultados estatisticamente significantes entre características linguísticas e os fabricantes de aeronaves, por exemplo, significam que a variação linguística encontrada nos manuais pode ser atribuída aos fabricantes, ou seja, manuais de diferentes fabricantes possuem características linguísticas distintas. Essa verificação foi feita por meio da análise de variância (ANOVA), conforme descrito na subseção 2.2.2.2.5. As variáveis independentes adicionadas foram:

- (1) Fabricantes de aeronave: cinco, a saber:

³⁷ Disponível em <http://www.ibm.com/software/analytics/spss/>. A versão 20.0 em inglês foi a utilizada nesta pesquisa.

- a. *Airbus*
- b. ATR
- c. *Boeing*
- d. *Cessna*
- e. Embraer.

(2) Modelos de aeronave: 15, a saber:

- a. A320
- b. A330
- c. ATR42
- d. ATR45
- e. ATR72600
- f. ATR75
- g. B737NG
- h. B767
- i. B777
- j. CE650
- k. CE680
- l. CE750
- m. E190
- n. MD11
- o. PHENOM

(3) Registro: nove, divididos de acordo com o nome e tipo de manual ao qual pertencem, sendo eles:

- a. *AMM Procedure* (Práticas e Procedimentos)
- b. *AMM Descriptive* (Descritivo)
- c. *AMM Proc Descr* (Práticas e Procedimentos)
- d. *FCOM Descriptive* (Treinamento – Operacional)
- e. *FIM Troubleshoot (Troubleshooting)*
- f. *M MEL Checklist (Checklist – Manutenção)*
- g. *QRH Checklist (Checklist – Operacional)*
- h. *TSM Troubleshoot (Troubleshooting)*
- i. *TTM Descriptive* (Treinamento – Manutenção)

Em seguida, os escores médios (*mean scores*) do *corpus* como um todo em cada uma das cinco dimensões foram calculados, a partir dos escores de fator de cada texto do *corpus*. Esse procedimento é feito pela opção “Analyze” > “Descriptive Statistics” > “Descriptives”. As variáveis escolhidas foram as cinco dimensões (dim1, dim2, dim3, dim4 e dim5) (Figura 5):

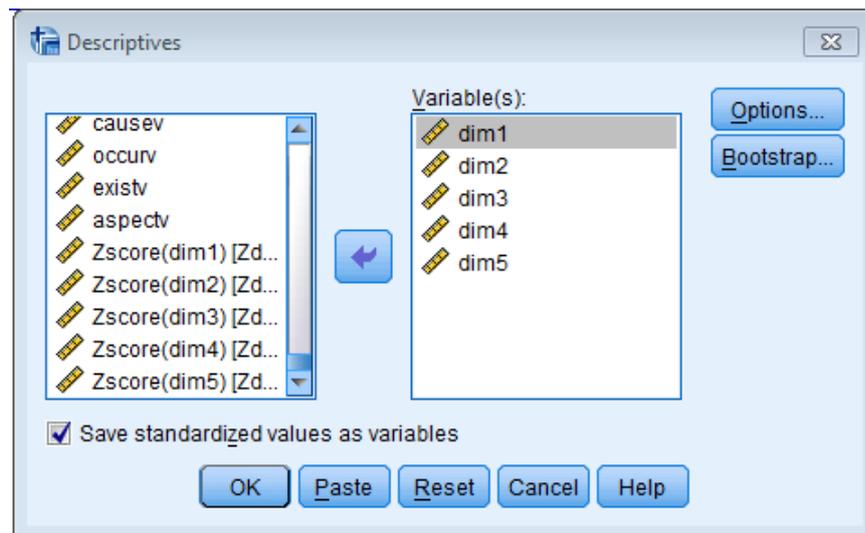


Figura 5. SPSS 20 – Estatísticas descritivas – Dimensões de 1988

Esses escores médios foram usados para localizar onde o *corpus* CAM se situa entre os textos usados por Biber na pesquisa de 1988 (retirados dos *corpora* LOB e LLC). Os escores médios do *corpus* CAM como um todo foram inseridos em gráficos de barra contendo os registros analisados na pesquisa de 1988, os quais foram fornecidos por Douglas Biber. A análise foi feita por meio da comparação dos escores médios do CAM com os escores médios dos registros usados na pesquisa de 1988.

2.2.2.2 SPSS e análise qualitativa: Identificação das dimensões do CAM

O segundo processamento do *corpus* foi realizado apenas nesta pesquisa, isto é, independentemente dos resultados da pesquisa de Biber (1988), a fim de identificar as dimensões de variação próprias dos manuais aeronáuticos. Nesta subseção, descreve-se os procedimentos realizados para a análise fatorial completa necessária

para essa identificação, ou seja, os passos 8 a 13 da metodologia de AMD supracitada.

2.2.2.2.1 Redução de dimensão: extração inicial

Nesta segunda etapa de processamento do *corpus* CAM, o mesmo processo de importação de dados da planilha em Excel resultante do *TagCount* para o SPSS 20 foi realizado, inclusive a inserção das mesmas variáveis independentes.

Com a planilha pronta, iniciou-se o procedimento estatístico. Conforme relatado no Capítulo Fundamentação Teórica, esse procedimento é feito por meio da análise fatorial. No programa, o item “*Factor...*” foi escolhido a partir do caminho “*Analyze*” > “*Dimension Reduction*” > “*Factor...*” (Figura 6):

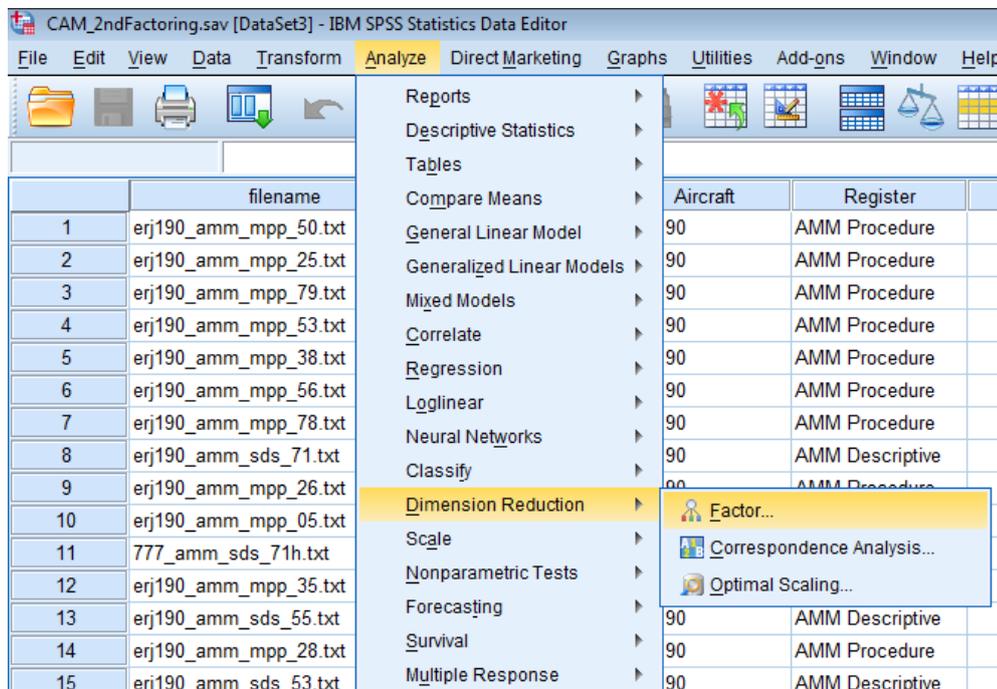


Figura 6. SPSS 20 – Redução de Dimensão – Análise Fatorial

Ao clicar em “*Factor...*” uma janela aparece para que as variáveis a serem incluídas na análise possam ser escolhidas (Figura 7). Conforme mencionado

anteriormente, as variáveis são as etiquetas linguísticas processadas pelos programas *Biber Tagger* e *TagCount*.

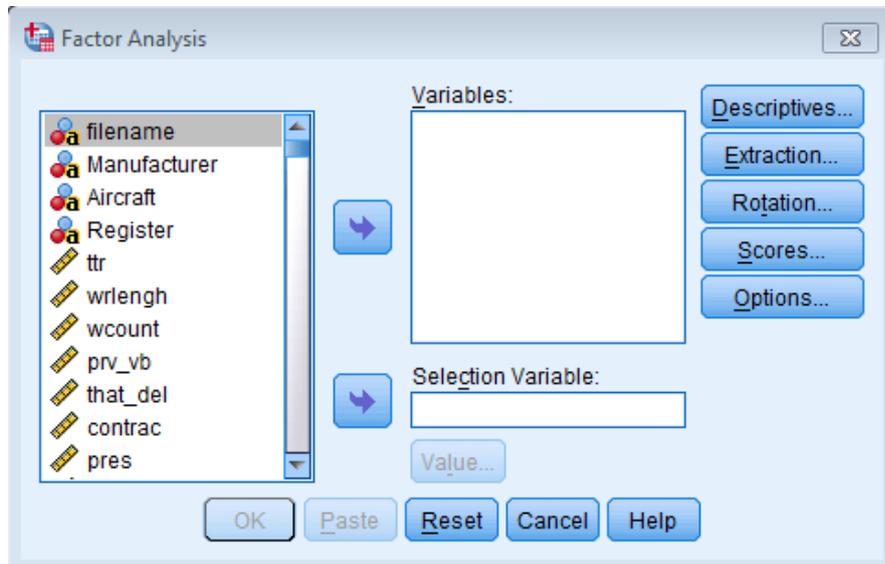


Figura 7. SPSS 20 – Análise Fatorial – Inserção de variáveis

Como pode ser visto na Figura 7, há configurações adicionais para a realização da análise fatorial. Para a AMD, as configurações “*Descriptives*” (Descritivos), “*Extraction*” (Extração) e “*Rotation*” (Rotação) devem ser ajustadas. Na configuração “*Descriptives*”, deve-se escolher, entre as opções, os tipos de teste estatístico (“*univariate descriptives*” ou “*initial solution*”) e matriz de correlação (“*Coefficients*”, “*Significance levels*”, “*Determinant*”, “*Inverse*”, “*Reproduced*”, “*Anti-image*” ou “*KMO and Bartlett’s test of sphericity*”) a serem aplicados. De acordo com a metodologia de AMD, deve-se escolher “*initial solution*” (solução inicial) como estatística e “*Anti-image*” (anti-imagem) e “*KMO and Bartlett’s test of sphericity*” (KMO e teste de esfericidade de Bartlett) (Figura 8). De acordo com Kauffmann (2005), esses testes gerais preliminares têm como objetivo verificar a adequação dos dados e a medida de adequação de amostragem.

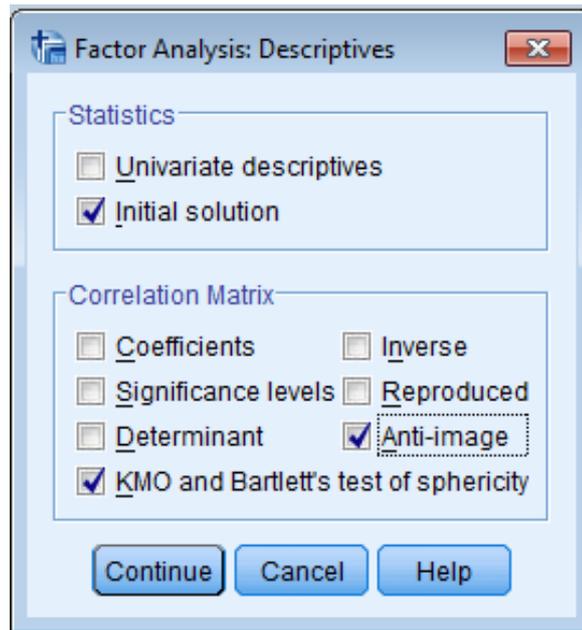


Figura 8. SPSS 20 – Análise Fatorial – Descritivos

Em “*Extraction*”, o pesquisador deve optar pelo método de extração de dados, o tipo de matriz de análise, como os dados serão mostrados, como a extração deve ser feita e o número máximo de iterações para convergência. As seguintes opções foram feitas (Figura 9):

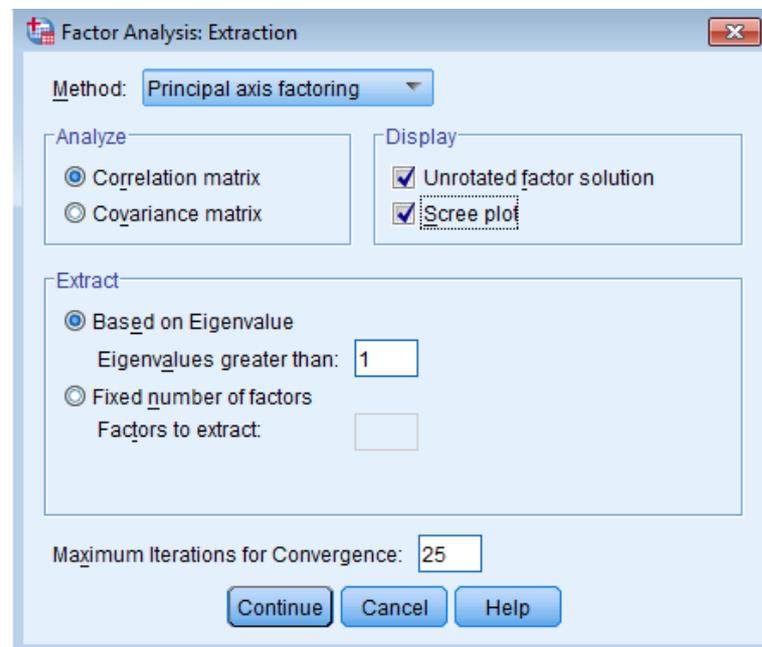


Figura 9. SPSS 20 – Análise Fatorial – Extração

Por ser uma análise linguística de cunho exploratório, o método de extração “*principal axis factoring*” (fatoração de eixo principal), padrão para pesquisas em AMD,

foi o escolhido. Essa escolha é fundamental pois permite a extração da variação máxima compartilhada pelas variáveis em cada fator. Ao contrário da extração por análise de componentes principais, em que não há distinção entre variação compartilhada e única, a fatoração de eixo principal apresenta apenas variação compartilhada pelas variáveis, evitando que as estimativas da variação encontrada sejam inflacionadas (COSTELLO e OSBORNE, 2005). Assim, o primeiro fator extraído apresenta a quantidade máxima de co-ocorrência entre todas as variáveis, o segundo fator apresenta a quantidade máxima de co-ocorrência entre as palavras restantes, e assim por diante (BIBER, 1988). A escolha da matriz de correlação foi feita para obter as comunalidades, ou seja, a “proporção da variância de cada variável explicada pelos fatores comuns” (SOUZA, 2012, p.158).

A extração inicial foi feita sem rotação para que o número máximo de fatores pudesse ser extraído. Desse modo, optou-se por “*unrotated factor solution*”, extração baseada em autovalores e visualização dos resultados em formato de gráfico de sedimentação (“*Scree plot*”). A quantidade de máxima de iterações para convergência de 25 é *default* do programa e foi mantida.

Os resultados da extração inicial exploratória tiveram que ser analisados para que os próximos passos da AMD pudessem ser realizados. Os primeiros resultados observados foram o teste Kaiser-Meyer-Olkin (KMO) e o teste de esfericidade de Bartlett. De acordo com Souza, o teste KMO “traz a medida de adequação da amostragem e testa se as correlações parciais entre as variáveis são pequenas ou não” (2012, p.159). Apesar de o valor desejável para pesquisas em AMD ser igual ou superior a 0,7, pode-se prosseguir com a análise se o valor for superior a 0,5, considerado satisfatório (cf. LEE, 2000; KAUFFMANN, 2005; SOUZA, 2012). O teste de esfericidade de Bartlett confirma se há correlação entre as variáveis, sendo que o nível de significância (Sig.) deve ser inferior a 0,05, conforme supracitado. Esses dois testes auxiliam o pesquisador a verificar se os dados são adequados para a realização da análise fatorial. Assim como o KMO e o teste de esfericidade de Bartlett, o teste de qui-quadrado é usado para avaliar a adequação da amostra na medida em que verifica se há desvio significativo ou não entre a frequência de um evento observado na amostra e a frequência esperada para esse evento.

A tabela a seguir apresenta os resultados do teste KMO e do teste de esfericidade de Bartlett para os dados desta pesquisa (Tabela 9):

Medida de adequação de amostragem de Kaiser-Meyer-Olkin		,602
Teste de esfericidade de Bartlett	Qui-quadrado aprox.	26765,877
	Df	6105
	Sig.	,000

Tabela 9. SPSS 20 – Análise Fatorial – Teste KMO e Bartlett

A tabela de “*communalities*” (comunalidades) foi analisada na sequência (vide recorte da tabela gerada pelo SPSS na Tabela 10). Comunalidades “indicam até que ponto uma determinada característica participa da variação total compartilhada contabilizada pela análise fatorial”³⁸ (BIBER, 2006, p.183). Desse modo, “uma variável que não tem uma variância exclusiva (ou variância aleatória) terá uma comunalidade de 1, enquanto que uma variável que nada compartilha de sua variância com qualquer outra variável terá uma comunalidade de 0” (FIELD, 2009, p.643 *apud* SOUZA, 2012, p.159). Em pesquisas em AMD, não há restrições para comunalidades iniciais; no entanto, variáveis com comunalidades finais abaixo de 0,15 podem ser excluídas da análise (cf. BIBER, 2006; LEE, 2000; KAUFFMANN, 2005; SOUZA, 2012). Nenhuma variável foi descartada por esse critério nesta pesquisa.

Comunalidades	
	Inicial
wrlengh	,965
wcount	,771
prv_vb	,940
that_del	,958
pres	,993
pro2	1,000
pro_do	,971
pdem	,900

Tabela 10. SPSS 20 – Análise Fatorial – Comunalidades (recorte)
(Método de extração: fatoraçoão de eixo principal)

Conforme supracitado, cada fator extrai a quantidade máxima de variância entre as variáveis presentes, sendo que o primeiro fator sempre apresentará a quantidade máxima de variância entre todas as variáveis incluídas na fatoraçoão, o

³⁸ “indicate the extent to which a given feature participates in the overall pool of shared variance accounted for by the factor analysis”

segundo fator entre as variáveis restantes do primeiro, o terceiro fator entre as variáveis restantes do segundo, e assim continuará até que 100% da variância seja explicada pelos fatores (cf. BIBER, 1988; KAUFFMANN, 2005). Os resultados dessa análise de variância são apresentados na tabela “*Total Variance Explained*” (“Variância Total Explicada”). A tabela traz informações que explicam a proporção de variância dos dados encontrada inicialmente em cada fator, por meio de “autovalores iniciais”. Autovalores “são índices diretos da quantidade de variação encontrada em cada fator”³⁹ (BIBER, 1988, p.82). Na coluna “% acumulada” do recorte apresentado na tabela abaixo (Tabela 11), nota-se que seis fatores respondem por quase 50% da variação presentes nos dados.

Fator	Autovalores iniciais		
	Total	% de variação	% acumulada
1	21,441	19,317	19,317
2	10,824	9,751	29,068
3	7,260	6,541	35,609
4	6,007	5,412	41,020
5	5,022	4,524	45,545
6	4,717	4,249	49,794
7	3,384	3,049	52,843
8	3,278	2,953	55,796
9	2,636	2,375	58,171
10	2,493	2,246	60,417

Tabela 11. SPSS 20 – Análise Fatorial – Variância Total Explicada (recorte)
(Método de extração: fatoração de eixo principal)

As informações da coluna “total” da tabela, cujo recorte está representado acima, também são mostradas por meio de um gráfico de sedimentação (“*Scree plot*”). Tanto a tabela quanto o *scree plot* são usados na determinação da quantidade de fatores a ser extraída. Em geral, há uma quebra ou curva natural característica no gráfico indicando o ponto a partir do qual fatores adicionais pouco contribuem para a análise dos dados. Portanto, o número ideal de fatores a serem extraídos encontra-se acima dessa quebra. Observe a quebra no *scree plot* resultante da extração inicial, a qual mostra claramente os seis primeiros fatores como os mais significativos nesta pesquisa

(Gráfico 1).

³⁹ “are direct indices of the amount of variance accounted for by each factor.”

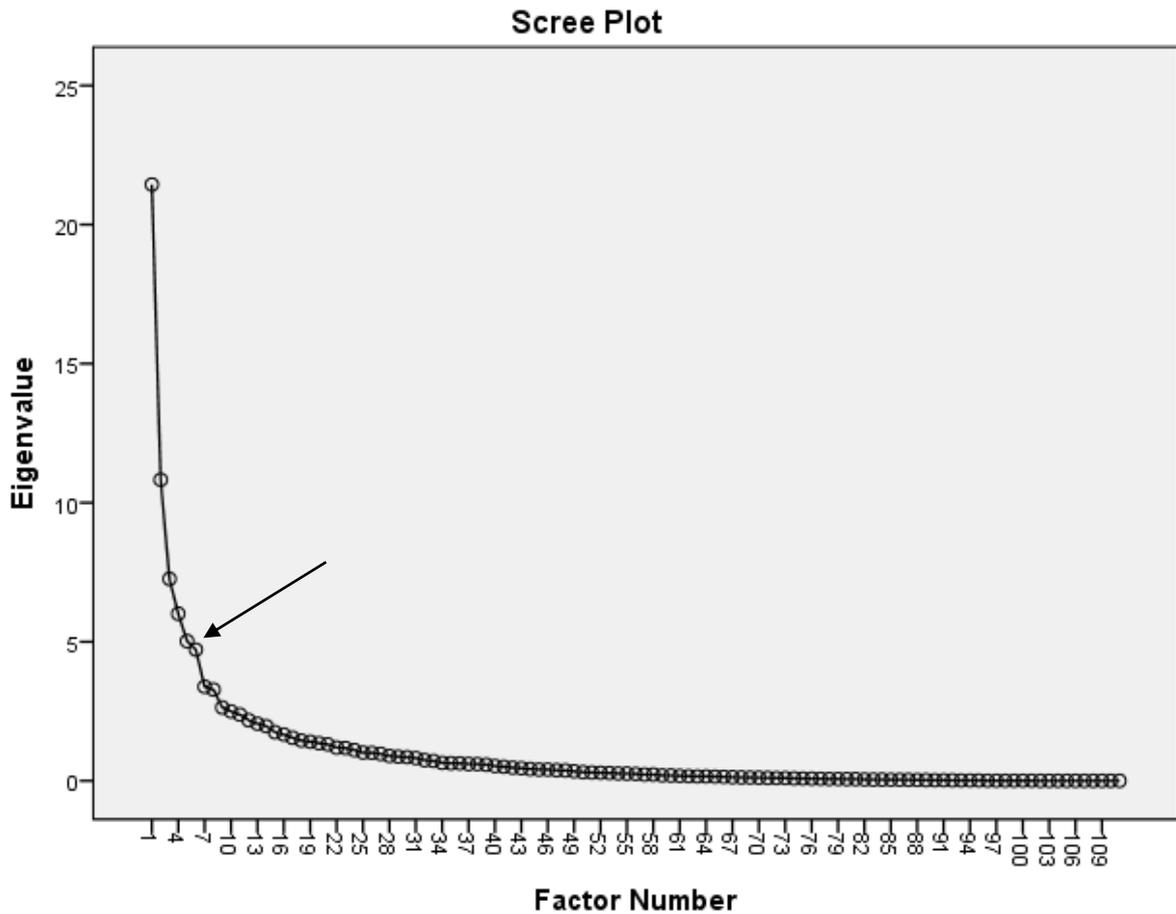


Gráfico 1. SPSS 20 – Análise Fatorial – *Scree Plot*

A partir dos dados apresentados acima, a segunda extração dos dados, ou seja, a análise fatorial com rotação, foi realizada. Baseado no *scree plot*, uma primeira extração rotacionada foi realizada com seis fatores. A interpretação dos fatores foi feita levando-se em consideração tanto as características linguísticas co-ocorrentes em cada fator quanto as diferenças e semelhanças existentes entre os registros relacionadas a cada grupo de características (BIBER, 1988 e 2006). A partir dessa análise, foi observado que as dimensões estavam muito semelhantes entre si, dificultando a interpretação da relação entre as características linguísticas e suas funções comunicativas em cada registro. Desse modo, novamente com base no *scree plot*, optou-se por repetir a extração rotacionada com três fatores, a qual provou ser adequada para a interpretação dos fatores como dimensões de variação. Portanto, decidiu-se por manter a extração rotacionada com três fatores, a qual é descrita na próxima subseção.

2.2.2.2.2 Redução de dimensão: extração rotacionada

Conforme supracitado, nenhuma das variáveis usadas na extração inicial foi descartada. Sendo assim, todas as variáveis originais foram incluídas na fatoração rotacionada. O procedimento segue os mesmos passos da extração inicial, com a diferença de que, nesse momento, optou-se por rotacionar a extração. O propósito da rotação é simplificar a estrutura de dados, pois com ela “cada fator é caracterizado pelas poucas características que são mais representativas de uma determinada proporção de variação compartilhada”⁴⁰ (BIBER, 1988, p.84), facilitando a interpretação. É importante reforçar que a rotação não altera a variância total dos dados, apenas a distribuição da variância entre os fatores é alterada (COSTELLO e OSBORNE, 2005; GÓMEZ, 2013).

Há diversos métodos para a realização da extração rotacionada, sendo *Varimax* e *Promax* os mais comuns. Esses dois métodos variam em termos de estrutura e pressupostos. O método *Varimax* possui estrutura ortogonal de rotação e parte do pressuposto de que não há correlação entre os fatores. Já o método *Promax* possui estrutura oblíqua de rotação e parte do pressuposto de que existe correlação entre os fatores (COSTELLO e OSBORNE, 2005). O método de extração rotacionada recomendado para pesquisas em AMD é a rotação *Promax* (Figura 9), uma vez que em estudos da linguagem e nas ciências humanas e sociais “geralmente espera-se que haja alguma correlação entre os fatores, já que comportamentos raramente são divididos em unidades empacotadas que operam isoladamente”⁴¹ (COSTELLO e OSBORNE, 2005, p.3). Os valores padrão do SPSS para a rotação *Promax* foram mantidos.

⁴⁰ “each factor is characterized by the few features that are most representative of a particular amount of shared variance.”

⁴¹ “we generally expect some correlation among factors, since behavior is rarely partitioned into neatly packaged units that function independently of one another.”

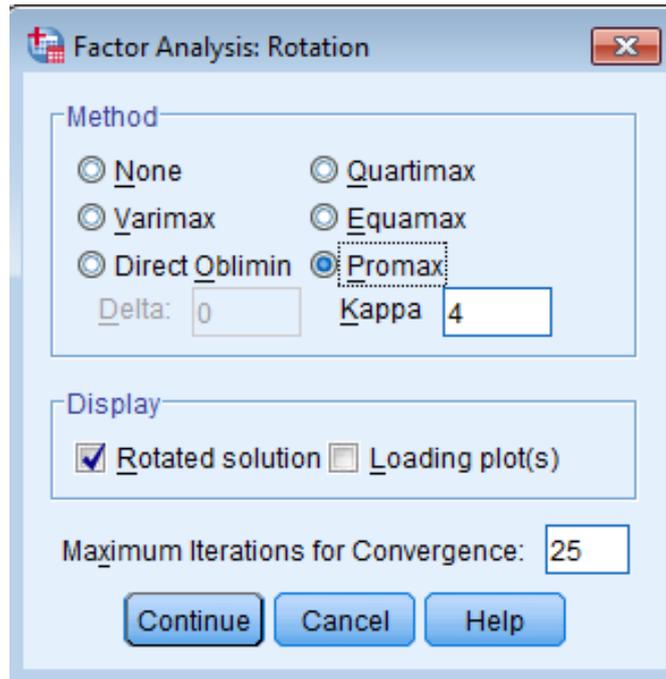


Figura 10. SPSS 20 – Análise Fatorial – Rotação

Além disso, no menu de configuração “*Options*” (“Opções”), optou-se por suprimir pesos com valores absolutos inferiores a 0,30 (Figura 10). Essa recomendação é feita porque em AMD variáveis com pesos inferiores a esse valor são consideradas irrelevantes para a análise de dados (BIBER, 1988). A opção “*sorted by size*” também foi selecionada para que os resultados fossem apresentados em ordem numérica.

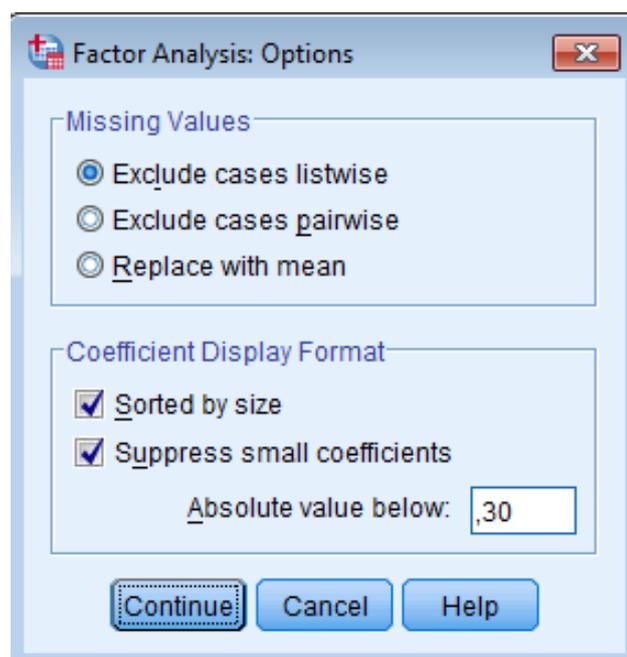


Figura 11. SPSS 20 – Análise Fatorial – Opções

Como não houve mudança nas variáveis incluídas na extração rotacionada, os resultados do teste de significância e da tabela de comunalidades permaneceram iguais. A tabela de variância explicada mostrou, além dos autovalores iniciais, a proporção de variação após a rotação (Veja um recorte na Tabela 12):

Fator	Autovalores iniciais			Somadas rotacionadas das cargas quadráticas
	Total	% de variação	% acumulada	
1	21,441	19,317	19,317	18,082
2	10,824	9,751	29,068	16,332
3	7,260	6,541	35,609	6,698
4	6,007	5,412	41,020	
5	5,022	4,524	45,545	
6	4,717	4,249	49,794	

Tabela 12. SPSS 20 – Análise Fatorial – Variância Total Explicada com rotação (recorte) (Método de extração: fatoração de eixo principal)

Além disso, outras matrizes foram incluídas nos resultados da extração rotacionada. A matriz utilizada para análise de resultados em pesquisas em AMD é a “*Pattern Matrix*” (Matriz de padrões). Observe abaixo um recorte dessa matriz (Tabela 13), a qual será apresentada por completo no capítulo Apresentação e Discussão dos Resultados.

	Fator		
	1	2	3
all_vto	,828		
all_to	,795		
efrt_vto	,773		
nonf_vth	,762		
all_vth	,758		
spl_aux	,750		
prep	,720		
prd_mod	,712		
inf	,464	-,362	,398
factadvl	,459		
rel_obj	,451	,303	
ttr	,444		

Tabela 13. SPSS 20 – Análise Fatorial – Matriz de padrões (recorte) (Método de extração: fatoração de eixo principal / Método de rotação: *Promax* com normalização Kaiser)

Nessa matriz, é possível observar o peso de cada variável (negativo ou positivo) em cada fator extraído. Esses pesos foram usados para calcular os escores de cada fator. Note que algumas variáveis possuem pesos em mais de um fator. Nesses casos, o peso com maior valor absoluto, ou seja, independente de ser positivo ou negativo, define o fator no qual a variável será contabilizada. Por exemplo, a variável *inf* foi contabilizada somente no escore do fator 1.

2.2.2.2.3 Cálculo dos escores dos fatores

O cálculo dos escores de fator é realizado com o propósito de caracterizar cada texto em relação a cada um dos fatores (BIBER, 1988). Esse cálculo é feito a partir dos valores absolutos (sem considerar o sinal de positivo ou negativo) dos pesos de cada variável em seu respectivo fator, de acordo com os resultados da matriz de padrões.

O primeiro passo consiste na padronização dos pesos das variáveis. Na padronização, os valores absolutos das variáveis são convertidos em escores Z (*Z scores*) para remover a influência de uma unidade de medida ou escala. Esse procedimento deve ser feito com todas as variáveis incluídas na extração. O processo de padronização foi feito por meio da opção: “*Analyze*” > “*Descriptive Statistics*” > “*Descriptives*”. A opção “*save standardized values as variables*” foi selecionada para que as variáveis padronizadas sejam criadas na tabela (Figura 11).

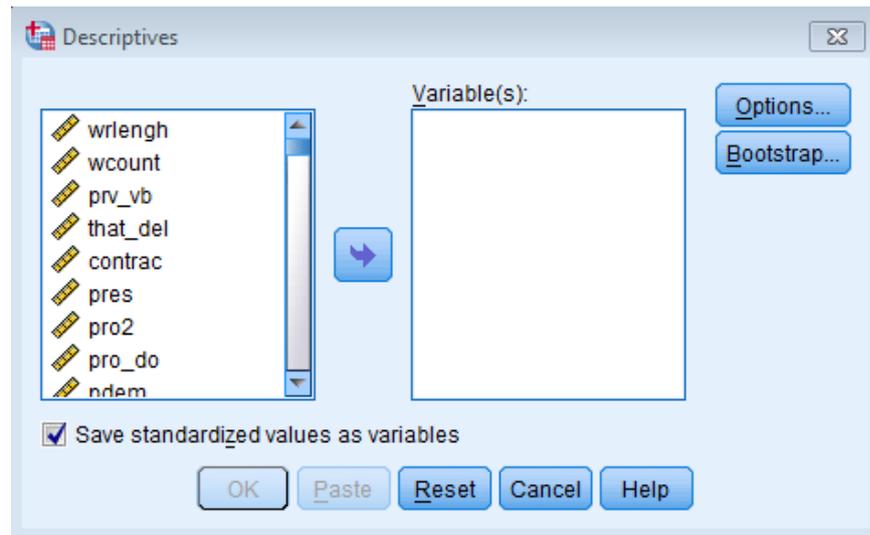


Figura 12. SPSS 20 – Estatísticas Descritivas – Padronização

As novas variáveis são diferenciadas pela inserção da letra Z na frente do nome da variável original. Por exemplo, a variável padronizada criada a partir da variável original *contrac* foi nomeada como *Zcon*.

A partir dessas novas variáveis, o cálculo dos escores de fator pôde ser feito por meio de uma expressão numérica. Esse procedimento foi realizado com a opção de calcular variável: “*Transform*” > “*Compute variable*” (Figura 12).

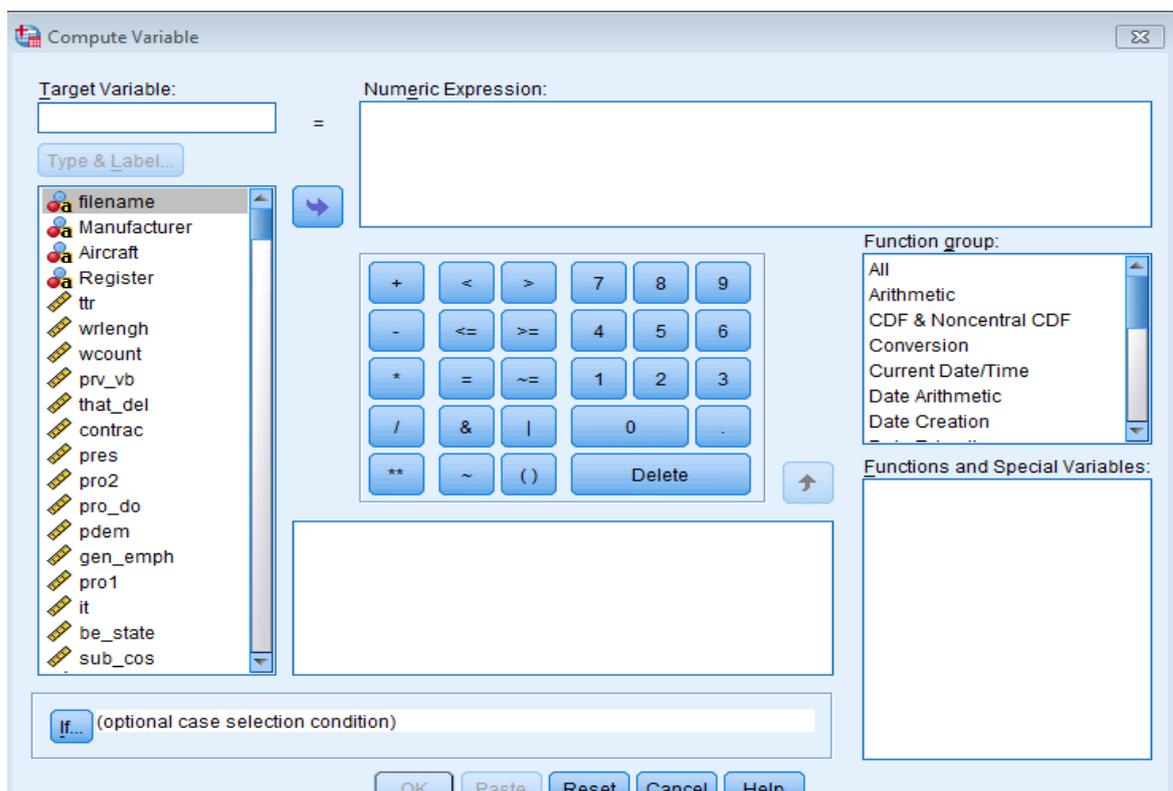


Figura 13. SPSS 20 – Análise Fatorial – Cálculo do escore de fator

O nome da nova variável foi inserido em “*Target variable*”. (ex: Factor02). Para cada fator, a expressão numérica inserida consiste na soma das variáveis de peso positivo menos a soma das variáveis de peso negativo atribuídas àquele fator. Observe, por exemplo, a expressão numérica feita para o Fator 2 no processamento do *corpus* CAM (Quadro 2):

$$\text{Factor02} = (\text{Zbe_state} + \text{Zpos_mod} + \text{Zp_and} + \text{Zo_and} + \text{Zallmodal} + \text{Zpasttense} + \text{Zpred_adj} + \text{Zgroupn} + \text{Zallverb} + \text{Zsua_vb} + \text{Zagls_psv} + \text{Zallpasv} + \text{Zpub_vb} + \text{Zconjuncts} + \text{Zcausev} + \text{Zdowntone} + \text{Zpres} + \text{Zalladj} + \text{Zthat_del} + \text{Zwhiz_vbn} + \text{Zwrlengh} + \text{Zmentalv}) - (\text{Zn} + \text{Zallpro} + \text{Zall_jth} + \text{Zjcmp} + \text{Zpro2} + \text{Zfact_jth} + \text{Zall_th})$$

Quadro 2. Exemplo de expressão numérica para cálculo do escore de fator

As expressões numéricas usadas para o cálculo de todos os escores de fatores, assim como suas respectivas interpretações, constam no capítulo Apresentação e Discussão dos Resultados.

2.2.2.2.4 Análise qualitativa das dimensões

Conforme mencionado no capítulo Fundamentação Teórica e no início do presente capítulo, a AMD contempla análises quantitativas e qualitativas. As etapas quantitativas da análise encerraram-se com o cálculo dos escores de fator, dando-se início então à etapa de análise qualitativa.

A etapa de análise qualitativa consiste na interpretação dos fatores identificados na análise quantitativa. Nessa interpretação, procura-se por uma dimensão funcional que subjaza o padrão de co-ocorrência das variáveis identificadas em cada um dos fatores, ou seja, funções comunicativas subjacentes à linguagem, refletidas no conjunto de características linguísticas que compõem o fator, são identificadas (BIBER, 1988). Segundo Biber, “um agrupamento de características co-

ocorre frequentemente em textos porque elas estão servindo a uma função em comuns nesses textos” (1988, p.91).⁴²

A interpretação deve levar em consideração ambos os agrupamentos positivos e negativos de um fator, pois, conforme supracitado, a dimensão consiste em uma escala e a relação entre os polos positivos e negativos não pode ser descartada. Considerando o fator 3 de Biber (1988) como exemplo, observa-se as variáveis positivas e como elas podem estar relacionadas: oração *WH-* em posição de objeto, oração *WH-* em posição de sujeito, oração *WH-* com preposição inicial, conjunção coordenada e nominalização. Em termos funcionais, apesar de terem funções individuais diferentes, quando agrupadas no mesmo fator, orações *WH-* em posição de objeto, orações *WH-* em posição de sujeito e orações *WH-* com preposição inicial podem ser instrumentos de identificação de referentes explícitos ou elaborados no texto. Além disso, ao considerar a co-ocorrência desses tipos de oração relativa com conjunções coordenadas e nominalização, nota-se que essa referência explícita também tende a ser integrada e informacional. Portanto, pode-se dizer que textos com escore médio positivo nesse fator destacam-se por conter referências explícitas. Assim, quanto mais positivo o escore, mais o texto é caracterizado pelas referências explícitas. O mesmo procedimento de análise é realizado para as variáveis que compõem o polo negativo do fator. No polo negativo, as variáveis com peso mais expressivo são advérbio de tempo, advérbio de lugar e todos os advérbios. Advérbios de tempo e lugar podem ser usados como referência interna no texto ou, mais comumente, como referência a lugares e tempo além do texto em si. Dessa forma, dependem da inferência do ouvinte/leitor, interna ou externamente ao texto. Portanto, pode-se dizer que textos com escore médio negativo nesse fator destacam-se por conter referências dependentes da situação.

A interpretação dos fatores do *corpus* CAM seguiu o mesmo formato de análise descrito acima: cada fator foi observado como uma dimensão; suas características linguísticas com pesos positivos e negativos foram consideradas individualmente, em conjunto e em relação de oposição; e os textos compostos pelas principais características do fator, positivas e negativas, foram analisados a fim de verificar suas funções comunicativas, situacionais e sociais. A fim de melhor identificar os exemplos

⁴² “a cluster of features co-occur frequently in texts because they are serving some common function in those texts.”

das características linguísticas identificadas em cada fator, linhas de concordância foram geradas usando como palavra nódulo a etiqueta (código) morfossintática ou semântica de cada característica linguística analisada. Desse modo, foi possível observar com mais precisão os exemplos de cada característica linguística presente em um determinado fator e os padrões formados por esses exemplos. Além disso, ao analisar o contexto a partir dessas linhas de concordância, pôde-se observar as funções comunicativas e situacionais exercidas no registro em questão. Por exemplo, as linhas de concordância permitiram observar e identificar todos os verbos de probabilidade e fato que controlam orações complementares com *to* em manuais aeronáuticos. Esse tipo de precisão só foi possível por meio desse tipo de análise. Esse procedimento foi realizado para as principais características positivas e negativas de cada um dos três fatores identificados.

Dessa forma, como em Biber (1988), as variações de cada fator puderam ser descritas e a nomeação de cada dimensão foi feita. Após a identificação e análise das dimensões, a análise de variância das variáveis independentes foi realizada, conforme descrito a seguir.

2.2.2.2.5 Análise de variância – ANOVA

Com base nos resultados da fatoração e da subsequente rotulação das dimensões, foi possível realizar um teste estatístico adicional a fim de analisar a significância estatística das variáveis independentes (fabricantes de aeronaves, modelos de aeronaves e registros presentes no *corpus*) em relação às dimensões do CAM. Para isso, foi utilizado o teste para análise de variância (ANOVA), o qual tem por objetivo a comparação de um fator fixo com uma variável dependente para observar se há diferença significativa entre eles. Nesse caso, a ANOVA compara a variação entre as médias de vários grupos com a variação dentro dos grupos. Nesta pesquisa, os fatores fixos são as variáveis independentes, ou seja, os fabricantes de aeronaves, modelos de aeronaves e os tipos de manuais, e as variáveis dependentes

são os três fatores identificados na análise fatorial. A análise de variância foi feita por meio da opção: “Analyze” > “General Linear Model” > “Univariate” (Figura 12).

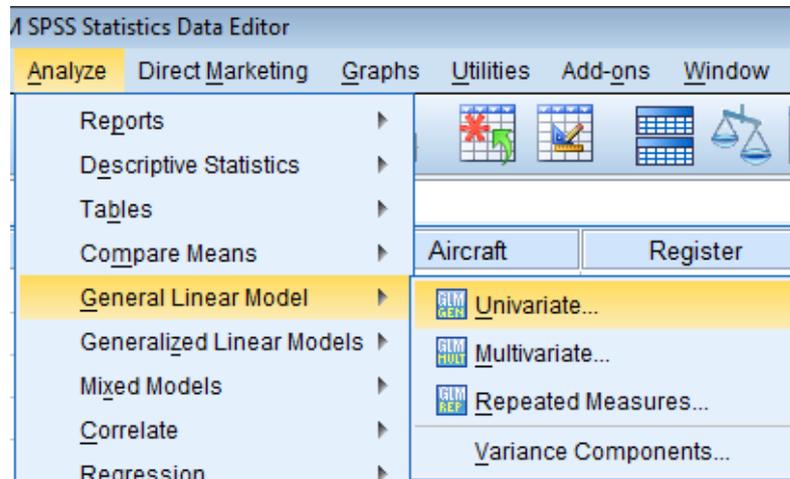


Figura 13. SPSS 20 – ANOVA

Ao clicar em “Univariate...” uma janela aparece para que a variável independente e o fator fixo possam ser escolhidos (Figura 13). Conforme mencionado anteriormente, as variáveis dependentes são os três fatores do CAM e os fatores fixos são as variáveis independentes. O mesmo procedimento foi realizado para cada uma das variáveis independentes e fatores fixos. Todas as opções padrão do programa foram mantidas.

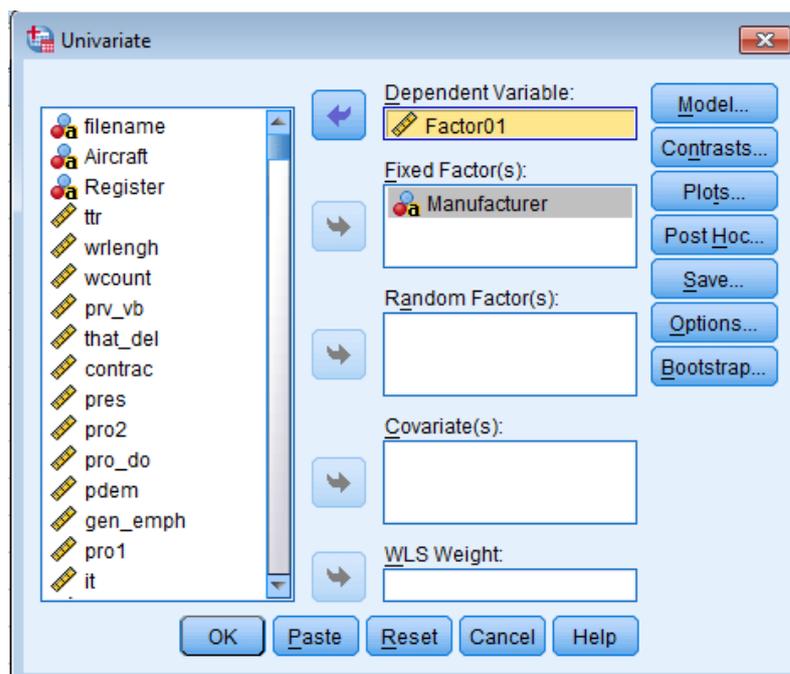


Figura 14. SPSS 20 – ANOVA – Exemplo de cálculo

Em seguida, os resultados da análise de variância foram consolidados em tabelas criadas para cada fator fixo. As tabelas foram feitas a partir da consolidação dos resultados da ANOVA de cada variável dependente (fatores 1 a 3) para cada fator fixo (fabricantes de aeronaves, modelos de aeronaves e os tipos de manuais). Por exemplo, para o fator fixo “fabricantes de aeronaves”, foram geradas três tabelas de resultados de ANOVA, uma para cada fator do CAM. Os resultados dessas três ANOVAS foram consolidados em uma única tabela a fim de facilitar a sua interpretação. Cada tabela consolidada contém os seguintes resultados: F , p , df e R^2 ajustado. Observe o exemplo de uma tabela de ANOVA consolidada (Tabela 14):

Fator	F	p	df	R^2 Ajustado
1	34,119	0,000	4	0,464
2	8,719	0,000	4	0,168
3	11,835	0,000	4	0,221

Tabela 14. SPSS 20 – ANOVA – Exemplo de tabela

Como mencionado, a tabela contém os valores de F , p , df e R^2 ajustado para cada fator do CAM. A estatística F (razão F) avalia a proporção da diferença entre os grupos em função do tamanho da variação dentro de cada grupo. Para se obter a razão F , divide-se a variância entre grupos pela variância dentro de cada grupo. Em uma análise de variância, a razão F é a estatística que indica se a hipótese nula pode ser rejeitada, ou seja, ela indica se é possível afirmar que a variação encontrada não é devida ao acaso. Assim, “quanto maior a razão, maior a chance de que as diferenças entre as medidas sejam em decorrência de algo que não o acaso, ou seja, de efeitos reais”⁴³ (GOMÉZ, 2013, p.54). O valor da razão F considerado estatisticamente significativo é aquele igual ou superior ao valor dos “graus de liberdade” (df , *degrees of freedom*).

Os graus de liberdade são representados pela sigla df (*degrees of freedom*) na tabela ANOVA. Os graus de liberdade representam “a quantidade de valores de uma variável que estão livres para variar”⁴⁴ (GOMÉZ, 2013, p.49). O df é calculado por meio da fórmula $n - 1$, em que n representa o tamanho da amostra. Por exemplo, se há cinco fatores fixos (ex: 5 fabricantes de aeronaves), o df é igual a 4.

⁴³ “The larger the ratio, the greater the likelihood that the differences between the measures are due to something other than chance, namely real effects”

⁴⁴ “the number of values of the variable which are free to vary”

Conforme mencionado na seção 2.2.2.1, p indica o nível de significância da estatística. São considerados significantes valores de p igual ou inferiores a 0,05, o que representa uma chance de menos de 5% de erro, ou seja, pode-se afirmar com 95% de confiança que a variação não é devida ao acaso e sim aos fatores fixos. Assim, quanto menor o valor de p , mais significativo é o resultado estatístico.

A última estatística indicada na tabela é o R^2 ajustado. O R^2 , também chamado de coeficiente de determinação, indica a proporção de variação que pode ser explicada pelo modelo, ou seja, quanto da variação existente na amostra é em razão do fator fixo. Em pesquisas de AMD, o valor de R^2 indica o quanto da variação linguística em uma determinada dimensão pode ser explicada pelo fator fixo (variável independente) sendo considerado. Usando a tabela 14 como exemplo, o valor de R^2 para o fator 1 é de 0,464 (ou 46,4%) e $p = 0,000$, ou seja, pode-se afirmar que 46,4% da variação encontrada nesse fator pode ser explicada pelo fator fixo sendo considerado.

A análise de variância é a etapa final no processo metodológico desta pesquisa. Conforme explicitado neste capítulo, a metodologia empregada seguiu os passos da AMD, iniciando pela coleta, limpeza e anotação do *corpus* CAM, seguido do mapeamento do *corpus* CAM nas dimensões da língua inglesa estabelecidas por Biber (1988), da análise fatorial e subsequente interpretação dos fatores para nomeação das dimensões de variação próprias dos manuais aeronáuticos e terminando com a análise de variância para explicar estatisticamente a variação encontrada.

3 Apresentação e Discussão dos Resultados

Neste capítulo, os resultados das análises fatoriais descritas no Capítulo Metodologia são apresentados e discutidos em detalhes. Primeiramente, são apresentados e discutidos os resultados do mapeamento do *corpus* CAM nas dimensões da língua inglesa estabelecidas por Biber (1988). Em seguida, as dimensões de variação próprias dos manuais aeronáuticos e o impacto das variáveis independentes no *corpus* são apresentados.

3.1 Mapeamento do *Corpus* CAM nas dimensões da língua inglesa (Biber, 1988)

Nesta subseção são apresentados os resultados do mapeamento do *Corpus* CAM nas dimensões de variação do Inglês (BIBER, 1988, 2009). Conforme descrito nos capítulos Fundamentação Teórica e Metodologia, o mapeamento do *corpus* CAM foi realizado de acordo com a pesquisa de 1988, isto é, com as mesmas variáveis em cada fator e as cinco dimensões estabelecidas por Biber (1988). Retomando as dimensões consideradas nessa análise, de acordo com a nomenclatura de Biber (1988) (Tabela 15):

Dimensão	Nome (Biber, 1988)
1	Produção com interação <i>versus</i> informacional
2	Preocupações narrativas <i>versus</i> não-narrativas
3	Referências explícitas <i>versus</i> dependentes da situação
4	Persuasão explícita <i>versus</i> não-explícita
5	Informação abstrata <i>versus</i> não-abstrata

Tabela 15. Dimensões de 1988 consideradas no mapeamento do CAM

As variáveis linguísticas consideradas na pesquisa de Biber (1988) que compõem os cinco fatores identificados e mantidos são (Tabela 16):

Dimensão 1 - Produção com envolvimento versus produção informacional	
Polo Positivo	Polo negativo
Verbo de cognição	Adjetivo atributivo
Omissão de <i>that</i> em oração subordinada	Preposição
Contração	Relação entre item e ocorrência
Pronome em 2ª pessoa	Tamanho de palavra
Verbo no presente	Substantivo
Verbo <i>do</i> como substituto de outro verbo ou sintagma verbal / verbo vicário /pró-verbo	(Voz passiva sem agente)
Negação analítica	(Advérbio de lugar)
Pronome demonstrativo	
Advérbio ou palavra quantificador(a) enfático(a)	
Pronome em 1ª pessoa	
Verbo <i>to be</i> indicativo de estado	
Pronome <i>it</i>	
Partícula do discurso	
Subordinação causativa	
Pronome indefinido	
Advérbio delimitador/atenuador	
Advérbio qualificador/amplificador	
Pronome relativo	
Pronome <i>WH-</i> usado em perguntas	
Verbo modal de possibilidade	
Conjunção coordenada - conectivo clausal	
Oração com pronome <i>WH-</i>	
Preposição desacompanhada (Advérbios)	
Dimensão 2 - Preocupações narrativas versus não-narrativas	
Polo positivo	Polo negativo
Passado	(Adjetivo atributivo)
Aspecto perfeito	(Verbo no presente)
Pronome em 3ª pessoa	
Verbo <i>dicendi</i>	
Negação sintética	
Oração reduzida de gerúndio	
Dimensão 3 - Referências explícitas versus dependentes da situação	
Polo positivo	Polo negativo
Oração <i>WH-</i> em posição de objeto	Advérbio de tempo
Oração <i>WH-</i> com preposição inicial	Advérbio de lugar
Oração <i>WH-</i> em posição de sujeito	Advérbios
Nominalização no singular	
Conjunção coordenada - conectivo frasal	

Dimensão 4 - Persuasão explícita versus não-explícita
Polo positivo (único)
Infinitivo
Verbo modal preditivo
Verbo de persuasão
Conjunção subordinativa condicional
Verbo modal de necessidade
Advérbio usado entre verbo auxiliar e verbo principal (Verbo modal de possibilidade)
Dimensão 5 - Informação abstrata versus não-abstrata
Polo positivo (único)
Conjunção
Voz passiva sem agente
Orações adverbiais reduzidas de participio
Voz passiva com agente e preposição <i>by</i>
Modificador pós-nominal da voz passiva
Outros advérbios/conjunções usadas em orações subordinadas

Tabela 16. Variáveis linguísticas dos Fatores 1 a 5 (cf. Biber, 1988)

O mapeamento consiste em localizar o escore médio de fator calculado para o *corpus* CAM em cada dimensão da língua inglesa entre os registros que compõem essas dimensões. Vale lembrar que os escores de fator são medidas que possibilitam caracterizar cada texto em relação a cada uma das dimensões. O escore médio do *corpus* CAM é representado pela média estatística do escore de fator de todos os textos que perfazem o *corpus*.

O *corpus* CAM (escore médio = - 21) apresenta o escore negativo mais alto na dimensão 1, produção com envolvimento *versus* produção informacional, indicando uma produção altamente informacional. Essa posição corrobora com o que é conhecido sobre manuais técnicos: seu conteúdo tende a ser específico e não interativo. As principais características linguísticas negativas, que fazem parte dessa dimensão e foram identificadas nos manuais são: substantivos, tamanho das palavras, razão forma/item, preposições e adjetivos atributivos. A alta frequência dessas características está relacionada a conteúdos com alta carga de informação e a apresentação cuidadosa da informação no texto, já que substantivos são fundamentais para referências de significados no texto e indicam a alta densidade de informações. Além disso, o tamanho das palavras e a razão forma/item marcam escolhas lexicais precisas a fim de apresentar o conteúdo informacional de forma exata, sem ambiguidades.

O gráfico a seguir (Gráfico 2) mostra onde o *corpus* CAM se situa nessa dimensão. Em todos os gráficos desta subseção, o *corpus* CAM aparece grafado em caixa-alta. Os números na base dos gráficos representam a escala dos escores médios dos registros em cada dimensão. As barras são representações visuais do escore médio de cada registro e quanto maior a barra, maior o escore do registro a que ela se refere, seja positivo ou negativo. O valor do escore médio de cada registro encontra-se ao lado da barra correspondente.

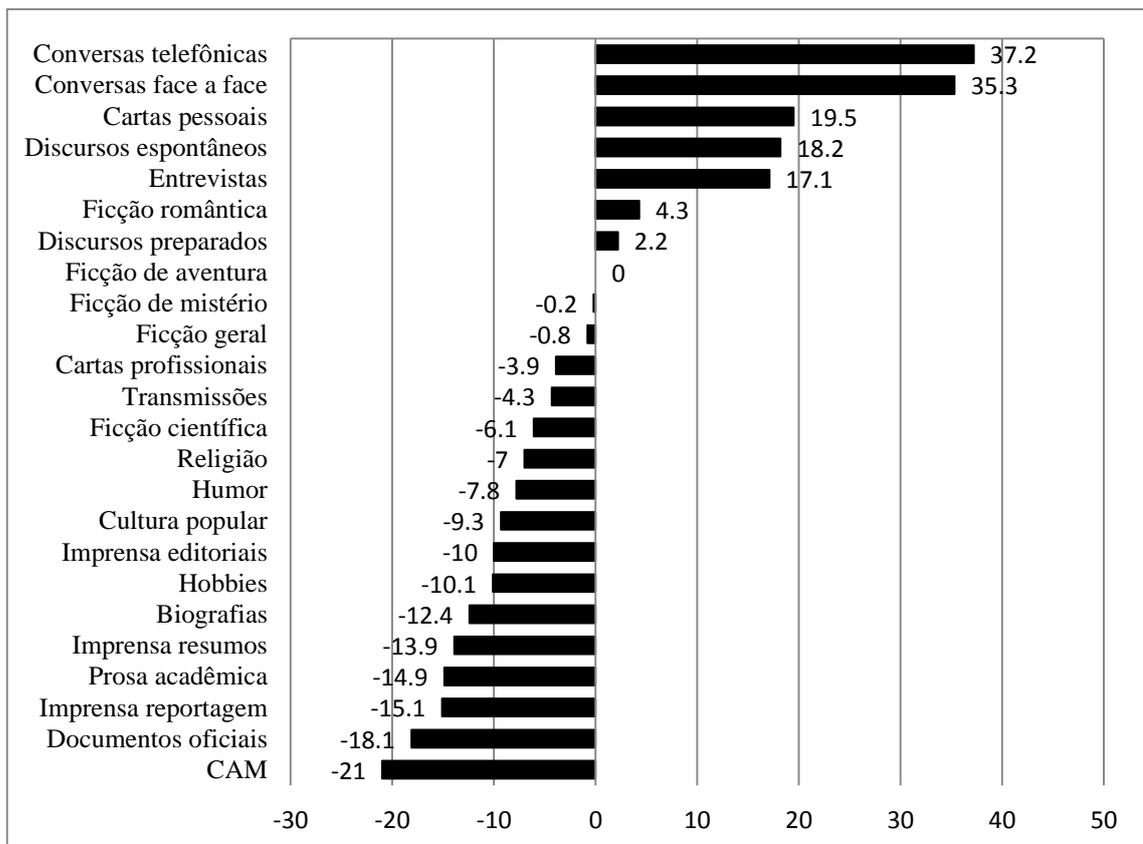


Gráfico 2. Localização do *corpus* CAM na Dimensão 1 – Mapeamento CAM/Biber

O lado direito do gráfico corresponde ao polo positivo, ou seja, ao polo “produção com envolvimento”, da dimensão. O lado esquerdo do gráfico corresponde ao polo negativo, ou seja, ao polo “produção informacional”, da dimensão. Note que o *corpus* CAM está localizado abaixo de registros que apresentam grande quantidade de informação e precisão, tais como prosa acadêmica, reportagem e documentos oficiais. Observe o trecho de um texto com alto escore negativo na Dimensão 1, em que há presença marcante de substantivos, adjetivos atributivos e preposições (Quadro 3):

A. Zones are used to identify different areas of the aircraft. Each zone contains three numbers. The first number identifies the primary areas or major zones of the aircraft. They are the lower fuselage, upper fuselage, empennage, engines, wings, landing gears, and the doors. Each major zone is divided into submajor zones: the flight compartment, for example. These submajor zones are identified by the second number of the zone number. The submajor zones, in turn, are divided into zones: the radome, for example. They are identified by the third number in the zone number.

MD11 – AMM – Dimensions & Areas

Quadro 3. Trecho de texto – Dimensão 1 – Mapeamento CAM/Biber

O valor médio dos escores dos manuais aeronáuticos que compõem o *corpus* CAM na dimensão 2, preocupações narrativas *versus* não narrativas, consta no Gráfico 3. Assim como na dimensão 1, o CAM também apresenta o valor negativo mais alto na dimensão 2, indicando conteúdo não narrativo. Nessa dimensão, o *corpus* CAM se situa abaixo de documentos oficiais e transmissões. O posicionamento do *corpus* CAM tão próximo ao registro falado “transmissões” pode causar estranhamento. No entanto, o registro “transmissões” é composto por textos escritos para serem falados no rádio, sendo eles transmissões de eventos de esporte e comentários sobre temas não relacionados ao esporte. Essas transmissões descrevem fatos e opiniões no tempo presente e não narrativas sobre eventos passados. Isso ajuda a esclarecer a posição do *corpus* CAM. Em geral, as características presentes na dimensão 2 associadas a narração, como verbos no tempo passado, pronomes pessoais em terceira pessoa, verbos *dicendi*, orações reduzidas e negações sintéticas, não são tão marcadas nos manuais quanto as características com peso negativo, como verbos no tempo presente e adjetivos atributivos. Essas características negativas têm peso maior na dimensão 1, e sua co-ocorrência pode indicar o uso mais frequente de referências nominais elaboradas em discursos não narrativos.

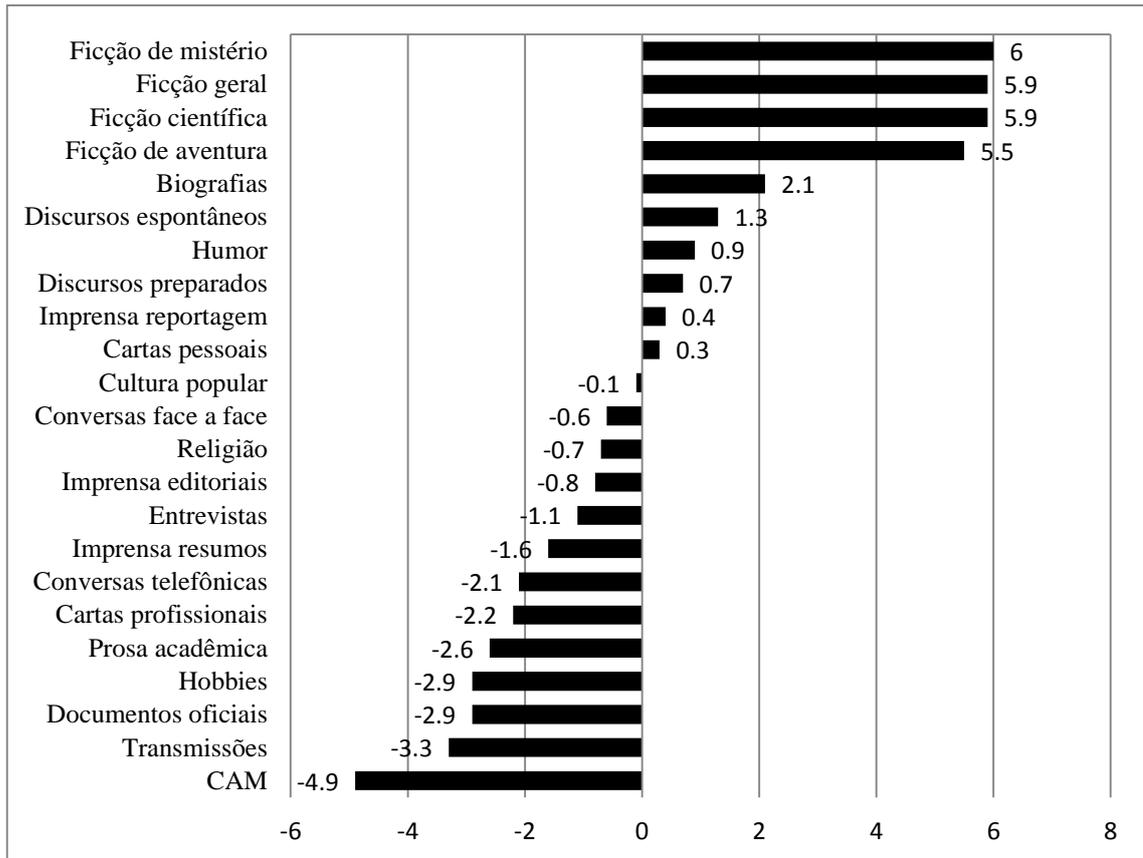


Gráfico 3. Localização do corpus CAM na Dimensão 2 – Mapeamento CAM/Biber

Observe no quadro (Quadro 4) a seguir um trecho de texto com escore médio -4,98, o qual não apresenta nenhum traço de narração e contém alta carga de verbos no presente.

Make sure that you follow the instructions to prevent injury to persons and damage to other airplanes, equipment or structures.
 The engine suction can pull a person or loose material into the air inlet.
 The engine exhaust is very hot and moves at high speeds. The fan air is very strong and can move persons or material a long distance.
 When you operate the thrust reverser, the fan air goes to the front of the engine.

B777 – AMM P&P – 71H

Quadro 4. Trecho de texto – Dimensão 2 – Mapeamento CAM/Biber

Representada no Gráfico 4, a posição do *corpus* CAM na dimensão 3, referências explícitas *versus* referências dependente do contexto, apresenta escore médio positivo, o que sugere a apresentação de informações com referências explícitas. As principais características com peso positivo nessa dimensão são três tipos de orações subordinadas relativas (com 'que' na posição de sujeito, com 'que'

na posição de objeto e *pied-piping* – orações *wh-* com preposição inicial), coordenação sintagmática e nominalizações. A co-ocorrência dessas características indica que um texto com referências explícitas também tende a ser integrado e informacional. No polo negativo, as características marcantes são os advérbios de tempo e lugar, os quais são usados geralmente para referenciar lugares e tempos externos ao texto. Novamente, o *corpus* CAM localiza-se abaixo de prosa acadêmica e documentos oficiais, textos com conteúdo altamente explicitado para evitar ambiguidades.

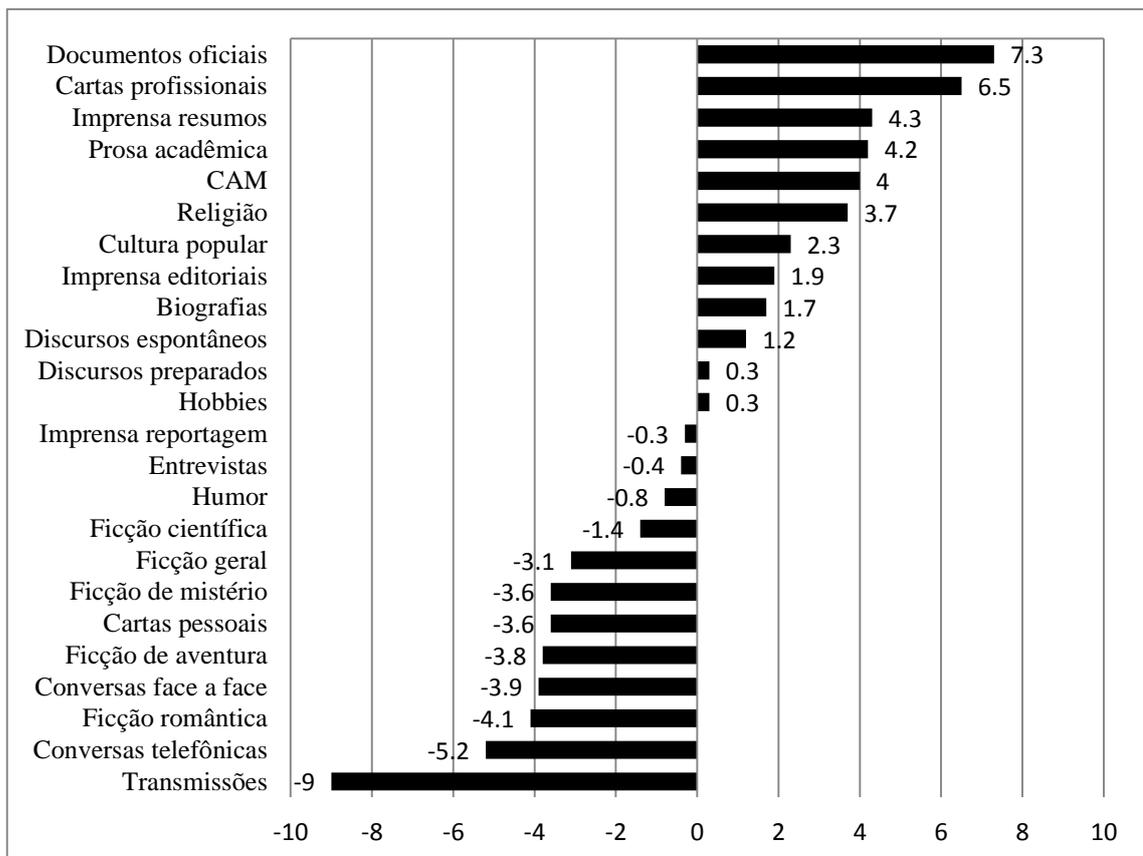


Gráfico 4. Localização do corpus CAM na Dimensão 3 – Mapeamento CAM/Biber

O trecho abaixo retirado de um texto do CAM com score médio 4,25 ajuda a ilustrar a tendência de textos contendo referências explícitas serem mais integrados e informacionais (Quadro 5).

The engine is a two spool, axial flow, high bypass ratio turbofan engine. Its compression system features a single stage fan, a four stage booster, and a ten stage High Pressure Compressor (HPC). The Low Pressure Compressor (LPC) is driven by a five stage Low Pressure Turbine (LPT) and the HPC by a two stage High Pressure Turbine (HPT). The HPT also drives a gearbox which, in turn, drives the engine and aircraft mounted accessories. The two shafts are supported by five main bearings. The engine incorporates a full authority digital Electronic Engine Control (EEC). The control system governs all engine functions, including power management. Reverse thrust for braking the aircraft after landing is supplied by an integrated system which acts on the fan discharge airflow. Borescope accesses are provided for inspection purposes.

A320 – AMM – 72

Quadro 5. Trecho de texto – Dimensão 3 – Mapeamento CAM/Biber

Na dimensão 4, expressão explícita de persuasão *versus* não explícita, o *corpus* CAM apresenta escore negativo, ou seja, os textos dos manuais aeronáuticos, em geral, não apresentam características linguísticas de persuasão. Essa dimensão possui características somente com peso positivo: infinitivos, modais de predição, verbos de persuasão (*agree, ask, order, demand, insist, instruct, propose, recommend, suggest*), subordinação condicional, modais de necessidade, auxiliares '*split*' e modais de possibilidade. Essas características tendem a mostrar a intenção do autor/falante de convencer o leitor/ouvinte por meio de argumentação. Portanto, valores negativos indicam a escassez dessas características e uma tendência à expressão não explícita de persuasão. O Gráfico 5 apresenta onde o *corpus* CAM se situa nessa dimensão, aparecendo perto de textos de ficção e resumos de imprensa.

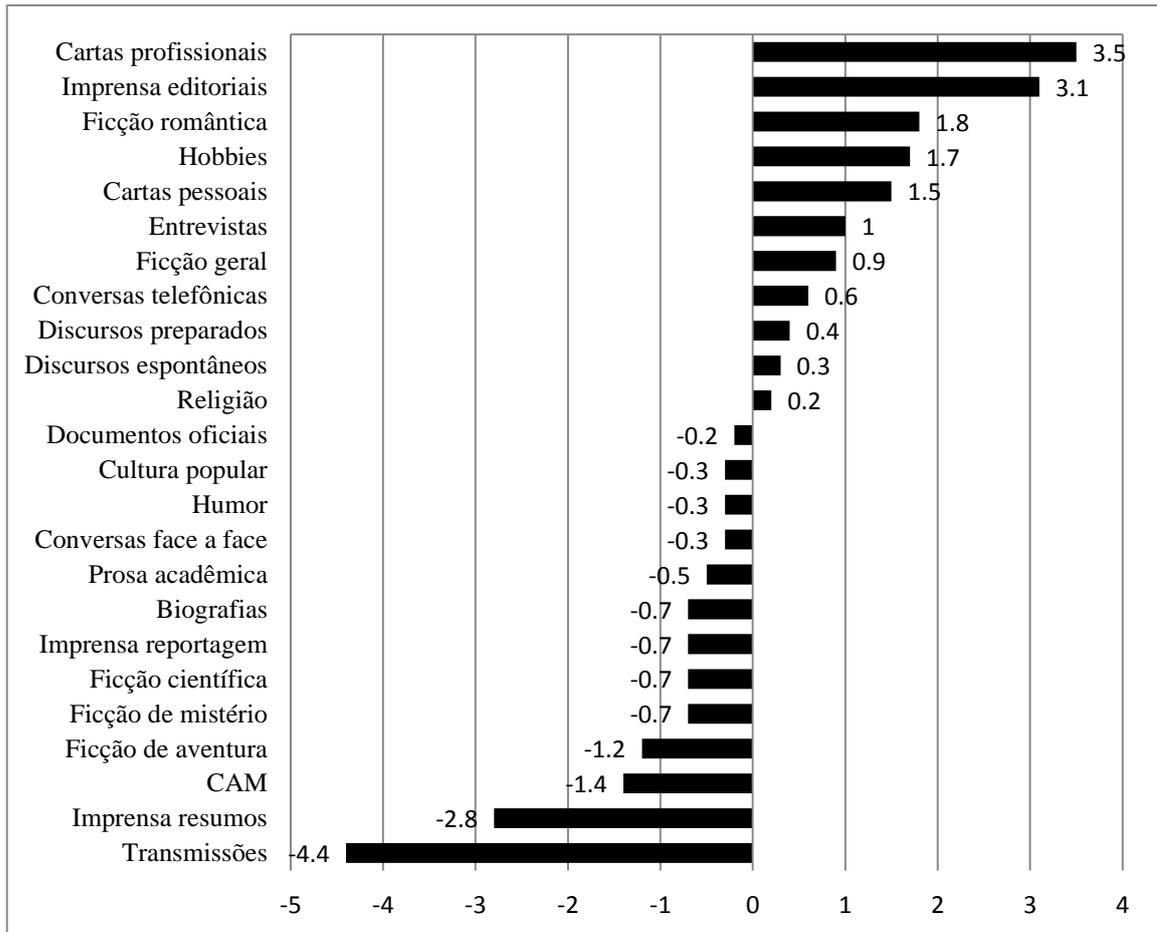


Gráfico 5. Localização do corpus CAM na Dimensão 4 – Mapeamento CAM/Biber

A posição do *corpus* CAM, junto a textos não marcados para persuasão/argumentação era esperada devido ao seu caráter técnico e por conter textos contendo instruções que devem ser seguidas sem questionamento. No entanto, é possível notar em determinados textos do CAM que modais de possibilidade atuam com a função de informar os danos à pessoa ou à aeronave que podem ocorrer caso o procedimento não seja feito de forma correta. Nesse caso, pode-se dizer que funcionam como forma de chamar a atenção para possíveis consequências danosas e convencer o leitor a realizar a tarefa cuidadosamente. Observe o trecho de um texto com score médio -1,55 que ilustra esse ponto (Quadro 6):

CAUTION:

DO NOT LET THE DU HANG BY THE FIBER-OPTIC CABLE. THE WEIGHT OF THE DU CAN CAUSE DAMAGE TO THE FIBER-OPTIC CONNECTOR.

Remove the DU as follows:

- a. Hold the DU and carefully pull the DU away from the console until you can get access to the DU connector.
- b. Disconnect the cable connector from the DU connector.
- c. Put a protective cover on the front of the DU.

NOTE:

If provided, you can use the display cover from the new DU.

- d. Carefully lift the DU out of the console.
- e. Put dust caps on the DU connector and the cable connector

B777 – AMM P&P – 46

Quadro 6. Trecho de texto – Dimensão 4 – Mapeamento CAM/Biber

O Gráfico 6 apresenta a localização do *corpus* CAM na dimensão 5, informação abstrata *versus* não abstrata. Essa dimensão apresenta características linguísticas com pesos positivos e negativos. As principais características linguísticas com peso positivo são: conjunções, passivas com e sem agente, orações com particípio passado, orações com particípio passado e apagamento do pronome relativo, outras orações subordinadas adverbiais e adjetivos predicativos. Nessa dimensão, quase não há características com alto peso negativo, com exceção da razão forma/item, ou seja, a distribuição de grande variedade lexical, refletida na razão forma/item, é inversamente proporcional à presença de passivas, conjunções, etc. Em geral, os textos dessa dimensão são marcados por conteúdo informacional abstrato, técnico e formal em contrapartida a outros tipos de textos. O *corpus* CAM localiza-se em uma posição praticamente neutra, entre resumos e reportagens de imprensa, textos não marcados por conteúdo técnico ou informacional. Isso surpreende, pois conforme visto nas outras dimensões, os textos do CAM podem ser considerados altamente informacionais.

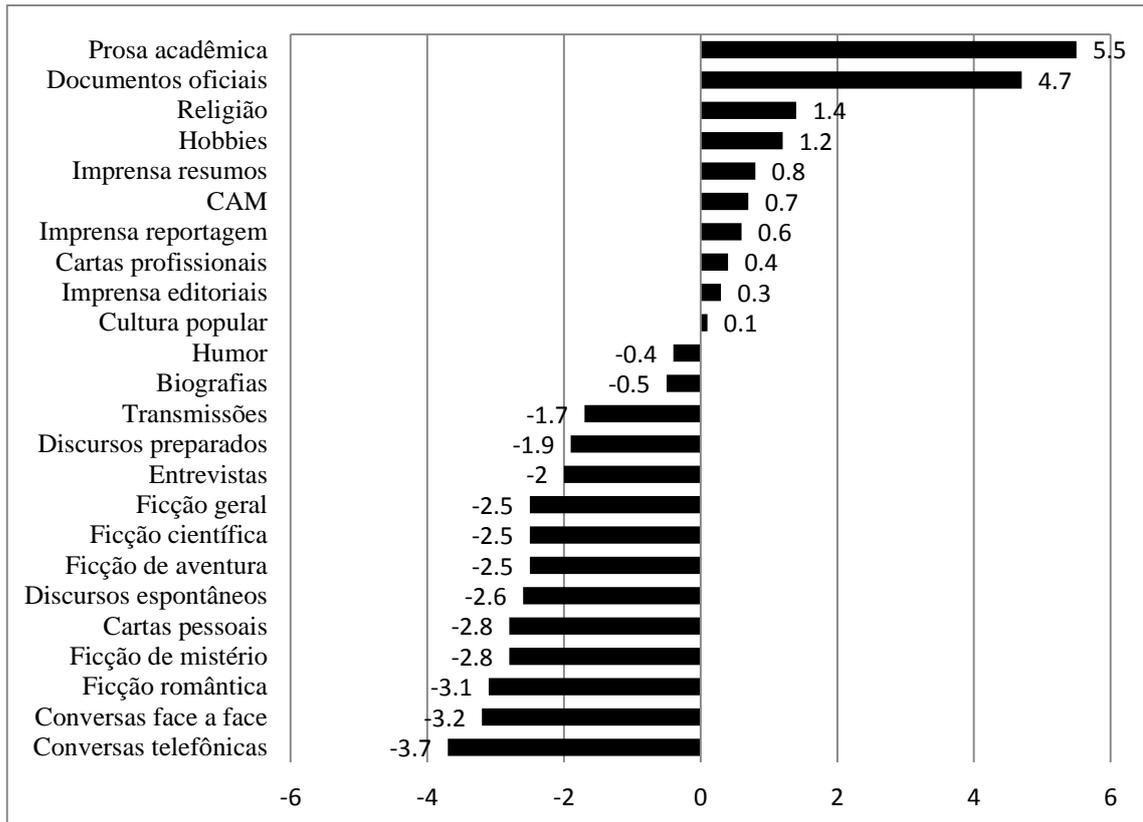


Gráfico 6. Localização do corpus CAM na Dimensão 5 – Mapeamento CAM/Biber

Observe no quadro abaixo (Quadro 7), um trecho de texto de escore médio 1,01:

Components
HORIZONTAL STABILIZER
 The horizontal stabilizer is an all-metal, three spar construction hinged to the rear fuselage, designed as a movable control surface for longitudinal trim. The stabilizer consists of two sections attached to each other at the aircraft center line, and can be removed as a single unit. Each section has a one-section leading edge and a tip fairing.

ELEVATOR
 The elevators are non-balanced flight controls hinged to the THS - Trimmable Horizontal Stabilizer by means of five hinge fittings. Each elevator is actuated by two hydraulic actuators and its movement is controlled by two LVDT - Linear Variable Differential-Transducer.

E190 – AMM SDS – 55

Quadro 7. Trecho de texto – Dimensão 5 – Mapeamento CAM/Biber

A análise do mapeamento do *corpus* CAM nas dimensões de variação da língua inglesa (BIBER, 1988) sugere que os textos de manuais aeronáuticos tendem a ser essencialmente informacionais e não narrativos, contendo referências explícitas, expressão não explícita de persuasão e informações abstratas. Além disso, o

mapeamento apontou que os manuais, em geral, estão próximos a registros mais informativos, tais como prosa acadêmica e documentos oficiais.

A seguir as análises fatoriais e a subsequente interpretação e nomeação dos fatores identificados para o *corpus* CAM são apresentadas.

3.2 As dimensões do *Corpus* CAM

A análise do *corpus* CAM para a identificação das dimensões de variação próprias dos manuais aeronáuticos foi realizada conforme os procedimentos descritos no capítulo Metodologia.

Após a coleta e a limpeza e a subsequente anotação dos textos e contagem de frequências realizadas pelos programas *Biber Tagger* e *TagCount*, a análise do *corpus* CAM foi iniciada. A planilha de dados, produzida pelo *TagCount*, forneceu os valores de frequência já normalizados para as 127 variáveis linguísticas. O processamento dessas variáveis foi feito, conforme supracitado, pelo programa SPSS versão 20 a partir da opção “Redução de Dimensão”, a qual permite a realização de análise fatorial, conforme descrito na Metodologia.

Primeiramente, foi realizada uma fatoraçoão exploratória de eixo principal (*Promax*) sem rotaçoão. Vale ressaltar que a fatoraçoão de eixo principal é a mais usada em análises linguísticas, pois permite a extraçoão da variaçoão máxíma compartilhada pelas variáveis em cada fator. Os primeiros testes estatísticos, teste KMO e teste de esfericidade Bartlett, tinham a função de validar o processo e contaram com valor acima de 0,5 para o KMO (KMO=0,602) e nível de significância igual a zero ($p=0,000$), conforme descrito a seguir (Tabela 17).

Medida de adequação de amostragem de Kaiser-Meyer-Olkin		,602
Teste de esfericidade de Bartlett	Qui-quadrado aprox.	26765,877
	Df	6105
	Sig.	,000

Tabela 17. CAM – Teste KMO e teste de esfericidade de Bartlett

Esses resultados forneceram a validação necessária para dar prosseguimento à análise fatorial. A seguir, a tabela de comunalidades foi analisada para verificação da existência de variáveis com valores inferiores a 0,15, ponto de corte para pesquisas em AMD, conforme explicitado no capítulo Metodologia. Nenhuma variável apresentou valor inferior a 0,15, sendo que o menor valor presente na tabela de comunalidades era de 0,771. Portanto, todas as variáveis foram mantidas para a realização da análise fatorial rotacionada.

Em seguida, o gráfico de sedimentação (*scree plot*) foi analisado. O *scree plot* apresenta os autovalores (*eigenvalues*) para cada fator no eixo vertical (eixo *y*) e o número de fatores (*factor number*) no eixo horizontal (eixo *x*). Conforme mencionado anteriormente na Metodologia, a definição do número ideal de fatores a serem extraídos é baseada na indicação de uma quebra no gráfico, mostrando o ponto a partir do qual fatores adicionais pouco contribuem para a análise dos dados. O *scree plot* a seguir mostra essa quebra após os seis primeiros fatores (Gráfico 7).

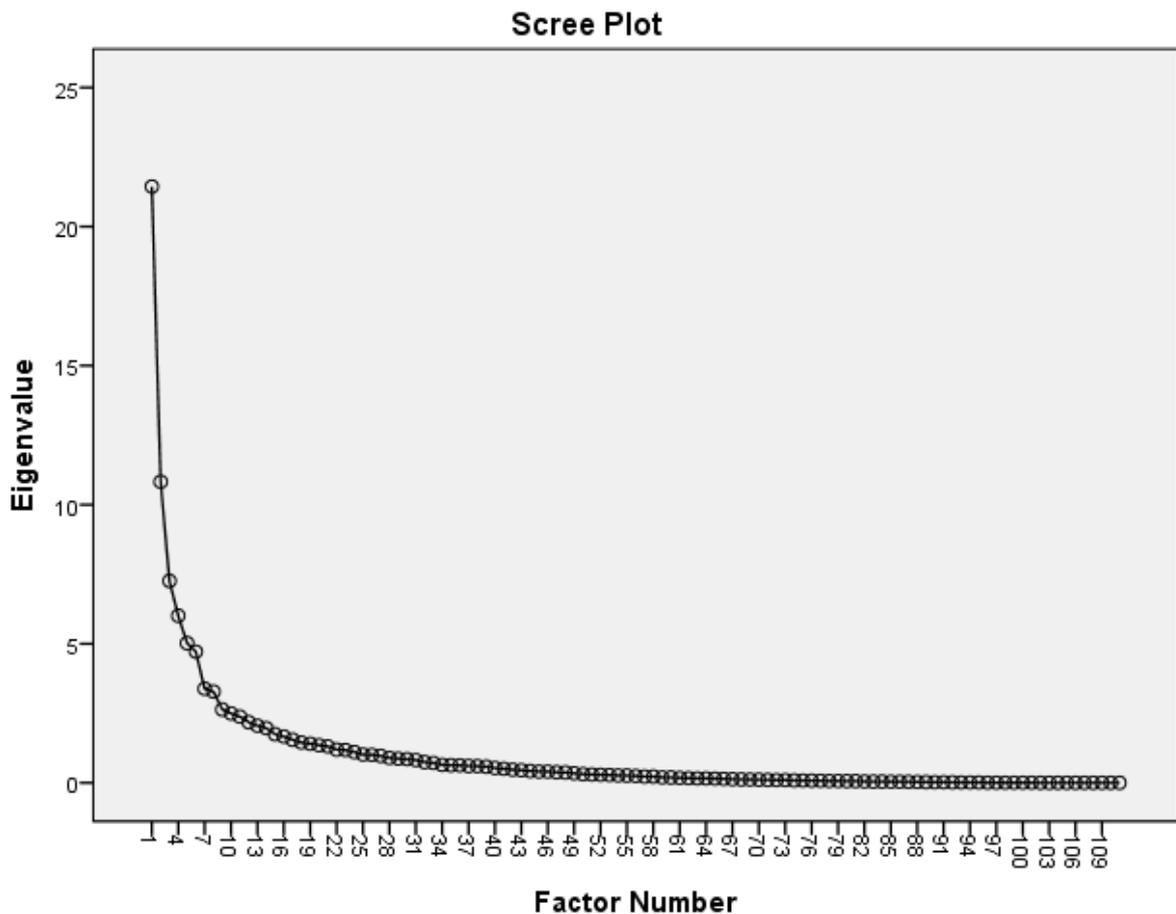


Gráfico 7. CAM – Scree Plot

No entanto, ao considerar o espaçamento dos autovalores até o quinto fator e a interpretação preliminar da fatora o rotacionada com seis fatores, optou-se por realizar a fatora o de eixo principal rotacionada com tr s fatores para an lise final. Al m disso, a an lise com tr s fatores mostrou-se a mais “interpret vel”, conforme explicitado na Metodologia.

A tabela de Vari ncia Total Explicada (vide recorte na Tabela 18) mostrou que seis fatores respondem por 50% da varia o acumulada e tr s fatores respondem por 36% da varia o acumulada, valores considerados significativos que validam os achados resultantes da an lise do *scree plot*. Essa redu o na porcentagem de varia o capturada quando uma solu o com menos fatores   escolhida   normal. Pode-se dizer que uma propor o menor da varia o existente nos textos do *corpus*   explicada pela solu o com tr s fatores. No entanto, tamb m   poss vel afirmar que esses tr s fatores concentram as principais caracter sticas subjacentes   varia o. A tabela completa de Vari ncia Total Explicada encontra-se no Anexo C.

Fator	Autovalores iniciais			Soma da extração de cargas quadráticas			Somas rotacionadas das cargas quadráticas
	Total	% de Variação	% acumulada	Total	% de Variação	% acumulada	Total
1	21,441	19,317	19,317	21,043	18,957	18,957	14,712
2	10,824	9,751	29,068	10,469	9,432	28,389	13,998
3	7,260	6,541	35,609	6,835	6,157	34,547	14,495
4	6,007	5,412	41,020	5,473	4,930	39,477	11,167
5	5,022	4,524	45,545	4,617	4,160	43,636	6,503
6	4,717	4,249	49,794	4,248	3,827	47,463	7,327
7	3,384	3,049	52,843				
8	3,278	2,953	55,796				
9	2,636	2,375	58,171				
10	2,493	2,246	60,417				
11	2,380	2,144	62,562				
12	2,180	1,964	64,526				
13	2,052	1,848	66,374				
14	1,969	1,774	68,148				
15	1,746	1,573	69,721				

Tabela 18. CAM – Variância Total Explicada

A partir desses dados, uma nova fatoração de eixo principal (*Promax*) foi realizada, porém agora rotacionada com três fatores fixos. A primeira matriz analisada após a fatoração rotacionada é a Matriz de Correlação entre Fatores. Essa matriz avalia a extensão da dependência estatística existente entre os fatores. Embora em análises fatoriais, o primeiro fator extraído apresente a quantidade máxima de co-ocorrência entre todas as variáveis, o segundo fator apresente a quantidade máxima de co-ocorrência entre as palavras restantes, e assim por diante, conforme supracitado, a fatoração de eixo principal considera que pode haver variação compartilhada entre os fatores. Portanto, espera-se que haja alguma correlação entre os fatores, porém de baixa densidade. A Tabela 19 apresenta a correlação entre os três fatores do CAM.

Fator	1	2	3
1	1,000		
2	0,303	1,000	
3	0,007	0,012	1,000

Tabela 19. CAM – Matriz de Correlação entre Fatores

Nota-se que há certa codependência entre os três fatores, sendo a dependência mais expressiva entre os fatores 1 e 2 (30%). Os valores de correlação entre os fatores 1 e 3 e os fatores 2 e 3 aproximam-se de zero, o que indica que esses fatores são praticamente independentes.

Dando prosseguimento à análise, a Matriz de Padrões foi analisada. Essa é a principal matriz a ser analisada em pesquisas de AMD, pois é a partir de seus dados que as variáveis pertencentes a cada fator são identificadas. Isto é, a Matriz de Padrões destaca quais variáveis possuem maior valor absoluto (ou seja, sem considerar se é positivo ou negativo) em cada um dos fatores. Observe a tabela a seguir (Tabela 20) com os valores das variáveis em cada fator.

	Fator		
	1	2	3
all_vto	,828		
all_to	,795		
efrt_vto	,773		
nonf_vth	,762		
all_vth	,758		
spl_aux	,750		
prep	,720		
prd_mod	,712		
allwhrel	,690		
rel_subj	,676		
colorj	,673		
abstrcn	,669		
occurv	,655		
existv	,628		
gen_emph	,605		
n_nom	-,604		
vprogrsv	,586		
by_pasv	,564		
quann	,541		
att_vth	,531		

sub_othr	,529		
vcmo	,518		
nec_mod	,485		
rel_pipe	,477		
sizej	,472		
x1_jto	,470		
inf	,464	-,362	,398
factadvl	,459		
rel_obj	,451	,303	
ttr	,444		
advs	,441		
topicj	,432		
relatnj	,411		
lkly_vth	,394	,323	
dsre_vto	,383		
prob_vto	,347		
cognitn	,346		
wcount	-,340		
all_nth	,339		
all_nto	,326		
amplifr	,320		
actv	,314		
all_advl	,310		
spch_vto			
act_tpv			
x2_jto			
nonfadvl			
sub_cos			
pl_adv			
humann			
pany			
pro1			
timej			
mntl_vto			
be_state	-,397	,975	
pos_mod	-,358	,969	
p_and	-,344	,922	
o_and		,848	
allmodal		,822	
pasttse	-,301	,804	
pred_adj	-,304	,758	
groupn		,709	

allverb		,708	
sua_vb		,696	
agls_psv	,357	,690	
allpasv	,410	,681	
pub_vb		,677	
conjuncts		,676	
causev	,342	,639	
downtone		,626	
pres		,624	
alladj		,598	,480
that_del		,499	
n		-,444	
allpro	-,320	-,435	
whiz_vbn		,427	
all_jth	-,327	-,424	,365
wrlengh		,408	,343
jcmp	-,338	-,407	,348
mentalv		,396	
pro2	-,344	-,374	
fact_jth	-,318	-,371	
all_th		-,371	,345
placen			
pdem			
act_ipv			
have			
finlprep			
lkly_jth			
x4_jto			
sub_cnd			-,793
aspectv			-,747
tm_adv			-,623
allconj		,463	-,579
att_jth		-,319	,546
x5_jto		-,326	,545
pro_do	-,446		-,530
pro3			,510
all_jto			,492
commv			-,474
adj_attr			,407
tccnrt			-,366
prcessn			-,357
that_rel			,335

concrtn			,311
perfects			
it			
evalj			
prv_vb			
prtcle			
fact_vth			

Tabela 20. CAM – Matriz de padrões (Método de extração: fatoraçoão de eixo principal / Método de rotaçoão: proporçoão máxíma com normalizaçoão Kaiser)

Ao analisar a Tabela 20, nota-se que há variáveis com valores em todos os fatores e variáveis sem valor em todos os fatores. Para classificar uma variável com valor em todos os fatores em um determinado fator, considerou-se o maior valor absoluto, ou seja não foi levado em cosideraçoão se esse valor era positivo ou negativo. Por exemplo, a variável *inf* possui valores em todos os fatores (0,464 no fator 1; -0,362 no fator 2; e 0,398 no fator 3), mas foi considerada para análise como pertencente ao fator 1 por ser o de maior valor. A presença de variáveis sem valor (por exemplo, *it*) é devida à exclusão de variáveis com valores absolutos inferiores a 0,30, seguindo o critério estabelecido por Biber (1988).

Para cada fator (1 a 3), foram listadas as variáveis de maior valor. Em seguida, conforme descrito na Metodologia as variáveis foram padronizadas para obtençoão do *Zscore*. Os valores das variáveis padronizadas (que passaram a ter a letra *Z* na frente, por exemplo, *Zinf*) foram usados para calcular os escores de fator para cada um dos textos que perfazem o *corpus* CAM. Esse cálculo foi feito por meio da inserçoão de expressões numéricas para cada fator no SPSS, as quais continham a soma das variáveis padronizadas que foram identificadas para cada um dos fatores. As expressões numéricas usadas foram:

- (a) **Fator 1:** $(Zall_vto + Zall_to + Zefrt_vto + Znonf_vth + Zall_vth + Zspl_aux + Zprep + Zprd_mod + Zallwhrel + Zrel_subj + Zcolorj + Zabstrcn + Zoccurv + Zexistv + Zgen_emph + Zvprogrsv + Zby_pasv + Zquann + Zatt_vth + Zsub_othr + Zvcmo + Znec_mod + Zrel_pipe + Zsizej + Zx1_jto + Zinf + Zfactadvl + Zrel_obj + Ztr + Zadvsv + Ztopicj + Zrelatnj + Zkly_vth + Zdsre_vto + Zprob_vto + Zcognitn + Zall_nth + Zall_nto + Zamplifr + Zactv + Zall_advl) - (Zn_nom + Zwcount)$

- (b) **Fator 2:** (Zbe_state + Zpos_mod + Zp_and + Zo_and + Zallmodal + Zpasttse + Zpred_adj + Zgroupn + Zallverb + Zsua_vb + Zagls_psv + Zallpasv + Zpub_vb + Zconjuncts + Zcausev + Zdowntone + Zpres + Zalladj + Zthat_del + Zwhiz_vbn + Zwrlengh + Zmentalv) - (Zn + Zallpro + Zall_jth + Zjcmp + Zpro2 + Zfact_jth + Zall_th)
- (c) **Fator 3:** (Zatt_jth + Zx5_jto + Zpro3 + Zall_jto + Zadj_attr + Zthat_rel + Zconcrtn) - (Zsub_cnd + Zaspectv + Ztm_adv + Zallconj + Zpro_do + Zcommv + Ztccnrt + Zprcessn)

De posse das variáveis que compõem os polos positivos e negativos de cada um dos fatores, deu-se início a análise qualitativa das características linguísticas por essas variáveis. A análise qualitativa foi realizada por meio da observação e análise de exemplos primeiramente das características identificadas em si e em seguida do contexto em que elas apareciam, ou seja, foram analisados exemplos de textos com maior (polo positivo) e menor (polo negativo) escore médio em cada fator.

Essa interpretação minuciosa levou à nomeação dos fatores, transformando-os em dimensões de variação. Os nomes das três dimensões de variação identificadas estão elencados na tabela a seguir (Tabela 21). Os nomes das dimensões foram elaborados em português e em inglês para fins de divulgação da pesquisa.

Dimensão	Nome
1	Ampla descrição de sistemas (<i>Broad system descriptions</i>)
2	Discurso de tomada de decisões <i>versus</i> Discurso de procedimentos (<i>Decision-making discourse versus Procedural discourse</i>)
3	Discurso de solução de problemas <i>versus</i> Descrição de partes específicas (<i>Problem-solving discourse versus Specific parts descriptions</i>)

Tabela 21. Dimensões do *corpus* CAM

As dimensões e suas características linguísticas e funcionais, assim como exemplos de textos do CAM com maior e menor polaridade em cada delas, são apresentadas a seguir.

3.2.1 Dimensão 1

A primeira dimensão, “ampla descrição de sistemas”, parece ser marcada por discurso com alta densidade lexical e que reflete descrições amplas. A dimensão conta com a maioria das características no polo positivo, sendo que apenas duas características linguísticas estão presentes no polo negativo. Sendo assim, optou-se por nomear a dimensão a partir das características do polo positivo, ou seja, os textos dessa dimensão podem ser caracterizados em um continuum contendo mais ou menos descrições amplas. Observe as características linguísticas presentes nessa dimensão (Tabela 22).

Polo Positivo	
Etiqueta	Característica Linguística
all_vto	todas as orações complementares com <i>to</i> controlada por verbos
all_to	todas as orações complementares com <i>to</i>
efrt_vto	<i>to</i> usado em oração controlada por verbos de modalidade, causalidade e esforço
nonf_vth	<i>that</i> usado em oração complementar controlada por verbo não factivo
all_vth	todas as orações complementares com <i>that</i> controlada por verbos
spl_aux	advérbio usado entre verbo auxiliar e verbo principal
prep	preposição
prd_mod	verbo modal preditivo
allwhrel	todas as orações relativas com pronome <i>WH-</i>
rel_subj	oração <i>WH-</i> em posição de sujeito
colorj	adjetivo – cor
abstrcn	substantivo relacionado a coisas ou processos
occurv	verbo relacionado a evento/ocorrência
existv	verbo relacionado a existência ou relacionamento
gen_emph	advérbio ou palavra quantificador(a) enfático(a)
vprogrsv	presente contínuo
by_pasv	voz passiva com agente e preposição <i>by</i>
quann	substantivo – quantidade
att_vth	<i>that</i> usado em oração complementar controlada por verbo atitudinal
sub_othr	outros advérbios/conjunções usadas em orações subordinadas
vcmo	<i>that</i> usado em oração complementar controlada por verbo
nec_mod	verbo modal de necessidade
rel_pipe	oração <i>WH-</i> com preposição inicial
sizej	adjetivo – tamanho
x1_jto	<i>to</i> usado em oração complementar controlada por adjetivos de certeza
inf	infinitivo
factadvl	advérbio factivo
rel_obj	oração <i>WH-</i> em posição de objeto

ttr	relação entre item e ocorrência
advs	advérbio
topicj	adjetivo relacionado a tópicos
relatnj	adjetivo – relacionamentos
lkly_vth	<i>that</i> usado em oração complementar controlada por verbo de probabilidade
dsre_vto	<i>to</i> usado em oração complementar controlada por verbos de desejo, intenção e decisão
prob_vto	<i>to</i> usado em oração complementar controlada por verbos de probabilidade e fato
cognitn	substantivo cognitivo
all_nth	todas as categorias de orações complementares com <i>that</i> controladas por substantivos
all_nto	<i>to</i> usado em oração controlada por substantivos de posicionamento
amplifr	advérbio qualificador/amplificador
actv	verbo relacionado a atividade/ação
all_advl	todos os advérbios de posicionamento
Polo Negativo	
Etiqueta	Característica Linguística
n_nom	nominalização no singular
wcount	quantidade de palavras

Tabela 22. CAM – Variáveis do Fator 1

A análise das características do polo positivo mostra a presença de orações complementares com *to* e *that* controladas por diversas categorias de verbo, tais como verbos de modalidade, causalidade e esforço (ex: *try, permit, attempt, fail, enable*); probabilidade (ex: *indicate*); desejo, intenção e decisão (ex: *decide, want, need, choose, prepare*); e probabilidade e fato (ex: *appear, tend*), que a princípio sugerem um discurso com função de direcionar o leitor para uma ação ou informação (BIBER, 2006). A presença de verbos modais de predição (ex: *will*) e necessidade (ex: *must*) parecem corroborar com essa avaliação.

Além disso, há a presença de quatro tipos de orações relativas e voz passiva com agente, as quais refletem discurso elaborado com referências explícitas para evitar ambiguidades, assim como categorias semânticas de adjetivos (cor, tamanho, relacional etc.) e substantivos (cognitivo, quantidade, abstrato), que indicam expansão e elaboração das informações apresentadas no texto. Verbos relacionados a existência (ex: *contain, stay, remain, include*), que reportam um estado ou relações lógicas entre entidades, e relacionados a ocorrências (ex: *occur, become, develop*),

que reportam eventos ou acontecimentos físicos ocorridos sem agentes, também são indicativos de textos descritivos e informacionais (BIBER, 1988; BIBER *et al.*, 2002). A análise dos textos com escores mais positivos evidenciou que a função comunicativa mais saliente é a descrição ampla de sistemas da aeronave. Observe um trecho do texto com escore médio mais positivo que apresenta essas características (Quadro 8).

System test is accomplished **<by_pasv>** by **<prep>** turning the cockpit rotary test switch to **<prep>** Fire Warn. Proper system operation is indicated **<existv>** **<by_pasv>** by **<prep>** illumination of **<prep>** the Apu Fire indicating **<existv>** light illumination of **<prep>** the Master Warning light , and the appearance of **<prep>** the red **<colorj>** Eicas message Apu Fire . Baggage Compartment Smoke Detection The baggage compartment smoke detection system is comprised **<actv>** of **<prep>** an optical sensor type detector, mounted on **<prep>** the left side of **<prep>** the baggage compartment in **<prep>** a protective housing. The smoke detection system serves to notify **<inf>** the flight crew of **<prep>** smoke or fire in **<prep>** the tailcone baggage compartment. If smoke is detected, an electrical circuit is made **<actv>** and the sensor sends **<actv>** **<actv>** a signal to **<prep>** the Eicas system, which **<allwhrel>** displays a red **<colorj>** visual message Baggage Smoke, and sounds a double chime tone. The Master Warning light will **<prd_mod>** also **<spl_aux>** **<adv>** illuminate. The detection system is tested **<by_pasv>** by **<prep>** selecting Smoke on **<prep>** the rotary test switch in **<prep>** the cockpit. If the system tests good the red **<colorj>** Baggage Smoke Cas message will **<prd_mod>** appear **<existv>**. If the baggage smoke annunciator illuminates, the baggage compartment isolation Iso valve should **<nec_mod>** be closed and the airplane should **<nec_mod>** be landed as soon as **<prep>** possible. The valve will **<prd_mod>** not reopen until zero cabin differential pressure is achieved at **<adv>** low **<sizej>** cabin altitude. Fuel The fuel system is comprised **<actv>** of **<prep>** the storage, distribution, refueling and indicating **<existv>** **<vprogrsv>** systems. The storage system is made **<actv>** up **<adv>** of **<prep>** a set of **<prep>** integral tanks in **<prep>** each wing , and a center wing tank which **<allwhrel>** includes **<existv>** a forward **<adv>** fairing fuel tank . Each wing tank has a hopper tank which **<allwhrel>** is integral to **<prep>** it. The two wing tanks incorporate check valves and baffles allowing each wing tank complex to **<prep>** function as **<prep>** a single tank. Other integral components of **<prep>** the storage system are the gravity fuel fillers, drain valves, flapper check **<actv>** valves, vent system components, positive pressure relief valves and all **<adv>** of **<prep>** the associated system plumbing. Transfer capability is incorporated enabling **<vprogrsv>** all **<adv>** usable fuel to be **<inf>** available to **<prep>** either engine.

CE750 – TRAINING MANUAL - FUEL

Quadro 8. CAM – Trecho de texto do Fator 1 – Polo positivo

O polo negativo apresenta apenas duas características negativas, nominalização no singular e quantidade de palavras. Conforme mencionado anteriormente, essas duas características não são suficientes para dar nome a um polo da dimensão. Portanto, os textos com escore médio negativo na Dimensão 1 são marcados pela ausência das características do polo positivo. Observe a ausência dessas características no texto com escore médio mais negativo (Quadro 9).

Thrust Reverser Sleeve is not in Commanded Position
 A. Maintenance Messages
 This task is for maintenance messages:
 B. Fault Isolation Procedure
WARNING: DO THE DEACTIVATION PROCEDURE TO PREVENT THE OPERATION OF THE THRUST REVERSER. THE ACCIDENTAL OPERATION OF THE THRUST REVERSER CAN CAUSE INJURIES TO PERSONNEL AND DAMAGE TO EQUIPMENT.
 Deactivate the thrust reverser. This is the task: Thrust Reverser Deactivation For Ground Maintenance, AMM TASK
 Deactivate the leading edge slats.
 This is the task: Leading Edge Slat Deactivation, AMM TASK
 If the sleeve is in the retracted position, do these steps:
 a. Use the manual method to extend the thrust reverser.
 This is the task: Thrust Reverser Operation -Extend Manual Method, AMM TASK
 b. If the sleeve operates smoothly, do these steps:
 Replace the RVDT/Gear head Assembly.
 These are the tasks: Thrust Reverser Position Transducer RVDT Removal, Thrust Reverser Position Transducer RVDT Installation, AMM TASK
 Do the repair confirmation at the end of this task.
 c. If the sleeve does not operate smoothly, do these steps:
 Replace the thrust reverser.
 These are the tasks: Thrust Reverser Removal, Thrust Reverser Installation, AMM TASK
 Do the repair confirmation at the end of this task.
 If the sleeve is in the extended position, do these steps:
 a. Make sure there is no contamination in the main and auxiliary tracks.
 b. Use the manual method to retract the thrust reverser.
 This is the task: Thrust Reverser Operation -Retract Manual Method, AMM TASK
 c. If the sleeve operates smoothly, do these steps: Replace the RVDT/Gear head Assembly.

B777 – FIM – 78H

Quadro 9. CAM – Trecho de texto do Fator 1 – Polo negativo

A partir dessa análise, pode-se afirmar que a primeira dimensão é caracterizada por textos mais ou menos descritivos e informacionais de alta densidade lexical. Ao

considerar como os tipos de manuais se comportam nessa dimensão (Gráfico 8), nota-se que a primeira dimensão reflete a divisão entre o treinamento para operar uma aeronave e a operação/reparação dos sistemas em si. Nesse caso, pode-se dizer que a dimensão 1 reflete o mundo de trabalho do técnico de manutenção aeronáutica, o qual deve conhecer e estudar os manuais de treinamento dos sistemas da aeronave antes de operar ou fazer a manutenção desses sistemas. A linguagem nessa dimensão está atrelada à situação de uso “escola/treinamento” (polo positivo) com textos descritivos, contendo informações sobre os sistemas da aeronave, e caminha em direção à situação de uso “trabalho/operacional” (polo negativo) com textos marcados pela ausência de descrição e presença de procedimentos. Observe como isso se reflete no posicionamento dos tipos de manuais nessa dimensão (Gráfico 8).

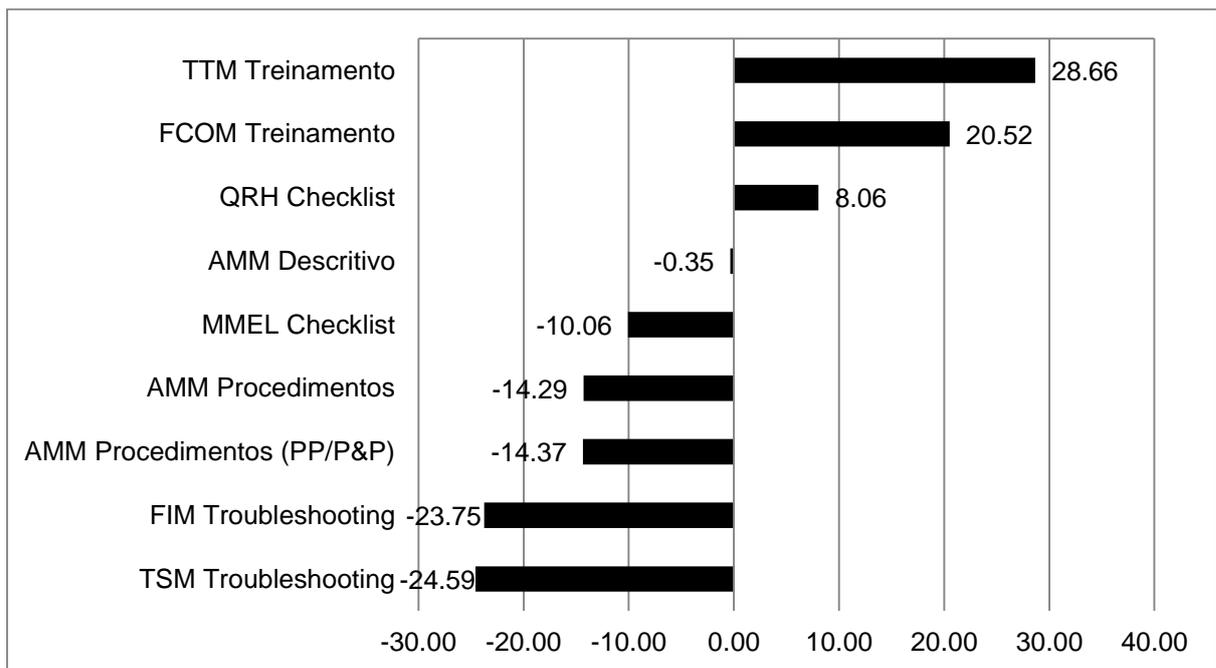


Gráfico 8. CAM – Distribuição dos registros no Fator 1

3.2.2 Dimensão 2

A segunda dimensão, nomeada “discurso de tomada de decisões *versus* discurso de procedimentos”, é notadamente marcada por discurso procedimental, envolvendo tomada de decisão antes da ação e realização de procedimentos em geral. A dimensão conta com número expressivo de características nos polos positivo e negativo. Sendo assim, a nomeação foi feita considerando a polaridade entre textos positivos e negativos. Observe as características linguísticas presentes nessa dimensão (Tabela 23).

Polo Positivo	
Etiqueta	Característica Linguística
be_state	verbo <i>to be</i> indicativo de estado
pos_mod	verbo modal de possibilidade
p_and	conjunção coordenada – conectivo frasal
o_and	conjunção coordenada – conectivo clausal
allmodal	todas as categorias de verbos modais
pasttense	passado
pred_adj	adjetivo predicativo
groupn	substantivo relacionado a grupo ou instituição
allverb	todas as categorias de verbos, excluindo verbos auxiliares
sua_vb	verbo de persuasão
agls_psv	voz passiva sem agente
allpasv	todos os usos de voz passiva
pub_vb	verbo <i>dicendi</i>
conjuncts	conjunção
causev	verbo causativo
downtone	advérbio suavizador
pres	verbo no presente
alladj	todas as categorias de adjetivos
that_del	omissão de <i>that</i> em oração subordinada
whiz_vbn	modificador pós-nominal da voz passiva
wrlengh	tamanho de palavra
mentaltv	verbo relacionado a atividade mental
Polo Negativo	
Etiqueta	Característica Linguística
n	substantivo
allpro	todas as categorias de pronomes
all_jth	todas as categorias de orações complementares com <i>that</i> controladas por adjetivos
jcmp	<i>that</i> usado em oração complementar controlada por adjetivo
pro2	pronome em 2ª pessoa

fact_jth	<i>that</i> usado em oração complementar controlada por adjetivo factivo
all_th	soma das orações complementares com <i>that</i>

Tabela 23. CAM – Variáveis do Fator 2

O polo positivo apresenta textos com características impessoais e discurso baseado em tomada de decisões. As principais características do polo positivo são classes especiais de verbos, verbos no passado, verbos modais de possibilidade, voz passiva sem agente e adjetivos predicativos (adjetivos que não fazem parte do sintagma nominal). Verbos modais de possibilidade (ex: *may, can*) evidenciam possibilidade lógica e voz passiva sem agente está relacionada à apresentação mais abstrata e estática de informações (BIBER, 1988). Os verbos pertencentes a classes especiais são os verbos *dicendi* (ex: *maintain, state, verify, insure, submit*), de persuasão (ex: *require, specify, recommend, provide*) e causativos (ex: *let, make, get*), os quais estão atrelados a ações e/ou processos ou indicam intenções futuras. A análise de linhas de concordâncias mostrou que, em geral, esses verbos estão no passado e/ou fazem parte da construção da voz passiva sem agente. Levando em consideração essas características aliada à análise dos exemplos de texto mais representativos dessa dimensão, foi possível constatar que os textos possuem características que marcam a necessidade de tomada de decisão para realização dos procedimentos. Observe um trecho do texto com escore médio mais positivo que apresenta essas características (Quadro 10).

Gaspar Valves Repair Category. Number Installed <allverb> <allpasv> <whiz_vbn>. Number Required <pasttense> <allverb> For Dispatch. Remarks <allverb> <pub_vb> And Exceptions Mfd Indications not addressed <pasttense> <allverb> elsewhere in the Mmel may <pos_mod> <allmodal> be <allverb> <pres> inoperative <alladj>.

O M May <pos_mod> <allmodal> be <allverb> <pres> inoperative <alladj> provided <conjuncts>:

a. Mfd Ecs Synoptic <alladj> is <allverb> operative <alladj>, b. Ecs Switch command to Ram <allverb> Air Valve is tested <allverb> <agls_psv> <allpasv> before each flight <groupn>, c. Ecs Switch is set <alladj> to the opposite <alladj> side for flight <groupn>, d. Affected <allverb> <allpasv> <whiz_vbn> Fcsov is confirmed <allverb> <agls_psv> <allpasv> <mentalv> Closed <pasttense> <allverb> and Deactivated, and <o_and> e. Airplane is operated <allverb> <agls_psv> <allpasv> at or below Fl .

C O May <pos_mod> <allmodal> be <allverb> <pres> inoperative <alladj> provided <conjuncts>:

a. Mfd Ecs Synoptic <alladj> is <allverb> operative <alladj>, b. Ecs Switch command to Ram <allverb> Air Valve is tested <allverb> <agls_psv> <allpasv> before each flight <groupn>, c. Ecs Switch is set <alladj> to the opposite <alladj> side for flight <groupn>, d. Prsov of affected <alladj> side is kept <allverb> <agls_psv> <allpasv> Closed <pasttense> <allverb>, e. Cross Bleed <allverb> <pres> Valve is kept <allverb> <agls_psv> <allpasv> Closed <pasttense> <allverb>, f. Airplane is operated <allverb> <agls_psv> <allpasv> at or below Fl , and <o_and> g. Operations are <pres> not conducted <pasttense> <allverb> in known <alladj> or <p_and> forecast <alladj> icing <alladj> conditions.

PHENOM – MMEL-FAA

Quadro 10. CAM – Trecho de texto do Fator 2 – Polo positivo

Ao avaliar os exemplos dos textos com escore médio mais positivos, pôde-se observar que, além das possíveis dificuldades linguísticas enfrentadas pelos técnicos de manutenção aeronáutica ao manusear o tipo de manual MMEL *Checklist* (tipo de manual com o escore médio mais alto nessa dimensão), a organização visual do texto é complexa, baseada em forma de *outline* com o uso de letras maiúsculas e minúsculas para definir itens e subitens, e, conseqüentemente, exigindo um nível mais alto de letramento e treinamento para compreender essas informações.

O polo negativo é marcado pela presença maciça de substantivos, em geral como parte de longas sequências nominais. Há também a presença de pronomes, em que se destaca o pronome em segunda pessoa *you*. O uso do pronome em segunda pessoa exige um destinatário específico, no caso o técnico de manutenção da

aeronave, e indica um alto grau de envolvimento com esse destinatário. A análise mostrou que a função desse pronome nos manuais com escore mais negativos está relacionada a alertar o técnico para riscos de danos à aeronave ou lesões à sua pessoa. A análise das linhas de concordância das orações complementares com *that* controladas por adjetivos presentes no polo negativo da dimensão 2 mostrou que todas elas representam o mesmo padrão: *Make sure that*, evidenciando algo que deva ser realizado ou verificado ou de forma a garantir que algo tenha sido feito. Observe no quadro a seguir (Quadro 11) a presença dessas características no trecho do texto com escore médio mais negativo nessa dimensão.

Do only the procedure <n> related to the installation <n>. This maintenance <n> technician <n> is necessary to do this task <n>: - powerplant <n> and Apu <n>
 B. References <n> Reference <n> Designation <n> Amm <n> Task <n> - a Disconnection <n> of the external Ac power <n> supply <n> from the aircraft <n> - aircraft <n> Configuration <n> Amm <n> Task <n> - a Fan <n> Cowl <n> - close Amm <n> Task <n> - a Oil <n> Tank <n> - removal <n> Emm <n> Task <n> Engine <n> On Wing <n> Emm <n> Task <n> Oil <n> Tank <n> Replacement <n> C. Zones <n> and Accesses <n> Zones <n> Location <n> Reference <n> Access <n> D. Circuit <n> Breaker <n> List <n> Type <n> Designation <n> Location <n> Bus <n> Location <n> Tip <n> Mcd� <n> Page <n> Sspc <n> E Start Vlv <n> Spda <n> Dc Ess <n> Bus <n> Spda <n> - m
 E. Job <n> Set - up Subtask <n> - a Make sure <fact_jth> that <jcmp> the aircraft <n> is in the same configuration <n> as it <allpro> was at the end <n> of the removal <n> task <n> Amm <n> Task <n> - a. F. Oil <n> Tank <n> - installation <n> Subtask <n> - a Install <n> the lube <n> oil <n> tank <n>. Refer to the latest revision <n> of the Geae <n> Cf E Engine <n> Manual Emm <n> Task <n>. G. Job <n> Close - up Subtask <n> - a
 Caution <n>: Examine All The Work <n> Areas <n> To Make Sure That <jcmp> You <allpro> <pro2> Removed Tools <n> And Equipment <n> After You <allpro> <pro2> Completed The Work <n>.
 If You <allpro> <pro2> Do Not Obey This Procedure <n>, Damage <n> To The Aircraft <n> Can Occur. Remove <n> all tools <n>, equipment <n> and unwanted material <n> from the work <n> area <n>.
 Subtask <n> - a Close <n> the fan <n> cowl <n> Amm <n> Task <n> - a. Subtask <n> - a On the Mcd� <n>, use <n> the sequence <n> in parentheses <n> to make sure <fact_jth> that <jcmp> the Sspcs <n> are closed: E Start Vlv <n> - spda <n> Cbmenu <n> By Sys <n> E Start Vlv <n> - spda <n> Cbmenu <n> By Sys <n> Subtask <n> - a Do the necessary tests <n> after the replacement <n> or repair <n> of the lube <n> oil <n> tank <n> Emm <n> Task <n> Subtask <n> - a Disconnect <n> the external Ac power <n> supply <n> to the aircraft <n>
 E190 – AMM PP – 79

Quadro 11. CAM – Trecho de texto do Fator 2 – Polo negativo

Ao analisar como os tipos de manuais se encaixam nessa dimensão (Gráfico 9), observa-se que o tipo de manual MMEL (*Checklist*) define o polo positivo da dimensão. Conforme visto no exemplo acima, a linguagem desse tipo de texto é caracterizada pela necessidade de tomada de decisão por parte do técnico de manutenção (ex: *May be inoperative provided...*), que deve avaliar se as condições foram atendidas para que a aeronave possa sair do solo ou não. Por outro lado, todos os tipos de manuais de procedimentos e *troubleshooting* encontram-se no polo negativo, o que corrobora o nome dado ao polo negativo, "discurso de procedimentos". Observe o posicionamento dos tipos de manuais na dimensão 2 (Gráfico 9).

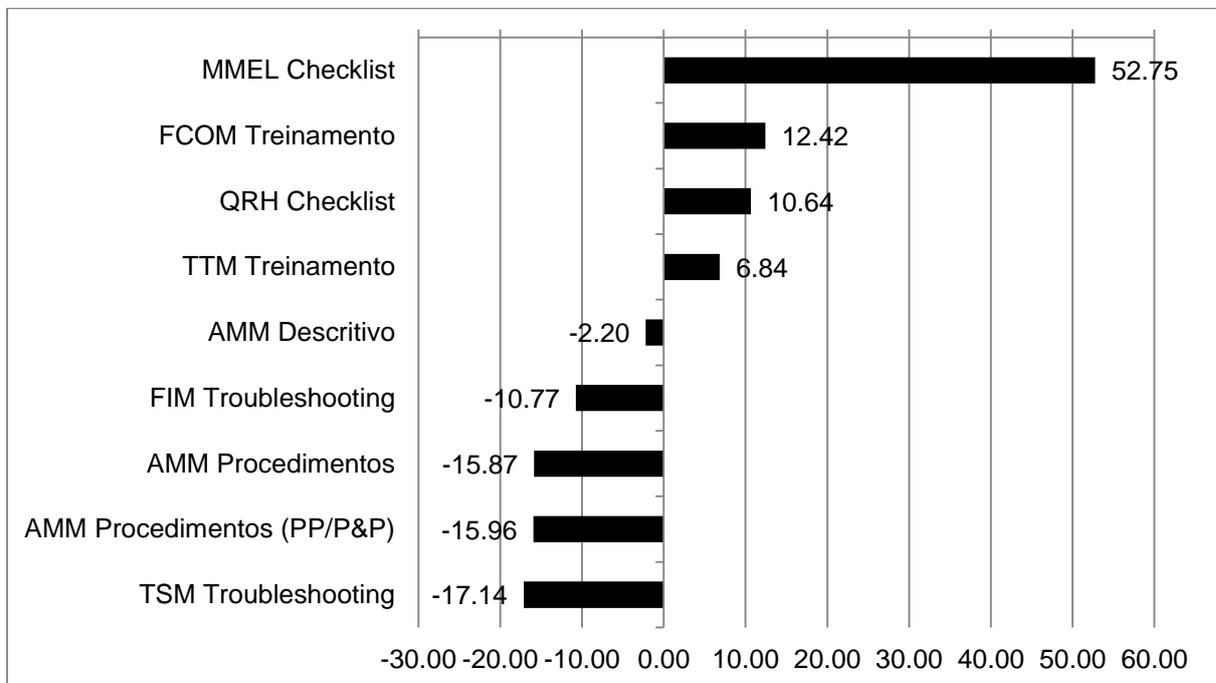


Gráfico 9. CAM – Distribuição dos registros no Fator 2

3.2.3 Dimensão 3

A dimensão 3, "discurso de solução de problemas *versus* descrição de partes específicas", parece ser marcada, por um lado, por discurso relacionado a identificação e resolução de problemas e, por outro, por descrições detalhadas e

específicas de partes da aeronave ou partes de sistemas da aeronave. Durante a análise fatorial constatou-se que os textos do polo negativo continham os escores médios absolutos mais altos na dimensão. Assim, optou-se inverter a posição dos polos para fins de nomear da dimensão. Em geral, inicia-se o nome pelo polo positivo, porém, na dimensão 3, a primeira parte do nome (discurso de solução de problemas) refere-se ao polo negativo; e a segunda parte (descrição de partes específicas), ao polo positivo. É importante ressaltar que como as dimensões não representam dicotomias e sim um continuum, não é necessário que os nomes estejam em relação de oposição. A rotulação das dimensões refletem as funções comunicativas exercidas pelas características linguísticas identificadas nos dois polos, ou seja, textos marcados para discurso de solução de problemas terão menos ou nenhuma descrição de partes específicas e textos marcados para descrição de partes específicas terão menos ou nenhum discurso de solução de problemas. Observe as características linguísticas presentes nessa dimensão (Tabela 24).

Polo Positivo	
Etiqueta	Característica Linguística
att_jth	<i>that</i> usado em oração complementar controlada por adjetivo atitudinal
x5_jto	<i>to</i> usado em oração complementar controlada por adjetivos de avaliatividade
pro3	pronome em 3ª pessoa
all_jto	todas as categorias de orações complementares com <i>to</i> controladas por adjetivos
adj_attr	adjetivo atributivo
that_rel	<i>that</i> em orações subordinadas com pronome relativo
concrtn	substantivo concreto
Polo Negativo	
Etiqueta	Característica Linguística
sub_cnd	conjunção subordinativa condicional
aspectv	verbo acurativo/determinativo
tm_adv	advérbio de tempo
allconj	todas as categorias de conjunções
pro_do	verbo <i>do</i> como substituto de outro verbo ou sintagma verbal / verbo vicário / pró-verbo
commv	verbo relacionado a comunicação
tccnrt	substantivo relacionado a assuntos técnicos ou concretos
prcssn	substantivo relacionado a processos

Tabela 24. CAM – Variáveis do Fator 3

Observa-se no polo negativo a presença marcante de conjunção subordinativa condicional (ex: *if*), advérbio de tempo (ex: *then*), substantivos relacionados a assuntos técnicos ou concretos e a processos (ex: *flow, task*) e verbos acurativos, que estão associados a etapas de um processo (ex: *start, stop, complete, continue*). Essas características, quando analisadas no contexto dos manuais, evidenciam o um discurso de problema e solução, em que problemas são identificados por meio da realização de processos e solucionados de acordo com os procedimentos descritos. Durante a análise, notou-se que as orações condicionais indicam o problema seguido do procedimento para solucioná-lo (ex: *If the water flow is satisfactory, then do the Fault Isolation Procedure below*). Observe um trecho do texto com escore médio mais negativo que apresenta essas características (Quadro 12).

Of Task Faucet Water Flow not Sufficient - fault Isolation
 A. Initial Evaluation Open the faucet for seconds. a. If **<sub_cnd> <allconj>** the flow **<prcessn>** is not satisfactory, then **<tm_adv>** do these steps: Note: Make sure that all applicable water shutoff valves and isolation valves are open before **<allconj>** you start **<aspectv>**. Go to one of the other lavatories and then **<tm_adv>** open the faucet for seconds. If **<sub_cnd> <allconj>** the water flow **<prcessn>** is satisfactory, then **<tm_adv>** do the Fault Isolation Procedure below. If **<sub_cnd> <allconj>** the water flow **<prcessn>** is not satisfactory, then **<tm_adv>** do these steps: a. Do this task **<prcessn>**: Potable Water System Flow not Sufficient - fault Isolation, Task b. If **<sub_cnd> <allconj>** the flow **<prcessn>** is satisfactory, then **<tm_adv>** there was an intermittent fault.
 B. Fault Isolation Procedure Do a check of the level of the water in the water tanks.
 a. If **<sub_cnd> <allconj>** the level of the water is low, then **<tm_adv>** do these steps: Tasks Fault Isolation Manual Do this task **<prcessn>**: Potable Water Tank - fill, Amm Task Open the faucet for seconds. If **<sub_cnd> <allconj>** the water flow **<prcessn>** is satisfactory, you corrected the fault.
 b. If **<sub_cnd> <allconj>** the level of the water is not low, then **<tm_adv>** continue **<aspectv>**. Do a check of the shutoff valve for the water supply.
 a. If **<sub_cnd> <allconj>** the shutoff valve is closed, then **<tm_adv>** open the valve. Open the faucet for seconds. If **<sub_cnd> <allconj>** the water flow **<prcessn>** is satisfactory, you corrected the fault.
 b. If **<sub_cnd> <allconj>** the shutoff valve is open, then **<tm_adv>** continue **<aspectv>**. Do a check of the faucet for the correct operation **<prcessn>**.
 a. If **<sub_cnd> <allconj>** the faucet does not operate correctly, replace **<commv>** the water faucet. These are the tasks **<prcessn>**: Faucet Assembly Removal Adams Rite, Amm Task Faucet Assembly Installation Adams Rite, Amm Task b. Open the faucet for seconds. c. If **<sub_cnd> <allconj>** the water flow **<prcessn>** is satisfactory, you corrected the fault.

B777 – FIM – 38

Os textos do polo positivo nessa dimensão possuem características que refletem descrições detalhadas de partes específicas da aeronave. As orações subordinadas com pronome relativo parecem atuar na especificação dos substantivos concretos, corroborando essa interpretação. Adjetivos atributivos são indicativos de um alto grau de integração das informações. Além disso, a análise das linhas de concordância dos exemplos de pronomes em terceira pessoa (ex: *it, they, them*) revelou que, em nenhuma instância, esses pronomes se referem a pessoas nos manuais aeronáuticos. Ao contrário, nos manuais, os pronomes em terceira pessoa se referem a objetos em geral, sistemas e/ou partes específicas. Observe no quadro a seguir (Quadro 13) como essas características se apresentam no trecho do texto com escore médio mais positivo nessa dimensão.

Electrical **<adj_attr>** Harnesses The engine **<concrtn>** electrical **<adj_attr>** harnesses connect electrical **<adj_attr>** components, located inside the nacelle, to the pylon. They **<pro3>** distribute the power required by the aircraft electrical **<adj_attr>** system and transmit signals for nacelle subsystems and for engine **<concrtn>** control and monitoring **<adj_attr>** functions. Air Intakes The powerplant **<adj_attr>** air inlet cowl is a fixed interchangeable **<adj_attr>** aerodynamic **<adj_attr>** fairing. It **<pro3>** supplies the inlet airflow to the fan and core sections of the engine **<concrtn>**. Engine Drains The engine **<concrtn>** drain **<concrtn>** system consists of tubes **<concrtn>** and openings that **<that_rel>** convey overboard any waste fluids from the engine **<concrtn>**, accessories, nacelle cavities, and pylon **<adj_attr>** cavities that **<that_rel>** can occur during any abnormal **<adj_attr>** operating **<adj_attr>** condition or failure condition from engine **<concrtn>** and pylon components . The main powerplant components are listed below: Geae Cf E Nacelle Thrust reverser: Aircraft Maintenance Manual **<adj_attr>** Mounts and Ebu Engine Buildup Unit Operation All the interfaces between the cockpit and the engine **<concrtn>** are electrical. The engine **<concrtn>** cockpit primary **<adj_attr>** indication is provided by the Eicas. The Mfd Multi - function Display may be used as a backup for engine **<concrtn>** indication. The powerplant **<adj_attr>** system also gives: Electric **<adj_attr>** power generation. Hydraulic **<adj_attr>** power generation. Ecs Environmental **<adj_attr>** Control System bleed air. Anti - ice **<concrtn>** bleed air. Pressurization bleed air. The Air turbine starter operates with air from the Apu Auxiliary **<adj_attr>** Power Unit, pneumatic **<adj_attr>** external **<adj_attr>** power, or crossbleed from the other operating **<adj_attr>** engine **<concrtn>**. Training Information Points If you install a new **<adj_attr>** fan cowl door **<concrtn>**, you must have to adjust the latches Amm Task - a. Use a dolly **<adj_attr>** Dolly - inlet Transport Gse to hold and move the air inlet module in the hangar after you remove it **<pro3>**. The figure **<concrtn>** Powerplant - powerplant Component Location And Operating Interfaces provides further data on the preceding **<adj_attr>** text.

E190 – AMM SDS – 71H

Considerando a disposição dos tipos de manuais na dimensão 3, é possível afirmar que as características e funções identificadas nos exemplos de textos estão refletidas no posicionamento dos manuais nessa dimensão (Gráfico 10). Ressaltando que o gráfico a seguir está com a polaridade invertida, ou seja, o lado direito representa o polo negativo e o lado esquerdo, o positivo. Observa-se que, no polo negativo, os textos mais marcados são os do tipo *troubleshooting*. Isso corrobora a análise das funções comunicativas, pois esse tipo de manual é utilizado na identificação e subsequente solução de problemas. Já o polo positivo contém os manuais de práticas e procedimentos, o que faz sentido uma vez que a descrição de peças e/ou objetos a serem reparados/trocados deve ser a mais específica possível para que não haja enganos. Os manuais de treinamento não são marcados para as características dessa dimensão. Observe o posicionamento dos tipos de manuais na dimensão 3 (Gráfico 10).

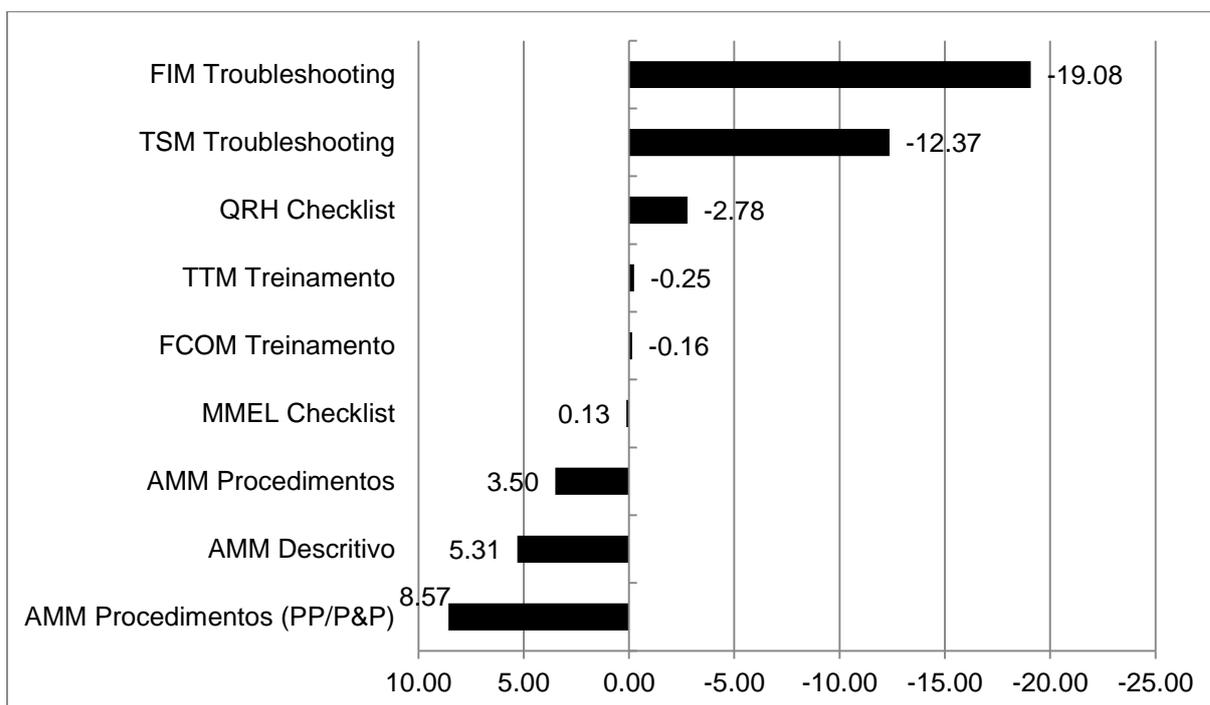


Gráfico 10. CAM – Distribuição dos registros no Fator 3

3.3 Variáveis independentes no *Corpus* CAM (ANOVAS)

Conforme descrito nos capítulos Fundamentação teórica e Metodologia, análises de variância (ANOVA) foram realizadas a fim de verificar se há diferenças estatisticamente significativas entre fabricantes e modelos de aeronaves e os tipos de manuais do CAM em relação às dimensões de variação próprias do *corpus* CAM. Além disso, a análise por meio de ANOVA objetiva observar se a variação encontrada nos textos do CAM pode ser resultante de variáveis independentes ao texto.

Nesse sentido, foram realizadas análises de variância (ANOVA) para três fatores fixos (variáveis independentes): (1) fabricantes de aeronaves; (2) modelos de aeronaves; e (3) Tipos de manuais.

Os resultados dessas análises são apresentados nas tabelas a seguir. No entanto, antes de apresentar os resultados, é importante retomar o que significam as estatísticas visíveis nas tabelas, as quais foram apresentadas no capítulo Metodologia. F equivale à razão F , p ao nível de significância, df aos graus de liberdade e R^2 ajustado à proporção de variação que pode ser decorrente do fator fixo. Todas as análises apresentaram resultados estatisticamente significantes ($p=0,000$). Os valores de R^2 ajustado são discutidos em termos de porcentagem (valor de R^2 ajustado multiplicado por 100).

A tabela 25 apresenta os resultados da ANOVA para os fabricantes de aeronaves (Airbus, ATR, Boeing, Cessna e Embraer). Considerando os valores do R^2 ajustado, pode-se afirmar que fabricantes de aeronaves respondem por 46,4% da variação encontrada na dimensão 1. Isso quer dizer que, se o fabricante de uma aeronave é conhecido, pode-se prever quase 47% do conteúdo referente a ampla descrição de sistemas dessa aeronave. Os resultados nas dimensões 2 e 3 não são tão expressivos: 16,8% e 22,1%, respectivamente. Observe (Tabela 25).

Dimensão		F	p	df	R ² Ajustado
1	Ampla descrição de sistemas	34,119	0,000	4	0,464
2	Discurso de tomada de decisões <i>versus</i> Discurso de procedimentos	8,719	0,000	4	0,168
3	Discurso de solução de problemas <i>versus</i> Descrição de partes específicas	11,835	0,000	4	0,221

Tabela 25. ANOVA – Fabricantes de Aeronaves

Na análise de variância (ANOVA) para modelos de aeronaves (A320, A330, ATR42, ATR45, ATR72600, ATR75, B737NG, B767, B777, CE650, CE680, CE750, E190, MD11, PHENOM), foi possível observar um leve aumento na proporção de variação explicada em todas as dimensões. Valores de R² ajustado para as dimensões 1, 2 e 3 foram de 54,4%, 36,5% e 25,7%, respectivamente. Isso significa que modelos de aeronaves são responsáveis por mais da metade da variação encontrada na dimensão 1, aproximadamente um terço da variação encontrada na dimensão 2 e ¼ da variação encontrada na dimensão 3. Observe (Tabela 26).

Dimensão		F	p	df	R ² Ajustado
1	Ampla descrição de sistemas	13,171	0,000	15	0,544
2	Discurso de tomada de decisões <i>versus</i> Discurso de procedimentos	6,859	0,000	15	0,365
3	Discurso de solução de problemas <i>versus</i> Descrição de partes específicas	4,525	0,000	15	0,257

Tabela 26. ANOVA – Modelos de Aeronaves

A ANOVA realizada com os tipos de manuais como fator fixo foi a que apresentou os melhores resultados em todas as dimensões. De acordo com os valores de R² ajustado, os tipos de manuais são responsáveis por 62,9%, 90,2% e 76% da variação encontrada nas dimensões 1, 2 e 3, respectivamente. Esses resultados validam esta pesquisa, pois mostram que se o tipo de manual é conhecido, pode-se prever até 90% do seu conteúdo linguístico. Observe (Tabela 27).

	Dimensão	F	p	df	R² Ajustado
1	Ampla descrição de sistemas	33,377	0,000	8	0,629
2	Discurso de tomada de decisões <i>versus</i> Discurso de procedimentos	176,973	0,000	8	0,902
3	Discurso de solução de problemas <i>versus</i> Descrição de partes específicas	61,579	0,000	8	0,760

Tabela 27. ANOVA – Tipos de Manuais

As análises acima encerram este capítulo, que apresentou o mapeamento do *corpus* CAM nas dimensões de variação da língua inglesa sugeridas por Biber (1988), a identificação das dimensões de variação próprias dos manuais aeronáuticos, e a verificação de diferenças estatísticas significantes entre fabricantes, modelos de aeronaves e tipos de manuais em relação às dimensões identificadas. As considerações finais a seguir encerram a dissertação.

Considerações Finais

As considerações finais são apresentadas neste capítulo, iniciando com a retomada dos pontos principais da pesquisa, apontando limitações e fazendo sugestões de pesquisas futuras e possíveis aplicações pedagógicas dos resultados obtidos.

O inglês na aviação permeia diversas esferas, sendo a linguagem escrita dos manuais aeronáuticos apenas uma delas. Apesar disso, essa linguagem é primordial para o mundo da aviação, pois é por meio dos manuais que pilotos e técnicos de manutenção aeronáutica são treinados e realizam suas tarefas profissionais. Conforme mencionado na Introdução, mecânicos de aeronaves passam até 40% do tempo de serviço lidando com manuais em inglês (FAA, 2005); e sabe-se que há carência de conhecimento em língua inglesa por parte desses profissionais.

O ensino de inglês para técnicos de manutenção aeronáutica no Brasil é limitado a um número reduzido de institutos de idiomas e professores particulares especializados, os quais utilizam materiais próprios desenvolvidos, em sua maioria, sem embasamento teórico. Além disso, há poucos estudos sobre as características linguísticas dos textos utilizados por esses profissionais na sua rotina de trabalho.

Esta pesquisa buscou preencher essa lacuna ao investigar as características linguísticas mais salientes dos manuais de manutenção aeronáutica, de forma quantitativa e qualitativa, por meio dos princípios teórico-metodológicos da Linguística de *Corpus* e, mais especificamente, da Análise Multidimensional. A pesquisa em Linguística de *Corpus* facilitou o uso e manuseio de um *corpus* expressivo, algo que não seria possível em outra área. A Análise Multidimensional permitiu caracterizar linguística e funcionalmente os manuais aeronáuticos, além de proporcionar a validação estatística dos dados. Além disso, a metodologia une a análise quantitativa à qualitativa, ou seja, macro e micro análises foram realizadas, permitindo maior segurança na interpretação dos dados.

Uma das limitações da pesquisa em AMD para muitos pesquisadores é a exigência de conhecimento relativamente grande de estatística e informática. Além disso, para a realização das análises quantitativas, é necessário o uso de *software*

especializado para a limpeza, anotação e etiquetagem e análise estatística, os quais não estão disponíveis em larga escala ou são muito caros. Muitas vezes, como no caso desta pesquisa, pequenos programas (*scripts*) personalizados precisam ser desenvolvidos, o que torna essa área proibitiva para muitos. O suporte para esta pesquisa veio do Grupo de Estudos em Linguística de *Corpus* (GELC), o qual possui as ferramentas, infraestrutura e conhecimento específico para dar apoio a pesquisas nessa área.

Os resultados e análises de resultados apresentados no capítulo Apresentação e Discussão dos Resultados procuraram responder as três perguntas de pesquisa. A primeira consistiu em mapear o *corpus* CAM nas dimensões de variação da língua inglesa estabelecidas por Biber (1988). Os resultados mostraram que o *corpus* CAM tende a ser essencialmente informacional e não narrativo, conter referências explícitas e informações abstratas e não ser marcado para persuasão/argumentação.

A segunda pergunta de pesquisa procurou identificar as características linguísticas próprias dos manuais aeronáuticos, independente das dimensões da língua inglesa de 1988. Para isso, foi feita análise fatorial completa do *corpus* CAM, seguida da interpretação dos fatores em dimensões de variação, de acordo com os passos descritos na Metodologia. Foram extraídos três fatores, conduzindo à nomeação de três dimensões de variação: (1) Ampla descrição de sistemas (*Broad system descriptions*); (2) Discurso de tomada de decisões *versus* Discurso de procedimentos (*Decision-making discourse versus Procedural discourse*); e (3) Discurso de solução de problemas *versus* Descrição de partes específicas (*Problem-solving discourse versus Specific parts descriptions*). Assim como em pesquisas anteriores em AMD, a primeira dimensão concentra maior variação.

Observou-se que os textos dos manuais aeronáuticos podem ser classificados, em termos retóricos, em: (1) descrição, que responde a pergunta “o que é?”, com textos representados nas dimensões 1 e 3; (2) procedimentos, que responde a pergunta “como agir?”, com textos na dimensão 2; e (3) problema e solução, que identifica e diz como agir, com textos na dimensão 3.

Após a identificação das dimensões, análises de variância (ANOVA) foram realizadas para responder a terceira pergunta de pesquisa: verificar a proporção das diferenças entre as dimensões e as variáveis independentes (fatores fixos). Observou-

se que fabricantes de aeronaves são responsáveis por 46,4%, 16,8% e 22,1% da variação encontrada no *corpus* CAM nas dimensões 1, 2 e 3, respectivamente; modelos de aeronaves, por 54,4%, 36,5% e 25,7%, respectivamente; e tipos de manuais, por 62,9%, 90,2% e 76,0%, respectivamente. A análise de variância mostrou que o conhecimento do tipo de manual, modelo e fabricante de aeronave pode ajudar a prever até 90% do conteúdo linguístico dos manuais com relação às suas dimensões de variação.

As análises de variância indicam que é possível prever em proporções diferentes as características que estarão presentes em um determinado manual aeronáutico se o fabricante, modelo ou tipo de manual for conhecido. Esse achado é importante especialmente para o ensino de inglês para aviação, pois permite que, a partir dos dados resultantes desta pesquisa, materiais didáticos possam vir a ser elaborados, por professores ou pesquisadores, com mais facilidade, aplicabilidade e fidelidade à caracterização dos manuais usados pelos técnicos de manutenção aeronáutica.

Por fim, espera-se que esta pesquisa contribua para a área de inglês para aviação, em especial, para os professores que queiram substanciar o desenvolvimento de materiais próprios e instigue outros pesquisadores a desenvolver pesquisas em dimensões de variação, seja em língua geral ou específica.

BIBLIOGRAFIA

AGÊNCIA NACIONAL DE AVIAÇÃO CIVIL. Regulamento Brasileiro de Homologação Aeronáutica nº 65. 2001. Disponível em: <<http://www.anac.gov.br/biblioteca/rbha/rbha065.pdf>>. Acesso em: 25 jun. 2012.

ALMEIDA, Dilso Corrêa.; PRADO, Malila Carvalho de Almeida. Desenvolvendo o conteúdo programático de um curso de inglês para mecânicos de aeronaves com base em um corpus DIY: um estudo de caso. **Aviation in Focus**, v.2, n.2, p. 6-20, 2011.

ATKINS, Sue; CLEAR, Jeremy; OSTLER, Nicholas. Corpus Design Criteria. **Literary and Linguistic Computing**, v.7, n.1, p. 1-16, 1992.

BERBER SARDINHA, Tony. Análise Multidimensional. **D.E.L.T.A.**, São Paulo, v.16, n.1, 2000.

_____. Linguística de Corpus: histórico e problemática. **D.E.L.T.A.**, São Paulo, v.16, n.2, 2000.

_____. Informatividade, interatividade e narratividade na reunião de negócios – Análise Multidimensional e palavras-chave. **Direct Papers** 52, 2004a. Disponível em: <http://www2.lael.pucsp.br/direct/direct_papers.htm>. Acesso em 02 out. 2011.

_____. **Linguística de Corpus**. Barueri, SP: Manole, 2004b.

_____. Variação entre registros da Internet. In: SHEPHERD, Tania G.; SALIÉS, Tânia G. (Org.) **Linguística da Internet**. São Paulo: Editora Contexto, 2013.

BERNSTEIN, Basil. **Class, codes, and control**. Volume 1: Theoretical studies towards a sociology of language. London: Routledge & Kegan Paul, 1970.

_____. A Sociolinguistic approach to socialization; with some reference to Educability. In: GUMPERZ, John J.; HYMES, Dell. (Eds.). **Directions in sociolinguistics**. New York: Holt, Rinehart and Winston, 1972.

BÉRTOLI-DUTRA, Patrícia. **Linguagem da música popular anglo-americana de 1940-2009**. Tese de Doutorado, LAEL, PUCSP, São Paulo, SP, 2010.

BIBER, Douglas. **Variation Across Speech and Writing**. Cambridge: Cambridge University Press, 1988.

_____. **University Language: A Corpus-Based Study of Spoken and Written Registers**. Amsterdam/Philadelphia: John Benjamins, 2006.

_____. Multi-dimensional approaches. In: LÜDELING, Anke; KYTÖ, Merja (Orgs.). **Corpus Linguistics: An International Handbook**. v.2. Berlin: Walter de Gruyter, 2009, p. 822-855.

_____. Representativeness in corpus design. **Literary and Linguistic Computing**. v.8, n.4, p. 243-257, 1993a.

_____. The multi-dimensional approach to linguistic analyses of genre variation: an overview of methodology and findings. **Computers and the Humanities**. v.26, p. 331-345, 1993b.

BIBER, Douglas; CONRAD, Susan. **Register, genre, and style**. Cambridge: Cambridge University Press, 2009.

BIBER, Douglas; JONES, James K. Quantitative Methods in Corpus Linguistics. In: LÜDELING, Anke; KYTÖ, Merja (Orgs.). **Corpus Linguistics: An International Handbook**, v. 2. Berlin: Walter de Gruyter, 2009, p. 1286-1304.

BIBER, Douglas; HARED, Mohamed. Linguistic correlates of the transition to literacy in Somali: Language adaptation in six press registers. In: BIBER, Douglas; FINEGAN, Edward (Eds.), **Sociolinguistic Perspectives on Register**. Oxford: Oxford University Press, 1994. p. 182-216.

BIBER, Douglas et. al. **Corpus Linguistics: Investigating Language Structure and Use**. 2. ed. Cambridge: Cambridge University Press, 2000.

_____. Spoken and written register variation in Spanish: A multi-dimensional analysis. **Corpora**, 1(1), 2006. p. 1-37.

BIBER, Douglas; JOHANSSON, Stig; LEECH, Geoffrey; CONRAD, Susan; FINEGAN, Edward. **Longman Grammar of Spoken and Written English**. Harlow: Longman, 1999.

BIBER, Douglas; CONRAD, Susan; LEECH, Geoffrey. **Student Grammar of Spoken and Written English**. Harlow: Longman, 2002.

BOCORNY, Ana. **Descrição das unidades especializadas poliléxicas nominais no âmbito da aviação: subsídios para o ensino de inglês para fins específicos (ESP)**. 2008. 230 f. Tese de Doutorado em Estudos da Linguagem – Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2008.

BOCORNY, Ana. Panorama dos estudos sobre a linguagem da aviação. **RBLA**, Belo Horizonte, v.11, n.4, p.963-986, 2011.

CONDE, Helena Maria de Andrade. **Escolhas léxico-gramaticais em composições de alunos avançados de inglês originários de instituições de ensino bilíngües e monolíngües - um estudo multidimensional baseado em corpus**. MA Thesis, LAEL, PUCSP, São Paulo, SP, 2002.

CONRAD, Susan. Corpus linguistics, language variation, and language teaching. In: SINCLAIR, John M. (Ed.), **How to Use Corpora in Language Teaching**. Philadelphia: John Benjamins, 2004. p. 67-88.

COSTELLO, Anna B.; OSBORNE, Jason W. Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. **Practical Assesment, Research & Evaluation**, v.10, n.7, 2005.

CROSSLEY, Scott.; LOUWERSE, Max. Multi-dimensional register classification using bi-grams. **International Journal of Corpus Linguistics**, 12(4), 2007. p. 453-478.

DE MÖNNINK, Inge; BROM, Niek; OOSTDIJK, Nelleke. Using the MF/MD method for automatic text classification. In: GRANGER, Sylviane; PETCH TYSON, Stephanie (Eds.), **Extending the scope of corpus based research: new applications new challenges**. Amsterdam: Rodopi, 2003. p. 15-25.

DELEGÁ-LÚCIO, Denise. **A variação entre textos argumentativos e o material didático de inglês: aplicações da análise multidimensional e do Corpus Internacional de Aprendizagem de Inglês (ICLE)**. Tese de Doutorado em Linguística Aplicada e Estudos da Linguagem – LAEL, Pontifícia Universidade Católica de São Paulo, São Paulo, 2013.

DRURY, Colin; MA, Jiao. **Do Language Barriers Result in Aviation Maintenance Errors?** Human Factors and Ergonomics Society 47th Annual Meeting Proceedings, Denver, Colorado, Oct. 13-17, 2003. Disponível em: <<http://www.hf.faa.gov/docs/508/docs/Maint%20-%20Language%20SUNY.pdf>>. Acesso em: 20 out. 2011

_____. **Language Errors in Aviation Maintenance: Final Report**, Reports to William J. Hughes Technical Center, the Federal Aviation Administration, 2005. Disponível em: <https://www2.hf.faa.gov/HFPortalNew/Search/DOCs/maint_Language_final.pdf>. Acesso em: 20 out. 2011.

ERVIN-TRIPP, Susan. On Sociolinguistic rules: alternation and co-occurrence. In: GUMPERZ, John J.; HYMES, Dell. (Ed.) **Directions in sociolinguistics**. New York: Holt, Rinehart and Winston, 1972.

FEDERAL AVIATION ADMINISTRATION. **Operator's Manual: Human Factors in Aviation Maintenance**, 2005. Disponível em: <<http://www.hf.faa.gov/opsmanual>>. Acesso em: 01 nov. 2011.

FEDERAL AVIATION ADMINISTRATION. **Technical Documentation Challenges in Aviation Maintenance: a proceedings report**, November, 2012. Disponível em: <http://www.faa.gov/data_research/research/med_humanfacs/oamtechreports/2010s/media/201216.pdf>. Acesso em: 01ago. 2013.

GÓMEZ, Pascual Cantos. **Statistical Methods in Language and Linguistic Research**. Sheffield: Equinox, 2013.

HOEY, Michael. **Patterns of Lexis in Text**. Oxford: Oxford University Press, 1991.

HUNSTON, Susan. **Corpora in Applied Linguistics**. Cambridge: Cambridge University Press, 2002.

HYMES, Dell. **Foundations in Sociolinguistics**. Philadelphia: University of Pennsylvania Press, 1974.

INTERNATIONAL CIVIL AVIATION ORGANIZATION. **Manual on the Implementation of ICAO Language Proficiency Requirements**. Doc 9835. 2nd edition. Montreal: ICAO, 2010.

IRVINE, Judith. Formality and informality in communicative events. In: BAUGH, John; SCHERZER, Joel. (Ed.) **Language in use: readings in sociolinguistics**. Englewood Cliffs, NJ: Prentice Hall, 1984/1979.

KAUFFMANN, Carlos. **O corpus do jornal: variação lingüística, gêneros e dimensões da imprensa diária escrita**. Master's thesis, LAEL, PUCSP, São Paulo, SP, 2005.

KIM, Yong-Jin.; BIBER, Douglas. A corpus-based analysis of register variation in Korean. In: BIBER, Douglas; FINEGAN, Edward (Eds.), **Sociolinguistic Perspectives on Register**. Oxford: Oxford University Press, 1994. p. 157-181.

LEE, David Y. W. **Modelling Variation in Spoken and Written Language: the Multi-Dimensional Approach Revisited**. Tese de doutoramento, Department of Linguistics and Modern English Language, Lancaster University, Uk, 2000.

McENERY, Tony; HARDIE, Andrew. **Corpus Linguistics: Method, Theory and Practice**. Cambridge: Cambridge University Press, 2011, p. 29-34.

O'KEEFFE, Anne. **Corpora, Pragmatics and Language Teaching**. Trabalho apresentado no IX Encontro de Linguística de Corpus, Porto Alegre, out. 2010.

O'KEEFFE, Anne; McCARTHY, Michael; CARTER, Ronald. **From Corpus to Classroom: language use and language teaching**. Cambridge: Cambridge University Press, 2007.

OLIVEIRA, Lúcia Pacheco de. **Variação intercultural na escrita: contrastes multidimensionais em inglês e português**. Tese de Doutorado em Linguística Aplicada e Estudos da Linguagem – LAEL, Pontifícia Universidade Católica de São Paulo, São Paulo, 1997.

PARODI, Giovanni. Variation across registers in Spanish: Exploring the El-Grial PUCV Corpus. In: PARODI, Giovanni (Ed.), **Working with Spanish Corpora**. London: Continuum, 2007. p. 11-53.

PINTO, Marcia Veirano. **A linguagem dos filmes norte-americanos ao longo dos anos: uma abordagem multidimensional**. Tese de Doutorado em Linguística Aplicada e Estudos da Linguagem – LAEL, Pontifícia Universidade Católica de São Paulo, São Paulo, 2013.

SANTOS, Valéria Branco Moreira Pinto dos. As características léxico-gramaticais de um manual de gestão a partir da análise multidimensional de Biber. **The ESpecialist**, 24(2), 2003. p. 201-227.

SARMENTO, Simone. **O uso dos verbos modais em manuais de aviação em inglês: um estudo baseado em corpus**. Tese de Doutorado em Estudos da Linguagem – Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2008.

_____. Linguística de corpus e o desenvolvimento de material didático para Inglês. In: SARMENTO, Simone; FREITAS, Ana Luiza P. de (Org.) **O ensino do Inglês como língua estrangeira: estudos e reflexões II**. Porto Alegre: ediPUCRS, 2009.

SIGNORINI, Inês. Do residual ao múltiplo e ao complexo: o objeto da pesquisa em LA. In: SIGNORINI, Inês; CAVALCANTI, Marilda C. **Linguística Aplicada e Transdisciplinaridade: questões e perspectivas**. Campinas: Mercado de Letras, 1998.

SHAWCROSS, Philip. **English for Aircraft Maintenance**. Paris: Belin, 1992.

SINCLAIR, John M. **Corpus, concordance, collocation**. Oxford: Oxford University Press, 1991.

_____. From theory to practice. In: LEECH, Geoffrey et al (org.). **Spoken English on Computer – Transcription, mark-up and application**. London: Longman, 1995.

_____. **Trust the text: language, corpus and discourse**. London/ New York/ Canada: Routledge, 2004.

_____. Corpus and text: basic principles. In: WYNNE, Martin. (Org.). **Developing linguistic corpora: a guide to good practice**. Oxford: Oxbow Books, 2005. p. 1-16.

SOUZA, Renata Condi de. **A revista TIME em uma perspectiva multidimensional**. Tese de Doutorado em Linguística Aplicada e Estudos da Linguagem – LAEL, Pontifícia Universidade Católica de São Paulo, São Paulo, 2013.

STUBBS, Michael. On texts, corpora and models of language. In: HOEY, Michael *et al.* **Text, Discourse and Corpora: theory and analysis**. London: Continuum, 2007.

ZUPPARDO, Maria Carolina. A linguagem da aviação: um estudo de manuais aeronáuticos baseado na análise multidimensional. **ReVEL**, v.11, n.21, 2013. p.6-25. Disponível em: < <http://revel.inf.br/files/4b416887c9e8c51b14c95dacc4f39d5.pdf> >

ANEXO A – Lista de Etiquetas

Lista de etiquetas contabilizadas pelo programa *TagCount*. (Adaptada de SOUZA, 2012)

Etiqueta	Descrição
<abstracn>	substantivo relacionado a coisas ou processos (<i>abstract/process noun</i>)
<act_ipv>	<i>phrasal verb</i> intransitivo relacionado a atividade (<i>activity - intransitive phrasal verb</i>)
<act_tpv>	<i>phrasal verb</i> transitivo relacionado a atividade (<i>activity - transitive phrasal verb</i>)
<actv>	verbo relacionado a atividade/ação (<i>activity verb</i>)
<adj_attr>	adjetivo atributivo (<i>attributive adjective</i>)
<adv>	advérbio (<i>adverb</i>)
<agls_psv>	voz passiva sem agente (<i>agentless passive</i>)
<all_adv<	todos os advérbios de posicionamento (<i>sum of stance adverbs</i>)
<all_jth>	todas as categorias de orações complementares com <i>that</i> controladas por adjetivos (<i>sum stance that complement clauses controlled by adjectives</i>)
<all_jto>	todas as categorias de orações complementares com <i>to</i> controladas por adjetivos (<i>sum stance to complement clauses controlled by adjectives</i>)
<all_nth>	todas as categorias de orações complementares com <i>that</i> controladas por substantivos (<i>sum stance that complement clauses controlled by nouns</i>)
<all_nto>	<i>to</i> usado em oração controlada por substantivos de posicionamento (<i>to complement clause controlled by stance nouns</i>)
<all_th>	soma das orações complementares com <i>that</i> (<i>sum stance that complement clauses</i>)
<all_to>	todas as orações complementares com <i>to</i> (<i>sum stance to complement clauses</i>)
<all_vth>	todas as orações complementares com <i>that</i> controlada por verbos (<i>sum stance that complement clauses controlled by verbs</i>)
<all_vto>	todas as orações complementares com <i>to</i> controlada por verbos (<i>sum stance to complement clauses controlled by verbs</i>)
<alladj>	todas as categorias de adjetivos (<i>all adjectives</i>)
<allconj>	todas as categorias de conjunções (<i>all conjunctions</i>)
<allmodal>	todas as categorias de verbos modais (<i>all modals</i>)
<allpasv>	todos os usos de voz passiva (<i>all passives</i>)
<allpro>	todas as categorias de pronomes (<i>all pronouns</i>)
<allverb>	todas as categorias de verbos, excluindo verbos auxiliares (<i>all verbs</i>)
<allwh>	todas as palavras iniciadas por <i>WH-</i> (<i>all WH- words</i>)
<allwhrel>	todas as orações relativas com pronome <i>WH-</i> (<i>all WH- relative clauses</i>)
<amplifr>	advérbio qualificador/amplificador (<i>amplifier</i>)
<aspectpv>	<i>phrasal verb</i> acurativo/determinativo (<i>aspectual verb – phrasal verb</i>)

<aspectv>	verbo acurativo/determinativo (<i>aspectual verb</i>)
<atadvl>	advérbio atitudinal (<i>attitudinal adverb</i>)
<att_jth>	<i>that</i> usado em oração complementar controlada por adjetivo atitudinal (<i>that complement clause controlled by attitudinal adjective</i>)
<att_nth>	<i>that</i> usado em oração complementar controlada por substantivo atitudinal (<i>that complement clause controlled by attitudinal noun</i>)
<att_vth>	<i>that</i> usado em oração complementar controlada por verbo atitudinal (<i>that complement clause controlled by attitudinal verb</i>)
<be_state>	verbo <i>to be</i> indicativo de estado (<i>be state</i>)
<by_pasv>	voz passiva com agente e preposição <i>by</i> (<i>BY- passive</i>)
<causev>	verbo causativo (<i>causative verb</i>)
<cognitn>	substantivo cognitivo (<i>cognitive noun</i>)
<colorj>	adjetivo – cor (<i>color adjective</i>)
<commpv>	<i>phrasal verb</i> transitivo relacionado a comunicação (<i>communication – transitive phrasal verb</i>)
<commv>	verbo relacionado a comunicação (<i>communication verb</i>)
<concrtn>	substantivo concreto (<i>concrete noun</i>)
<conjuncts>	conjunção (<i>conjunction</i>)
<contrac>	contração (<i>contraction</i>)
<copulapv>	<i>phrasal verb</i> de ligação/copula (<i>copular phrasal verb</i>)
<downtone>	advérbio suavizador (<i>downtoner</i>)
<dsre_vto>	<i>to</i> usado em oração complementar controlada por verbos de desejo, intenção e decisão (<i>to complement clauses controlled by verbs of desire, intention, and decision</i>)
<efrt_vto>	<i>to</i> usado em oração controlada por verbos de modalidade, causalidade e esforço (<i>to complement clause controlled by verbs of modality, causation and effort</i>)
<evalj>	adjetivo avaliativo (<i>evaluative adjective</i>)
<existv>	verbo relacionado a existência ou relacionamento (<i>existence verb</i>)
<fact_jth>	<i>that</i> usado em oração complementar controlada por adjetivo factivo (<i>that complement clause controlled by factive adjective</i>)
<fact_vth>	<i>that</i> usado em oração complementar controlada por verbo factivo (<i>that complement clause controlled by a factive verb</i>)
<factadvl>	advérbio factivo (<i>factive adverb</i>)
<fct_nth>	<i>that</i> usado em oração complementar controlada por substantivo factivo (<i>that complement clause controlled by a factive noun</i>)
<finlprep>	preposição desacompanhada (<i>stranded preposition</i>)
<gen_emph>	advérbio ou palavra quantificador(a) enfático(a) (<i>general emphatics</i>)
<gen_hdg>	advérbio delimitador/atenuador (<i>general hedges</i>)
<groupn>	substantivo relacionado a grupo ou instituição (<i>group / institution noun</i>)
<have>	verbo <i>have</i> (<i>have</i>)
<humann>	substantivo animado (<i>animate noun</i>)

<inf>	infinitivo (<i>infinitive</i>)
<it>	pronome <i>it</i> (<i>pronoun it</i>)
<jcmp>	<i>that</i> usado em oração complementar controlada por adjetivo (<i>that complement clause controlled by adjective</i>)
<lkly_jth>	<i>that</i> usado em oração complementar controlada por adjetivo de probabilidade (<i>that complement clause controlled by adjective of likelihood</i>)
<lkly_nth>	<i>that</i> usado em oração complementar controlada por substantivo de probabilidade (<i>that complement clause controlled by noun of likelihood</i>)
<lkly_vth>	<i>that</i> usado em oração complementar controlada por verbo de probabilidade (<i>that complement clause controlled by verb of likelihood</i>)
<lklyadvl>	advérbio de probabilidade (<i>likelihood adverb</i>)
<mentalpv>	<i>phrasal verb</i> transitivo relacionado a atividade mental (<i>mental – transitive phrasal verb</i>)
<mentlv>	verbo relacionado a atividade mental (<i>mental verb</i>)
<mntl_vto>	<i>to</i> usado em oração complementar controlada por verbo de cognição (<i>to complement clause controlled by verb of cognition</i>)
<n_nom>	nominalização no singular (<i>noun nominalization</i>)
<n>	substantivo (<i>noun</i>)
<nec_mod>	verbo modal de necessidade (<i>necessity modal</i>)
<nfct_nth>	<i>that</i> usado em oração complementar controlada por substantivo não factivo (<i>that complement clause controlled by non-factive noun</i>)
<nonf_vth>	<i>that</i> usado em oração complementar controlada por verbo não factivo (<i>that complement clause controlled by non-factive verb</i>)
<nonfadvl>	advérbio não factivo (<i>non-factive adverb</i>)
<o_and>	conjunção coordenada – conectivo clausal (<i>coordinating conjunction – clausal connector</i>)
<occurpv>	<i>phrasal verb</i> intransitivo relacionado a evento/ocorrência (<i>occurrence – intransitive phrasal verb</i>)
<occurv>	verbo relacionado a evento/ocorrência (<i>occurrence verb</i>)
<p_and>	conjunção coordenada – conectivo frasal (<i>coordinating conjunction - phrasal connector</i>)
<pany>	pronome indefinido (<i>indefinite pronoun</i>)
<pasttense>	passado (<i>past tense</i>)
<pdem>	pronome demonstrativo (<i>demonstrative pronoun</i>)
<perfects>	aspecto perfeito (<i>perfect aspect</i>)
<pl_adv>	advérbio de lugar (<i>place adverb</i>)
<placen>	substantivo – lugar (<i>place noun</i>)
<pos_mod>	verbo modal de possibilidade (<i>modal of possibility</i>)
<prcessn>	substantivo relacionado a processos (<i>process noun</i>)
<prd_mod>	verbo modal preditivo (<i>modal of prediction</i>)
<pred_adj>	adjetivo predicativo (<i>predicative adjective</i>)
<prep>	preposição (<i>preposition</i>)

<pres>	verbo no presente (<i>present</i>)
<pro_do>	verbo <i>do</i> como substituto de outro verbo ou sintagma verbal / verbo vicário / pró-verbo (<i>pro-verb do</i>)
<pro1>	pronome em 1a pessoa (<i>1st person pronoun</i>)
<pro2>	pronome em 2a pessoa (<i>2nd person pronoun</i>)
<pro3>	pronome em 3a pessoa (<i>3rd person pronoun</i>)
<prob_vto>	<i>to</i> usado em oração complementar controlada por verbos de probabilidade e fato (<i>to complement clause controlled by verbs of probability and simple fact</i>)
<prtcle>	partícula do discurso (<i>discourse particle</i>)
<prv_vb>	verbo de cognição (<i>private verb</i>)
<pub_vb>	verbo <i>dicendi</i> (<i>public verb</i>)
<quann>	substantivo – quantidade (<i>quantity noun</i>)
<rel_obj>	oração <i>WH-</i> em posição de objeto (<i>WH- relative clause on object position</i>)
<rel_pipe>	oração <i>WH-</i> com preposição inicial (<i>pied-piping construction</i>)
<rel_sub>	oração <i>WH-</i> em posição de sujeito (<i>WH- relative clause on subject position</i>)
<relatnj>	adjetivo – relacionamentos (<i>relational adjective</i>)
<sizej>	adjetivo – tamanho (<i>size adjective</i>)
<spch_vto>	<i>to</i> usado em oração complementar controlada por verbos de atos de fala (<i>to complement clause controlled by speech act verbs</i>)
<spl_aux>	advérbio usado entre verbo auxiliar e verbo principal (<i>split auxiliary</i>)
<sua_vb>	verbo de persuasão (<i>suasive verb</i>)
<sub_cnd>	conjunção subordinativa condicional (<i>conditional subordination</i>)
<sub_cos>	subordinação causativa (<i>causative subordination</i>)
<sub_othr>	outros advérbios/conjunções usadas em orações subordinadas (<i>other adverbial subordinators</i>)
<tcncrtn>	substantivo relacionado a assuntos técnicos ou concretos (<i>technical / concrete nouns</i>)
<that_del>	omissão de <i>that</i> em oração subordinada (<i>that deletion</i>)
<that_rel>	<i>that</i> em orações subordinadas com pronome relativo (<i>that relative clauses</i>)
<timej>	adjetivo – tempo (<i>time adjective</i>)
<tm_adv>	advérbio de tempo (<i>time adverb</i>)
<topicj>	adjetivo relacionado a tópicos (<i>topical attributive adjective</i>)
<typetokn>	relação entre item e ocorrência (<i>type token ratio</i>)
<vcmp>	<i>that</i> usado em oração complementar controlada por verbo (<i>that complement clause controlled by verb</i>)
<vprogrsv>	presente contínuo (<i>present progressive</i>)
<wh_cl>	oração com pronome <i>WH-</i> (<i>WH- clauses</i>)
<wh_ques>	pronome <i>WH-</i> usado em perguntas (<i>WH- questions</i>)
<whiz_vbn>	modificador pós-nominal da voz passiva (<i>passive postnominal modifier</i>)

<wordcnt>	quantidade de palavras (<i>word count</i>)
<wrldngth>	tamanho de palavra (<i>word length</i>)
<x1_jto>	to usado em oração complementar controlada por adjetivos de certeza (<i>to complement clause controlled by adjectives of certainty</i>)
<x2_jto>	to usado em oração complementar controlada por adjetivos de habilidade/desejo (<i>to complement clause controlled by adjectives of ability/willingness</i>)
<x3_jto>	to usado em oração complementar controlada por adjetivos de afeição pessoal (<i>to complement clause controlled by adjectives of personal affect</i>)
<x4_jto>	to usado em oração complementar controlada por adjetivos de facilidade/dificuldade (<i>to complement clause controlled by adjectives of ease/difficulty</i>)
<x5_jto>	to usado em oração complementar controlada por adjetivos de avaliatividade (<i>to complement clause controlled by evaluative adjectives</i>)

ANEXO B – Descrições das Etiquetas do *Biber Tagger*

As descrições das etiquetas usadas pelo programa *Biber Tagger* estão disponíveis na internet, no site do *American National Corpus* (<http://www.americannationalcorpus.org/SecondRelease/Biber-tags.txt>). Algumas descrições contam com exemplos de palavras que podem receber as respectivas etiquetas.

:+clp+++	colon + clause punctuation
;+clp+++	semi-colon + clause punctuation
?+clp+++	question mark + clause punctuation
!+clp+++	exclamation mark + clause punctuation
,++++	comma
-++++	dash
"++++	double quote mark
+ ++++	single quote mark
(++++	left parenthesis
)++++	right parenthesis
\$++++	dollar sign
%++++	percent sign
&fo++++	formula symbols
&fw++++	foreign word
abl++++	pre-qualifier (rather, such)
abn++++	pre-quantifier (all, half)
abx++++	pre-quantifier/double conjunction (both)
ap++++	post-determiner (many, more, most, only, other, own, same, ...)
aps++++	(others)
at++++	singular indefinite article (a, an)
ati++++	singular definite article (the, no)
cc++++	coordinating conjunction (and, but, or)
cc+cls+++	coordinating conjunction + clausal connector
cc+phrs+++	coordinating conjunction + phrasal connector
cc"++++	multi-word coordinating conjunction (as well as)
cc++neg++	coordinating conjunction + + negation (nor)
cd++++	cardinal number (2, 3, 4, two, three, four, hundred, ...)
cd+date+++	cardinal number + date (year only)
cd1++++	cardinal number: 1, one
cd1s++++	cardinal number: ones
cds++++	cardinal plural (tens, hundreds, thousands)
od++++	ordinal number (1st, 2nd, first, second, ...)
cs+cnd+++	subordinating conjunction + conditional (if, unless)
cs+con+++	subordinating conjunction + concessive (although, though)
cs+coss+++	subordinating conjunction + causative (because)

cs+who+++	subordinating conjunction + WH word (whether)
cs+sub+++	subordinating conjunction + other (as, except, until, ...)
cs"++++	multi-word subordinating conjunction (in that, so that, ...)
dt+dem+++	determiner + demonstrative (this,that,these,those modifying N)
dt+pdem+++	determiner + demonstrative pronoun (this, that, these, those)
dti++++	singular or plural determiner (any, enough, some)
dt++++	other singular determiner (another, each)
dtx++++	determiner/double conjunction (either)
ex+pex+++	existential there
in++++	preposition
in+ppvb+++	preposition + prepositional verb (account for, join in, ...)
in+pl+++	preposition + place marker (above, behind, beside, ...)
in"++++	multi-word preposition (as to, away from, instead of, ...)
in+strn+++	preposition + stranded
jj+atrb+++	adjective + attributive function
jj+atrb++xvbg+	adjective + attributive function + + -ing form
jj+atrb++xvbn+	adjective + attributive function + + past participle form
jj+pred+++	adjective + predicative function
jj++++	adjective + indeterminate function
jjb+atrb+++	attributive-only adjective + attributive (chief, entire)
jjr+atrb+++	comparative adjective + attributive function
jjr+pred+++	comparative adjective + predicative function
jjt+atrb+++	superlative adjective + attributive function
md+nec+++	modal + necessity (ought, should, must)
md+pos+++	modal + possibility (can, may, might, could)
md+prd+++	modal + prediction (will, would, shall)
md"++pmd"++	modal + + multi-word periphrastic modal (e.g., be going to)
nn++++	singular common noun
nn+nom+++	singular noun + nominalization
nvbg+++xvbg+	singular noun + + + -ing form
nn+++xvbn+	singular noun + + + past participle form
nns++++	plural common noun
nns+nom+++	plural noun + nominalization
nnu++++	unit of measurement (lb, kg, ...)
np++++	singular proper noun
nps++++	plural proper noun
npl++++	locative noun
npt++++	singular titular noun
npts++++	plural titular noun
nr++++	singular adverbial noun (east, west, today, home, ...)
nrs++++	plural adverbial noun
pp1a+pp1+++	first person subject pronoun + first person pronoun
pp1a+pp1+++0	first person subject pronoun + 1st person pro. + contracted

pp1o+pp1+++	first person object pronoun + first person pronoun
pp\$+pp1+++	possessive determiner + first person pronoun (my, our)
ppl+pp1+++	singular reflexive pronoun + first person pronoun (myself)
ppls+pp1+++	plural reflexive pronoun + first person pronoun (ourselves)
pp2+pp2+++	second person pronoun + second person pronoun
pp\$+pp2+++	possessive determiner + second person pronoun (your)
ppl+pp2+++	singular reflexive pronoun + second person pronoun (yourself)
pp3a+pp3+++	third person subject pronoun + third person personal pronoun
pp3o+pp3+++	third person object pronoun + third person personal pronoun
pp3+pp3+++0	third person pronoun + 3rd person personal pro. + contracted
pp\$+pp3+++	possessive + 3rd pers. personal pro. (his, her, their)
ppl+pp3+++	sg. reflexive pronoun + 3rd pers. personal pro. (her/himself)
ppls+pp3+++	pl. reflexive pronoun + 3rd pers. personal pro. (themselves)
pp3+it+++	third person pronoun + third person impersonal pronoun (it)
pp\$+it+++	possessive determiner + third person impersonal pronoun (its)
pp\$\$++++	possessive pronoun (mine, yours, ...)
pn"++++	multi-word nominal pronoun (no one, ...)
pn++++	nominal pronoun (someone, everything, ...)
ql++++	qualifier + (as, less, more, too)
ql+amp+++	qualifier + amplifier (very)
ql+emph+++	qualifier + emphatic (most)
qlp++++	post-qualifier (enough, indeed)
rb++++	general adverb
rb"++++	multi-word adverb (at last, in general)
rb+cnj+++	adverb + conjunct (however, therefore, thus, ...)
rb++neg++	neither
rb+amp+++	adverb + amplifier (absolutely, completely, entirely, ...)
rb+down+++	adverb + downtoner (nearly, only, merely, ...)
rb+emph+++	adverb + emphatic (just, really, so, ...)
rb+hdg+++	adverb + hedge (almost, maybe, ...)
rb"+hdg"+++	multi-word adverb + hedge (kind of, sort of)
rb+phrv+++	adverb + phrasal verb (get in, wrap up, ...)
rb+pl+++	adverb + place marker (abroad, ahead, far, upstream, ...)
rb+tm+++	adverb + time marker (afterwards, again, immediately, ...)
rb+dspt+++	adverb + discourse particle (anyway, well, ...)
rbr++++	comparative adverb (better, quicker)
rbr+tm+++	comparative adverb + time marker (earlier, later, sooner, ...)
rn+pl+++	nominal adverb + place marker (here, there)
rn+tm+++	nominal adverb + time marker (now, then)
rn+dspt+++	nominal adverb + discourse particle (now)
rp++++	adverbial particle (back, in, round, up, ...)
rp+pl+++	adverbial particle + place marker (away, behind, out, ...)
tht+jcmp+++	that as dependent clause head + adjective complement

tht+ncmp+++	that as dependent clause head + noun complement
tht+vcmp+++	that as dependent clause head + verb complement
tht+rel+++	that as dependent clause head + relative clause
tht+rel+obj++	that as dep. clause head + relative clause + object position
tht+rel+subj++	that as dep. clause head + relative clause + subject position
to++++	infinitive marker
to"++++	multi-word infinitive marker (in order to)
uh++++	interjection/filler (hey, oh, ok, yes, erm ...)
vb++++	base form of verb, excluding verbs in infinitive clauses (uninflected present tense, imperative)
vb+++xvbn+	base form of verb + + + past participle form (e.g., cut, hit, hurt, ...)
vb+be+aux++	base form of verb + be + auxiliary verb
vb+be+vrb++	base form of verb + be + main verb
vb+bem+aux++	verb + am + auxiliary verb
vb+bem+aux++0	verb + am + auxiliary verb + + contracted ('m)
vb+bem+vrb++	verb + am + main verb
vb+bem+vrb++0	verb + am + main verb + + contracted ('m)
vb+ber+aux++	verb + are + auxiliary verb
vb+ber+aux++0	verb + are + auxiliary verb + + contracted ('re)
vb+ber+vrb++	verb + are + main verb
vb+ber+vrb++0	verb + are + main verb + + contracted ('re)
vb+do+aux++	verb + do + auxiliary verb
vb+do+vrb++	verb + do + main verb
vb+hv+aux++	verb + have + auxiliary verb
vb+hv+aux++0	verb + have + auxiliary verb + + contracted ('ve)
vb+hv+vrb++	verb + have + main verb
vb+hv+vrb++0	verb + have + main verb + + contracted ('ve)
vb+seem+++	base form of verb + seem / appear
vb+vprv+++	base form of verb + private verb (believe, feel, think, ...)
vb+vprv+tht0++	base form of verb + private verb + that deletion **
vb+vpub+++	base form of verb + public verb (assert, complain, say, ...)
vb+vpub+tht0++	base form of verb + public verb + that deletion **
vb+vsua+++	base form of verb + suasive verb (ask, command, insist, ...)
vbd+++xvbn+	past tense verb + + + past participle form
vbd+bed+aux++	past tense verb + were + auxiliary verb
vbd+bed+vrb++	past tense verb + were + main verb
vbd+bedz+aux++	past tense verb + was + auxiliary verb
vbd+bedz+vrb++	past tense verb + was + main verb
vbd+dod+aux++	past tense verb + did + auxiliary verb
vbd+dod+vrb++	past tense verb + did + main verb
vbd+hvd+aux++	past tense verb + had + auxiliary verb
vbd+hvd+vrb++	past tense verb + had + main verb
vbd+seem++xvbn+	past tense verb + seem/appear

vbd+vprv++xvbn+	past tense + private verb (believe, feel, think, ...)
vbd+vprv+tht0+xvbn+	past tense + private verb + that deletion **
vbd+vpub++xvbn+	past tense + public verb (assert, complain, say, ...)
vbd+vpub+tht0+xvbn+	past tense + public verb + that deletion **
vbd+vsua++xvbn+	past tense + suasive verb (ask, command, insist, ...)
vbg+++xvbg+	present progressive verb + + + -ing form
vbg+beg++xvbg+	present progressive verb + being
vbg+beg+aux+xvbg+	present progressive verb + being + auxiliary verb
vbg+hvg++xvbg+	present progressive verb + having
vbg+vprv++xvbg+	pres. prog. + private verb (believe, feel, think, ...)
vbg+vprv+tht0+xvbg+	present progressive + private verb + that deletion
vbg+vpub++xvbg+	pres. prog. + public verb (assert, complain, say, ...)
vbg+vpub+tht0+xvbg+	present progressive + public verb + that deletion
vbg+vsua++xvbg+	pres. prog. + suasive verb (ask, command, insist, ...)
vwbg+++xvbg+	present progressive postnominal modifier
vwbg+beg++xvbg+	present progressive postnominal modifier + being
vwbg+hvg++xvbg+	present progressive postnominal modifier + having
vwbg+vprv++xvbg+	present prog. postnom. modifier + private verb
vwbg+vpub++xvbg+	present prog. postnom. modifier + public verb
vbi++++	base form of verb in infinitive clause
vbi+vprv+++	infinitive verb + private verb (believe, feel, think, ...)
vbi+vprv+tht0++	infinitive verb + private verb + that deletion **
vbi+vpub+++	infinitive verb + public verb (assert, complain, say, ...)
vbi+vpub+tht0++	infinitive verb + public verb + that deletion **
vbi+vsua+++	infinitive verb + suasive verb (ask, command, insist, ...)
vbz++++	3rd person singular verb
vbz+bez+aux++	3rd person sg. verb + is + auxiliary verb
vbz+bez+aux++0	3rd person sg. + is + auxiliary verb + + contracted ('s)
vbz+bez+vrb++	3rd person sg. verb + is + main verb
vbz+bez+vrb++0	3rd person sg. + is + main verb + + contracted ('s)
vbz+doz+aux++	3rd person sg. verb + does + auxiliary verb
vbz+doz+vrb++	3rd person sg. verb + does + main verb
vbz+hvz+aux++	3rd person sg. verb + has + auxiliary verb
vbz+hvz+vrb++	3rd person sg. verb + has + main verb
vbz+seem+++	3rd person sg. verb + seem/appear
vbz+vprv+++	3rd person sg. + private verb (believe, feel, think, ...)
vbz+vprv+tht0++	3rd person sg. + private verb + that deletion
vbz+vpub+++	3rd person sg. + public verb (assert, complain, say, ...)
vbz+vpub+tht0++	3rd person sg. + public verb + that deletion
vbz+vsua+++	3rd person sg. + suasive verb (ask, command, insist, ...)
vprf+++xvbn+	perfect aspect verb + + + past participle form
vprf++tht0+xvbn+	perfect aspect verb + + that deletion
vprf+ben+aux+xvbn+	perfect aspect verb + been + auxiliary verb

vprf+ben+vrb+xvbn+	perfect aspect verb + been + main verb
vpsv++agls+xvbn+	main clause passive verb + + agentless passive
vpsv++by+xvbn+	main clause passive verb + + by passive
wbn+++xvbn+	passive postnominal modifier + + + past participle form
wbn+vprv++xvbn+	passive postnominal modifier + private verb
wbn+vpub++xvbn+	passive postnominal modifier + public verb
wbn+vsua++xvbn+	passive postnominal modifier + suasive verb
wdt+who+++	WH determiner + WH word (what, whatever, whichever, ...)
wdt+who+whcl++	WH determiner + WH word + WH clause
wdt+who+whq++	WH determiner + WH word + WH question
whp+rel+obj++	WH pronoun + relative clause + object position
whp+rel+pied++	WH pronoun + relative clause + object position with prepositional fronting ('pied piping')
whp+rel+subj++	WH pronoun + relative clause + subject position
whp+who+++	WH pronoun + WH word (not a relative clause)
whp+who+whq++	WH pronoun + WH word + WH question
wrb+who+++	WH adverb (how, when, where, ...) + WH word
wrb+who+whcl++	WH adverb + WH word + WH clause
wrb+who+whq++	WH adverb + WH word + WH question
xnot++not++	not + + negation
xnot++not++0	not + + negation + + contracted form (n't)
xvbn+++xvbn+	past participle form -- indeterminate grammatical function
xvbg+++xvbg+	present participle form -- indeterminate grammatical function
zz++++	letter of the alphabet

ANEXO C – CAM – Tabela de Variância Total Explicada

Fator	Autovalores iniciais			Soma da extração de cargas quadráticas			Somas rotacionadas das cargas quadráticas
	Total	% de Variação	% acumulada	Total	% de Variação	% acumulada	Total
1	21,441	19,317	19,317	21,043	18,957	18,957	14,712
2	10,824	9,751	29,068	10,469	9,432	28,389	13,998
3	7,260	6,541	35,609	6,835	6,157	34,547	14,495
4	6,007	5,412	41,020	5,473	4,930	39,477	11,167
5	5,022	4,524	45,545	4,617	4,160	43,636	6,503
6	4,717	4,249	49,794	4,248	3,827	47,463	7,327
7	3,384	3,049	52,843				
8	3,278	2,953	55,796				
9	2,636	2,375	58,171				
10	2,493	2,246	60,417				
11	2,380	2,144	62,562				
12	2,180	1,964	64,526				
13	2,052	1,848	66,374				
14	1,969	1,774	68,148				
15	1,746	1,573	69,721				
16	1,662	1,497	71,217				
17	1,548	1,395	72,612				
18	1,447	1,304	73,916				
19	1,403	1,264	75,180				
20	1,352	1,218	76,397				
21	1,309	1,179	77,577				
22	1,192	1,074	78,651				
23	1,181	1,064	79,715				
24	1,108	,998	80,713				
25	1,015	,914	81,627				
26	1,005	,906	82,533				
27	,967	,872	83,405				
28	,893	,805	84,209				
29	,869	,783	84,993				
30	,849	,765	85,758				
31	,818	,737	86,495				
32	,744	,670	87,165				

33	,713	,642	87,807			
34	,643	,579	88,387			
35	,633	,570	88,957			
36	,626	,564	89,521			
37	,607	,547	90,067			
38	,603	,543	90,610			
39	,581	,523	91,134			
40	,529	,477	91,610			
41	,507	,457	92,068			
42	,474	,427	92,494			
43	,456	,411	92,905			
44	,424	,382	93,287			
45	,405	,365	93,652			
46	,404	,364	94,016			
47	,379	,342	94,358			
48	,373	,336	94,694			
49	,344	,310	95,004			
50	,312	,281	95,285			
51	,298	,268	95,553			
52	,292	,263	95,816			
53	,280	,252	96,068			
54	,264	,238	96,306			
55	,259	,234	96,540			
56	,244	,220	96,760			
57	,226	,204	96,964			
58	,226	,204	97,168			
59	,195	,175	97,343			
60	,189	,171	97,514			
61	,180	,162	97,676			
62	,171	,154	97,830			
63	,163	,147	97,976			
64	,160	,144	98,120			
65	,153	,138	98,258			
66	,144	,130	98,388			
67	,130	,118	98,506			
68	,120	,109	98,614			
69	,115	,104	98,718			
70	,113	,102	98,820			
71	,104	,094	98,914			
72	,102	,092	99,006			

73	,097	,088	99,094			
74	,088	,079	99,173			
75	,081	,073	99,246			
76	,079	,071	99,317			
77	,073	,066	99,383			
78	,068	,061	99,444			
79	,065	,058	99,502			
80	,062	,056	99,558			
81	,060	,054	99,612			
82	,050	,045	99,657			
83	,045	,040	99,697			
84	,044	,040	99,737			
85	,040	,036	99,773			
86	,037	,033	99,806			
87	,034	,031	99,837			
88	,029	,026	99,863			
89	,027	,024	99,887			
90	,024	,022	99,909			
91	,022	,019	99,928			
92	,017	,015	99,943			
93	,013	,012	99,955			
94	,012	,011	99,966			
95	,011	,010	99,976			
96	,007	,006	99,982			
97	,005	,005	99,987			
98	,004	,004	99,991			
99	,003	,003	99,993			
100	,002	,002	99,996			
101	,002	,002	99,998			
102	,001	,001	99,999			
103	,001	,001	99,999			
104	,000	,000	100,000			
105	,000	,000	100,000			
106	,000	,000	100,000			
107	7,428E-005	6,692E-005	100,000			
108	3,142E-005	2,831E-005	100,000			
109	7,879E-006	7,098E-006	100,000			

110	7,111E-006	6,406E-006	100,000				
111	5,145E-006	4,635E-006	100,000				

Método de extração: Fatoração de eixo principal.

ANEXO D – CAM – ESTATÍSTICAS DESCRITIVAS

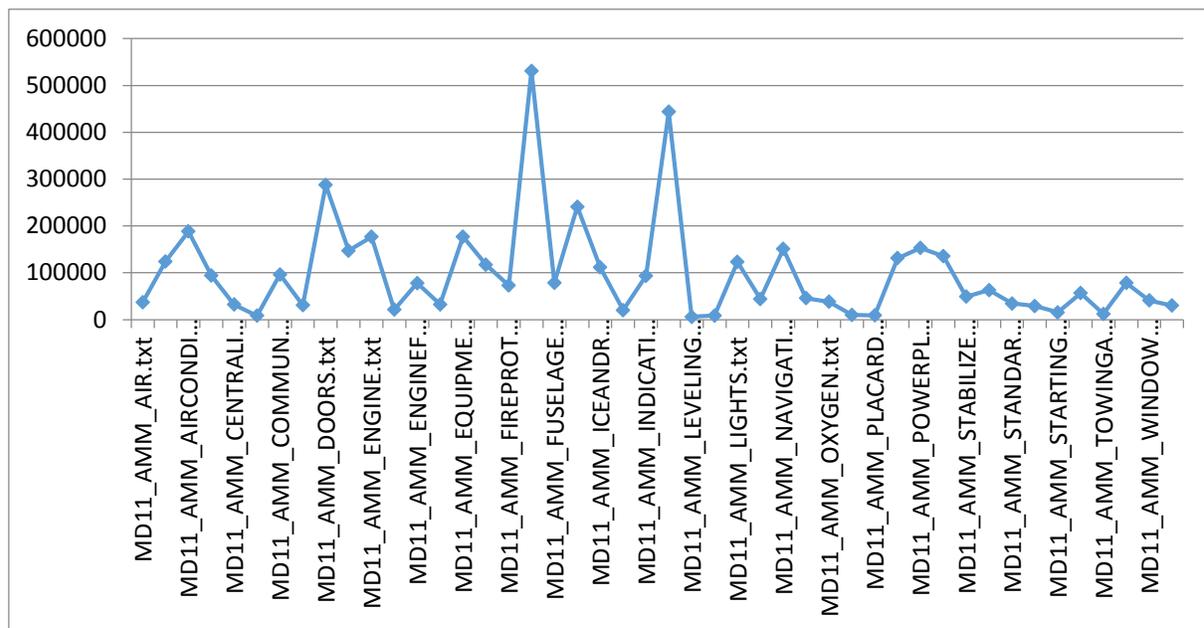
	N	Mínimo	Máximo	Média	Desvio Padrão
ttr	154	9,0000	37,8000	25,635065	6,0089969
wrlengh	154	4,0000	6,0000	4,819481	,3437734
wcount	154	782	94957	27134,56	22857,259
prv_vb	154	,0000	26,8000	5,457143	4,3940521
that_del	154	,0000	3,0000	,292857	,4329723
pres	154	28,2000	117,6000	72,429221	14,7634327
pro2	154	,0000	16,7000	2,872078	3,7731680
pro_do	154	,0000	24,8000	3,277922	5,2686053
pdem	154	,0000	9,3000	1,095455	1,2351254
gen_emph	154	,0000	1,6000	,190909	,3087744
pro1	154	,0000	8,4000	,607143	1,2267235
lt	154	,0000	7,8000	2,618831	1,5546586
be_state	154	,0000	32,2000	2,738961	7,2576515
sub_cos	154	,0	1,8	,089	,1949
prtcle	154	,0	1,9	,071	,2337
pany	154	,0000	4,5000	,469481	,6423332
amplifr	154	,0000	2,3000	,407143	,4179169
pos_mod	154	,0000	34,7000	4,807143	7,7598607
o_and	154	,3000	17,4000	4,122078	3,5659106
finlprep	154	,0000	6,9000	2,195455	1,5089424
n	154	331,0000	652,0000	458,328571	55,3361135
prep.	154	45,5000	125,9000	77,327922	15,4969299
adj_attr	154	16,5000	116,9000	51,992857	14,6580817
pasttNSE	154	1,7000	30,1000	10,216883	5,2739191
pro3	154	,0000	3,3000	,824675	,7449448
perfects	154	,0000	4,9000	1,301299	,9881644
pub_vb	154	,0000	9,7000	1,211688	1,7721132
rel_obj	154	,0	1,0	,134	,2256
rel_subj	154	,0000	7,0000	,829221	1,1985790
rel_pipe	154	,0	1,3	,112	,2149
p_and	154	,0000	9,5000	1,776623	1,7879331
n_nom	154	,0000	118,4000	61,199351	28,3007206
tm_adv	154	,0000	18,7000	2,142857	3,2114146
pl_adv	154	,0000	21,1000	3,673377	3,3867107
advs	154	3,6000	33,9000	14,013636	4,7486918

Inf	154	,0000	14,1000	5,121429	3,0162948
prd_mod	154	,0000	20,4000	2,279870	3,6611791
sua_vb	154	,0000	9,1000	1,558442	2,0952390
sub_cnd	154	,0000	30,6000	6,381169	6,6814715
nec_mod	154	,0000	11,4000	1,598052	1,7098129
spl_aux	154	,0000	6,9000	1,160390	1,1282133
conjnts	154	,0000	14,1000	,933766	1,5465425
agls_psv	154	,7000	37,1000	12,002597	9,1442467
by_pasv	154	,0000	8,5000	1,418182	1,6875730
whiz_vbn	154	,0000	11,8000	2,638312	2,2585263
sub_othr	154	,0000	6,9000	2,044156	1,2909432
vcmo	154	,0000	2,5000	,375325	,4276875
downtone	154	,0000	9,5000	1,528571	1,5500248
pred_adj	154	,1000	18,8000	3,520779	3,0893381
allmodal	154	,0000	37,8000	8,692208	8,9336299
allconj	154	2,2000	37,2000	14,417532	7,3144153
allpasv	154	,8000	44,6000	16,056494	11,4740339
allwhrel	154	,0000	9,0000	1,082468	1,4748051
alladj	154	35,2000	161,1000	78,234416	21,0975679
allpro	154	,0000	18,9000	4,305844	4,1595034
have	154	,0000	2,8000	,337013	,3978621
allverb	154	39,4000	144,6000	94,213636	18,7153883
vprogrsv	154	,0000	9,1000	1,576623	1,6000732
that_rel	154	,0000	5,7000	1,374026	1,1898155
jcmp	154	,0000	8,5000	1,244156	1,9201371
nonf_vth	154	,0000	1,5000	,266883	,3760919
att_vth	154	,0000	2,1000	,150000	,2874840
fact_vth	154	,0	,7	,093	,1377
lkly_vth	154	,0000	1,7000	,142208	,2774257
att_jth	154	,0000	1,9000	,262338	,4346252
fact_jth	154	,0000	7,8000	1,001299	1,6345129
lkly_jth	154	,0	,6	,053	,1049
spch_vto	154	,0000	1,1000	,035065	,1180077
mntl_vto	154	,0	1,3	,025	,1139
dsre_vto	154	,0	1,4	,140	,2578
efrt_vto	154	,0000	6,2000	,459091	,7906821
prob_vto	154	,0	,3	,014	,0458
x1_jto	154	,0	,7	,058	,1327
x2_jto	154	,0000	1,1000	,102597	,2032250

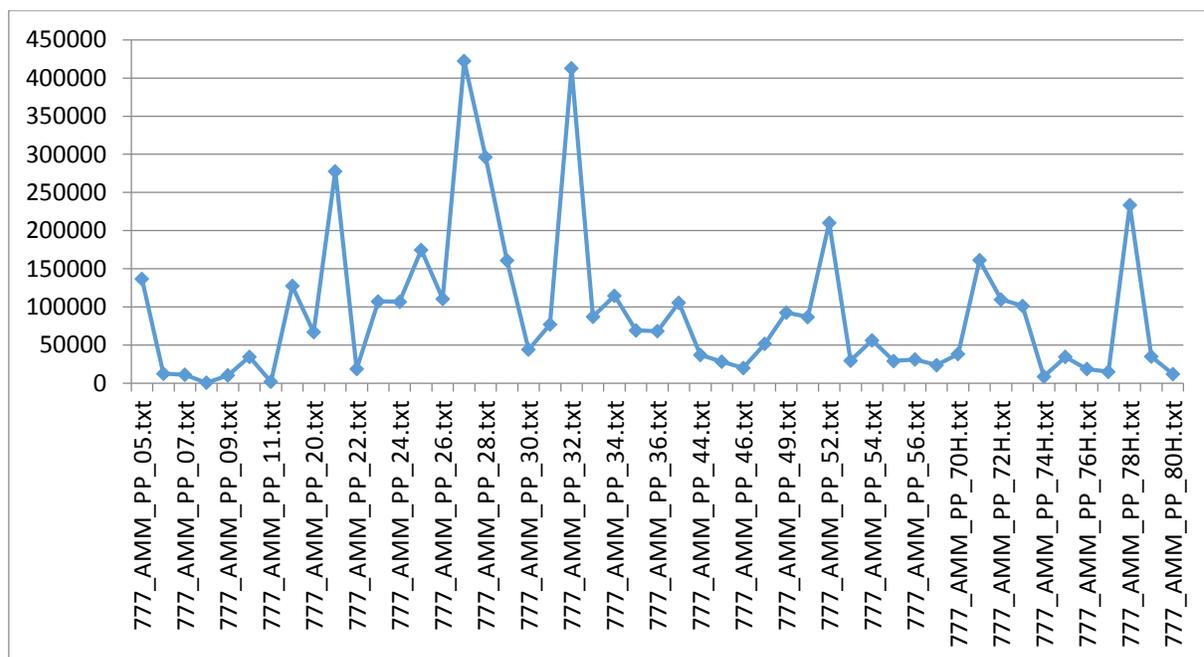
x4_jto	154	,0	,4	,029	,0683
x5_jto	154	,0000	1,9000	,227922	,4227975
all_nto	154	,0000	3,2000	,229870	,5086352
nonfadvl	154	,0000	1,5000	,081818	,2017747
factadvl	154	,0	,9	,058	,1197
all_vth	154	,0000	3,8000	,649351	,7483530
all_jth	154	,0000	8,1000	1,320130	1,7996143
all_nth	154	,0	,3	,011	,0451
all_th	154	,0000	8,3000	1,983766	1,6781748
all_vto	154	,0000	6,9000	,674026	,9389749
all_jto	154	,0000	1,9000	,418831	,4660738
all_to	154	,0000	6,9000	1,330519	1,2352305
all_advl	154	,0000	1,9000	,166883	,2908305
act_ipv	154	,0000	7,0000	,635714	1,3036238
act_tpv	154	,0000	1,1000	,096104	,2076583
humann	154	,0000	3,9000	,540260	,6646557
prcessn	154	1,9000	29,7000	11,534416	4,9466749
cognitn	154	,0000	3,9000	,597403	,7172887
abstrcn	154	6,2000	75,2000	22,629870	12,2520220
concrtn	154	,0000	67,4000	18,125974	10,6380609
tccnrt	154	,5000	38,4000	7,863636	7,0643894
quann	154	,0000	15,3000	4,730519	3,1420265
placen	154	,2000	14,7000	3,677273	2,5092645
groupn	154	,0000	9,2000	1,759091	2,0044595
sizej	154	,0000	6,4000	,703896	1,0402918
timej	154	,0	3,6	,221	,4475
colorj	154	,0000	2,1000	,288312	,4219389
evalj	154	,0000	16,8000	1,020779	1,8327691
relatnj	154	,0000	5,2000	1,357143	1,1823792
topicj	154	,0000	4,9000	,539610	,7781677
actv	154	4,5000	36,5000	15,596104	6,0566809
commv	154	,0000	11,1000	1,823377	2,0201531
mentlv	154	,0000	7,2000	2,937013	1,7654251
causev	154	,0000	11,0000	2,981169	2,6648712
occurv	154	,0000	6,2000	1,033766	,9363551
existv	154	,0000	7,9000	2,212987	1,6281712
aspectv	154	,0000	8,7000	1,930519	1,8794147

ANEXO E – Gráficos de distribuição

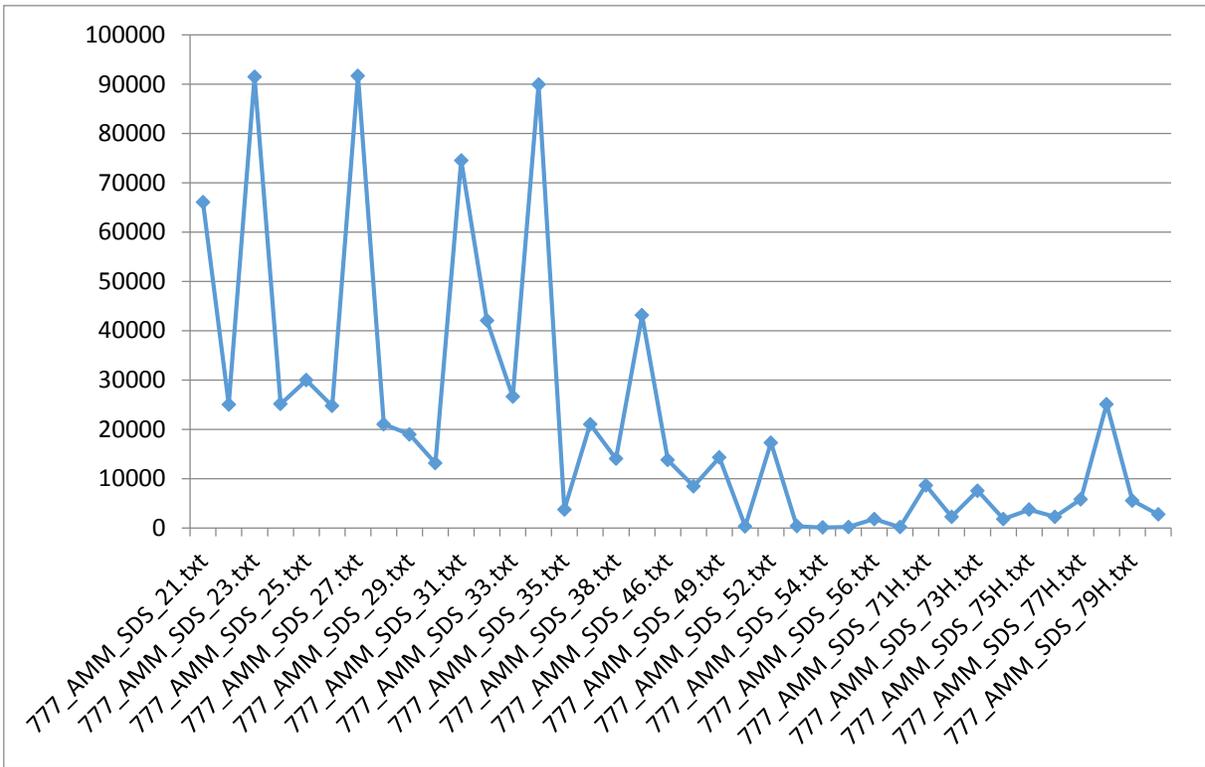
Os gráficos de distribuição utilizados para definir limites de corte estão elencados abaixo, de acordo com o modelo de aeronave e tipo de manual.



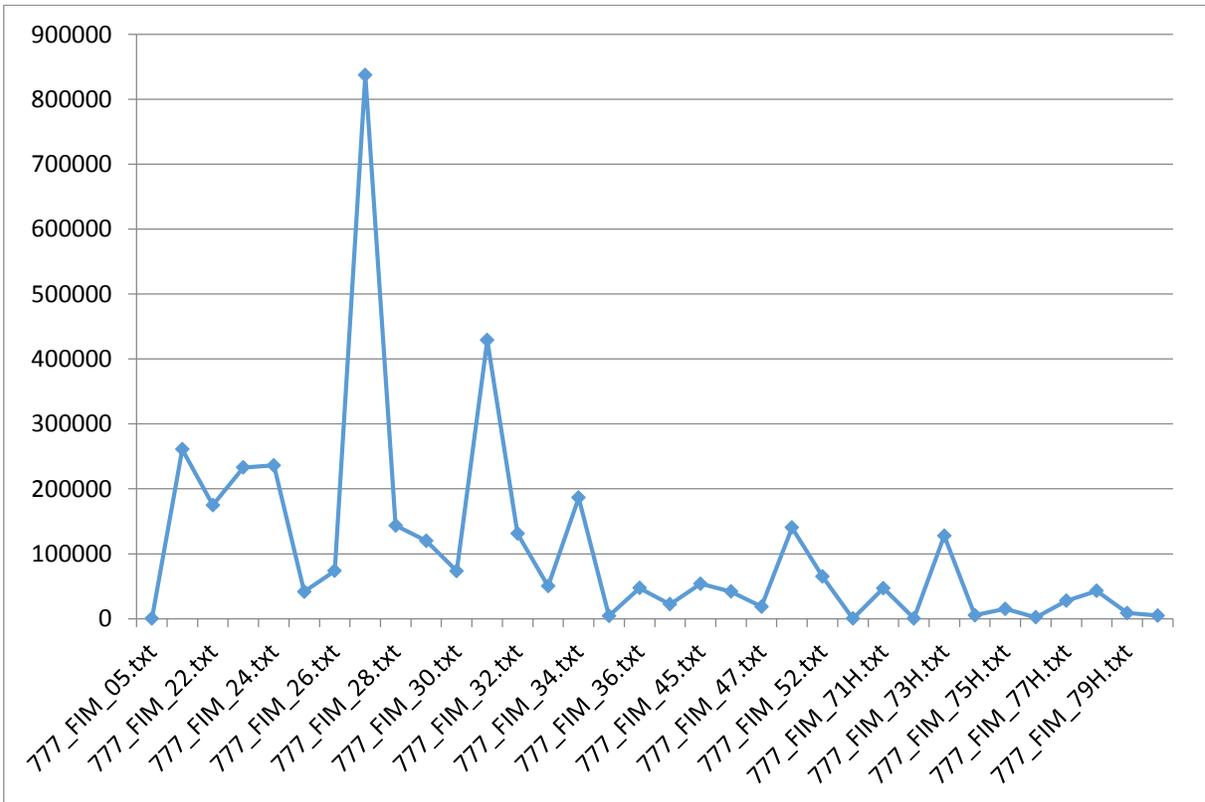
Fabricante: Boeing / Modelo de Aeronave: MD11 / Tipo de Manual: AMM



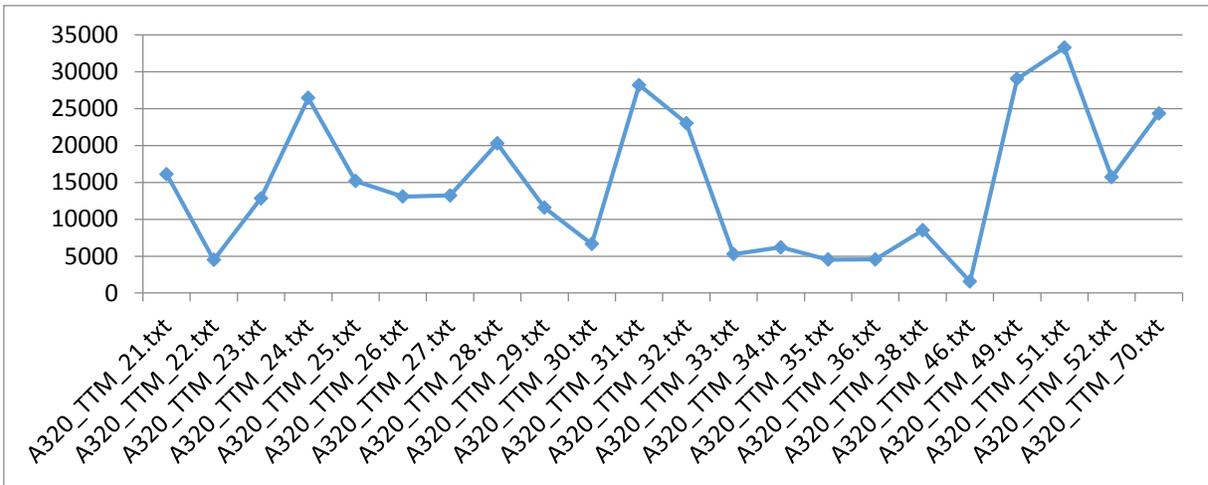
Fabricante: Boeing / Modelo de Aeronave: B777 / Tipo de Manual: AMM PP



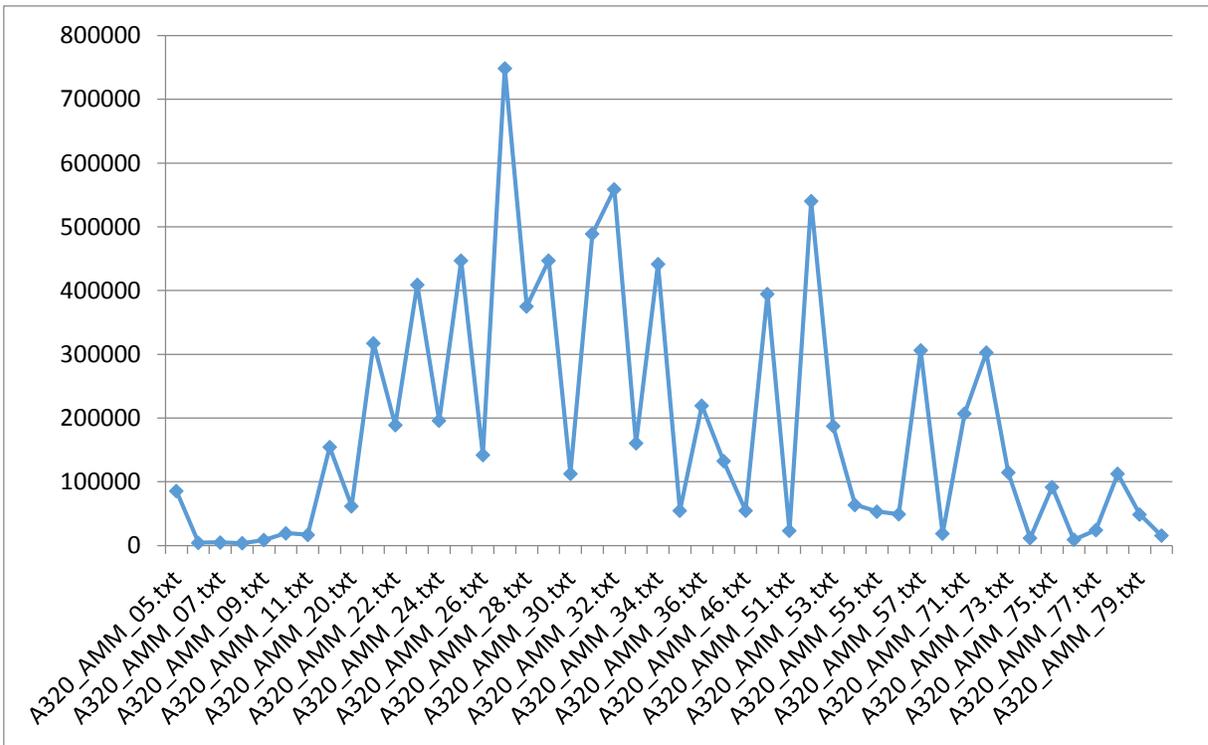
Fabricante: Boeing / Modelo de Aeronave: B777 / Tipo de Manual: AMM SDS



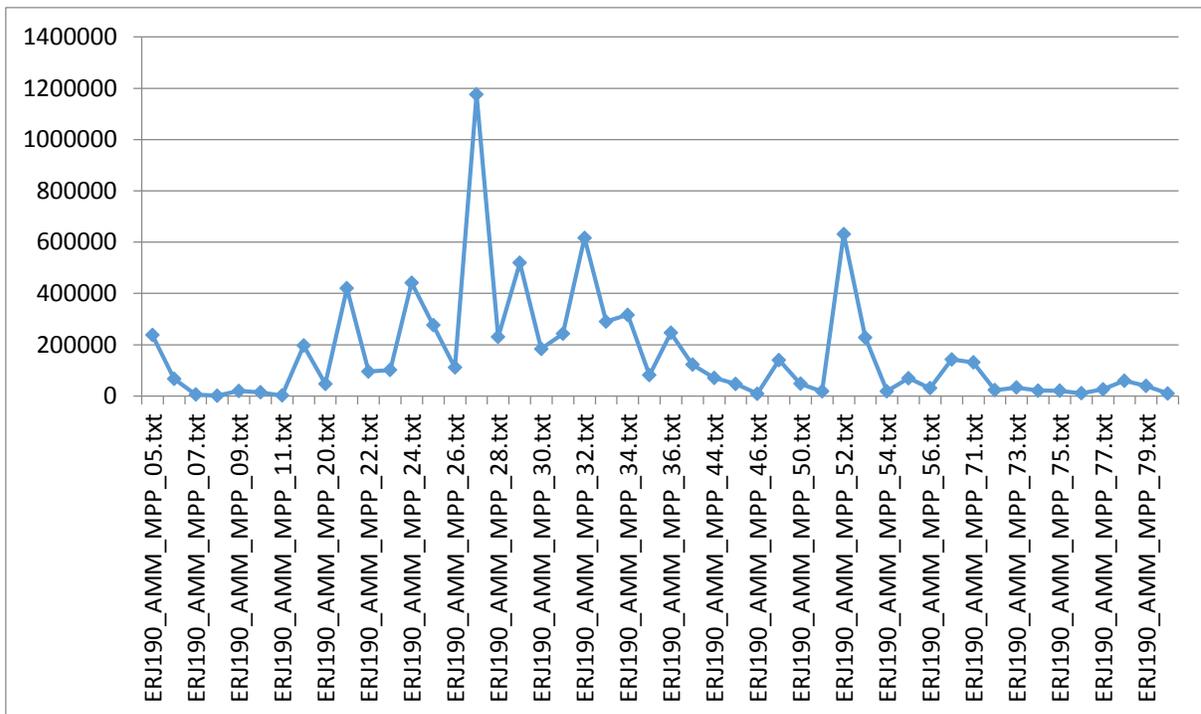
Fabricante: Boeing / Modelo de Aeronave: MD11 / Tipo de Manual: FIM



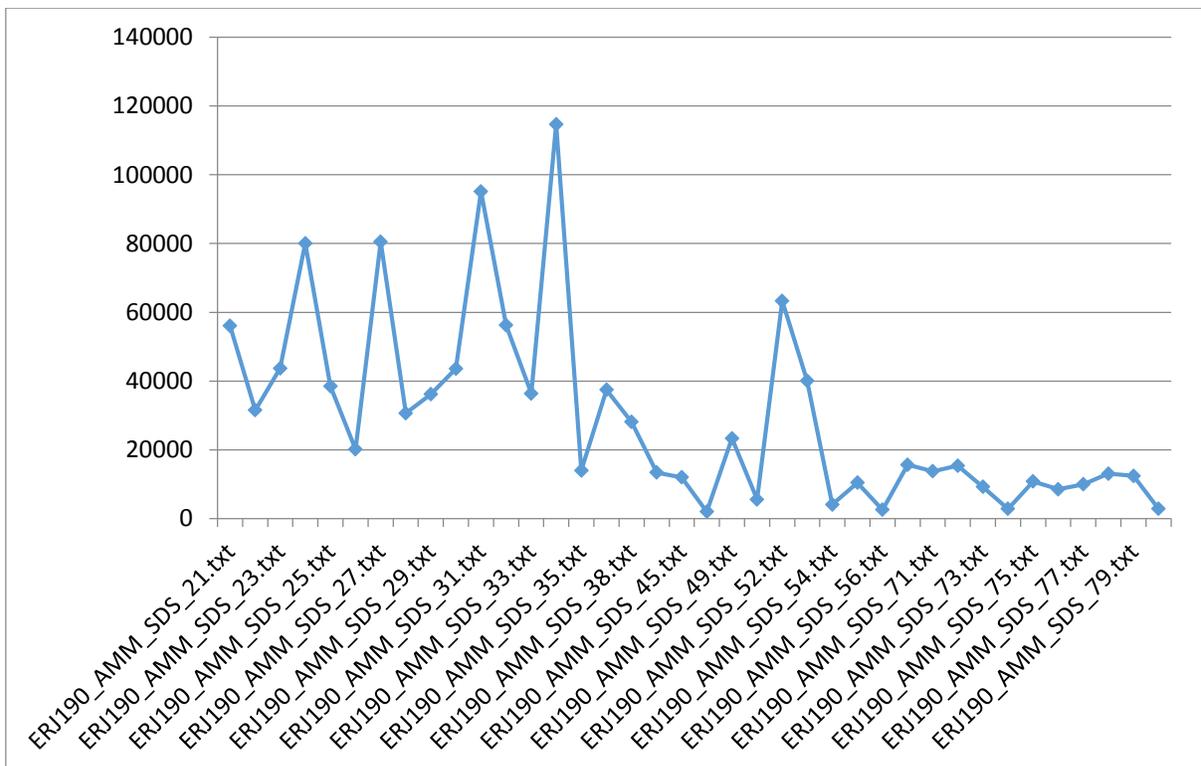
Fabricante: Airbus / Modelo de Aeronave: A320 / Tipo de Manual: TTM



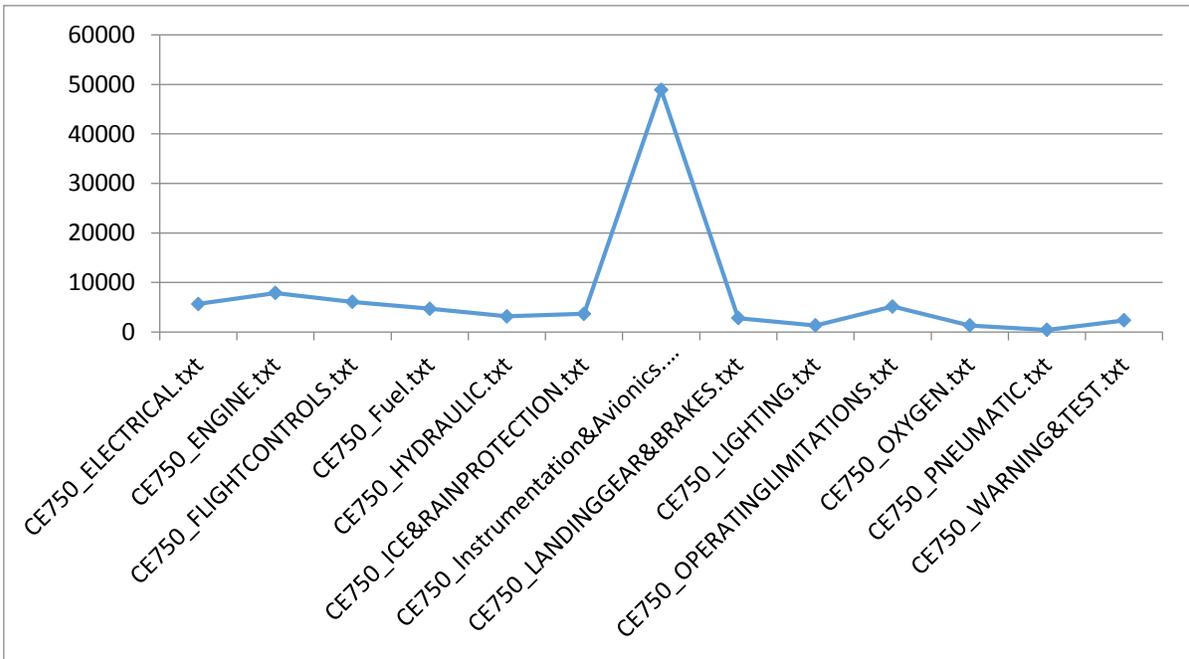
Fabricante: Airbus / Modelo de Aeronave: A320 / Tipo de Manual: AMM



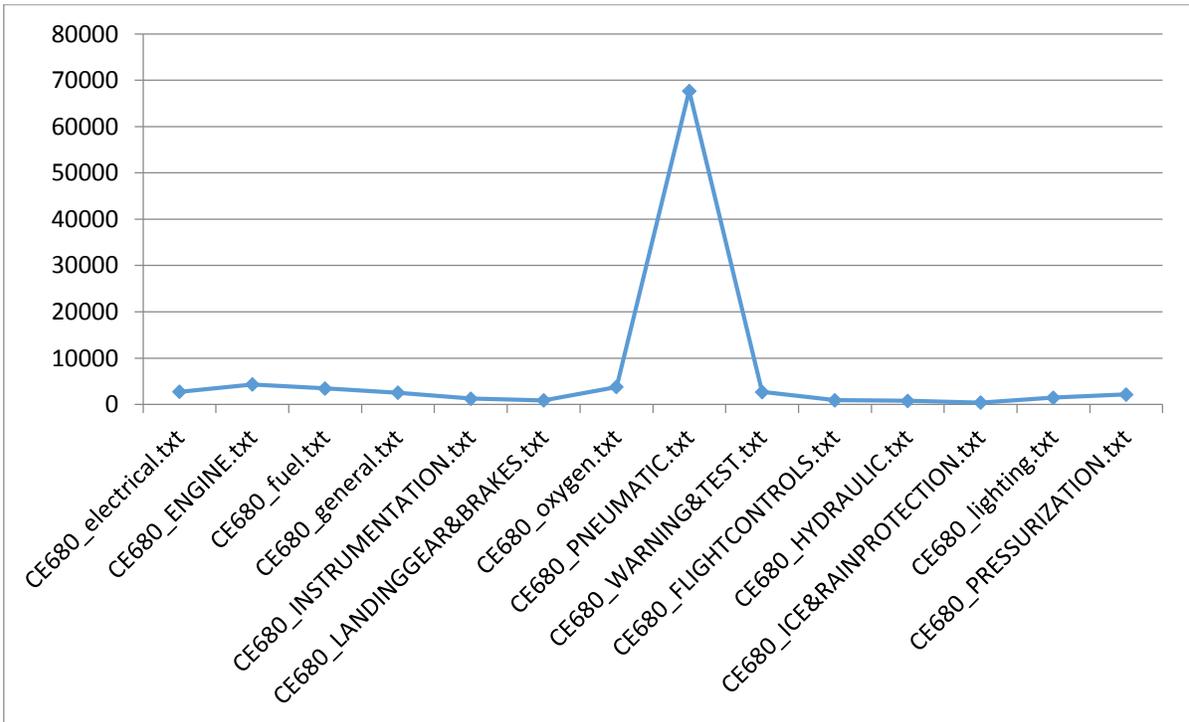
Fabricante: Embraer / Modelo de Aeronave: ERJ190 / Tipo de Manual: AMM MPP



Fabricante: Embraer / Modelo de Aeronave: ERJ190 / Tipo de Manual: AMM SDS



Fabricante: Cessna / Modelo de Aeronave: Citation X / Tipo de Manual: Training Manual



Fabricante: Cessna / Modelo de Aeronave: Citation Sovereign / Tipo de Manual: Training Manual