

UNIVERSIDADE FEDERAL DE MINAS GERAIS

FACULDADE DE LETRAS

Programa de Pós-graduação em Estudos Linguísticos

Luiza Francisca Ferreira da Silva

**A MARCAÇÃO DA QUEBRA DE EXPECTATIVA NO PORTUGUÊS BRASILEIRO
E NO PORTUGUÊS MOÇAMBICANO: protagonismo de *mas* e *só que***

Belo Horizonte

2025

Luiza Francisca Ferreira da Silva

**A MARCAÇÃO DA QUEBRA DE EXPECTATIVA NO PORTUGUÊS BRASILEIRO
E NO PORTUGUÊS MOÇAMBICANO: protagonismo de *mas* e *só que***

Tese de doutorado apresentada ao Programa de Pós-Graduação em Estudos Linguísticos da Faculdade de Letras da Universidade Federal de Minas Gerais, como requisito parcial para a obtenção do título de Doutora em Linguística.

Orientador: Prof. Dr. Eduardo Tadeu Roque Amaral

Área de concentração: Linguística Teórica e Descritiva

Linha de pesquisa: Estudo da Variação e Mudança Linguística

**Belo Horizonte
2025**

S586m Silva, Luiza Francisca Ferreira da.
A marcação da quebra de expectativa no português brasileiro e no português moçambicano [manuscrito] : protagonismo de *mas* e *só que* / Luiza Francisca Ferreira da Silva. – 2025.
1 recurso online (178 f. : il., tabs., p&b.) : pdf.

Orientador: Eduardo Tadeu Roque Amaral.

Área de concentração: Linguística Teórica e Descritiva.

Linha de Pesquisa: Estudo da Variação e Mudança Linguística.

Tese (doutorado) – Universidade Federal de Minas Gerais, Faculdade de Letras.

Bibliografia: f. 151-158.

Apêndices: f. 159.

Anexos: f. 160-178.

Exigências do sistema: Adobe Acrobat Reader.

1. Língua portuguesa – Variação – Teses. 2. Sociolinguística – Teses. 3. Língua portuguesa – Português falado – Brasil – Teses. 4. Língua portuguesa – Português falado – Moçambique – Teses. 5. Língua portuguesa – Conjunções – Teses. I. Amaral, Eduardo Tadeu Roque. II. Universidade Federal de Minas Gerais. Faculdade de Letras. III. Título.

CDD : 469.798



UNIVERSIDADE FEDERAL DE MINAS GERAIS
FACULDADE DE LETRAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTUDOS LINGUÍSTICOS

FOLHA DE APROVAÇÃO

**A marcação da quebra de expectativa no português brasileiro e no português moçambicano:
protagonismo 'de mas' e 'só que'**

LUIZA FRANCISCA FERREIRA DA SILVA

Tese submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em Estudos Linguísticos, como requisito para obtenção do grau de Doutora em Estudos Linguísticos, área de concentração Linguística Teórica e Descritiva, linha de pesquisa Estudo da Variação e Mudança Linguística.

Aprovada em 11 de dezembro de 2025, pela banca constituída pelos membros:

Prof. Dr. Eduardo Tadeu Roque Amaral - Orientador
Universidade Federal de Minas Gerais

Prof. Dr. David Alberto Seth Langa
Universidade Eduardo Mondlane

Prof. Dr. Alexandre António Timbane
Universidade da Integração Internacional da Lusofonia Afro-Brasileira

Prof. Dra. Juliana Bertucci Barbosa
Universidade Federal do Triângulo Mineiro

Prof. Dr. Lorenzo Teixeira Vitral
Universidade Federal de Minas Gerais

Belo Horizonte, 11 de dezembro de 2025.



Documento assinado eletronicamente por **Alexandre Antonio Timbane, Usuário Externo**, em 15/01/2026, às 11:57, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Juliana Bertucci Barbosa, Usuária Externa**, em 15/01/2026, às 13:31, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Eduardo Tadeu Roque Amaral, Subcoordenador(a)**, em 19/01/2026, às 00:32, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **David Alberto Seth Langa, Usuário Externo**, em 03/02/2026, às 01:36, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Lorenzo Teixeira Vitral, Professor do Magistério Superior**, em 03/02/2026, às 10:43, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **4883348** e o código CRC **D8286B89**.

Ao meu Papicito.

AGRADECIMENTOS

Minha eterna gratidão ao professor Eduardo Tadeu Roque Amaral, que me resgatou de onde eu não queria estar e gentilmente se aventurou comigo nos mares da quebra de expectativa.

Meu agradecimento também ao meu amigo João Victor Rocha, por me ajudar com a parte de programação e com as ferramentas tecnológicas.

Obrigada, Tai, por acreditar mais em mim do que eu mesma. Te amo pra sempre!

Obrigada, ainda, queridos familiares, por entenderem minhas ausências e fugas, sem me pressionarem ou cobrarem mais do que eu pudesse dar. Foram dias difíceis e sombrios, SÓ QUE, sobrevivemos!

Amo vocês!

“Todo estudo da língua é um estudo da língua em uso” (Ilari, 1979).

RESUMO

Neste trabalho, analisamos, dentro de uma perspectiva sociolinguística, o fenômeno da quebra de expectativa em duas variedades da língua portuguesa, quais sejam: o português brasileiro (PB) e o português moçambicano (PM). A quebra de expectativa, também chamada de adversatividade na tradição gramatical, refere-se às estruturas coordenadas capazes de cancelar uma pressuposição anteriormente estabelecida, rompendo o que é esperado ou tido como socialmente habitual e/ou previsto (Neves, 2000). É o que ocorre, por exemplo, em “o vizinho fechou a casa, mas os ladrões roubaram a TV” (Perini, 2016, p. 246), em que a segunda oração, que informa sobre o roubo, quebra a expectativa de uma suposta segurança informada na primeira oração, posto que, mesmo fechada, a casa tenha sido roubada. Em nosso estudo, abordamos esse fenômeno dentro do panorama variacionista. Segundo Labov (1978), o fenômeno linguístico variável ocorre quando duas ou mais variantes possuem um mesmo valor referencial/representacional, o que se nota, por exemplo, com a conjunção adversativa *mas*, presente no exemplo anterior, e a locução conjuntiva adversativa *só que*, ambas marcadoras da quebra de expectativa, vista, por exemplo, em “o vizinho fechou a casa, só que os ladrões roubaram a TV”. Buscamos verificar quais são as variantes mais utilizadas para expressar a quebra de expectativa nas duas variedades e a influência de fatores internos e externos na variação em questão. Utilizamos amostras de fala dos *corpora* C-Oral Brasil I (Raso; Mello, 2012), para a variedade brasileira, e Corporaport (Vieira; Pissurno, 2016), para a moçambicana. Como fatores externos à língua, estratificamos tais amostras conforme a faixa etária, o sexo e a escolaridade dos falantes, consoante preveem vastos estudos sociolinguísticos. Como fatores internos, analisamos a posição do marcador de quebra de expectativa dentro do período composto por coordenação e a precedência ou não de outros marcadores adversativos. As ocorrências foram submetidas ao programa estatístico Goldvard 3.0, a fim de obtermos uma análise quantitativa dos dados, o que não nos impediu de estabelecermos também uma análise qualitativa dos números apresentados. Como principais resultados encontrados, observamos um comportamento bastante semelhante para a quebra de expectativa nas duas variedades, com predomínio, por exemplo, das conjunções e locuções conjuntivas adversativas em posição intermediária, conectando as orações. Podemos destacar também o predomínio dos marcadores *mas* e *só que* na expressão da quebra de expectativa tanto em PB quanto em PM. Por fim, caberia expor o maior uso da variante inovadora, *só que*, pelas mulheres; pela faixa etária mediana, dos falantes de 26 a 40 anos; e pelo grupo mais escolarizado, dos falantes de nível universitário, o que poderia indicar um princípio de mudança linguística.

PALAVRAS-CHAVE: Sociolinguística. Variação linguística. Quebra de expectativa.

ABSTRACT

In this work, we analyze, from a sociolinguistic perspective, the phenomenon of the breach of expectation in two varieties of the Portuguese language, namely Brazilian Portuguese (BP) and Mozambican Portuguese (MP). The breach of expectation, also called adversativity in the grammatical tradition, refers to coordinated structures capable of canceling a previously established presupposition, breaking what is expected or considered socially habitual and/or predicted (Neves, 2000). This occurs, for example, in “o vizinho fechou a casa, mas os ladrões roubaram a TV” (Perini, 2016, p. 246), in which the second clause, reporting the theft, breaks the expectation of supposed security conveyed in the first clause, since, even though the house was locked, it was still robbed. In our study, we address this phenomenon within the variationist framework. According to Labov (1978), a variable linguistic phenomenon occurs when two or more variants share the same referential or representational value, as seen, for example, with the adversative conjunction *mas*, present in the previous example, and the adversative conjunctive phrase *só que*, both of which mark a breach of expectation, as in “o vizinho fechou a casa, só que os ladrões roubaram a TV”. We seek to determine which variants are most commonly used to express the breach of expectation in both varieties and the influence of internal and external factors on this variation. We used speech samples from the C-Oral Brasil I *corpus* (Raso; Mello, 2012) for the Brazilian variety and the Corporaport *corpus* (Vieira; Pissurno, 2016) for the Mozambican variety. As external factors, we stratified the samples according to speakers age group, gender, and educational level, as established by extensive sociolinguistic studies. As internal factors, we analyzed the position of the expectation-breaking marker within the compound coordinated clause and whether or not it was preceded by other adversative markers. The occurrences were processed using the statistical program GoldVarb 3.0 in order to obtain a quantitative analysis of the data, which also allowed us to establish a qualitative interpretation of the results. The main results show very similar behavior for the breach of expectation in both varieties, with a predominance of conjunctions and adversative conjunctive phrases in intermediate position, connecting the clauses. We also highlight the predominance of the markers *mas* and *só que* in both BP and MP. Finally, we note a greater use of the innovative variant *só que* among women, speakers aged 26 to 40, and those with a university education, which may indicate the beginning of a linguistic change.

KEYWORDS: Sociolinguistics. Linguistic variation. Breach of expectation.

RESUMEN

En este trabajo analizamos, desde una perspectiva sociolingüística, el fenómeno de la ruptura de expectativa en dos variedades de la lengua portuguesa: el portugués brasileño (PB) y el portugués mozambiqueño (PM). La ruptura de expectativa, también llamada adversatividad en la tradición gramatical, se refiere a las estructuras coordinadas capaces de cancelar una presuposición previamente establecida, rompiendo lo que se espera o se considera socialmente habitual y/o previsto (Neves, 2000). Esto ocurre, por ejemplo, en “o vizinho fechou a casa, mas os ladrões roubaram a TV” (Perini, 2016, p. 246), en la que la segunda oración, que informa sobre el robo, rompe la expectativa de una supuesta seguridad expresada en la primera oración, ya que, a pesar de estar cerrada, la casa haya sido robada. En nuestro estudio abordamos este fenómeno dentro del enfoque variacionista. Según Labov (1978), el fenómeno lingüístico variable se produce cuando dos o más variantes tienen el mismo valor referencial o representacional, como se observa, por ejemplo, con la conjunción adversativa *mas*, presente en el ejemplo anterior, y la locución conjuntiva adversativa *só que*, ambas marcadoras de la ruptura de expectativa, como en “o vizinho fechou a casa, só que os ladrões roubaram a TV”. Buscamos verificar cuáles son las variantes más utilizadas para expresar la ruptura de expectativa en las dos variedades y la influencia de factores internos y externos en dicha variación. Utilizamos muestras de habla de los *corpora* C-Oral Brasil I (Raso; Mello, 2012), para la variedad brasileña, y Corporaport (Vieira; Pissurno, 2016), para la mozambiqueña. Como factores externos al idioma, estratificamos dichas muestras según la edad, el sexo y el nivel educativo de los hablantes, tal como lo prevén numerosos estudios sociolingüísticos. Como factores internos, analizamos la posición del marcador de ruptura de expectativa dentro del período coordinado y la precedencia o no de otros marcadores adversativos. Las ocurrencias se sometieron al programa estadístico GoldVarb 3.0, con el fin de obtener un análisis cuantitativo de los datos, lo que no nos impidió establecer también una interpretación cualitativa de los resultados. Como principales resultados, observamos un comportamiento muy similar en ambas variedades, con predominio, por ejemplo, de las conjunciones y locuciones conjuntivas adversativas en posición intermedia, conectando las oraciones. También podemos destacar el predominio de los marcadores *mas* y *só que* en la expresión de la ruptura de expectativa tanto en PB como en PM. Por último, cabe señalar el mayor uso de la variante innovadora *só que* por parte de las mujeres, del grupo etario medio (26 a 40 años) y del grupo con mayor nivel educativo (universitario), lo que podría indicar un principio de cambio lingüístico.

PALABRAS CLAVE: Sociolingüística. Variación lingüística. Ruptura de expectativa.

LISTA DE FIGURAS

Figura 1 - Construção A mas B, segundo Ilari (2008)	54
Figura 2 - Estilos contextuais hierarquizados por Labov (1966)	71
Figura 3 - Fatores e pontuações que compõem a classe social, segundo Labov (1966)	79
Figura 4 - Percurso histórico da formação da perífrase só que	143

LISTA DE TABELAS

Tabela 1 - Quantificação dos marcadores de quebra de expectativa encontrados nos corpora de PB e PM	114
Tabela 2 - Contabilização dos marcadores adversativos por cada variável no corpus de PB	118
Tabela 3 - Contabilização dos marcadores adversativos por cada variável no corpus de PM	120
Tabela 4 - Ocorrências de mas e só que quanto à faixa etária nas duas variedades	125
Tabela 5 - Ocorrências de mas e só que quanto à escolaridade nas duas variedades	126
Tabela 6 - Ocorrências de mas e só que quanto ao sexo nas duas variedades	126
Tabela 7 - Ocorrências de mas e só que quanto ao corpus	127
Tabela 8 - Ocorrências de mas e só que quanto à posição no período composto nas duas variedades	127
Tabela 9 - Influência do fator faixa etária no uso da variante inovadora só que	128
Tabela 10 - Grupo de fatores e respectivos valores de significância	129
Tabela 11 - Percentuais de mas e só que em relação aos grupos de fatores no corpus brasileiro	131
Tabela 12 - Percentuais de mas e só que em relação aos grupos de fatores no corpus moçambicano	132
Tabela 13 - Mas e só que no PB e no PM	134
Tabela 14 - Comparativo percentual das ocorrências de mas em PB e PM	144

LISTA DE QUADROS

Quadro 1 - Propostas classificatórias para os subtipos de orações adversativas consoante Flamenco García (1999, p. 3855-3872) e Moura (2004, p. 168-169)	31
Quadro 2 - Nomenclaturas para a quebra de expectativa segundo cada autor pesquisado	32
Quadro 3 - Possibilidades sintáticas para a quebra de expectativa consoante Pezatti e Longhin-Thomazi (2008)	39
Quadro 4 - Valores semântico-pragmáticos para a quebra de expectativa, consoante Pezzatti e Longhin-Thomazi (2008)	47
Quadro 5 - Valores semânticos de <i>mas</i> coordenando sintagmas, orações ou enunciados, segundo Neves (2000, p. 755-770)	49
Quadro 6 - Valores semânticos de <i>mas</i> em início de turnos, segundo Neves (2000, p. 755-770)	52
Quadro 7 - _Mostras da fórmula A <i>mas</i> B retiradas de Raso e Mello (2012) e de Vieira e Pissurno (2016)	55
Quadro 8 - Regras do método sociolinguístico segundo Moreno Fernández (1990, p. 27-38)	91
Quadro 9 - Alguns casos de variação linguística entre PB, PM e PE, consoante Timbane (2017)	102
Quadro 10 - Critérios extralinguísticos do C-Oral Brasil I e do Corporaport	111
Quadro 11 - Estratificação adotada neste trabalho	112
Quadro 12 - Marcadores adversativos presentes e ausentes nos <i>corpora</i>	117
Quadro 13 - Sistematização dos marcadores menos frequentes em ambos os <i>corpora</i>	124

LISTA DE SIGLAS

CB: classe baixa

CMA: classe média alta

CMB: classe média baixa

CT: classe trabalhadora

MAP: Amostra Maputo

PB: português brasileiro

PP: português de Portugal

PM: português moçambicano

RMBH: Região Metropolitana de Belo Horizonte

LISTA TIPOGRÁFICA

(1); (2): enumeração de exemplos no corpo do texto e nas notas de rodapé e enumeração de alguns tópicos teóricos.

(I); (II): enumeração de princípios teóricos ou metodológicos utilizados.

(a); (b): enumeração de tópicos teóricos que se subdividem.

Itálico: palavras estrangeiras, palavras em menção ou outros destaques no corpo do texto.

Aspas (“ ”): citações diretas.

/: pausa prosódica não concluída.

//: fim da elocução na transcrição; pausa prosódica concluída.

<>: sobreposição na transcrição.

*: início de turno.

:: prolongamento do som vocálico.

§: itens que não são palavras, como sinais de pontuação e números.

SUMÁRIO

INTRODUÇÃO	16
1 O FENÔMENO DA QUEBRA DE EXPECTATIVA	22
1.1 Aspectos conceituais	28
1.2 Aspectos sintáticos	32
1.3 Aspectos semântico-pragmáticos	39
1.4 Possibilidades de expressão da quebra de expectativa	58
2 A SOCIOLINGUÍSTICA VARIACIONISTA	63
2.1 Sociolinguística: aspectos gerais	64
2.2 Sociolinguística: aspectos terminológicos e conceituais	67
2.3 Sociolinguística: aspectos metodológicos	85
2.4 Variação e mudança linguística	91
3 PROCEDIMENTOS METODOLÓGICOS	94
3.1 Procedimentos metodológicos: aspectos gerais	94
3.2 Algumas considerações teóricas sobre o PB e o PM	99
3.3 Os <i>corpora</i>	107
3.4 Algumas especificidades sobre os marcadores de quebra de expectativa	114
4 RESULTADOS E ANÁLISE DOS DADOS	118
4.1 Resultados iniciais	121
4.2 <i>Entretanto, no entanto e porém</i> nos <i>corpora</i>	122
4.3 Resultados obtidos após rodagem dos dados de ambos os <i>corpora</i>	124
4.4 Resultados obtidos após rodagem dos dados em relação a cada variedade	130
4.5 Análise qualitativa dos dados	133
4.6 Os casos de <i>mas só que</i>	137
4.7 Convergências e divergências entre PB e PM	140
CONSIDERAÇÕES FINAIS	147
REFERÊNCIAS	152
APÊNDICE	159
APÊNDICE 1 – Códigos atribuídos às variáveis analisadas neste trabalho	159
ANEXOS	160
ANEXO 1 – Arquivo <i>res</i> , resultados obtidos após rodagem do Goldvarb 3.03 em ambos os <i>corpora</i>	160
ANEXO 2 – Arquivo <i>res</i> , resultados obtidos após rodagem do Goldvarb 3.0 no <i>corpus</i> de PB	168

INTRODUÇÃO

Neste trabalho, nos propomos a analisar comparativamente o português brasileiro (PB) e o português moçambicano (PM) quanto à quebra de expectativa ou adversatividade. A quebra de expectativa é um fenômeno linguístico que ocorre quando uma segunda informação quebra uma expectativa criada anteriormente, gerando uma ideia de contraste, ressalva, oposição. Utilizando uma metodologia variacionista, temos como objetivo geral ampliar os estudos comparativos das duas variedades linguísticas – sobretudo quanto à quebra de expectativa – e estabelecer aspectos convergentes e divergentes de PB e PM relativamente a esse fenômeno. Como objetivos específicos, almejamos esmiuçar como se dá a marcação da quebra de expectativa explícita – quando aparecem conectores adversativos – e verificar um possível protagonismo dos marcadores adversativos *mas* e *só que* no quadro da quebra de expectativa nas duas variedades, utilizando, para tal, *corpora* de língua falada das duas variedades.

Segundo Neves (2001), os indivíduos são capazes não só de codificar e decodificar expressões, mas também de usar e interpretar tais expressões de modo interacionalmente satisfatório. Assim, tem-se a competência comunicativa, que é, *grosso modo*, a habilidade de utilizar a língua de modo a fazer-se dela um uso social, adequado às necessidades comunicativas. Isso a torna uma atividade cooperativa, executada entre sujeitos inseridos em práticas e comunidades específicas. Assim, veiculam-se informações, conectando a situação de fala, a intencionalidade dos falantes e o próprio contexto linguístico. Nesse sentido, a autora recupera as noções de tema, que seria o ponto de partida da mensagem, e rema, que seria a mensagem em si. Logo, o tema seria a informação dada, compartilhada, enquanto o rema seria a informação nova.

Considerando isso, neste trabalho, nos propomos a analisar o fenômeno da quebra de expectativa, que seria o fato de uma informação nova quebrar o que é esperado em relação a algo anteriormente dito. Logo, na quebra de expectativa, a segunda informação, nova, rompe com o paradigma que é esperado na informação dada primeiramente, gerando uma ideia de contraste, ressalva, oposição. Para a tradição gramatical, esse fenômeno recebe o nome de adversatividade. Vejamos um exemplo:

- (1) O vizinho fechou a casa, *mas* os ladrões roubaram a TV (Perini, 2016, p. 246).

Em (1), a primeira informação está no fato de o vizinho ter fechado a casa. Essa mesma informação pode gerar uma expectativa socialmente partilhada de que uma casa fechada estaria segura. Entretanto, essa expectativa de segurança é quebrada com a segunda informação, sobre o fato de os ladrões terem levado a TV. Isto é, o roubo da TV quebra a expectativa anteriormente posta de uma casa fechada estar protegida. Como dito, essa quebra de expectativa é chamada na tradição de adversatividade, e a oração responsável pela quebra é chamada oração coordenada adversativa, sendo o *mas* a conjunção adversativa. Nesse sentido, para Perini (2016, p. 446), *mas*, chamado por ele de coordenador, tem o papel de “negar uma expectativa”. Para Pezatti e Longhin-Thomazi (2008, p. 929),

embora *mas* e *só que* sejam os conectivos mais frequentemente utilizados para estabelecer relações de adversidade, o português dispõe de um conjunto de elementos e sintagmas de classificação fluida – *contudo, todavia, no entanto, entretanto, porém* – que tangenciam as categorias de advérbio e conjunção (Pezatti; Longhin-Thomazi, 2008, p. 929).

Neves (2000, p. 756) afirma que, nas construções adversativas, os coordenadores evidenciarão “exterioridade” entre dois segmentos. Nesse sentido, os falantes vão valer-se de seus conhecimentos de mundo, das experiências compartilhadas e socialmente normatizadas. Em (1), por exemplo, para alcançar a quebra de expectativa, o locutor precisa lançar mão de conceitos e valores coletivos, como o fato de uma casa fechada inspirar certa segurança e com o fato de uma TV ser um objeto de valor, cobiçado socialmente e, portanto, passível de roubo. Assim, organiza contextualmente as ideias de modo que a segunda informação, a do roubo, rompa a expectativa inicialmente criada sobre a casa fechada estar segura. Nessa relação, Neves (2000) afirma que a primeira informação é mantida, em graus variáveis, para depois ser negada, recusada, contrariada, também em graus variáveis.

Para Pezatti e Longhin-Thomazi (2008), uma construção coordenada compõe-se de dois ou mais membros com equivalência funcional, ligados por mecanismos de conexão, combinados em mesmo nível estrutural. Assim, as estruturas de quebra de expectativa seriam pares, conectados ou não por conjunções ou locuções conjuntivas, como *mas* e *só que*, em que uma ideia contesta a outra. Explicando o conceito das autoras supracitadas, elas consideram que, nas estruturas coordenadas, nenhum dos membros pode estar subordinado aos demais ou dependente em relação a eles. Logo, aplica-se o conceito da equivalência funcional, pois, nesse tipo de construções, os membros devem ter funções semântica, sintática e pragmática idênticas. Vejamos a seguir dois exemplos de nosso *corpora* a fim de ilustrarmos não só a equivalência

funcional das estruturas coordenadas adversativas, sobre as quais nos debruçamos neste trabalho, como também a fim de observarmos alguns usos de *mas* e *só que*, dois marcadores adversativos bastante proeminentes em duas variedades do português, quais sejam: o português brasileiro (PB) e o português moçambicano (PM).

- (2) Tornar essa língua mais poderosa e vamos dominá-la *só que* isto foi se diluindo ao longo do tempo (Vieira; Pissurno, 2016).
- (3) Não pedindo desculpa *mas* assim falando que a próxima vai ser diferente em tais e tais aspectos e tal (Raso; Mello, 2012).

Em (2) e (3), vemos dois pares de informações separados pelos marcadores adversativos *só que* e *mas*, respectivamente. No primeiro exemplo, tem-se a ideia de tornar a língua poderosa e dominá-la e, num segundo momento, essa ideia de dominação é quebrada em função de ter-se diluído com o tempo. Em (3), a primeira informação é a que se nega o pedido de desculpa para, num segundo momento, trazer uma ideia de suavizar tal pedido, assumindo-se o compromisso de que se fará diferentemente em outras oportunidades. Nesse sentido, vemos, em (2), que, por um lado, a ideia da dominação da língua, e, por outro, a ideia da diluição desse projeto são estruturas equivalentes, pois são ambas coordenadas, que não competem entre si, não existe uma superioridade entre os dois membros que compõem o período em questão. Em (3), por sua vez, o não pedido de desculpa equivale em termos estruturais ao compromisso de ser diferente na próxima vez. Mais uma vez não se nota uma desproporção, mas sim um equilíbrio entre os pares.

Coadunando os autores supracitados, Vitral (2017), define que a adversatividade ocorre quando “a segunda oração da coordenação exprime um contraste ou uma quebra de expectativa em relação ao que era previsto acontecer na primeira oração” (Vitral, 2017, p. 340). Considerando, por exemplo, o período composto “Nós fomos ver o filme *A ilha do medo*, mas não conseguimos entrar”, o autor explica que, com a primeira oração, a expectativa é que os falantes tenham ido ao cinema e efetivamente tenham visto o filme. Mas, com a segunda oração, há a quebra dessa expectativa, a partir do momento em que se informa que os falantes não conseguiram entrar na sala de exibição.

Outro ponto importante de nossa análise é o que expusemos anteriormente sobre o fato de as estruturas coordenadas adversativas poderem vir ou não conectadas por meio de mecanismos explícitos de ligação. Isso implica dizer que, não necessariamente, para uma ideia quebrar a outra, é necessário que haja evidente uma conjunção ou perífrase conjuncional explicitando o rompimento de ideias. Conforme salientam Pezatti e Longhin-Thomazi (2008, p. 867), “no processo de coordenação, o mecanismo de ligação pode não estar exposto, e nesses casos a coordenação ocorre por justaposição”. Vejamos, a seguir, mais um exemplo de nosso *corpora*, em que se notam contraste e oposição, elementos típicos da quebra de expectativa, sem necessariamente haver um marcador adversativo explícito.

- (4) Existem aqueles nossos professores antigos... que agora já estão meio velhinhos daquele jeito que nós sabemos que estão/eh:: aquele trabalho é a profissão deles (Vieira; Pissurno, 2016).

Em (4), a quebra de expectativa está no fato de, no primeiro momento, ao se mencionarem os professores “velhinhos”, uma possível expectativa criada é que, pelo fator idade, eles já estarem incapacitados de exercer suas profissões. Entretanto, quebra-se essa ideia da impossibilidade com o fato de afirmar-se “aquele trabalho é a profissão deles”, gerando uma noção de desempenho e determinação por parte desses profissionais. Essa interpretação da quebra de expectativa ancora-se também no fato de que poderíamos ter colocado entre os dois elementos que compõem a ideia global de (4) algum marcador adversativo capaz de evidenciar a quebra por nós identificada, o que geraria algo como “já estão meio velhinhos, mas aquele trabalho é a profissão deles”. Assim, percebemos que em (4) há um exemplo de coordenação por justaposição, conforme Pezatti e Longhin-Thomazi (2008).

Para melhor nos atermos ao fenômeno da quebra de expectativa, neste trabalho, observamo-la em duas variedades da língua portuguesa já mencionadas anteriormente, quais sejam, o português brasileiro (PB) e o português moçambicano (PM). Para tanto, utilizaremos dois *corpora* de língua falada, o C-oral Brasil I (Raso; Mello, 2012)¹, para a primeira variedade; e o Corporaport (Vieira; Pissurno, 2016)², para a segunda. Como metodologia adotada, lançaremos mão da Sociolinguística Variacionista. Dentro dessa perspectiva, como aspectos

¹ Disponível em: <https://www.c-oral-brasil.org/>. Acesso em: 30 jun. 2025.

² Disponível em: <https://corporaport.lettras.ufjf.br/corpora/portugues-mocambique/>. Acesso em: 30 jun. 2025.

extralinguísticos, comparamos o comportamento da quebra de expectativa nas duas variedades quanto aos critérios: idade, gênero e escolaridade, e, como fatores internos, verificamos o fenômeno observando a posição das conjunções e locuções conjuntivas adversativas dentro das sentenças coordenadas e a presença ou não de alguma outra locução ou conjunção adversativa precedendo o marcador observado, como se nota em:

- (5) Eu mudaria muita coisa né *mas só que* você dá a sua opinião e deixa o resto com o cliente que quiser, né? (Raso; Mello, 2012).

Em (5), vemos duas fórmulas de quebra de expectativa ocorrendo na mesma oração coordenada, quais sejam: *mas* e *só que*. Vale lembrar que vamos discutir essas ocorrências mais detalhadamente no capítulo 4 deste trabalho. Cabe-nos, neste momento, expressar que nossas hipóteses iniciais são a de que a presença de duas conjunções seguidas, como exemplificado em (5), mostre que *só que*, por ser uma variante mais recente na língua, ainda não esteja totalmente encaixada no sistema linguístico, precisando apoiar-se em formas mais estabilizadas, como *mas*, ou mesmo que este, por ser mais frequente, esteja se desgastando e apoiando-se em *só que*, segunda variante mais usada. Além disso, conforme exposto, nosso segundo fator linguístico é a posição que as locuções e conjunções adversativas ocupam dentro do período composto. Nossa hipótese é que as locuções e conjunções tendam a aparecer mais no meio do período, ligando as duas orações, pois, tratando-se de orações coordenadas, em que há independência sintática, esperamos que os coordenadores apareçam em sua maioria centralizados, de modo a conectar as ideias contrapostas, evidenciando a quebra de expectativa. Inclusive, para Perini (2016, p. 446), “os coordenadores ligam dois constituintes da mesma classe, formando um constituinte maior, igualmente da mesma classe”. Ora, se os coordenadores têm o papel de ligar, é de se esperar que apareçam, em sua maioria, numa posição central, de modo a enlaçar as orações coordenadas.

Por fim, quanto aos fatores externos – as variáveis sociais idade, gênero e escolaridade – justificamos o uso dessa metodologia embasados sobretudo na tradição laboviana, que apresenta vastos estudos que se utilizam desses três critérios extralinguísticos. Feitas, pois, essas considerações introdutórias, vejamos, a seguir, no capítulo 1, intitulado “O fenômeno da quebra de expectativa”, como se dá a quebra de expectativa ou adversatividade na língua portuguesa. Cabe explicitar, inclusive, que este trabalho compõe-se também do capítulo 2 “A Sociolinguística Variacionista”, no qual dissertamos sobre aspectos gerais ligados à da Teoria

da Variação e Mudança; do capítulo 3 “Procedimentos metodológicos”, em que expomos a metodologia utilizada em nossa pesquisa; do capítulo 4 “Resultados e análise dos dados”, no qual nos propomos a debater os dados de fala que compõem os *corpora* e, finalmente, das “Considerações finais”, em que tecemos algumas conclusões.

1 O FENÔMENO DA QUEBRA DE EXPECTATIVA

A quebra de expectativa é um fenômeno produtivo nas línguas em geral. Trata-se de uma manifestação linguística em que o que é dito de um lado é imediatamente quebrado, cancelado em outro. Isso se dá, dentre outros fatores, pelo fato de os falantes compartilharem aspectos comuns na sociedade, o que se reflete também na linguagem que usam. Esse pressuposto coaduna-se, por exemplo, com a noção de que as manifestações linguísticas podem ser influenciadas por fatores sociais. Nesse sentido, o comportamento linguístico pode ser determinado pela identidade social dos indivíduos. Isso sugere que, quanto mais semelhantes sejam as características sociais de dois ou mais sujeitos, mais próximo será o uso que fazem da linguagem que os permeia (Castro; Muñoz; Esquivel, 2016, p. 40).

Para Heine, Claudi e Hünne Meyer (1991), a noção de expectativa presente nas línguas está no fato de os falantes diferenciarem as normas socialmente compartilhadas e aceitas das normas que se desviam do socialmente aceito. Nesse sentido, o interlocutor precisa marcar aquilo que crê que se oponha ao que o ouvinte espera como socialmente compartilhado e admitido. Nesse sentido, a diferença está entre o que é dito, de um lado, e o que é considerado normal, de outro; entre o que é afirmado e o que é assumido como norma. Vejamos um exemplo, de própria autoria:

(6) Eu fui à festa, *mas não me diverti*.

Em (6), quando se menciona a festa, a expectativa criada e socialmente compartilhada é que, em geral, festas são ambientes de diversão e descontração, portanto, quem as frequenta normalmente se diverte. Entretanto, quando se afirma, por outro lado, não ter se divertido nesse ambiente, quebra-se o esperado, a expectativa levantada com a associação entre festa e diversão é rechaçada.

Halliday e Hasan (1995 [1976]), ao refletirem sobre a coesão, consideram-na sob três vieses: o do significado (nível semântico), o da forma (nível léxico-gramatical) e o da expressão (nível fonológico-ortográfico). Desse modo, para os autores, os significados são realizados como formas e as formas são recodificadas como expressões, resultando num processo múltiplo e interdependente em que o significado é colocado na palavra e a palavra é colocada no som ou na escrita. Assim, a interpretação e o entendimento dos elementos constituintes dos textos,

sejam eles falados ou escritos, são interdependentes. Os pressupostos envolvidos, pelo lado de quem produz o texto e pelo lado de quem o recebe, integram-se no resultado final. Para esclarecer a noção de pressuposto, os autores apresentam a situação a seguir: “imagine que encontremos a seguinte instrução num livro de culinária: ‘lave e tire o caroço de seis maçãs cozidas. Coloque-as em um prato à prova de fogo’” (Halliday; Hasan, 1995 [1976], p. 2). Explicitando a estrutura anafórica, os autores ilustram-na por meio do pronome *as*, que retoma o termo *maçãs*, anteriormente mencionado. É nesse ponto que entra o conceito de pressuposição. Para eles, não basta apenas o uso do elemento de coesão, no caso o pronome *as*. É necessário que haja uma relação entendível entre esse pronome e o termo a que ele se refere. A essa relação chamam pressuposição. Nesse sentido, Halliday e Hasan (1995 [1976], p. 3) afirmam que “não basta que haja uma pressuposição; a pressuposição deve ser satisfeita”, ou seja, plenamente estabelecida e compreendida por quem produz e por quem recebe a comunicação. Logo, podemos definir o pressuposto como uma informação compartilhada e compreendida pelos participantes da comunicação.

Nesse sentido, Dik (1978) afirma que o falante, munido de intenção comunicativa, antecipa a informação pragmática do seu destinatário, formando a expressão linguística. Essa expressão é reconstruída pelo destinatário, que a interpreta. Assim, a comunicação dá-se como uma operação cooperativa e estruturada. Estruturada, porque envolve regras de funcionamento, como o estabelecimento de pressuposições possíveis e aceitas pelos lados envolvidos; e cooperativa, porque se dá por meio da interação social entre duas ou mais pessoas. Conforme sistematiza Neves (2001 [1997]), de um lado, temos: um falante intencionalmente pronto para comunicar-se e informar algo, que antecipa, no ato comunicativo, o que pensa que o destinatário irá interpretar do que diz; e, de outro lado, um destinatário que conjectura sobre essa intencionalidade e a processa à sua maneira. Marcuschi (2007) complementa essa ideia, afirmando que o sentido abstraído de uma comunicação é um fenômeno social, produzido por meio da interação e necessariamente com ela, o que confirma a visão de que só há comunicação se houver também interação, duas ou mais partes envolvidas na construção de significados. Para ele, “a significação é trabalho social” (Marcuschi, 2007, p. 77), que só se dá na presença do outro, pois o ser linguístico passa essencialmente pelo ser social. Cunha (2009) também corrobora essa visão, afirmando que a motivação para os fatos da língua está nos interlocutores, em seus propósitos e no contexto que compartilham. Desse modo, a comunicação e a língua em si não são autônomas, mas sim refletem a adaptação que o falante faz delas para satisfazer as suas necessidades comunicativas.

Ilari (2008) argumenta que um princípio consensual da pressuposição é que ela possui um mecanismo de projeção. Nesse sentido, são inerentes às sentenças às quais se incorporam, expandindo-as. Antunes (2014), defendendo uma linguística interacional, considera que o uso da linguagem é uma ação conjunta, que envolve dois ou mais agentes, que exercem, entre si, mútuas influências. O que resulta desse processo é algo coletivo, construído na troca. Retomando a noção dos pressupostos, anteriormente dita, Antunes (2014) nos conceitua que:

no exercício de nossas atividades comunicativas, vamos pressupondo o que o outro já sabe, ou o que lhe interessa conhecer ou recordar; vamos pressupondo suas posições, seus pontos de vista e, assim, vamos calculando o melhor jeito de o abordar, vamos nos antecipando às suas possíveis discordâncias, e, se for o caso, acrescentando informações, comentários, ressalvas. Vamos, na verdade, tentando ajustar-nos às condições do outro e promovendo o que temos chamado de uma ação conjunta, resultado de uma troca de saberes, de informações, de propósitos e de mútuas influências. [...] De qualquer maneira, com maior ou menor evidência, ação da linguagem, sob esse prisma da interação, supõe encontro, supõe reciprocidade e disposição colaborativa. Daí que, em última instância, toda ação de linguagem é dialógica; quer dizer, é realizada em conjunto ou reciprocamente (Antunes, 2014, p. 19).

Para concluir seu argumento, esse autor defende que, além de a comunicação basear-se na troca e ser construída na interatividade, ela é também ativa, não se resumindo a um mero intercâmbio de ideias. Além disso, implica, de um lado, a disposição para interferir, e, de outro, a disposição para se deixar interferir, o que resulta em total reciprocidade.

Halliday e Hasan (1995 [1976]) também defendem a relevância da interação no processo comunicativo, quando afirmam que há vários estágios envolvidos. Há os significados que cada um dos falantes atribui a si mesmo como participante da situação como um todo, a representação de conteúdos da realidade externa, as representações da própria marca dos falantes na situação, suas escolhas de papel de fala e canal retórico, suas atitudes, seus julgamentos e assim por diante.

Nesse sentido, Burgoon, Denning e Roberts (2002), ao refletirem sobre os aspectos persuasivos da linguagem, afirmam que, sendo a linguagem regida por regras e normas, as pessoas que a usam desenvolvem expectativas e preferências em relação às suas escolhas linguísticas a fim de atenderem às necessidades comunicativas que possuem. Portanto, a noção de expectativa constrói-se também por meio do que é socialmente compartilhado, na interação comunicativa. Nesse sentido, retomemos o que afirmam Heine, Claudi e Hünemeyer (1991). Para eles, a

expectativa linguística constrói-se com base naquilo que os falantes creem que seja compartilhado e socialmente aceito como tal. Entretanto, quando se crê que o dito destoa desse padrão compartilhado, tem-se a quebra de expectativa.

Para Neves (1984, p. 22), a quebra de expectativa expressa uma noção de desigualdade, estabelecendo uma desproporção de ideias entre dois enunciados. Nesse caso, haveria uma relação entre dois elementos desiguais, sendo cada um deles não só externo ao outro – como se dá em todas as cláusulas coordenadas – mas também divergentes entre si. Assim, na quebra de expectativa ou adversatividade, reconhece-se uma entidade para posterior rechaçamento dela. Endossando essa visão, Martelotta (1998, p. 41) afirma que, na comunicação, os falantes lidam com os conhecimentos de mundo, que são compartilhados e de onde emergem normas e expectativas. Nesse contexto, alguns fatores, como “idade, sexo, grupo social, nível de instrução, ou mesmo nível de conhecimento do assunto que está em pauta” são cruciais e determinantes. Portanto, quando uma das partes crê que a outra não compactua com o que é socialmente esperado, recorre às estruturas de quebra de expectativa.

A fim de ilustrarmos nosso raciocínio, consideremos o seguinte exemplo, retirado do *corpus* C-Oral Brasil I (Raso; Mello, 2012), que, como dito na Introdução deste trabalho, representa o falar espontâneo da Região Metropolitana de Belo Horizonte (RMBH):

- (7) GIL: <é> concreto armado / né // aí vai //
 VER: é // *mas aquilo <ali> foi muito <malfeito>* (Raso; Mello, 2012).

Quando, no primeiro momento, menciona-se o concreto armado, a informação socialmente compartilhada e esperada é que se trata de uma estrutura forte e resistente. Entretanto, no segundo momento, quando se menciona que esse mesmo concreto armado foi malfeito, quebra-se a expectativa inicial de ser um material forte e resistente, pois algo malfeito, comumente falando, indica fragilidade e pouca resistência. É nesse rompimento daquilo que é socialmente compartilhado e aceito como esperado, normal, que reside a quebra de expectativa, chamada na tradição gramatical de adversatividade. Nesse quadro normativo, as estruturas de quebra de expectativa são chamadas de adversativas e apresentam noções de “oposição, contraste, ressalva, compensação” (Cegalla, 2005, p. 290).

Importante lembrar que, conforme salienta Rocha (2007), a expectativa é construída pelos falantes e pelo que depreendem da situação comunicativa na qual se inserem. Desse modo, entram na comunicação as concepções subjetivas e coletivas em relação a cada um dos envolvidos no processo. Retomando o exemplo (7), vemos que, na prática, nada impede que uma construção de concreto armado seja frágil. No mundo conversacional, entretanto, esse choque pode se dar consoante as expectativas particulares envolvidas. Sobre isso, Lakoff (1971) considera uma “hierarquia de naturalidade” presente na quebra de expectativa. Nesse sentido, algumas associações seriam mais aceitáveis e universais, como “John é alto, *mas* não é bom no basquete” (itálico nosso); enquanto outras seriam mais idiossincráticas, como “John detesta sorvete, *mas* eu também” (itálico nosso). Por conseguinte, a autora nos incentiva a analisar os exemplos sempre tentando reconstruir possíveis contextos que os satisfaçam minimamente.

Para Halliday e Hasan (1995 [1976]), o significado básico de uma relação adversativa está em contrariar uma expectativa. Esta, por sua vez, pode estar no conteúdo do que se diz ou no próprio processo comunicativo, no contexto a que falante e ouvinte pertencem. Desse modo, para eles, a adversatividade pode se dar no plano interno ou externo. Para exemplificar essa relação no plano externo à língua, os autores dão o seguinte exemplo:

- (8) Todos os números estavam corretos; eles foram verificados. *No entanto, o total saiu errado* (Halliday; Hasan, 1995 [1976], p. 250 – tradução nossa)³.

Em (8), nota-se que a contrariedade dada à expectativa depende da situação em que os falantes se encontram, situação esta relativa à contagem e à verificação dos números sobre os quais se referem especificamente naquele contexto. O significado básico da relação continua sendo o de uma contrariedade à expectativa, mas a fonte dessa expectativa estaria na configuração falante-ouvinte, alcançada no processo de comunicação.

Um possível exemplo para adversatividade interna, dado pelos autores, seria:

- (9) Ela falhou. *No entanto, ela tentou o seu melhor* (Halliday; Hasan, 1995 [1976], p. 252 – tradução nossa)⁴.

³ All the figures were correct; they'd been checked. Yet the total came wrong (Halliday; Hasan, 1995 [1976], p. 250).

⁴ She failed. However, she's tried her best. (Halliday; Hasan, 1995 [1976], p. 252).

Em (9), ainda que o contexto seja importante para o melhor entendimento do que se diz, o contraste explícito entre *falhar* e *dar o seu melhor* é suficiente para evidenciar a quebra de expectativa ou adversatividade.

Sweetser (1991), retomando os estudos de Lakoff (1971), argumenta que as formas linguísticas projetam-se em três domínios da linguagem: (1) o do conteúdo, que se refere ao mundo físico-espacial; (2) o epistêmico, que se refere ao mundo mental, e (3) o conversacional, que se refere aos atos de fala. Dessa forma, para a autora, a quebra de expectativa não se daria no nível do conteúdo, apenas no epistêmico e no conversacional. Isso porque o contraste não se dá devido às expressões linguísticas em si, mas sim devido às informações que elas suscitam nos falantes. Consideremos a seguinte situação dada pela autora:

- (10) John mantém seis caixas de mistura de panqueca estocadas, *mas nunca come panquecas* (Sweetser, 1991, p. 102 – tradução nossa)⁵.

No mundo real, nada impede que uma pessoa estoque panquecas e não as coma, por isso considera-se que a quebra de expectativa não se dá nesse nível, o do conteúdo. O choque se dá porque, dentro do nível epistêmico (mental), não faz muito sentido estocar-se algo sobre o que não se fará um uso. Desse modo Sweetser considera a adversatividade fora do nível do conteúdo. Quanto ao nível conversacional, que também permite a adversatividade, Sweetser nos dá a seguinte situação.

- (11) John é rico, *mas estúpido* (Sweetser, 1991, p. 102 – tradução nossa)⁶.

Se considerarmos um contexto em que mãe e filha conversam sobre casamento, usando (11), a mãe poderia estar sugerindo à filha que não se case com John, pois o considera estúpido. Mais uma vez, vemos a adversatividade fora do conteúdo, pois, na prática, riqueza e estupidez não necessariamente se opõem. A fim de ampliarmos o anteriormente dito, vejamos, na seção a seguir, alguns detalhamentos a respeito das definições do fenômeno em questão, para que possamos desenvolvê-lo a posteriori.

⁵ John keeps six boxes of pancake mix stocked, but never eats pancakes (Sweetser, 1991, p. 102).

⁶ John is rich but stupid (Sweetser, 1991, p. 102).

1.1 Aspectos conceituais

Na gramática descritiva *Nova gramática do português brasileiro*, de Ataliba Castilho, obra em que, nas palavras do próprio autor, procura-se “olhar o que se esconde por trás das classificações, identificando os processos criativos do português brasileiro” (Castilho, 2010, p. 31), as adversativas são denominadas contrajuntivas. Para o autor, trata-se de estruturas em que o que é dito no segundo termo contraria as expectativas criadas no primeiro. É como se a conjunção de contrajunção bloqueasse uma possível união entre os dois elementos encadeados. Isso posto, o autor dá o seguinte exemplo:

(12) Pensei que ia dar certo, *mas me enganei* (Castilho, 2010, p. 351).

Em (12), a ideia de *dar certo* é bloqueada pela ideia de *enganar-se*, havendo, portanto, a contrajunção, a impossibilidade de a expectativa criada efetivar-se. Martelotta (1998), por sua vez, ao desenvolver seu estudo sobre as cláusulas utilizadas por falantes quando estes querem expressar noções opostas entre si, chama esse fenômeno opositivo de cláusulas contrastivas. Para esse autor, haveria, nas cláusulas contrastivas, uma pressuposição que o falante faz a respeito do posicionamento do ouvinte. Ele acrescenta ainda o fato de que nem sempre as expectativas entre falantes e ouvintes são iguais, dada a diversidade de aspectos atuantes, como a idade e o gênero dos interlocutores, o grau de instrução deles, o nível de conhecimento que possuem sobre a pauta em questão, o conhecimento de mundo de que dispõem, as ideologias que carregam etc. No exemplo a seguir, podemos visualizar esse jogo contrastivo:

(13) *ONO: <fortão> igual <ocês> //
 *TIT: <bebe> //
 *TIT: <bebe> //
 *TIT: *mas trabalhador* //
 *ONO: mas n' é [/2] n' é ruim nada / <não> //
 *CAR: <ele> é [/1] ele faz o quê //
 *CAR: qual que é a
 *TIT: soldador //
 *ONO: e' trabalha soldador //
 *ONO: ele é bom // (Raso; Mello, 2012).

Em (13), temos a ideia de beber sendo oposta à ideia de trabalhar. Assim, quando o falante expõe primeiramente o fato de a pessoa a quem se refere ser ébria, cria a expectativa socialmente partilhada de que essa mesma pessoa, por beber em excesso, poderia deixar de cumprir adequadamente suas tarefas profissionais. Entretanto, ao afirmar que é trabalhador, inclusive um bom soldador, quebra essa expectativa inicial, colocando-o como uma pessoa laboriosa e produtiva, apesar de seu vício. Logo, vemos, ilustradamente, a capacidade que os falantes têm de jogar com os conhecimentos de mundo, partilhados ou não, para construir sentenças contrastivas.

Na *Gramática descriptiva de la lengua española*, um projeto coletivo que resultou em 78 capítulos e mais de cinco mil páginas dedicadas a expor o comportamento de categorias gramaticais, suas diretrizes e as relações morfológicas, sintáticas, semânticas e discursivas dessa língua (Soto, 1999), Flamenco García (1999) considera a adversatividade uma relação de coordenação em que dois elementos estão ligados por uma conjunção específica, sendo que esta nos indica a existência de contraste entre tais elementos. Ele separa essa relação de contraste em dois campos, de acordo com o grau total ou parcial expresso pela oposição. Assim, considera dois tipos de adversatividade: a restritiva e a exclusiva. Na restritiva, o segundo elemento limita ou restringe o alcance semântico do primeiro elemento, como em:

- (14) Pepe é feiíssimo, *mas simpático* (Flamenco García, 1999, p. 3855 – tradução nossa)⁷.

Nesse caso, a simpatia não anula por completo a falta de beleza de Pepe, mas sim a suaviza, a atenua, por isso apenas a restringe. Na adversatividade exclusiva, os dois enunciados são vistos como totalmente incompatíveis, de modo que o segundo não apenas reduz o primeiro, mas também o rechaça por completo. Para ilustrar seu raciocínio, Flamenco García (1999, p. 3855) dá o seguinte exemplo:

- (15) A baleia não é um peixe, *mas sim um mamífero* (Flamenco García, 1999, p. 3855 – tradução nossa)⁸.

⁷ “Pepe es feísimo, pero simpático” (Flamenco García, 1999, p. 3855).

⁸ “La ballena no es um pez, sino um mamífero” (Flamenco García, 1999, p. 3855).

Em (15), não é o caso de o mamífero, segundo elemento, reduzir o campo semântico dos peixes, primeiro elemento, mas trata-se de domínios absolutamente diferentes em termos classificatórios. Não há um rebaixamento ou limitação de algum dos termos, há uma completa exclusão entre ambos.

Vale lembrar, por fim, nossa escolha por citarmos essa obra, apesar de ela não abordar especificamente a língua portuguesa. Isso se dá pelo fato de, como dito, tratar-se de um projeto grandioso e muito bem estruturado, cujos autores buscaram aproveitar os avanços de diversos campos linguísticos, como o gerativo, o funcionalista, o cognitivo, o lexical, o semântico-formal e o pragmalinguístico, construindo uma obra em que um mesmo tópico gramatical é abordado sob diferentes perspectivas, na intenção de detalhar o fenômeno e permitir ao leitor uma visão mais integral e complexa do objeto estudado (Soto, 1999). Desse modo, aproveitamo-nos do fato de as línguas portuguesa e espanhola terem a mesma matriz, qual seja o latim, para utilizarmos em nossa pesquisa teórica a *Gramática descriptiva de la lengua española*.

Moura (2004, p. 168), na *Gramática do português actual*, obra que busca ver o português como uma língua românica de comunicação, tendo em conta a inter-relação da linguagem com o agir humano, posiciona a adversatividade no português de Portugal (PP) como as estruturas capazes de exprimir:

- (a) oposição: Tenho relógio, *mas está avariado*;
- (b) retificação: Não pagou, *mas tinha dinheiro*;
- (c) concessão: Ele é meu amigo, *mas seja sportinguista (sic)*⁹;
- (d) neutralização de uma inferência: São gente pobre, *mas vivem com honestidade*;
- (e) viragem no rumo dos acontecimentos: Já sofreu muito, *mas agora vive mais aliviado*.

Vejamos, no Quadro 1, as propostas classificatórias de Flamenco García (1999) e Moura (2004) para os subtipos de orações adversativas.

⁹ Torcedor do time de futebol português Sporting.

Quadro 1 - Propostas classificatórias para os subtipos de orações adversativas consoante Flamenco García (1999, p. 3855-3872) e Moura (2004, p. 168-169)

Autor	Proposta	Exemplo¹⁰
Flamenco García (1999)	Restritiva: o segundo elemento restringe o alcance semântico do primeiro elemento.	Pepe é feiíssimo, <i>mas simpático</i> .
	Exclusiva: os dois enunciados são vistos como incompatíveis.	A baleia não é um peixe, <i>mas sim um mamífero</i> .
Moura (2004)	Oposição: na oposição, os conectores substitutivos de <i>mas</i> são <i>porém</i> e <i>só que</i> .	Tenho relógio, <i>mas está avariado</i> .
	Retificação: na retificação, os conectores substitutivos de <i>mas</i> são <i>contudo</i> e <i>e no entanto</i> .	Não pagou, <i>mas tinha dinheiro</i> .
	Concessão: na concessão, os conectores substitutivos de <i>mas</i> são <i>apesar de</i> e <i>embora</i> .	Ele é meu amigo, <i>mas seja sportinguista</i> .
	Neutralização de inferência: na neutralização de inferência, os conectores substitutivos de <i>mas</i> são <i>todavia</i> , <i>contudo</i> e <i>e apesar disso</i> .	São gente pobre, <i>mas vivem com honestidade</i> .
	Viragem no rumo dos acontecimentos: na viragem do rumo dos acontecimentos, os conectores substitutivos de <i>mas</i> são <i>todavia</i> e <i>no entanto</i> .	Já sofreu muito, <i>mas agora vive mais aliviado</i> .

Fonte: elaborado pela autora, com base em Flamenco García (1999) e Moura (2004).

Também sobre o PE, Matos (2003, p. 566), na *Gramática da língua portuguesa*, obra revista e ampliada para ter “pendor descritivo, estilo menos tecnicista e cobertura linguística ampla” (Mateus *et al.*, 2003), por sua vez, aponta as adversativas – também chamadas por ela de contrajuntivas – como aquelas que “exprimem prototipicamente um contraste entre os membros coordenados”.

A fim de finalizarmos esta seção de aspectos conceituais, reforçamos nosso intuito de buscarmos fazer um percurso de algumas gramáticas das línguas românicas que julgamos serem obras consistentes e robustas, que nos auxiliassem em nossa revisão teórica. A fim de sistematizarmos essa revisão, observemos o Quadro 2, a seguir, no qual sintetizamos as nomenclaturas adotadas sobre o fenômeno da quebra de expectativa pelos autores pesquisados e abordados nesta seção.

¹⁰ Exemplos dos autores.

Quadro 2 – Nomenclaturas para a quebra de expectativa segundo cada autor pesquisado

Autor	Nomenclatura adotada para as orações adversativas
FLAMENCO GARCÍA (1999)	Adversativas
MATOS (2003)	Contrajuntivas ou Adversativas
MOURA (2004)	Adversativas
MARTELOTTA (1998)	Contrastivas
CASTILHO (2010)	Contrajuntivas

Fonte: elaborado pela autora.

Interessante notar uma maior tendência de os autores europeus – Flamenco García (1999), Matos (2003) e Moura (2004) – adotarem uma nomenclatura mais tradicional e ligada ao latim, classificando as orações como “adversativas”, e os autores brasileiros – Martelotta (1998) e Castilho (2010) – adotarem uma conceituação mais atual, denominando as orações de *contrastivas* e *contrajuntivas*, respectivamente, o que pode demonstrar uma tendência mais tradicionalista quanto às denominações por parte dos estudiosos europeus. Neste trabalho, endossamos a terminologia usada pelos estudiosos brasileiros, pois cremos que o prefixo *contra*, usado na denominação de Martelotta (1998) e Castilho (2010) corrobora a definição do fenômeno da quebra de expectativa, que, conforme mencionado anteriormente, prevê oposição e contraste entre duas informações. No Dicionário de dificuldades da língua portuguesa, de Domingos Paschoal Cegalla (2009), o autor define *contra* como “preposição que exprime oposição” (Cegalla, 2009, p. 102). No Novíssimo Aulete dicionário contemporâneo da língua portuguesa (2011), o verbete *contrastivo* é definido como “que contrasta; referente a ou em que há contraste (ideias contrastivas)” (Aulete, 2011, p. 394). Desse modo, cremos que tanto o termo *contrastivo* como o termo *contrajuntivo*, de Martelotta (1998) e Castilho (2010), respectivamente, por possuírem a ideia de contraste e oposição, conforme prevê a quebra de expectativa, estariam mais de acordo com o posicionamento assumido neste trabalho.

Após essa breve definição da quebra de expectativa, observemos, a seguir, alguns fatores sintáticos ligados ao fenômeno.

1.2 Aspectos sintáticos

Para abordarmos a perspectiva sintática da quebra de expectativa, cabe-nos, inicialmente, retomar a necessária preocupação de Ilari (2008) sobre a categoria das conjunções. Para esse autor, uma questão central para analisar essa classe gramatical seria colocar uma “visão crítica sobre um conjunto de fatos que a tradição gramatical nos acostumou a agrupar sob um mesmo

grande título, sem que ficassem claras as regularidades que elas compartilham” (Ilari, 2008, p. 809). Assim, de antemão, parece-nos importante começar esta seção reconhecendo o nosso intuito, coadunando os autores que nela serão citados, de que as conjunções são uma classe complexa, que merece ser estudada com critério, sendo este o objetivo de nosso trabalho também.

Um aspecto sintático característico da adversatividade, apontado por Martelotta (1998), dentre outros autores, é o fato de as sentenças que se encaixam nesse quadro possuírem rigidez posicional, enquanto outras estruturas que também possuem aspecto opositivo, como as concessivas, poderem mover-se no período. Para melhor ilustrar esse posicionamento, Martelotta (1998) dá os seguintes exemplos:

- (16) a. A família, *apesar de ter ficado chocada*, não processou o médico.
 b. *Apesar de ter ficado chocada*, a família não processou o médico.
 c. A família não processou o médico, *apesar de ter ficado chocada*.
 d. A família nem processou o médico, *mas era caso de processo*.
 e. **Mas era caso de processo*, a família nem processou o médico (Martelotta, 1998, p. 44).

Em todos os exemplos, vemos sentenças que expressam ideias contrastivas entre si, pois o choque familiar poderia ser motivo para processar o médico em questão, o que não ocorre, daí, portanto, o rompimento de uma expectativa anteriormente criada. O que se nota, entretanto, é que as orações adverbiais concessivas de (16a), (16b) e (16c) podem mover-se dentro do período, o mesmo não podendo ocorrer com as coordenadas adversativas de (16d) e (16e). Para finalizar esse argumento, Martelotta (1998) afirma ser essa impossibilidade de locomoção uma marca diferenciadora entre coordenação e hipotaxe adverbial.

Matos (2003, p. 553) também aponta a pouca mobilidade desfrutada por estruturas coordenadas adversativas: “as frases coordenadas frásicas não podem ser facilmente deslocadas”. Essa autora ilustra tal rigidez do seguinte modo:

- (17) a. Eles partiram para o Algarve *mas não foram de férias*.
 b. **Mas não foram de férias*, eles partiram para o Algarve (Matos, 2003, p. 553).

Ela acrescenta ainda que essa dificuldade de movimento se verifica não só em orações, mas também em estruturas não frásicas contrajuntivas, como em:

- (18) a. Poucos professores, *mas muitos alunos* compareceram à reunião geral.
 b. **Mas muitos alunos*, poucos professores compareceram à reunião geral
 (Matos, 2003, p. 553).

Analisando os itens adversativos sintaticamente, Castilho (2010, p. 354) considera que os operadores argumentativos de contrajunção têm seu escopo à esquerda, “ligando constituintes da direita para a esquerda, contrapondo o segundo elemento ao primeiro”, como em:

- (19) Ela está lá, *mas não funciona* (Castilho, 2010, p. 354).

Perini (1995, p. 145) julga *porém*, *entretanto* e *contudo* itens adverbiais, por se deslocarem livremente na sentença, contrapondo-os ao coordenador *mas*, que não tem mobilidade.

- (20) a. Titia adormeceu; *porém* *vovó continuou a cantar*.
 b. Titia adormeceu; *vovó, porém, continuou a cantar*.
 c. Titia adormeceu; *vovó continuou, porém, a cantar*.
 d. Titia adormeceu; *vovó continuou a cantar, porém* (Perini, 1995, p. 145).

Para esse autor, ainda, o *mas* não pode coordenar qualquer elemento coordenável. Ele só se liga a adjetivos ou verbos, mas não a SNs. Além disso, só pode coordenar duas orações, e não mais que duas. Vejamos como se dá isso em (21), por exemplo:

- (21) Carolina chegou, desembarcou, *mas não fez declarações* (Perini, 1995, p. 145).

Em (21), o *mas* não é capaz de coordenar as três orações, e sim duas, as mais próximas entre si, quais sejam: *desembarcou* e *mas não fez declarações*. A interpretação semântica de (21) inclusive nos confirma isso, posto que não há quebra de expectativa entre *chegar* e *desembarcar*. Além disso, conforme previsto pelo autor, a adversatividade se dá entre os verbos *desembarcar* e *fazer*, e não em relação a sintagmas nominais. Importante lembrar que Neves (2000) e Pezatti e Longhin-Thomazi (2008) também mencionam a impossibilidade de

recursividade de *mas*, considerando-o restrito a dois segmentos e resultando sempre em uma construção binária.

Endossando Perini (1995), Bagno (2012) também conceitua *porém*, *contudo*, *todavia*, *entretanto* e *no entanto* como advérbios.

Embora a TGP continue a dizer que entre as conjunções adversativas se incluem as formas *porém*, *contudo*, *todavia*, *entretanto*, *no entanto*, os estudos contemporâneos rejeitam essa classificação e incluem esses itens, como fizemos, na classe dos advérbios. Assim procedem, por exemplo, Perini (1996:45), Bechara (1999: 322), Neves (2001: 241), Azeredo (2008: 306) e Castilho (2010: 354). Ao contrário de *mas*, esses itens, por serem advérbios, admitem uma ampla mobilidade no interior da sentença (Bagno, 2012, p. 891, *itálicos* nossos).

Para Neves (2000, p. 756-757), a relação adversativa pode se dar entre (22a) sintagmas, (22b) orações e (22c) enunciados. Por exemplo:

- (22) a. *Ângela riu, fraca mas ostensivamente.*
 b. *O garçom tem cara de mentecapto, mas isto não me afeta grande coisa.*
 c. *Só se enganaria se quisesse. Mas, certamente, não era o seu desejo* (Neves, 2000, p. 756-757).

Sobre a coordenação adversativa entre sintagmas, a autora dá um acréscimo às ideias de Perini (1995), salientando, primeiramente, que, em se tratando de sintagmas nominais, eles só podem ser coordenados por *mas* se o primeiro apresentar uma negação, como em:

- (23) ... não o menino, *mas a mãe* (Neves, 2000, p. 756).

Em segundo lugar, para Neves (2000), a coordenação entre sintagmas pode se dar entre sintagmas oracionais ou não oracionais, desde que os dois elementos coordenados possuam estatuto sintático equivalente, como em:

- (24) *Devem ser preferidas as bananas e as laranjas, ricas em todas as vitaminas, mas que não precisam ser absorvidas em quantidades excessivas* (Neves, 2000, p. 757).

Em (24), vemos que, apesar de *ricas em todas as vitaminas* ser um sintagma não-oracional e *não precisam ser absorvidas em quantidades excessivas* ser um sintagma oracional, ambos possuem o mesmo status adjetival, relacionando-se a *bananas e laranjas*.

Pezatti e Longhin-Thomazi (2008) consideram três possibilidades sintáticas que incluiriam a quebra de expectativa, quais sejam: (1) coordenação simples, (2) coordenação múltipla e (3) coordenação simultânea. Dentro da (1) coordenação simples, haveria, por sua vez, a (1.1) coordenação de sentenças independentes; a (1.2) coordenação de sentenças dependentes; a (1.3) coordenação de termos; a (1.4) coordenação de predicados; a (1.5) coordenação de modificadores e a (1.6) coordenação de operadores. Detalhemos, a seguir, esse primeiro grande grupo, o da (1) coordenação simples.

A (1) coordenação simples ocorre quando um único constituinte “se multiplica” em outros, mantendo as mesmas características. É como se houvesse um espalhamento desse constituinte, posto que as diferentes informações coordenadas ocupam a mesma posição na estrutura da sentença. No caso da (1.1) coordenação de sentenças independentes, esses elementos “multiplicados” são sentenças, como em:

- (25) Eu acho cebola uma coisa imprescindível, *mas acho horrível o gosto puro da cebola em si* (Pezatti; Longhin-Thomazi, 2008, p. 868, exemplo adaptado).

No caso da (1.2) coordenação de sentenças dependentes, as sentenças estão encaixadas num predicado que lhes dá origem, chamado matriz. Esse encaixamento pode se dar por justaposição ou por coordenação explícita. Em (26), a seguir, vemos um exemplo explícito, dado que a conjunção adversativa está expressa.

- (26) Ele vai desejar que o aluno não fique apenas no nível de memorização, *mas que utilize essa informação em outras situações* (Pezatti; Longhin-Thomazi, 2008, p. 869, exemplo adaptado).

Em (1.3) coordenação de termos, é como se o falante replicasse diferentes informações, todas ligadas da mesma forma ao predicado. Um mesmo constituinte é multiplicado. É o que ocorre em (27), a seguir, em que “daquilo que o precede” é uma entidade replicada a partir de “o

aspecto do comer”, ambas ligadas ao verbo *é*. É como se a segunda entidade derivasse da primeira.

- (27) Não é tanto o aspecto do comer, *mas mais daquilo que precede* (Pezatti; Longhin-Thomazi, 2008, p. 869, exemplo adaptado).

A (1.4) coordenação de predicados ocorre quando só os predicados estão coordenados. Trata-se de estruturas com verbos de mesma valência, que selecionam argumentos de igual papel semântico, como *aparecer* e *desaparecer*, que exigem argumentos cujos papéis sejam de processado. Seria o caso de estruturas, como:

- (28) No vídeo, você aparece, *mas desaparece em seguida* (exemplo de própria autoria).

Na (1.5) coordenação de modificadores, há adjuntos coordenados, como em:

- (29) Um prato simples, *mas superdelicioso* (Pezatti; Longhin-Thomazi, 2008, p. 871, exemplo adaptado).

Por fim, na (1.6) coordenação de operadores, dá-se um tipo raro de coordenação, que ocorre entre palavras gramaticais, isto é, entre artigos, numerais, pronomes e preposições. Seria algo como em (30), exemplo de própria autoria, em que os pronomes indefinidos *tudo* e *nada* aparecem coordenados:

- (30) Ele escreveu tudo, *mas nada ao mesmo tempo* (exemplo de própria autoria).

No segundo grupo, o da (2) coordenação múltipla, coordenam-se diferentes constituintes, que podem ser, por exemplo, sujeito e complemento nominal, sujeito e objeto etc. No caso de (31), a seguir, *estilo* e *arte*, sujeitos, coordenam-se a *mundo* e *natureza*, que são objetos.

- (31) O estilo e a arte refletem uma maneira de considerar o mundo, *mas não a natureza* (Pezatti; Longhin-Thomazi, 2008, p. 872, exemplo adaptado).

Para finalizar, em (3) coordenação simultânea, a coordenação se aplica a pares, trios etc., como em (32), exemplo de própria autoria:

(32) Não é por você, *mas é pelo ansioso, mas é pelo desequilibrado, mas é pelo enfermo* etc. (exemplo de própria autoria).

Ponderemos sobre o efeito mais estilístico do que sintático de estruturas como (32). Vale lembrar que esse caso peculiar não contradiz o falar das autoras, quando afirmam que *mas* resulta em construções binárias. Em (32) não se vê tanto uma quebra de expectativa, mas mais um realce presente na repetição da conjunção.

Posto isso, antes de finalizarmos esta seção, relativa aos aspectos sintáticos da quebra de expectativa, vejamos, no Quadro 3, a sistematização dos conceitos de Pezatti e Longhin-Thomazi (2008) para a sintaxe da quebra de expectativa.

Quadro 3 – Possibilidades sintáticas para a quebra de expectativa consoante Pezatti e Longhin-Thomazi (2008)

Tipo de coordenação	Subtipo	Exemplo
Simples: um único constituinte se multiplica em outros.	Entre sentenças independentes: os elementos multiplicados são sentenças.	Eu acho cebola uma coisa imprescindível, <i>mas acho horrível o gosto puro da cebola em si.</i>
	Entre sentenças dependentes: as sentenças estão encaixadas num predicado que lhes dá origem.	Ele vai desejar que o aluno não fique apenas no nível de memorização, <i>mas que utilize essa informação em outras situações.</i>
	Entre termos: o falante replica diferentes informações, todas ligadas da mesma forma ao predicado.	Não é tanto o aspecto do comer, <i>mas mais daquilo que precede.</i>
	Entre predicados: coordenam-se predicados.	No vídeo, você aparece <i>mas desaparece em seguida.</i>
	Entre modificadores: coordenam-se adjuntos.	Um prato simples, <i>mas superdelicioso.</i>
	Entre operadores: coordenam-se palavras gramaticais.	Ele escreveu tudo, <i>mas nada ao mesmo tempo.</i>
Múltipla: coordenam-se diferentes constituintes.		O estilo e a arte refletem uma maneira de considerar o mundo, <i>mas não a natureza</i>
Simultânea: coordenam-se pares, trios etc.		Não é por você, <i>mas é pelo ansioso, mas é pelo desequilibrado, mas é pelo enfermo etc.</i>

Fonte: elaborado pela autora, com base em Pezatti e Longhin-Thomazi (2008).

Após nossas considerações a respeito do campo sintático, analisemos, neste momento, alguns aspectos semântico-pragmáticos da quebra de expectativa.

1.3 Aspectos semântico-pragmáticos

A quebra de expectativa é vista por Neves (2000) dentro de uma noção de desigualdade. Para essa autora, os marcadores adversativos estabelecem uma desproporção de ideias entre dois enunciados. Usando-os, o enunciador reconheceria a desigualdade presente e a consideraria para distribuir as unidades de informação e estruturar sua argumentação. Vejamos um exemplo dessa autora:

(33) Vocês servem mal, *mas a comida é ótima* (Neves, 2000, p. 755).

No exemplo acima, temos duas ideias desiguais, pois a noção de servir mal sugere queda de qualidade, e não a excelência sugerida pela palavra *ótima*. Portanto, a expectativa inicial de perda de qualidade é quebrada pela conjunção *mas* e a oração encabeçada por ela, quando se afirma que “a comida é ótima”. Conforme explica Neves, essas orações foram estruturadas de modo opositivo e assim distribuídas em dois enunciados ligados pela conjunção adversativa *mas*. Para essa autora, as construções adversativas evidenciam exterioridade entre os itens justapostos, em que o segundo elemento é colocado como destoante do primeiro, “especificando-se essa desigualdade conforme as condições contextuais” (Neves, 2000, p. 756). Assim, nota-se que o contexto, a questão pragmática, é fundamental para a existência e a compreensão da quebra de expectativa.

Importante considerar, também, que, nos exemplos dados neste trabalho, o que fazemos é uma tentativa de interpretação do antagonismo presente nas duas sentenças ligadas por adversatividade, sem necessariamente sabermos se a interpretação feita por nós a título de elucidar mais claramente o fenômeno corresponde ao que o falante ou mesmo os teóricos criadores dos exemplos quiseram de fato expressar. Vale lembrar, conforme prevê Neves (2001 [1997], p. 20), que “a relação entre a intenção do falante e a interpretação do destinatário é mediada, mas não estabelecida, pela expressão linguística”. Logo, a interpretação que se faz é apenas parcialmente embasada na informação dada pela expressão linguística, inclusive porque o destinatário já possui um conhecimento prévio, do qual utiliza-se para, por sua vez, fazer a sua interpretação do que lhe é dado, o que é bastante subjetivo.

Nesse sentido, Perini (2016) afirma que, na adversatividade ou quebra de expectativa, objetiva-se negar uma expectativa anteriormente posta. Isso mostra que, quando um constituinte não levanta expectativas, não pode haver adversatividade. Por isso, são possíveis exemplos como:

(34) Essa escola tem curso de Letras, *mas não confere diploma* (Perini, 2016, p. 446).

Nesse caso, é de se esperar que a escola emita diplomas. Se não o faz, quebra-se o anteriormente esperado. Entretanto, consideremos o exemplo a seguir:

(35) *O pai da Elza, *mas o vizinho* (Perini, 2016, p. 446).

Trata-se de um exemplo agramatical justamente pelo fato de não haver uma expectativa levantada na primeira ideia, a do pai de Elza. Como não há expectativa, não pode haver conjunções desse valor, como o *mas*. Daí a impossibilidade da ocorrência de (35).

Pezzatti e Longhin-Thomazi (2008) afirmam que as duas unidades ligadas pela adversatividade são opostas no campo das ideias, mas interdependentes no campo pragmático. Isso permite ao falante criar oposições mais ou menos subjetivas, como vemos em:

(36) Ela está aqui pra se defender / *mas ela <não consegue>* // (Raso; Mello, 2012).

Nesse caso, a oposição parece estar no fato de alguém precisar defender-se, mas não conseguir fazê-lo. Validando o que afirmam Pezzatti e Longhin-Thomazi (2008), por mais que haja uma oposição particular, subjetiva ao interlocutor, as ideias antepostas se complementam no campo pragmático, sendo interdependentes. Nesse sentido, a conjunção adversativa tem o papel de juntar as ideias que se contrapõem. Para as autoras, nas construções adversativas prototípicas, cada um dos elementos que as compõem constitui-se um ato de enunciação. Nesse caso, conforme a definição clássica de Dubois (1995[1973]), o *tema* seria o ponto de partida; e o *rema* o possuidor de maior carga informativa¹¹.

(37) A aula é gravada, *mas as perguntas podem ser feitas* (Pezzatti; Longhin-Thomazi, 2008, 918).

Em (37), o *rema*, que é o fato de as perguntas poderem ser feitas, traz a informação contextualmente nova, que ainda não havia sido considerada. Desse modo, o segundo elemento, ao trazer a quebra de expectativa, torna-se o foco, adquirindo saliência. Ao enunciar a primeira parte da quebra de expectativa, levantam-se várias informações possíveis, uma delas poderia ser o fato de, por a aula ser gravada, não permitir interrupções. Contudo, ao afirmar que as perguntas podem ser feitas, rompe-se a ideia inicial de que não pode haver interrupções. O primeiro enunciado dá, pois, o suporte para o rompimento do esperado. Além disso, o segundo elemento, para acrescentar o novo, precisa retomar de alguma forma o anteriormente dito. Nesse

¹¹ Tema: numa frase assertiva, chama-se tema o constituinte imediato (sintagma nominal) a respeito do qual se diz alguma coisa (Dubois, 1995 [1973], p. 581).

Rema: parte do enunciado que acrescenta algo de novo ao tema (Dubois, 1995 [1973], p. 118).

caso, aulas e perguntas, no mundo pragmático, são ideias que se relacionam, uma geralmente ocorre dentro da outra. Finalmente, o efeito resultante primeiramente da junção e aproximação entre os termos e a posterior dissociação deles será percebido como “negação, desigualdade, contrariedade, rejeição, não realização de uma relação de causa e efeito ou curso inesperado de eventos” (Pezzatti; Longhin-Thomazi, 2008, p. 919).

Perini (1995) considera que a coordenação aproxima-se mais do discurso que da sintaxe, chamando inclusive os operadores dessa relação não de conjunções, mas de coordenadores. Para esse autor,

em muitos casos, a coordenação se aproxima dos fenômenos discursivos, muito menos dependentes da estrutura interna das formas linguísticas e baseados, em vez disso, de preferência em fatores semânticos e cognitivos em geral. [...] Quando duas ou mais estruturas são unidas por coordenação, nenhuma delas exerce função sintática dentro de nenhuma outra (Perini, 1995, p. 143).

Assim, Perini (2016, p. 446) afirma que os coordenadores não “criam” novas categorias a partir de outras, como o fazem as conjunções, mas sim “ligam dois constituintes de mesma classe, formando um constituinte maior”, que mantém a classe originária. Em (37), por exemplo, o coordenador *mas* liga os constituintes de classe nominal *aula e perguntas*, formando a ideia maior de que, embora seja uma gravação, os estudantes podem interagir. Assim, não há a criação de uma nova categoria, e sim a conexão de suas categorias já pré-existentes.

Sobre a nomenclatura coordenadores, adotada por Perini (1995), é interessante retomar as palavras de Ilari (2008). Para ele, as conjunções são uma classe “reconhecidamente heterogênea e difícil de reduzir a uma definição unitária” (Ilari, 2008, p. 809).

Pezzatti e Longhin-Thomazi (2008), por sua vez, classificam as construções adversativas semântico-pragmaticamente em três subtipos principais, embora reconheçam que esses subtipos possuem limite fluido, não havendo entre eles uma separação rígida. Nesse sentido, um mesmo exemplo poderia se encaixar em mais de um grupo. São eles: (a) adversativas de conteúdo; (b) adversativas epistêmicas e (c) adversativas de atos de fala.

Nas (a) adversativas de conteúdo, há uma oposição explícita. E essa oposição se evidenciaria também pela presença, nos dois segmentos coordenados, de termos semanticamente antagônicos, como em:

- (38) Eu adivinho mímica muito fácil // *mas cê nunca vai adivinhar nenhuma minha*
(Raso; Mello, 2012).

Em (38), vemos o contraste entre *eu* e *você (cê)*, que são colocados em dois polos opostos: respectivamente o que tem facilidade em decifrar mímicas e o que tem dificuldade para tanto. Por outro lado, nas (b) adversativas epistêmicas, a adversatividade se dá através do cancelamento das expectativas, das pressuposições que os falantes creem representar o que é adequado no mundo ou o padrão socialmente aceito. É o que vemos, por exemplo, em:

- (39) Bom / <nũ era pa parar ali> / *mas tudo bem* // (Raso; Mello, 2012).

Quando se diz que não era para parar em determinado lugar, espera-se, como ouvinte, que seja uma proibição, algo taxativo. Entretanto, o segundo termo, ao explicitar *tudo bem*, cancela essa expectativa de proibição, revertendo-a.

As autoras subdividem as (b) adversativas epistêmicas em quatro diferentes valores, quais sejam: (ba) refutação, (bb) não-satisfação de condições, (bc) contra-argumentação e (bd) diferença.

Na (ba) refutação, o falante recusa o anteriormente dito e fornece uma nova informação, que julga ser a correta. É o que vemos, por exemplo, em:

- (40) O feijão com arroz lá tem o nome de baião de dois, *mas não é o feijão preto*
(Pezzatti; Longhin-Thomazi, 2008, p. 922, exemplo adaptado).

As autoras afirmam ser algo comum na refutação a presença de certa polifonia, pois um mesmo enunciado pode trazer diferentes pontos de vista, sejam do próprio falante, do senso comum, daquele que crê ser o do interlocutor etc. Em (39), por exemplo, o falante recusa a ideia de o baião de dois, apesar de ser composto por arroz e feijão, possuir feijão preto, tipo mais comum

da região a que pertence. Desse modo, assume implicitamente que esse prato até pode ter arroz e feijão, mas não da forma como se conhece mais comumente em sua região.

Na (bb) não-satisfação de condições, percebe-se que uma condição necessária para a concretude da ideia não é satisfeita. É o que ocorre em:

- (41) Eu gostaria de fazer, *mas não posso* (Pezzatti; Longhin-Thomazi, 2008, p. 922, exemplo adaptado).

Em (41), o poder fazer algo é condição necessária para efetivá-lo. Nesse caso, uma vez que não se pode, a ação não se realiza, quebra-se.

Na (bc) contra-argumentação, toma-se a ideia de que os dois enunciados que compõem a adversatividade conduzem a conclusões opostas. Nesse contexto, a conclusão do segundo enunciado é mais forte que a do primeiro, desfazendo-a. Vejamos um exemplo a seguir.

- (42) Ele é capaz, *mas não gosta de estudar* (Pezzatti; Longhin-Thomazi, 2008, p. 922, exemplo adaptado).

Consideremos o que ocorre em (42). No primeiro enunciado, quando se afirma ser ele capaz, gera-se a ideia de que ele conquistará grandes coisas, por exemplo. No segundo enunciado, quando se diz que não gosta de estudar, gera-se a ideia de que não se concretizarão esses grandes feitos que poderiam ser esperados. Portanto, a conclusão emitida no segundo momento, por não se gostar de estudar acaba por refutar a primeira conclusão, de que ele teria grandes realizações.

Vejamos, pois, neste momento, o último subgrupo, o da (bd) diferença. Nesse grupo, estabelece-se uma comparação de igualdade entre os dois elementos coordenados, para, em seguida, haver um cancelamento dessa similaridade inicialmente posta. Nesse processo, Pezatti e Longhin-Thomazi (2008) consideram que ocorrem três momentos. Primeiramente, o falante recorre a um modelo de informações, composto pela junção de ideias que julga serem compartilhadas por ele e por seu interlocutor, aproximando e comparando essas ideias. Em seguida, traça a diferença entre o que está comunicando e esse modelo anteriormente estabelecido, recorrendo à adversatividade, que traz um elemento novo, presente na ideia que

quer comunicar. Por fim, com a marcação da diferença, anula as expectativas que possam ter sido evocadas no interlocutor. Vejamos o seguinte exemplo:

- (43) A bicicleta é muito semelhante à motocicleta, *só que a unidade motriz dela é o próprio passageiro* (Pezzatti; Longhin-Thomazi, 2008, p. 925, exemplo adaptado).

Nesse exemplo, o falante estabelece a comparação entre a bicicleta e a motocicleta, podendo acessar vários conhecimentos de mundo a respeito das semelhanças entre os dois veículos, por exemplo, com o fato de ambos possuírem duas rodas, ambos terem guidão, dependerem do equilíbrio do condutor para serem conduzidos etc. Após considerar implicitamente esses conhecimentos socialmente compartilhados, o falante desfaz a comparação, introduzindo a informação nova de que, na bicicleta, o passageiro é o responsável pela propulsão.

Analisando ainda o exemplo (43), é interessante notarmos pelas autoras a menção à locução adversativa *só que*. A *Gramática do português culto falado no Brasil* (Ilari; Neves, 2008), obra em que o capítulo de Pezzatti e Longhin-Thomazi se situa, lança mão de amostras de língua falada do Projeto de Estudo da Norma Urbana Linguística Culta do Português Brasileiro (NURC). Nesse sentido, o uso de amostras reais da fala favoreceu que a variante inovadora *só que* fosse sistematicamente estudada, assim como as demais locuções e conjunções adversativas presentes nas gramáticas tradicionais abordadas na seção 1.1 Aspectos conceituais. Outro autor que também inclui a perífrase *só que* dentro do grupo das conjunções adversativas é Vitral (2017), na *Gramática inteligente do português do Brasil*. Nessa obra, o autor considera que essas conjunções introduzem “a interpretação da ocorrência de algo que é contrário às expectativas” (Vitral, 2017, p. 172), endossando nossa visão sobre o fato de essas conjunções ou locuções conjuntivas romperem com o que é esperado, marcando uma quebra de expectativa.

Após explicitarmos as adversativas de conteúdo (a) e as adversativas epistêmicas (b), que se subdividem nos valores semânticos de (ba) refutação, (bb) não-satisfação de condições, (bc) contra-argumentação e (bd) diferença, passemos agora às adversativas de atos de fala (c). Nelas, parece haver, por parte do falante, uma tentativa prévia de atenuar o que diz, mostrando polidez e cortesia. É o que ocorre, por exemplo, em:

- (44) Não quero ser indelicada, *mas você me deve alguns reais* (Pezzatti; Longhin-Thomazi, 2008, p. 926, exemplo adaptado).

Nesse exemplo, vemos que, antes mesmo de explicitar o aspecto desagradável da dívida em questão, o falante já se antecipa, prevenindo o interlocutor sobre não querer passar uma imagem grosseira. Desse modo, com a adversativa de atos de fala, o falante visa minimizar efeitos negativos da informação veiculada, buscando atenuar tais efeitos. É como se o primeiro enunciado fosse uma “preparação defensiva” para o segundo. E este trouxesse a informação que poderia ser indesejável para o interlocutor. Haveria, logo, um contraste de natureza discursivo-pragmática, em que um dos segmentos seria o atenuante do outro, como se marcasse uma polidez no discurso. É o que vemos em:

- (45) Ô / Marcelo // *cê vai me perdoar / bobo // mas eu nã tô dando <bola pr' ocê não>* // (Raso, Mello, 2012).

Em (45), no primeiro segmento, o falante antecipa uma possível frustração do ouvinte, desculpando-se inicialmente, para só no segundo momento verbalizar que não está *dando bola* para seu interlocutor. Esse contexto anterior em que se antecipa à frustração do outro é chamado “preparação defensiva” por Pezzatti e Longhin-Thomazi (2008, p. 926). Nesse caso, é como se houvesse uma contra-argumentação, em que o falante busca inibir uma resposta negativa de seu ouvinte.

Observemos, a seguir, no Quadro 4, a sistematização dos valores semânticos da quebra de expectativa para Pezzatti e Longhin-Thomazi (2008).

Quadro 4 – Valores semântico-pragmáticos para a quebra de expectativa, consoante Pezzatti e Longhin-Thomazi (2008)

Valor semântico-pragmático	Subtipo	Exemplos
Adversativa de conteúdo: a oposição está explícita.		eu adivinho mímica muito fácil, <i>mas</i> <i>cê nunca vai adivinhar nenhuma minha.</i>
Adversativa epistêmica: há o cancelamento de pressuposições.	Refutação: recusa-se o anteriormente dito e insere-se o novo.	O feijão com arroz lá tem o nome de baião de dois, <i>mas não é o feijão preto.</i>
	Não-satisfação de condições: uma condição necessária para a concretude da ideia não é satisfeita.	Eu gostaria de fazer, <i>mas não posso.</i>
	Contra-argumentação: os dois enunciados que compõem a adversatividade conduzem a conclusões opostas.	Ele é capaz, <i>mas não gosta de estudar.</i>
	Diferença: estabelece-se uma comparação de igualdade que posteriormente é desfeita.	A bicicleta é muito semelhante à motocicleta, <i>só que a unidade motriz dela é o próprio passageiro.</i>
Adversativa de atos de fala: tentativa prévia de atenuar o que diz, mostrando polidez e cortesia.		Não quero ser indelicada, <i>mas você me deve alguns reais.</i>

Fonte: elaborado pela autora, com base em Pezzatti e Longhin-Thomazi (2008).

Halliday e Hasan (1995 [1976], p. 237), explorando a conjunção *but* do inglês, que equivaleria ao nosso *mas*, consideram que ela contém, além do significado adversativo, uma relação aditiva. Para eles, além do significado adversativo, *mas* contém em si também o significado lógico de *e*, funcionando como uma espécie *e entretanto*¹². Daí o fato de usarmos a conjunção aditiva *e* em relações opositivas, como em:

(46) Vou ali, *e (mas)* já volto (exemplo de própria autoria).

Para os autores, o fato de *mas* conter implicitamente um *e* também é observado na impossibilidade de dizermos **e mas*, embora possamos dizer, por exemplo, *e no entanto* (Halliday; Hasan, 1995 [1976], p. 237). Bagno (2012, p. 891) explica essa impossibilidade pelo fato de *mas* ser uma conjunção e as demais, geralmente incluídas nesse escopo, como *porém*, *contudo*, *todavia*, *entretanto*, *no entanto*, serem, como exposto, advérbios. “Outra característica

¹² In addition to the meaning ‘adversative’, *but* contains within itself also the logical meaning of ‘and’ (Halliday; Hasan, 1995 [1976], p. 237).

dos advérbios, além da mobilidade dentro da sentença, é poderem vir antecidos da conjunção aditiva *e* [...]”, o que é impossível para a conjunção adversativa **e mas*.

Neves (2000) também faz um estudo pormenorizado dos valores semânticos adversativos, exemplificando-os por meio da conjunção *mas*. Como se trata de uma teoria robusta e complexa, a fim de melhor visualizarmos esses diferentes sentidos, elucidamo-la no Quadro 5.

Quadro 5 – Valores semânticos de *mas* coordenando sintagmas, orações ou enunciados, segundo Neves (2000, p. 755-770)

Função	Subtipo 1	Subtipo 2	Exemplo
1. Contraposição: o <i>mas</i> admite o elemento anterior, mas a ele se contrapõe.	1.1 Contraposição em direção oposta: o primeiro elemento é superior ou não inferior ao segundo.	1.1.1 Por contraste	O baiano sorria sem arrogância, <i>mas sem o menor temor.</i>
		1.1.2 Por compensação	Eu sei que vai ser difícil dormir hoje, <i>mas pelo menos tente.</i>
		1.1.3 Por restrição	Casou-se, <i>mas não foi com a Luizinha.</i>
		1.1.4 Por negação de inferência	A gente matava a fome, engolindo uma gororoba ruim, preta, <i>mas eu ia.</i>
	1.2 Contraposição em mesma direção: o segundo elemento é superior ou não inferior ao primeiro.		O sertão, para ele, não é uma coisa, <i>mas principalmente uma ideia e um sentimento.</i>
	1.3 Contraposição em direção independente: o segundo membro não foi ainda considerado, surge um novo argumento para consideração.		O assunto é polêmico, <i>mas o importante é deixar claro que toda relação estatística precisa ser discutida à luz de outros conhecimentos.</i>
2. Eliminação: o <i>mas</i> elimina o membro anterior.	2.1 Eliminação no tempo: elimina-se a sequência temporal natural.	2.1.2 Negada a subsequência, nada se põe no lugar.	Pensei em dizer mil coisas, <i>mas não consegui sequer abrir a boca.</i>
		2.1.3 Nega-se a subsequência, havendo recolocação.	Dirige-se ao bispo com os braços abertos, <i>mas o bispo ergue a mão num gesto de desprezo.</i>
	2.2 A eliminação não se refere a uma relação temporal.	2.2.1 O primeiro membro é negado.	Você pensa que sabe, <i>mas não.</i>
		2.2.2 O primeiro membro é visto como não oportuno.	Chego a me perguntar mesmo, <i>mas isso não importa muito aqui nesta conversa.</i>

Fonte: elaborado pela autora, com base em Neves (2000).

Nota-se, no quadro 5, que a autora separa a adversatividade em dois grandes grupos, quais sejam: o grupo da contraposição, em que a conjunção adversativa não elimina o anteriormente dito, mas sim cria uma contrariedade entre os elementos coordenados; e o grupo da eliminação, em que a conjunção é responsável por encadear um segundo elemento que elimina o primeiro. Vale lembrar que essa teorização se aproxima do que argumenta Flamenco García (1999), quando categoriza as adversativas em restritivas – que limitam a ideia anterior – e em exclusivas, que eliminam a ideia anterior.

Vejamos mais alguns exemplos a respeito desse aspecto, em que (47) ilustra o primeiro grande grupo e (48) ilustra o segundo.

(47) Obra que não se escreve com a pena, *mas que se realiza com a luta* (Neves, 2000, p. 758).

(48) Ele abriu a boca para responder à insolência, *mas conteve-se* (Neves, 2000, p. 765).

Em (47), vemos que a ideia de lutar não elimina a teoria, simbolizada pela *pena*, mas sim a confronta. Isto porque se afirma ser a luta a responsável pela obra, e não a pena, isto é, o exercício teórico. Em (48), por sua vez, nota-se a eliminação da resposta, pois há a contenção do gesto, ou seja, a eliminação da possibilidade de contestar-se à insolência mencionada.

Dentro do grupo (1) o da contraposição, conforme vemos no Quadro 5, a conjunção não anula por completo o elemento anterior, e sim o admite explícita ou implicitamente, para em seguida o confrontar. Nesse grupo, os subtipos vão se desenvolvendo no sentido de que mantenham ou não a direção semântica dos argumentos inicialmente colocados ou mostrem argumentos em outros direcionamentos. Além disso, esse grupo avalia a gradação dos argumentos, isto é, se apresentam certa crescente ou decrescente, e se há exclusão total ou parcial entre os membros.

Por outro lado, dentro do grupo (2) o da eliminação, em que a conjunção elimina o membro coordenado anterior, os subtipos vão se desenvolvendo conforme o elemento eliminado pelo segundo membro seja ou não substituído por outro elemento e conforme haja uma relação temporal natural entre os membros coordenados ou não. Nesse sentido, os subtipos avançam ainda caso se coloque um evento que substitua a sequência eliminada, como em:

- (49) Experimentou calcular se estaria perto ou longe daquilo que acontecia em algum lugar. Mas parava, e de novo o silêncio do Sol se refazia e o desorientava (Neves, 2000, p. 765).

Em (49), esboça-se o desejo de avaliar a distância em questão, e essa tentativa é substituída pelo fato de parar-se, que interrompe a possibilidade do cálculo anteriormente colocado.

Dando seguimento ao estudo de Neves (2000) sobre os aspectos semânticos da quebra de expectativa, observemos, a seguir, no Quadro 6, o comportamento da adversatividade em início de enunciados, principiando turnos.

Quadro 6 – Valores semânticos de *mas* em início de turnos, segundo Neves (2000, p. 755-770)

Função	Subtipo 1	Subtipo 2	Subtipo 3	Exemplos da autora
3. Iniciando turnos	3.1 Contraposição: <i>mas</i> se contrapõe ao enunciado anterior.	3.1.1 Em direção oposta: há restrição ao que foi enunciado.	3.1.1.1 Refutação de pressuposto	- Os bichos comem a gente. - <i>Mas a gente não é só isso.</i>
			3.1.1.2 Pedido de informação	Restaria entrar num daqueles botes e remar, içar velas. <i>Mas ir para onde?</i>
		3.1.2 Em direção independente: o argumento anterior, embora admitido, é tido como insuficiente.	3.1.2.1 Sugere-se novo argumento	- Agora, o resto é com vocês. - <i>Mas e se ele não quiser?</i>
			3.1.2 Mudança de foco da narrativa	- Elvira está ótima, não? - Felizmente, <i>mas de que é que estavam falando?</i>
			3.1.3 Introdução de novo tema	Falavam de roupas, de empregadas e do zelo com as crianças. <i>Mas os homens permaneciam no outro canto da sala.</i>
		3.2 Eliminação: <i>mas</i> elimina o enunciado anterior.	3.2.1 Sem recolocação	
	3.2.2 Com recolocação		3.2.2.1 Rejeição do ato de enunciação	- Era como as outras pessoas? - <i>Mas, pelo amor de Deus, minha filhinha, não me faça mais perguntas.</i>
		3.2.2.2 Rejeição do elemento da situação de enunciação	Mal fechei a porta, a dona espantou-se: - <i>Mas o senhor lá fora com um tempo destes!</i>	

Fonte: elaborado pela autora, com base em Neves (2000).

De acordo com o Quadro 6, iniciando turnos, a adversatividade também se dá por contraposição – em que o anteriormente dito não é totalmente eliminado pelo enunciado novo, sendo ideias que apenas se contrapõem – e por eliminação – em que o segundo enunciado suprime o anterior. Dentro desses dois modelos, os subtipos avançam conforme apresentem restrições ou eliminações ao anteriormente dito, requeira-se ou não um novo argumento para consideração, já que o anterior pode ser considerado insuficiente, e haja ou não mudança de foco. Vejamos o exemplo a seguir:

(50) – Terá sido mesmo? *Mas não, não pode ter sido* (Neves, 2000, p. 770).

Em (50), o segundo enunciado elimina o primeiro, não o restringindo, mas o rechaçando por completo. Depois de rechaçá-lo, não há a recolocação de outro enunciado que supra o que foi eliminado.

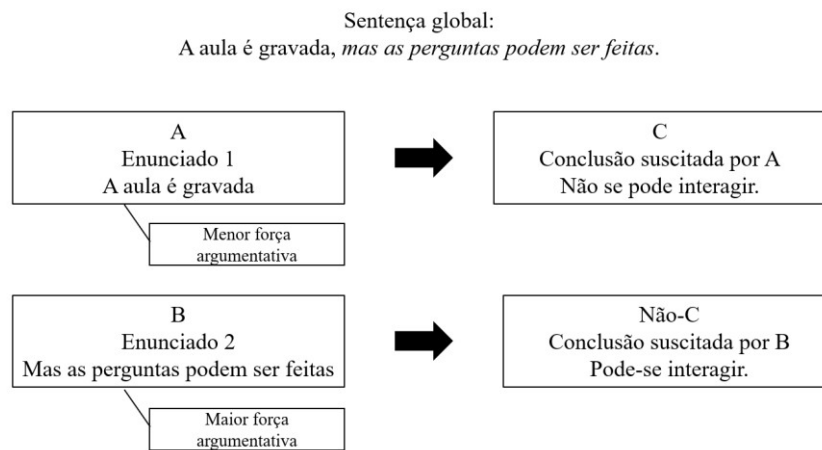
Ilari (2008) considera que, na coordenação adversativa, há a junção de sentenças por meio de uma conjunção, que desempenha a função de apontar a orientação e o peso dos argumentos presentes. As conclusões resultantes desse processo são os objetos de negociação verbal entre os interlocutores. Ilari também usa o *mas* para exemplificar essa argumentação semântica existente na adversatividade. Para ele, a construção *A mas B* expressa um contraste em que *A* e *B* apresentam argumentos que defendem duas conclusões opostas: *C* e não-*C*. Nessa construção, *B* é mais forte como argumento de não-*C* e *A* é mais fraco como argumento de *C*, mesmo que *C* e não-*C* estejam implícitos. Isso nos remete ao anteriormente exposto sobre o fato de as oposições presentes na adversatividade ou quebra de expectativa serem subjetivas e jogarem com pressupostos, que não necessariamente serão verbalizados explicitamente, posto que, muitas vezes, são compartilhados, subentendidos pelos falantes que interagem comunicativamente. O resultado de *A mas B*, como um todo, aponta para não-*C*. Vejamos um exemplo que ilustra essa semântica argumentativa:

(51) *A aula é gravada, mas as perguntas podem ser feitas* e devem... senão fica parecendo monólogo... nenhuma dúvida então? (Ilari, 2008, p. 843).

Em (51), temos como (A) *A aula é gravada* e como (B) *mas as perguntas podem ser feitas*. O fato de a aula ser gravada argumenta a favor de inúmeras conclusões subjetivas e particulares, entre as quais poderíamos citar que (C) sendo a aula gravada, trata-se de um modelo estático e

impermeável, dentro do qual não poderia haver intervenções dos alunos. Por outro lado, a sentença (B) traz (não-C), o argumento a favor de que se deve interagir com o professor por meio de perguntas, pois parece ser essa a solicitação feita pelo docente. Assim, quando considerado no todo, (51) trabalha para não-C, a favor da interação dos alunos com o professor, o argumento oposto a C, que mostra a aula gravada como algo impenetrável e fechado para perguntas. Vejamos essa fórmula de Ilari (2008) ilustrada na Figura 1.

Figura 1 - Construção A *mas* B, segundo Ilari (2008)



Fonte: elaborada pela autora, com base em Ilari (2008).

Observemos, no Quadro 7, mais algumas mostras da semântica argumentativa A *mas* B, presentes nos *corpora* C-Oral Brasil I (Raso; Mello, 2012) e Corporaport (Vieira; Brandão, 2016), objetos de nosso estudo.

Quadro 7 – Mostras da fórmula A *mas* B retiradas de Raso e Mello (2012) e de Vieira e Pissurno (2016)

Período	Enunciado A	Enunciado B	Argumento C	Argumento não-C
Esta geração que nasceu no pós-independência já fala um português língua materna <i>mas com influência grande das línguas moçambicanas.</i>	Esta geração que nasceu no pós-independência já fala um português língua materna.	<i>mas</i> com influência grande das línguas moçambicanas.	A geração pós-independência domina um português padrão.	A geração pós-independência domina um português misto, influenciado por outras línguas nativas.
Estão começando a faculdade aqui, <i>mas vêm de outros lugares.</i>	Estão começando a faculdade aqui.	<i>mas</i> vêm de outros lugares.	Os alunos que frequentam a faculdade local são nativos desse lugar.	Os alunos que frequentam a faculdade local não são nativos desse lugar.
Eu queria ver a comunidade dele, ver que que es falam entre si, <i>mas é fechado.</i>	Eu queria ver a comunidade dele, ver que que es falam entre si.	<i>mas</i> é fechado.	Vou buscar formas de penetrar na comunidade e ver o que falam.	Como a comunidade é fechada, não saberei o que falam.
É triste falar, <i>mas</i> <i>cê fala que é Futebol Arte</i> a galera começa a zoar já.	É triste falar.	<i>mas</i> <i>cê</i> fala que é Futebol Arte a galera começa a zoar já.	Já que é triste essa fala, não a direi.	Direi-a e caçoarão de mim.

Fonte: elaborado pela autora.

No Quadro 7, vemos, nas duas primeiras linhas, amostras do português moçambicano, e, nas duas últimas linhas, amostras do português brasileiro. Vale lembrar que pretendemos detalhar aspectos de cada uma das variedades no capítulo 4 deste trabalho, “Resultados e análise dos dados sobre a expressão da quebra de expectativa no PB e no PM”. Posto isso, outros detalhes importantes devem ser comentados sobre o Quadro 7 neste momento. Primeiramente, convém explicitar que o que estamos chamando de período, na primeira coluna, na verdade trata-se de um fragmento que possua minimamente um sentido completo. Não necessariamente foi retirado da transcrição como um período integral, e sim como um fragmento inteligível. Optamos por não colocar o período integralmente a fim de que o quadro não se estendesse demasiadamente e perdesse a dinamicidade visual, ficando excessivamente carregado de conteúdo textual. Por esse mesmo motivo, também optamos por eliminar as marcas indicativas da transcrição, como //, que significa, por exemplo, o fim da elocução. Como se trata de um quadro para observarmos a fórmula A *mas* B, não cremos que seja significativo manter os símbolos da transcrição,

embora estes sejam relevantes e serão melhor tratados no capítulo 3 deste trabalho, “Procedimentos metodológicos”. Por fim, convém-nos explicitar que as interpretações dos argumentos C e não-C foram por nós elaboradas, e não necessariamente correspondem às interpretações pretendidas pelos falantes no momento da interação. Como dissemos anteriormente, na semântica argumentativa, os argumentos são subjetivos e particulares aos falantes naquele contexto específico de comunicação. Portanto, o que fizemos no Quadro 7 foi uma tentativa de reconstrução, baseados no princípio de que as possibilidades da quebra de expectativas são subjetivas e inúmeras.

Assim como Ilari (2008), Koch (1995) também defende o fato de os elementos ligados pela adversatividade conduzirem a conclusões contrárias. Nesse caso, para a autora, a conjunção colabora para uma espécie de “suspense”. Isso porque o primeiro elemento conduz a um raciocínio, que, por sua vez, é quebrado pelo segundo elemento, rompendo o suspense anteriormente estabelecido.

Kehler (2002), por sua vez, considera dois valores possíveis na adversatividade, também exemplificada por ele através da conjunção *mas*. Segundo esse autor, haveria o *mas* de contraste e o *mas* de expectativa violada. No primeiro caso, pode ser parafraseado por *mas em contraste*; no segundo, pode ser parafraseado por *mas surpreendentemente* ou *mas contra a expectativa*. Vejamos a aplicação desses valores nos seguintes exemplos:

Mas de contraste

- (52) a. A sede do período colonial é a sede também atual... *mas... é um país no:vo...*
 b. A sede do período colonial é a sede também atual... *mas em contraste... é um país no:vo...* (Vieira; Pissurno, 2016).

Mas de expectativa violada

- (53) a. Eu tinha dezoito dezenove anos não sabia distinguir muito bem o que que é o certo o que que é o:/o/o errado... *mas também tem os aspectos bons não é?*
 b. Eu tinha dezoito dezenove anos não sabia distinguir muito bem o que que é o certo o que que é o:/o/o errado... *mas surpreendentemente também tem os aspectos bons não é?* (Vieira; Pissurno, 2016).

Nos exemplos anteriores, o contraste presente em (52) estaria no antagonismo entre o período colonial e o novo país. Em (53), a expectativa violada estaria no fato de o falante, apesar de considerar a própria imaturidade de quando era jovem, podendo gerar ao ouvinte a expectativa de que tenha sido uma fase ruim, na segunda parte da sentença, ressalta o lado bom da juventude, rompendo a expectativa inicial.

Retomando os pressupostos de Lakoff (1971), ele separa a adversatividade, também usando o *mas* como protótipo, em dois domínios: o semântico e o pragmático. Para essa autora, no domínio semântico, haveria uma oposição entre itens lexicais; no pragmático, haveria um contraste entre as expectativas e a realidade. Observemos os exemplos a seguir:

(54) John é rico, *mas* Bill é pobre (Lakoff, 1971, p. 134)¹³.

(55) John é rico, *mas* estúpido (Lakoff, 1971, p. 133)¹⁴.

Em (54), trata-se de um caso de adversatividade de domínio semântico, posto que haveria o antagonismo vocabular entre *rico* e *pobre*. Em (55), não há contraste lexical entre *rico* e *estúpido*, mas sim um rompimento de expectativas. Isso porque, no jogo prático social, espera-se que a pessoa rica acesse uma boa instrução, não sendo estúpida. Como no exemplo ocorre o contrário, houve a adversatividade de domínio pragmático, em que o contraste se dá muito mais pelas interpretações sociais feitas do que pelo léxico presente. Interessante notar que essa autora reconhece que a divisão da adversatividade nesses dois campos não é rígida, posto que os domínios podem se sobrepor. Para tanto, ela dá a seguinte explicação. Consideremos duas pessoas, Maria e Joana, na frase a seguir: “Maria é gorda, mas Joana é magra”. A princípio pode se tratar de um caso de nível semântico, dada a oposição entre *gorda* e *magra*. Entretanto, se Maria e Joana forem irmãs e os falantes souberem dessa informação, pode se tratar de um caso de domínio pragmático, pois, sendo ambas da mesma família, espera-se que possam compartilhar características físicas comuns. Isso não ocorrendo, há a quebra de expectativa pragmática.

¹³ John is rich, but Bill is poor (Lakoff, 1971, p. 134).

¹⁴ John is rich but stupid (Lakoff, 1971, p. 133).

Outro exemplo de adversatividade semântica pode ser visto em Neves (2000), que, como discutido, particulariza a relação de desigualdade existente entre as sentenças coordenadas usando a conjunção *mas*. Para essa autora, a conjunção *mas* seria capaz de demonstrar que o segundo elemento contraposto no encadeamento da adversatividade é evidentemente diferente do primeiro, e a conjunção seria a responsável por expor essas diferenças contextuais. Para ilustrar seu raciocínio, Neves (2000) expõe o seguinte exemplo:

(56) Vou bem, *mas* você vai mal (Neves, 2000, p. 758).

Nesse caso, bem e mal são ideias antagônicas, e a conjunção *mas* ressalta a contraposição entre os dois lados.

Após essa exposição sobre alguns aspectos semânticos da quebra de expectativa, vejamos, a seguir, algumas das diferentes possibilidades de expressão desse fenômeno.

1.4 Possibilidades de expressão da quebra de expectativa

Pezzatti e Longhin-Thomazi (2008) afirmam ser a conjunção *mas* a adversativa prototípica do PB (Pezatti; Longhin-Thomazi, 2008, p. 913). Essa prototipicidade também é vista em Lakoff (1971), como vimos na seção anterior e em Bagno (2012, p. 1007), quando afirma que “a única conjunção coordenativa adversativa é *mas*”, sendo os demais itens tradicionalmente elencados como tais - *porém, contudo, todavia, entretanto, no entanto* – considerados por ele advérbios, sem maiores justificativas.

Matos (2003) cita a conjunção *senão* como adversativa e considera *porém, contudo e todavia* conectores não-conjuncionais. Moura (2004), por sua vez, aborda, além de *mas, porém, todavia, contudo, ainda assim, não obstante, no entanto, entretanto, senão que e ao passo que*.

Para Castilho (2010, p. 354), as conjunções adversativas mencionadas nas gramáticas descritivas, como “*porém, contudo, todavia e entretanto*” estariam mais restritas à escrita, sendo rara a sua ocorrência na fala. Segundo esse autor, essas conjunções derivam de processos de gramaticalização ou de sintagmas preposicionados.

Para Pezzatti e Longhin-Thomazi (2008, p. 929), haveria, no português, um conjunto de conectivos geralmente acionados em relações de adversatividade, mas que possuem classificação complexa, tangenciando as classes adverbiais e as conjuncionais, como *contudo*, *todavia*, *no entanto*, *entretanto* e *porém*. Para essas autoras, essa complexidade em encaixar tais palavras e locuções em classes definidas estaria no processo de derivação sofrido por elas. Sendo os advérbios fontes para a maioria das conjunções existentes hoje, é natural que esses conectivos mantenham resquícios tanto dos advérbios que os originaram quanto das conjunções. Por isso, é possível vermos, por exemplo, palavras como *no entanto* serem capazes de coordenar termos, sentenças e parágrafos, como fazem as conjunções, e ainda manterem a mobilidade posicional, típica dos advérbios.

Bagno (2012) faz um interessante apontamento sobre essa falta de rigidez na classificação que envolve as conjunções, e, conseqüentemente, outras classes correlatas.

As conjunções são palavras que, tal como os advérbios e as preposições, constituem o que os linguistas vêm chamando de classe heterogênea, isto é, uma classe formada por indivíduos muito diferentes entre si na forma e na função. Na verdade, essas três classes parecem ser desdobramentos de alguma “sopa gramatical primitiva” e guardam tantas características comuns que resulta difícil separá-las completamente e inclui-las só numa das três categorias. É por isso também que, a cada nova investigação, os estudiosos retiram itens de uma categoria e os levam para outra, devolvem a uma categoria itens que tinham sido retirados dela para compor outra e assim por diante (Bagno, 2012, p. 881-882).

Ducrot e Vogt (1979, p. 318), ao explorar o processo de derivação sofrido por *mas* (*mais*, em francês), justificam o caráter opositivo dessa conjunção pelo fato de ela ser derivada de *magis*, advérbio latino usado na formação do grau comparativo. Tais autores explicam que, sobretudo na comparação por superioridade, eleva-se uma ideia para o conseqüente rebaixamento de outra, o que leva, por sua vez, a uma noção de desigualdade entre ambas. Eles prosseguem seu raciocínio, elucidando a seguinte comparação em língua latina:

(57) *A magis quam B* (A mais que B – tradução nossa).

Nesse caso, vemos que a superiorização de A implica necessariamente a inferiorização de B. E é justamente o advérbio *magis* o responsável por determinar essa ligação comparativa e conseqüentemente opositiva. Assim, por um processo de gramaticalização, o advérbio teria

passado a conjunção, diminuindo gradativamente a noção comparativa para assumir a opositiva, no seguinte processo descrito por Neves (1984, p. 23):

desigualdade pouco caracterizada (grau comparativo) → contraste → contrariedade →
oposição → negação → anulação → rejeição.

Bagno (2012, p. 890) considera que, nesse processo sofrido por *magis* até chegar ao *mas* adversativo que temos hoje, a comparação, inicialmente de superioridade, teria se expandido também a uma comparação de qualidade: “se x é mais do que y, então existe diferença entre x e y e, por conseguinte, existe oposição entre x e y”. Assim, a partir do momento em que a comparação explicitou inclusive uma diferenciação, *mas* foi absorvendo valor contrastivo.

Vejam os um exemplo do português arcaico, dado pelo autor, que o adapta ao PB contemporâneo, a fim de ilustrarmos esse processo:

(58) Nosso senhor não quer que os pecadores se percam, *mais que façam penitência*
(Bagno, 2012, p. 890, exemplo adaptado).

Temos, pois, duas ideias: a primeira é que Deus não quer que os pecadores se percam; e a segunda é que Ele quer que se penitenciem. A comparação de superioridade poderia ser vista no fato de que, *mais* do que não querer que se percam, Deus exige a penitência dos fiéis. Desse modo, se existe comparação, é porque também existe diferença entre os elementos comparados. Assim, a segunda ideia impõe-se sobre a primeira, gerando a contrajunção como a vemos hoje.

Além dos vocábulos e expressões de classificação fluida, que transitam entre os advérbios e as conjunções, como *porém* e *contudo*, e os processos de derivação experimentados por esses vocábulos, conforme descrito brevemente acima, haveria ainda, conforme Perini (1995, p. 143), a possibilidade de as relações coordenadas poderem se dar sem nenhuma marca explícita, simplesmente por meio da justaposição de sentenças. “Nesses casos, coloca-se sempre algum sinal de pontuação, como a vírgula”. Esse fenômeno é chamado pela tradição gramatical de orações coordenadas assindéticas. Vejam os o seguinte exemplo:

(59) Alguns animais iam hibernar, outros imigravam para lugares mais quentes
(Pezzatti; Longhin-Thomazi, 2008, 867).

Em (59), notam-se duas ideias sobrepostas. No primeiro momento, pode-se esperar que os animais ficariam em hibernação, mas essa ideia é quebrada com a informação, de que nem todos estariam hibernados e sim que poderiam ter-se deslocado para áreas mais quentes. Assim, nota-se que a quebra de expectativa é de fato um fenômeno bastante rico no português, podendo se dar explícita, por meio de coordenadores, consoante Perini (1995; 2010) ou implicitamente, nas chamadas junções sem marca, denominação também desse autor.

Para concluirmos, pois, este capítulo, cabe-nos recuperar um pouco de nossa Introdução. Como exposto, este trabalho visa analisar a quebra de expectativa, objeto deste capítulo, no PB e no PM. Vale lembrar, primeiramente, que os conectivos de quebra de expectativa aqui mencionados foram: *antes, ao passo que, contudo, e sim, em todo caso, entretanto, mas, não obstante, no entanto, porém, senão, só que e todavia*. Nesse sentido, pretendemos verificar quais as estruturas de adversatividade citadas neste capítulo ocorrem nas duas línguas em análise. Posteriormente, almejamos verificar como se dá a variante zero nas duas línguas, isto é, a justaposição das orações, em forma assindética, conforme Perini (1995). Esse autor afirma ainda que a lista dos coordenadores correntes no PB incluiria apenas os itens *e, ou e mas*, sendo, pois o *mas* o único representante da quebra de expectativa. Para ele, “há vários outros nas listas tradicionais de ‘conjunções coordenativas’, mas seu comportamento é tão variado que prefiro não os incluir aqui” (Perini, 2010, p. 316). Assim, nossa hipótese é a de que *mas* seja a variante mais usada nas duas variedades do português. Entretanto, cabe-nos expor ainda que essas e outras hipóteses serão propostas e analisadas no capítulo 4 deste trabalho, “Resultados e análise dos dados sobre a expressão da quebra de expectativa no PB e no PM”. Outro aspecto relevante dessa discussão e que pretendemos analisar também no capítulo 4 são as estruturas coordenadas pelos conectivos adversativos, observando se estes conectam sintagmas, orações ou enunciados, conforme prevê Neves (2000). Além disso, objetivamos verificar se os demais conectivos de quebra de expectativa compartilham, nas duas variedades em questão, a característica de *mas*, que não possui mobilidade dentro da sentença, conforme afirmado por Martelotta (1998), Matos (2003) e Perini (1995) e conforme vimos na seção 1.2 deste trabalho. Neste momento, competenos, logo, reforçar que a teorização da quebra de expectativa realizada neste capítulo tem como objetivo dar-nos subsídios teóricos para nos atermos especificamente às estruturas linguísticas utilizadas pelos falantes quando pretendem marcar o contraste entre o que é considerado normal e aquilo que fugiria desse padrão esperado, no português brasileiro (PB) e no português moçambicano (PM).

Vejamos, a seguir, nosso capítulo 2, “A Sociolinguística Variacionista”, em que expomos, em linhas gerais, a Teoria Variacionista, que também embasará nossa análise da expressão da quebra de expectativa nos falares da RMBH e moçambicano.

2 A SOCIOLINGUÍSTICA VARIACIONISTA

Sabe-se que as línguas são mutáveis e transformam-se, acompanhando a evolução das sociedades. Inclusive, retomando, ao mesmo tempo, nosso capítulo anterior, em que expusemos o fenômeno descrito pelos estudos funcionalistas como a quebra de expectativa e adentrando mais especificamente neste capítulo, em que dissertaremos sobre a teoria Sociolinguística, podemos recuperar uma fala crucial do programa elaborado por Weinreich, Labov e Herzog (2006 [1975], p. 147). Esses autores associam os dois campos que também tratamos neste trabalho, quais sejam: o Funcionalismo e a Teoria da Variação e Mudança Linguística. Para eles, “a associação entre os pressupostos variacionistas e pressupostos funcionalistas se apoia essencialmente num ponto de partida comum: o de que a língua só pode ser entendida nos seus variados contextos de uso”.

Para Trudgill (2000 [1974]), a linguagem não é simplesmente um meio de comunicar informações, mas sim um modo importante que as pessoas têm não só de se relacionarem entre si, mas também de manter tais relações. Nesse caso, para o autor, o mais relevante não são as palavras usadas numa troca comunicativa, mas o fato de ser possível que duas pessoas totalmente desconhecidas, por exemplo, possam trocar informações entre si e estabelecer comunicação. Isso mostra que existe, pois, uma estreita relação entre linguagem e sociedade. Desse modo, como a sociedade se reflete na linguagem, mudanças sociais podem produzir também mudanças linguísticas. Isso implica considerar a língua como um fenômeno social. Para Trudgill (2000 [1974], p. 21),

um estudo da linguagem sem referência ao contexto social inevitavelmente leva à omissão de alguns dos aspectos mais complexos e interessantes da linguagem e à perda de oportunidades para um maior progresso teórico. Um dos principais fatores que levam ao crescimento da pesquisa sociolinguística foi o reconhecimento da importância do fato de que a linguagem é um fenômeno muito variável que essa variabilidade pode ter tanto a ver com a sociedade quanto com a linguagem. Um idioma não é um código simples e único usado da mesma maneira por todas as pessoas em todas as situações. E os linguistas entendem que é possível e benéfico lidar com essa complexidade (Trudgill, 2000 [1974], p. 21).

Nesse sentido, para Trudgill (2000 [1974]), a Sociolinguística seria, pois, a parte da linguística que enxerga a língua como um fenômeno social e cultural. Como área do conhecimento, a Sociolinguística apresenta um vasto campo teórico e metodológico. Vejamos, a seguir, alguns

aspectos gerais que nos ajudam, inclusive, a defini-la mais claramente, sobretudo dentro da vertente variacionista, foco deste capítulo.

2.1 Sociolinguística: aspectos gerais

Para Wardhaugh (2006 [1986]), a Sociolinguística seria a ciência que se preocupa em investigar as relações entre linguagem e sociedade objetivando o entendimento da estrutura da linguagem e de como as línguas funcionariam na comunicação. Nela estuda-se, pois, “a linguagem e a sociedade para descobrir o máximo que pudermos sobre que tipo de coisa é a linguagem” (Wardhaugh, (2006 [1986], p. 13).

Coulmas (1997) considera a Sociolinguística como a área que investiga de que maneira a estrutura social influencia no modo como as pessoas usam a língua e como as variedades e os padrões de uso se correlacionam com os atributos sociais, como classe, gênero e idade.

Coelho *et al.* (2015), por sua vez, afirmam que a Sociolinguística é o resultado da língua que falamos relacionada à sociedade em que vivemos. Dado isso, é necessário, pois, que falemos em variação linguística. Esse conceito refere-se à possibilidade de haver, numa mesma língua, diferentes formas de se dizer a mesma coisa¹⁵. Portanto:

língua e variação são inseparáveis: a Sociolinguística encara a diversidade da linguística não como um problema, mas como qualidade constitutiva do fenômeno linguístico. Qualquer tentativa de buscar apreender apenas o invariável, o sistema subjacente – valendo-se de oposições como “língua e fala” ou “competência e desempenho” – significa uma redução na compreensão do fenômeno linguístico. O aspecto formal e estruturado do fenômeno linguístico é apenas parte do fenômeno total (Borin, 2010, p. 13).

Corroborando essa visão, Silva (2011, p. 51) afirma que a heterogeneidade a ser sistematizada pela Sociolinguística firma-se na análise das variantes linguísticas. “De forma simples e direta, podemos dizer que o objeto da Sociolinguística é o estudo da língua falada, observada, descrita e analisada em seu contexto social, isto é, em situações reais de uso”. Vale lembrar que questões como os conceitos acima mencionados serão desenvolvidas na próxima seção, intitulada “2.2 Sociolinguística: aspectos terminológicos e conceituais”.

¹⁵ Vale lembrar que o conceito de variação linguística será melhor abordado na seção 2.4 deste capítulo.

A teoria da variação e mudança linguística, também chamada de modelo variacionista, pressupõe a existência de uma linguística social, isto é, um modelo que assume a influência de fatores sociais na determinação de propriedades internas do sistema linguístico. Desse modo, perde-se a visão de que a língua está separada do uso social que dela se faz. Esse modelo de análise linguística é também chamado de “sociolinguística quantitativa”, por operar com números e tratamento estatístico dos dados coletados. Outro nome dado para a área a que vamos nos ater neste capítulo é Sociolinguística Laboviana, que recebe essa designação devido a seu principal expoente, o linguista norte-americano William Labov.

Inicialmente, devemos considerar os pressupostos que norteiam os estudos de variação e mudança linguísticas. O primeiro ponto que se pode destacar é o que William Labov chamou de sistematicidade do aparente caos. Contrariando visões anteriores de que era impossível estudar a fala e suas variações sistematicamente, já que se acreditava numa “variação livre” e imprevisível, Labov demonstrou que era possível “descrever a diferenciação ordenada numa língua que serve a uma comunidade.” (2008 [1972], p. 16). Nesse sentido, na Sociolinguística, não haveria “acidentes” ou incidentes, e sim fatos linguísticos observáveis, previsíveis e sistemáticos. Seria, pois, necessário reconhecer, primeiramente, que as mudanças ocorrem e se associam entre si de uma forma não acidental (Weinreich; Labov; Herzog, 2006 [1975], p. 172).

Coelho *et al.* (2015) inclusive afirmam que o principal foco da área é desvendar as regularidades encontradas, observando o comportamento dos fenômenos variáveis dentro e fora das línguas, em contato com a sociedade que as usa. Nesse sentido, para os autores, não haveria variação “aleatória”, posto que, apesar das mudanças, e dos movimentos que a língua faz e das variedades existentes, os falantes continuam por entenderem-se entre si. Isso prova que existem regras e pressupostos que regem tais movimentos.

Uma das principais contribuições de Labov foi passar a olhar para a variação social, especialmente urbana, já que os estudos anteriores se ocupavam principalmente da variação dialetal, com foco no espaço rural. Além disso, esse autor, que, por sua vez, seguiu os passos de seu professor, Uriel Weinreich, era a visão de que um fenômeno linguístico dizia respeito apenas a questões internas ou estruturais da língua, como se via, por exemplo, na teoria gerativa. Segundo os formalistas, não haveria heterogeneidade a ser sistematizada, posto que o objeto de estudo estaria centrado na competência linguística de um falante-ouvinte ideal, inserido em uma comunidade linguisticamente homogênea (Silva, 2011, p. 50). Nessa concepção, fatores sociais,

como o gênero e a escolaridade dos falantes, estariam fora do escopo da teoria linguística, uma vez que são aspectos externos à língua, e o enfoque estaria na capacidade desse falante ideal de assimilar e delimitar as regras combinatórias e articulatórias de sua língua (Etto; Carlos, 2017, p. 721). Embora não seja este o foco deste trabalho, cabe-nos reconhecer, contudo, o brilhante trabalho de Tarallo (1987), ao aproximar os campos da Variação Linguística e da Teoria Gerativa, afirmando que seria possível e até desejável compatibilizar os resultados obtidos em um ou outro campo, parametrizando-os, embora, claro, possuam domínios e interesses distintos.

Como bem mostrou Labov (2008 [1972]), por exemplo, em seus estudos sobre os falantes da ilha de Martha's Vineyard, no estado de norte-americano de Massachussets, aspectos extralinguísticos, como idade, gênero, ocupação, origem étnica e atitude podem ser determinantes para a escolha dos falantes por uma ou outra variante disponível em sua língua. Nesse caso especificamente, os nativos da língua, ao se sentirem excluídos em sua própria terra por causa da especulação imobiliária e do turismo de massa que se instalou na ilha, marcavam mais fortemente os ditongos /ay/ e /aw/, centralizando-os, em relação aos não nativos. Essa centralização, considerada destoante do inglês padrão, era, pois, um comportamento linguístico demarcador de identidade social. Os falantes que não a utilizavam, privilegiando o inglês padrão, que se aproximava do uso mais comum de veranistas e de pessoas de fora da ilha em geral, geralmente esboçavam seu desejo de deixar Marthas's Vineyard, por considerá-la atrasada e involuída em comparação com outros lugares do país. Isso demonstra, portanto, que a língua é um elemento social, que carrega características externas a si mesma, como fatores sociais e identitários. Meillet (2005 [1967], p. 21) também já havia identificado que a língua está sujeita aos usuários que dela se apropriam e aos fatores externos:

um indivíduo fala de maneiras perceptivelmente diferentes, dependendo de seu estado físico e mental em um determinado momento, das pessoas com quem está falando e do local, tempo e circunstâncias externas (Meillet, 2005 [1967], p. 21).

Após essa breve exposição acerca das questões conceituais que envolvem a Sociolinguística, vejamos, a seguir, alguns termos, igualmente importantes, ligados a essa área.

2.2 Sociolinguística: aspectos terminológicos e conceituais

Nesta seção, vamos discutir alguns termos ou conceitos cruciais para a Sociolinguística. O primeiro deles é a noção de condicionadores. Consoante Coelho *et al.* (2015), os condicionadores seriam as forças internas e externas à língua que determinariam a forma como falamos o que falamos; são impactantes em nossas escolhas linguísticas. Os condicionadores internos são também chamados de linguísticos, e estão restritos à língua. São questões como a posição dos termos na sentença, o ambiente silábico onde ocorre alguma variação, a classe de palavras envolvidas na variável analisada etc. Os condicionadores externos são o que chamamos também de extralinguísticos e referem-se a aspectos exteriores, como os já citados gênero dos falantes, idade, etnia, escolaridade, classe social etc. À junção dos condicionadores linguísticos e extralinguísticos damos o nome de variáveis independentes ou grupo de fatores. Ao fenômeno linguístico em estudo damos o nome de variável dependente.

Segundo Mollica e Braga (2003), chama-se o fenômeno variável de variável dependente, posto que, nesse fenômeno, o emprego de cada uma de suas variantes não é aleatório, mas influenciado por fatores internos e externos à língua, ou as chamadas variáveis independentes. As variáveis independentes são também denominadas grupos de fatores e são os aspectos que exercem pressão sobre os usos, influenciando ou não a frequência de cada uma das variáveis envolvidas na variação em análise. Para as autoras,

no conjunto de variáveis internas, encontram-se os fatores de natureza fonomorfo-sintáticos, os semânticos, os discursivos e os lexicais. Eles dizem respeito a características da língua em várias dimensões, levando-se em conta o nível do significante e do significado, bem como os diversos subsistemas de uma língua. No conjunto de variáveis externas à língua, reúnem-se os fatores inerentes ao indivíduo (como etnia e sexo), os propriamente sociais (como escolarização, nível de renda, profissão e classe social) e os contextuais (como grau de formalidade e tensão discursiva). Os do primeiro tipo referem-se a traços próprios aos falantes, enquanto os demais a características circunstanciais que ora envolvem o falante, ora o evento de fala (Mollica; Braga, 2003, p. 11).

Quanto ao termo variante, por sua vez, nota-se, como dito, que se trata de formas linguísticas diferentes que têm o mesmo valor referencial/representacional, intercambiáveis em um mesmo contexto linguístico. Recuperando a definição clássica e já bastante conhecida, diríamos que as variantes são as “diversas maneiras de se dizer a mesma coisa em um mesmo contexto e com o

mesmo valor de verdade” (Silva, 2011, p. 51). Duas ou mais variantes, posto que têm o mesmo significado, formam exemplos de um mesmo fenômeno, chamado de variável, mais especificamente, variável dependente. Conforme Labov (1978), dois enunciados que se referem ao mesmo estado de coisas com o mesmo valor referencial/representacional constituem-se como variantes de uma mesma variável. Estendendo um pouco essa noção, convém-nos mencionar a concepção de Lavandera (1984), segundo a qual a variação deveria ser vista também como duas ou mais formas equivalentes do ponto de vista funcional. Exemplos da variável *pronomes de tratamento informal do português brasileiro* seriam as formas *você, ocê* e *cê*, que seriam, por sua vez, *variantes* de tal fenômeno.

Pode haver variantes, por exemplo, no campo (a) fonético-fonológico, (b) morfossintático, (c) lexical etc. Estendendo essa classificação, Coelho *et al.* (2015, p. 23-32) abordam a variação em cinco modalidades, quais sejam: (1) lexical; (2) fonológica; (3) morfológica, morfofonológica ou morfossintática; (4) sintática e (5) discursiva. Primeiramente, a (1) variação lexical trata de variantes em nível lexical, de palavras, como se dá com os itens *abóbora* e *jerimum*, por exemplo (Coelho *et al.* 2015, p. 24), que predominam em diferentes regiões brasileiras. Um exemplo que resultou de estudos de variação lexical, entre outros, é o Atlas Linguístico do Brasil (Aguilera *et al.*, 2018), que faz um apanhado dessa variação nas diferentes regiões brasileiras, contando com 1100 informantes, espalhados por quase 9 mil km² de extensão.

Por sua vez, a variação (2) fonológica refere-se a variantes que apresentem diferentes fonemas. Um exemplo bastante comum desse tipo de variação é o fenômeno conhecido como rotacismo. Nele, há a troca da consoante [l] pela consoante [r], como em *pranta*, em lugar de *p/anta* (Coelho *et al.* 2015, p. 26).

Quanto à (3) variação morfológica, morfofonológica ou morfossintática, há alguma alteração no morfema da palavra ou há casos de interface, entre fonemas e morfemas ou morfemas e o campo sintático. Para os autores, um exemplo do primeiro caso seria a perda do morfema *-r* final marcador de infinitivo nos verbos, como em *parti*, em vez de *partir* (Coelho *et al.* 2015, p. 27). Por outro lado, um possível exemplo de variação morfofonológica seria a não realização da desinência verbal *-m*, marcadora de P6, como em *eles anda*, alternando com *eles andam* (Coelho *et al.* 2015, p. 27). Por fim, uma possibilidade de ocorrência da variação morfossintática, em que há interface entre morfemas e sintaxe, se daria, por exemplo, na relação

entre pronome e verbo em casos como *eles anda* (Coelho *et al.* 2015, p. 27), havendo a junção de um pronome P6 a um verbo de P3.

Com relação a (4) variação sintática, haveria uma alternância de nível apenas sintático, como na alteração da posição do clítico, em casos como *Eu vi-o no cinema* ou *Eu o vi no cinema* (Coelho *et al.* 2015, p. 28).

Por fim, quanto a (5) variação discursiva, em que a alternância se dá no nível do discurso, possíveis ocorrências seriam, por exemplo, em relação a expressões circunstanciais, como o item *aí*: “... eu tenho que dar baixa nessa carteira. *Aí* o cara falou...” (Coelho *et al.* 2015, p. 30). Nesse caso, o termo *aí* tem a função discursiva de introduzir a fala direta do *cara* em questão, precedendo o verbo *dicendi falou*. Nota-se, pois, que, segundo Coelho *et al.* (2015), a variação discursiva é também responsável pela organização dos turnos de fala, pela manutenção das estruturas de interação entre falantes e ouvintes e ainda pelo encadeamento coesivo das partes de um texto.

Vida Castro, Ávila Muñoz e Carriscondo Esquivel (2016) consideram a variação em quatro níveis, quais sejam: (1) fonológico; (2) sintático; (3) léxico-semântico e (4) estilístico.

Quanto a (1) variação fônica ou fonológica, os autores consideram que seja o nível em que se dão as condições mais propícias para a investigação variacionista, posto que forneceria “variáveis discretas facilmente identificáveis, segmentáveis e de uso muito frequente” (Vida Castro; Ávila Muñoz; Carriscondo Esquivel, 2016, p. 47). Esses autores consideram, ainda, que a aparição de variáveis fonológicas poderia estar relacionada a múltiplos fatores, como contextuais, situacionais e funcionais. Nesse sentido, a tarefa do pesquisador estaria em elucidar o grau de influência que esses fatores possuiriam na aparição ou na ausência dessas variáveis. Daí a necessidade de questionários “específicos e minuciosos, com a intenção de que nenhum desses fatores passe despercebido” (Vida Castro; Ávila Muñoz; Carriscondo Esquivel, 2016, p. 47).

Dando sequência aos níveis de variação estabelecidos por Vida Castro, Ávila Muñoz e Carriscondo Esquivel (2016), quanto à segunda classe (2) variação sintática, que, conforme sugere o próprio nome, seria observada no nível sintático, os autores consideram que, devido a um maior destaque que foi dado à variação fonológica, os sociolinguistas têm encontrado algumas dificuldades para investigar a variação nos outros níveis, o que incluiria o sintático.

Nesse sentido, consideram certa limitação da sintaxe descritiva para analisar os fenômenos, posto que eles, muitas vezes, se produziram devido a elementos extrassintáticos (Vida Castro; Ávila Muñoz; Carriscondo Esquivel, 2016, p. 49). Posto isso, os autores advertem que os métodos linguísticos não podem condicionar a teoria, devendo-se, então, adequar os modelos propostos, aprofundando-os, ou até mesmo procedendo a sua definitiva substituição. Para finalizar o seu argumento, os autores citam a tese de Lavandera (1975), na qual se comprovam os condicionadores sociais que afetam a distribuição das estruturas condicionais do espanhol, por exemplo, considerando-o como um trabalho de variação sintática que abre caminho às investigações da variação nesse nível linguístico.

Quanto à (3) variação léxico-semântica, os autores consideram que o léxico é um terreno privilegiado para verificar “as transformações sociais mais significativas de qualquer coletivo humano” (Vida Castro; Ávila Muñoz; Carriscondo Esquivel, 2016, p. 50). Ao tratar desse grupo, os autores consideram, além do léxico, as questões ligadas ao significado das expressões, por isso o denominam nível léxico-semântico. Nesse sentido, apontam que a base do significado deve situar-se “na articulação dos elementos linguísticos (palavras, morfemas, significados, fonemas etc.) em sistemas conectados e no conhecimento ou competência individual sobre o essencial (a gramática) dessa conexão” (Vida Castro; Ávila Muñoz; Carriscondo Esquivel, 2016, p. 51).

Por fim, quanto a (4) variação estilística, Vida Castro, Ávila Muñoz e Carriscondo Esquivel (2016) indicam que esse nível envolve desde estruturas relativamente complexas – típicas de estilos mais formais e controlados – a estruturas mais simplificadas, em que haja manifestações coloquiais e mais espontâneas. Assim, para os autores, ao linguista que se interessar por essa área caberá relacionar tais estruturas – sejam simples ou complexas – às variáveis sociais, mantendo em mente que as variações diafásicas submetem-se às diastráticas, pois, para eles, os grupos exercem maior pressão que as situações contextuais individuais.

Nesse ponto, tratando-se da questão estilística, podemos recuperar as 4 possibilidades hierarquizadas para observarem-se os estilos, possibilidades essas formuladas por Labov (1966). Interessante notar o fato de esse autor interpretar os estilos contextuais como um *continuum*, organizados conforme o grau de atenção dado pelo falante ao próprio discurso. Essa hierarquização vai do mais informal ao mais formal. Vejamo-la mais detalhadamente:

(1) discurso casual: pode ser observado na fala informal do dia a dia, em que se dê mínima atenção ao discurso. É geralmente carregado de emotividade.

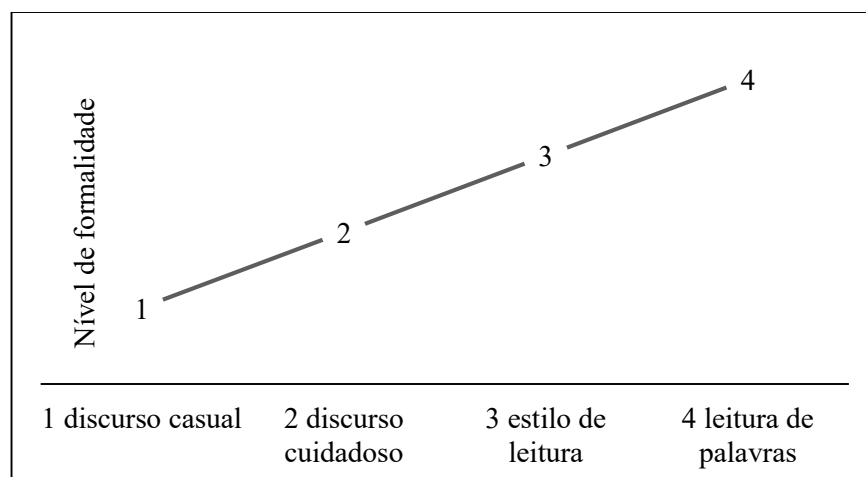
(2) discurso cuidadoso ou monitorado: seria um estilo intermediário, obtido sobretudo por meio de entrevista sociolinguística.

(3) estilo de leitura: com a leitura de textos, possibilita-se a coleta de amostras com alto nível de formalidade, posto que o informante/leitor presta bastante atenção em sua fala.

(4) leitura de palavras: segundo Labov (1966), esse procedimento permite a observação dos estilos mais formais do espectro sociolinguístico. É geralmente aplicado por meio da leitura de pares mínimos, que permitiriam ao pesquisador uma análise inclusive comparativa das variáveis propostas no método, que ficariam em evidência.

Observemos, a seguir, a figura 2, em que se ilustram os quatro estilos contextuais, hierarquizados por Labov (1966).

Figura 2 - Estilos contextuais hierarquizados por Labov (1966)



Fonte: elaborado pela autora.

Na figura 2, observamos que, conforme a linha avança de 1 a 4, mais formal será o estilo, pois, conseqüentemente, maior atenção será dada à fala pelo informante.

Aprofundando-se o tema, Shilling-Estes (2002) define a variação estilística como aquela que ocorre dentro da fala individual, e não em relação a um grupo de falantes. Assim, seria um contraponto entre *intra-speaker variation* e *inter-speaker variation* (Shilling-Estes, 2002, p. 375). A variação intrafalante envolve, por sua vez, outros diferentes tipos de variação, incluindo as variações dialetais – relativas a grupos específicos – e as variações de registro – relativas a situações particulares de uso. Para exemplificar seu raciocínio, a autora ilustra a variação de registro, por exemplo, quando um falante, para demonstrar formalidade, destaca a pronúncia padrão verbalizando o final dos gerúndios. Trazendo esse raciocínio para o português, teríamos a marcação *-ndo*, como em *caminhando*, em vez de *caminhano* (exemplos de própria autoria). A autora ressalta ainda que essa marcação padrão do gerúndio tende a ser feita pelo falante mais quando conversa com um colega de trabalho sobre assuntos profissionais do que quando fala trivialidades com um amigo. Quanto à variação dialetal, a autora a exemplifica com a situação de um falante que, ao conversar com um sulista mais velho, destaca a pronúncia carregada do *r* interiorano, “imitando” esse falante mais velho, e não o faz ao conversar com um sulista mais jovem, que claramente não usa esse *r* marcado.

Com relação às explicações para o fenômeno da ocorrência de variantes, pode-se destacar brevemente quatro possibilidades, que serão melhor exploradas futuramente, quando abordarmos as variáveis sociais.

- (1) Histórica: com o passar do tempo, uma ou outra variante cai em desuso.
- (2) Geográfica: diferentes regiões apresentam diferentes pronúncias, vocábulos e construções sintáticas.
- (3) Social: a classe social, a faixa etária, o gênero e o grau de escolaridade influem no modo de os falantes expressarem-se.
- (4) Estilística: cada falante possui um estilo próprio e o adapta à situação comunicativa em que se encontra.

Por sua vez, quanto ao conceito de variedade, Hudson (1996, p. 22) define-a como “um conjunto de itens linguísticos com distribuição semelhante”. É o que ocorre, por exemplo, com o português de Portugal, o brasileiro, o moçambicano etc., que se tratariam, pois, de variedades

da língua portuguesa. Ferguson (1972) aborda a variedade como padrões de fala mais ou menos homogêneos – como sons, palavras e aspectos gramaticais – passíveis de serem analisados sincronicamente pelas técnicas descritivas disponíveis no momento de análise. Nesse caso, seria necessário, ainda, que a variedade dispusesse de um escopo semântico suficiente para funcionar em todos os contextos formais de comunicação. Portanto, para esse autor, uma variedade ocorreria se pudesse se submeter a uma descrição sincrônica e se dispusesse de um leque semântico satisfatório dentro da comunidade em que é usada. Assim, exemplifica a variedade não só pensando em questões regionais, como a fala da cidade de Nova Iorque, mas também considerando aspectos sociais, que envolvem profissões, como o *juridiquês*, ou situações cotidianas, como as conversas de festas e *coquetéis* etc.

Coelho *et al.* (2015) conceituam a variedade como a fala que caracteriza determinado grupo. Para os autores, as variedades podem ser mais especificadas conforme se estabeleçam critérios, como o geográfico, o social, a profissão, os hábitos, isto é, qualquer característica que aproxime os falantes entre si, numa tentativa de unificá-los. Assim, seria possível falarmos, por exemplo, em variedade belo-horizontina, variedade das mulheres, variedade dos advogados, como o *juridiquês* acima mencionado etc. Além disso, pode-se, ainda, segundo os autores, combinar dois ou mais critérios para se alcançar determinada variedade mais específica, como, por exemplo, a variedade das jovens donas de casa de determinada cidade.

Saussure (2012 [1970]) define a língua como um “produto social da faculdade de linguagem e um conjunto de convenções necessárias, adotadas pelo corpo social para permitir o exercício dessa faculdade nos indivíduos”. (Saussure, 2012 [1970], p. 41). Dubois (1995 [1973]), por sua vez, define língua como um “instrumento de comunicação, um sistema de signos vocais específicos aos membros de uma mesma comunidade” (Dubois, 1995 [1973], p. 378). Haugen (1966) julga haver uma ambiguidade presente nas definições acerca de língua e crê que essa dificuldade provém da Antiguidade Grega. Para ele, o que entendemos hoje como a língua dessa época tratava-se, na verdade, de um grupo de variedades locais distintas, como o iônico, o dórico e o ático, que apresentavam usos literários particulares. Nesse sentido, o iônico ligava-se mais às narrativas épicas, o dórico às líricas e o ático às trágicas. Posteriormente, a língua comum ateniense, *koiné*, emerge como norma para a língua falada, posto que prevalecia no centro administrativo e cultural da época. Assim, a convergência entre iônico, dórico, ático e *koiné* vai gerar, futuramente, uma certa polissemia nas definições. Haugen (1966) considera, pois, que vem dessa época a mistura que se faz entre língua e dialeto, pois, naquele contexto,

língua poderia se referir tanto a uma única norma linguística quanto a um grupo de normas relacionadas; enquanto o dialeto referir-se-ia a uma dessas normas. Desse modo, não havia uma separação nítida entre os limites de cada definição, o que, para o autor, ter-se-ia arrastado ao longo dos séculos.

Wardhaugh (2006 [1986]) analisa que os conceitos de língua e dialeto perpassam pela noção de poder. Para ele, o poder envolve uma relação assimétrica e envolveria aspectos como *status*, dinheiro e influência. Assim a língua teria mais poder que o dialeto. Historicamente, o dialeto tem sido visto como algo fora da língua padrão, e fatores sócio-históricos são cruciais nessa definição. Ele considera, por exemplo, o hindi e o urdu, na Índia, que, apesar de serem praticamente a mesma língua, possuem fortes diferenças ligadas à política e à religião, pois passaram a marcar as desavenças político-ideológicas entre Índia e Paquistão. Embora as duas sejam consideradas línguas, cada um dos lados tende a considerar a sua como língua e a do outro como dialeto, na tentativa de rebaixar seu suposto adversário. Assim, indianos veem o hindi como língua e o urdu como dialeto; conseqüentemente, paquistaneses veem o urdu como língua e o hindi como dialeto. Apesar das sutilezas e da complexidade que envolve os termos sobre os quais nos detemos neste momento, quais sejam, língua, dialeto e variedade, vale lembrar que, neste trabalho, preferimos, pois, não entrar nas discussões, principalmente sobre o termo dialeto, traçadas por Haugen (1966), Wardhaugh (2006 [1986]) e Hudson (1996). Isso implica elucidarmos que, neste capítulo, não vamos considerar o termo dialeto como referente a uma variedade estigmatizada ou inferiorizada, mas sim como um falar característico de determinado grupo social ou regional.

Um ângulo interessante dessa discussão é também a noção de estandardização ou padronização, que se refere ao processo pelo qual uma língua foi codificada. Tal processo geralmente implica a criação de gramáticas, de convenções ortográficas, de dicionários e de literaturas, que recorrem a tal codificação. Wardhaugh (2006 [1986]) considera que esse processo é, em certa medida, arbitrário, pois requer definir o que pertence ou não a essa língua que se está codificando, levando, necessariamente, a exclusões e perdas: “o processo de padronização também é obviamente aquele que tenta reduzir ou eliminar a diversidade e a variedade” (Wardhaugh, 2006 [1986], p. 37). O autor inclusive discorre sobre o fato de os governos envolverem-se deliberadamente na estandardização, criando órgãos oficiais para regular o processo, que, por sua vez, vão efetivar as mudanças consideradas necessárias. Isso, mais uma vez, pode nos mostrar que a separação entre língua e dialeto não é tão clara, mas sim muitas

vezes tendenciosa. Essa separação tem sua complexidade aumentada pelo fato de que ambos os termos podem ser usados em sentido histórico. Por exemplo, teria havido uma única língua, a indo-europeia, cujos falantes, possuindo seus próprios dialetos, podem ter se espalhado por diferentes territórios, originando novas línguas que hoje consideramos pertencentes a essa matriz. Tentando chegar a uma definição, o autor finaliza seu argumento considerando que, se as línguas são entidades, o dialeto seria uma sub-variedade dessas entidades.

Sobre isso Trudgill (2000 [1974]) afirma que, uma vez que a língua é um fenômeno social, submetido aos sistemas de valores postulados pela sociedade que a usa e a padronizou, obviamente os dialetos serão avaliados de maneiras diferentes. Para ele, por exemplo, o inglês padrão possui muito mais prestígio e status que outros dialetos ingleses. Isso porque está associado a benefícios econômicos, sociais e políticos para quem o usa, enquanto os dialetos fora do padrão são desprestigiados e vistos como “errados, feios, corruptos e preguiçosos” (Trudgill, 2000 [1974], p. 8). Entretanto, como bem ressalta esse autor, não há nada entre os diferentes dialetos que nos permita considerá-los superiores ou inferiores entre si. Todos eles são igualmente bons como sistemas linguísticos, sendo, pois, os julgamentos de valor muito mais sociais do que linguísticos realmente.

Para Coelho *et al* (2015), a variedade culta, padronizada, é geralmente associada às esferas mais privilegiadas da estrutura social. Geralmente a dominam os falantes mais escolarizados, que ganham mais e que vivem melhor. “Essas pessoas, por seu status, comumente gozam de prestígio social, e esse prestígio é transferido para a sua fala” (Coelho *et al.*, 2015, p. 15). Entretanto, ainda que a variedade culta gere certa expectativa mais ou menos homogênea em relação àqueles que a usam, essa expectativa é também variável, posto, por exemplo, que há pessoas que moram em lugares considerados socialmente elevados, em grandes centros urbanos, mas que possuem pouca escolaridade, como há também profissionais de alta escolarização que recebem baixos salários. Nesse sentido, podemos nos falar não só de variedade culta, mas de variedades cultas. Além disso, convém-nos considerar, ainda, que, dentro da variedade culta, costumam-se apresentar as variantes mais conservadoras. Ao passo que, dentro das variedades não padronizadas, tendem a ocorrer as variantes mais inovadoras.

Avançando um pouco mais na discussão sobre o fenômeno da variação, podemos considerar que as variantes podem permanecer estáticas, de modo que as mesmas formas alternem-se entre si durante um período de tempo curto ou longo, ou que essas variantes podem sofrer mudança,

quando uma das formas cai em desuso. Neste caso, “as formas substituem outras que deixam de ser usadas, momento em que se configuram um fenômeno de mudança em progresso” (Mollica; Braga, 2003, p. 11). Portanto, a variação seria a coocorrência de duas ou mais variantes num mesmo período de tempo analisado. A mudança seria o predomínio de uma variante sobre outra até o apagamento desta. É como se uma das formas ganhasse a disputa na preferência dos falantes. Se não há esse predomínio de uma variante sobre outra, convencionasse dizer que se trata de variação estável.

Naro e Scherre (1991), por sua vez, consideram a complexidade relacionada aos dois termos em questão. Tais autores usam os termos *fluxo* e *contrafluxo* para melhor abordarem as nuances implicadas na variação e na mudança. Para eles,

o que está mudando para algumas pessoas pode estar estável para outras pessoas e o que está aumentando para alguns pode estar diminuindo para outros. Para algumas pessoas o mercado de trabalho pode ter efeito, enquanto o sistema escolar pode influenciar outras. (Naro; Scherre, 1991, p. 15).

Importante lembrar que o conceito de mudança linguística será melhor explorado na sessão “3.4 Variação e mudança linguística” deste trabalho.

Vejamos, neste momento, pois, outra parte conceitual também importante para a Sociolinguística, que são os fatores considerados externos à língua, chamados de variáveis sociais. Para Shaffer (1979), os três fatores que permitem ao indivíduo comunicar-se são o biológico, o cognitivo e o social. Os dois primeiros são considerados aspectos individuais. A parte biológica estaria ligada às questões orgânicas do indivíduo e a cognitiva relacionar-se-ia aos elementos intelectuais pessoais. Os fatores sociais seriam aqueles ligados a questões do indivíduo em relação a seu meio, à coletividade. Vida Castro, Ávila Muñoz e Carriscondo Esquivel (2016, p. 53) apontam que, ainda que se possa falar em variação social, cada língua oferece disparidades e particularidades, que variam conforme o contexto em que essas línguas são empregadas. Assim, fatores como a identidade dos falantes, seus antecedentes sociais, culturais e étnicos, sua procedência geográfica, suas atitudes, seu sexo, sua idade etc. são mais ou menos relevantes para cada língua em particular, o que demanda atenção especial do pesquisador. Isso implica considerar que, se dado fator social, como idade, é relevante para uma pesquisa, que observa determinado aspecto de uma língua x, não necessariamente será para outra pesquisa que se proponha a analisar outra língua, por exemplo. Quanto a isso, Labov

(2008 [1972]) afirma que: “dada pessoa ou grupo usa a língua x num contexto ou domínio social y” (Labov, 2008 [1972], p. 215).

Quanto à classe social, Ash (2002, p. 402) a considera um conceito central da pesquisa sociolinguística, apesar de ser uma variável que não possui rigidez. Isso porque não há um consenso sobre as variáveis independentes “concretas e quantificáveis” que, juntas, formariam o que chamamos de classe social. Para a autora, trata-se, pois, de uma variável “universalmente usada e extremamente produtiva, embora os linguistas possam reivindicar pouco para entendê-la” (Ash, 2002, p. 402). Labov (1966), por exemplo, defende o conceito de classe social como as oportunidades de vida que os indivíduos possuem a ponto de possibilitar-lhes acessar bens e serviços.

Uma definição clássica de classe social é a de Max Weber (*citado por* Ash, 2002, p. 403). Para esse teórico do século XX, o conceito multifatorial de classe social envolveria o grau de poder e controle desfrutado por um indivíduo por causa de suas posses materiais e o prestígio, a reputação e o *status* que esse indivíduo goza justamente devido a esses bens. Um tratamento interessante dado à classe social é o estudo de Labov (2008 [1972]) nas lojas de departamento da cidade de Nova Iorque. Para tanto, ele recupera a definição de estratificação social, dada por Barber (1957 citado por Labov, 2008 [1972]). Segundo Barber (1957), a estratificação seria o produto entre a diferenciação social dada aos indivíduos e pelos indivíduos e a avaliação social que recebem e realizam dos outros. Essa diferenciação gera hierarquias coletivamente compartilhadas, que produzem status e prestígio tácitos. Labov (1972 [2008]) observou que os profissionais das lojas socialmente mais prestigiadas tendiam a se apropriar do prestígio de seus clientes ou pelo menos se esforçavam nessa direção. Observou-se, inclusive, que altos salários não necessariamente estratificavam os empregados na mesma ordem, pois, mesmo funcionários que ganhavam menos, ao exercer atividades em lojas de maior status, sentiam-se mais prestigiados. Notou-se ainda que os colaboradores estavam dispostos a aceitar salários mais baixos se a loja possuísse maior prestígio. Esse estudo levou o autor a deparar-se, também, com o conceito de raça, observando que havia mais empregados negros na loja de menor status social.

Eckert (1989), por sua vez, analisa que muita atenção tem sido dada à categoria classe social nos estudos linguísticos. Para essa autora, há uma complexa relação entre as categorias que compõem a classificação socioeconômica dos indivíduos e a prática social dessas categorias.

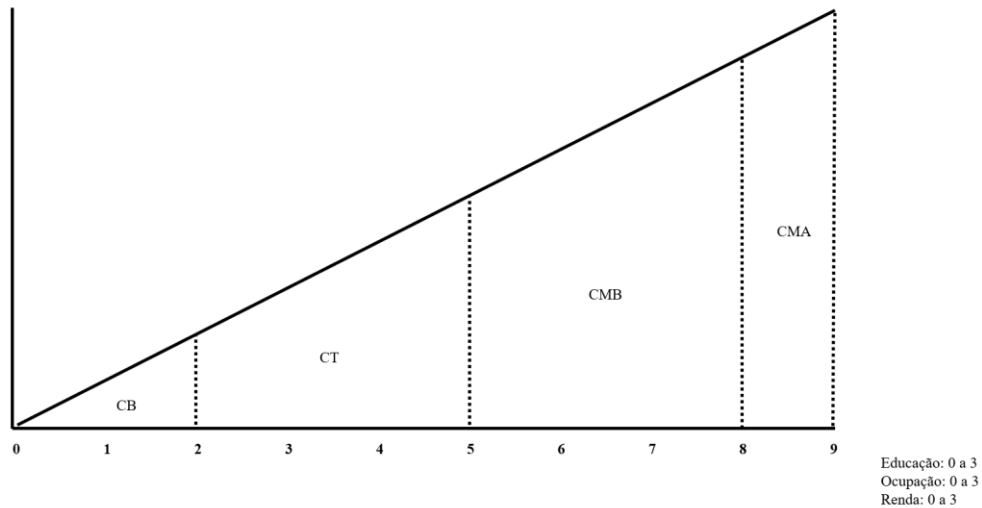
Assim não nos concentramos nos índices objetivos usados para medir classe (como salário, ocupação e educação) ao analisar correlações entre diferenças linguísticas de classe, mesmo quando a identificação de classe é baseada nesses índices. Em vez disso nos concentramos cada vez mais na relação do uso da linguagem com a prática cotidiana que constitui a participação social baseada em classe e a identidade dos falantes na comunidade (Eckert, 1989, p. 246).

Portanto, a autora reforça a visão da categoria classe social como um fator múltiplo, que possui parâmetros auxiliares, como o grau de instrução e a renda dos falantes, e que funciona a partir do uso que esses falantes fazem da linguagem e do modo prático como são vistos socialmente ao reunirem um ou mais parâmetros.

Para Ash (2002), posto que a classe social é um conceito múltiplo, um fator relevante para a sua estruturação seria a noção de ocupação. Mesmo em nações culturalmente díspares, como Japão e União Soviética, por exemplo, descobriu-se que o fator ocupação era mais ou menos convergente. Além dessa questão geográfica e cultural, a avaliação temporal da noção de ocupação também se manteve mais ou menos estável ao longo do tempo. Apesar disso, a autora salienta que a ocupação geralmente não é o único elemento a compor a variável social, justamente dada a complexidade que envolve tal conceito. Labov (2008 [1972]), por exemplo, ainda no estudo da variável (r) nas lojas de departamento nova-iorquinas, observou que, nas lojas Macy's e Saks, em que havia uma maior demarcação de ocupações, separando-se nitidamente chefes de seção, vendedores, caixas, reposidores, ascensoristas e recepcionistas, houve predominância da variante de prestígio.

A ocupação é também um dos critérios considerados por Labov (1966) para construir a noção de classe social, que, além desse fator, leva em consideração ainda o nível de estudo e a renda do indivíduo. Nessa percepção laboviana, cada sujeito é avaliado em até 9 pontos conforme cada critério supracitado, sendo dados até três pontos para o parâmetro educação; até três pontos para a renda e até três pontos para a ocupação. Embora essa avaliação seja ainda relativa, convencionou-se chamar o intervalo 0-2 de classe baixa (CB); o de 3-5 de classe trabalhadora (CT), o de 6-8 de classe média baixa (CMB) e a pontuação 9 é considerada a da classe média alta (CMA). A seguir, baseados em Vida Castro, Ávila Muñoz e Carriscondo Esquivel (2016, p. 58), ilustramos essa escala de Labov (1996).

Figura 3 – Fatores e pontuações que compõem a classe social, segundo Labov (1966)



Fonte: elaborado pela autora.

Moreno Fernández (1990), por sua vez, considera a classe social um conceito polêmico. Para ele, a classe estaria ligada não só a variáveis, como poder econômico e social, mas também à noção de pertencimento, à integração que os indivíduos demonstram em relação a ela. Portanto, não se trataria só de valores materiais, mas também de noções subjetivas, que têm a ver com o julgamento que se tem de fato sobre pertencer ou não a uma classe. Para o autor, essa percepção ocorre como em uma cadeia interligada. Nesse sentido, nota-se que pessoas de maior poder aquisitivo geralmente estudam por mais tempo e desenvolvem-se em ambientes que não só valorizam uma linguagem mais padronizada, mas também a incentivam.

Quanto à variável sexo, Cheshire (2002) afirma que, historicamente, nas pesquisas sociolinguísticas, o termo sexo tem sido usado no sentido apenas de apontar as diferenças biológicas envolvendo machos e fêmeas. Um princípio de mudança para o amadurecimento desse conceito foi a entrada do termo gênero, que, inicialmente, representou apenas uma mudança terminológica, uma mera troca de palavras. Entretanto, posteriormente, a categoria gênero foi ganhando uma maior elaboração. Nesse processo, começou a referir-se à diferença cultural e social existente entre homens e mulheres, o que ainda não era satisfatório, posto que

essa conceituação era reducionista, no sentido de restringir os papéis sociais, as oportunidades e as expectativas usufruídas por homens e mulheres nas diferentes sociedades.

Como princípios iniciais, Labov (1990) estabelece que:

(I) Numa estratificação sociolinguística estável, os homens usam uma frequência maior de formas não padronizadas do que as mulheres.

(II) Na maioria das mudanças linguísticas, as mulheres usam com maior frequência as formas inovadoras do que os homens.

O autor dá prosseguimento à sua análise afirmando que, embora esses princípios sejam já razoavelmente testados e aceitos, não há nada neles ou exteriormente que explique como a diferenciação sexual afeta a linguagem. Nesse momento, surgem, pois, quatro questões conceituais relevantes, que seriam: (1) o viés biológico; (2) a generalidade do gênero; (3) a inversão de papéis e (4) a diversificação íntima. Vejamos mais cuidadosamente cada um desses pontos, aventados por Labov (1990).

(1) O viés biológico: sobre esse tema, Labov (1990) reconhece que o sexo em si mesmo não é uma categoria apropriada para explicar o comportamento linguístico, pois o que determinaria as diferentes manifestações linguísticas não é a diferença de órgãos reprodutivos masculinos e femininos, mas sim os distintos papéis culturais assumidos por homens e mulheres nas sociedades.

Quanto ao segundo questionamento, (2) a generalidade do gênero, Labov (1990) retoma um princípio consensual, formulado por Meillet (1921), segundo o qual as mudanças linguísticas acompanham as mudanças sociais, alterando a estrutura das comunidades de fala. Nesse sentido, mudanças repentinas na língua tendem a ser fruto de grandes transformações sociais, como catástrofes iminentes. No entanto, as questões sexuais tendem a influenciar mais lentamente as mudanças sociais, seja na economia, na política, na demografia etc. Desse modo, observa-se que a diferenciação sexual é mais sutil na influência sobre a língua, logo, em relação ao gênero, notam-se movimentos mais locais e observáveis a mais longo prazo.

Quanto ao terceiro ponto, (3) a inversão de papéis, Labov (1990) observa a aparente contradição expressa nos princípios (I) e (II) anteriormente mencionados. A dualidade presente em tais princípios mostra-se porque, conforme o primeiro deles, as mulheres tendem a ser mais conservadoras, favorecendo variantes de maior prestígio social. No princípio (II), as mulheres demonstram menos conservadorismo ao usar formas não padronizadas e que podem muitas vezes ser estigmatizadas. O autor reconhece que tem havido esforços dos sociolinguistas para, de alguma forma, sistematizar essa aparente contradição, dando-lhe uma interpretação única, a de que mulheres são mais sensíveis a formas que confirmam maior status social.

Sobre isso, Eckert (1989) afirma que tem havido trabalhos que mostram a dualidade das inovações linguísticas sendo encabeçadas por homens e mulheres. Para exemplificar, ela cita os estudos de Cedergren (1973) no Panamá e os de Labov (1984) na Filadélfia, nos quais as mulheres foram apontadas como liderando a mudança sonora. Por outro lado, haveria, por exemplo, os trabalhos de Trudgill (1972a), em Norwich, e de Labov (1972), em Martha's Vineyard, estudos nos quais os homens foram os responsáveis por levar a mudança sonora adiante. Eckert (1989, p. 248) conclui, portanto, que, dados os diferentes estudos que mostram ambos os lados da questão, é-nos possível supor que “as diferenças de sexo podem variar de uma variável para outra”.

Nesse sentido, Trudgill (1972b citado por Eckert, 1989) argumentou que mulheres e homens respondem de modo oposto às normas. As mulheres alinham-se mais às normas de prestígio da linguagem padrão e os homens às normas vernaculares. Assim, as primeiras aproximam-se daquilo que é visto como refinado enquanto estes aproximam-se daquilo que é visto como rudimentar. Uma explicação para esse fato é a ainda dificuldade de ascensão social observada pelas mulheres na sociedade. Uma vez que elas se sentem mais impotentes ao aspirar a melhores posições, lançam mão de recursos simbólicos, como a linguagem. Logo, se não conseguem provar seu valor no mercado de trabalho, por exemplo, provam-no usando uma linguagem de maior prestígio. Assim, pode-se considerar que a situação social das mulheres influencia seu comportamento linguístico, e isso se torna ainda mais evidente em sociedades em que pouco espaço é dado a elas.

A polarização sexual do uso da língua, portanto, se dá em maior medida conforme descendemos na escala de estratificação socioeconômica e educacional (homens e mulheres prototípicos). Nos níveis mais baixos, as mulheres se encontram à mercê do poder de seus maridos, pais, irmãos,

parentes ou vizinhos e companheiros de trabalho masculinos. Por isso se reafirmam como mulheres e falam como tais (Vida Castro; Ávila Muñoz; Carriscondo Esquivel, 2016, p. 57).

Mas como explicar quando as mulheres são menos conservadoras e lideram mudanças em progresso em relação aos homens? Sobre isso, Labov (1990) observou que as mulheres tendem a afastar-se de variantes que, apesar de inovadoras, são suficientemente antigas para permitir que sejam vistas como estigmatizadas dentro da comunidade mais ampla. Nesse sentido, alinham-se a normas mais genéricas e prestigiadas. Paralelamente, quando ainda não percebem que alguma variável local possa ser estigmatizada na comunidade global, lideram seu uso, encabeçando mudanças em progresso. Nesse sentido, Eckert (1989) estabelece três tipos de prestígio diferentes. São eles, o global, que se vincula às normas da linguagem padrão; o prestígio dissimulado, que se opõe ao considerado padrão; e o local, que se refere à sensação de pertencimento a uma comunidade local.

Por fim, para discutir o último questionamento, (4) diversificação íntima, o autor recupera também a ideia da densidade local. Segundo essa noção, os atos comunicativos individuais acabam transferindo certa influência linguística ao meio, padronizando em certa medida a fala dos indivíduos presentes naquele ambiente. Importante lembrar que essa certa padronização inclusive exerce pressão a outras ambiências e locais, o que poderia explicar, por exemplo, a disseminação de mudanças linguísticas de comunidades maiores, mais expressivas, para comunidades menores, mais reduzidas.

Para Eckert (1989, p. 246), “o sexo é uma categoria biológica que serve como base fundamental para a diferenciação de papéis, normas e expectativas em todas as sociedades”. Evoluindo essa definição, ela passa a considerar o conceito de gênero, que seria a construção social realizada a partir da visão que se tem do sexo, com os papéis, as normas e as expectativas sociais envolvidos. Isso porque, inclusive, deve-se considerar que a sociedade está em constante movimento, e, nesse contexto, nota-se um avanço das mulheres nas mais diversas áreas, sobretudo profissionais, tornando os papéis de gênero cada vez mais questionados.

Nas pesquisas sociolinguísticas, essa autora reconhece que, enquanto o sexo é uma variável binária prontamente verificável, o gênero é uma variável multidimensional, observável através de uma análise criteriosa. Portanto, o gênero não deve ser visto de modo isolado, mas conjugado a outras dimensões, resultando em categorias compostas, como a de “mulher de classe média

baixa” (LABOV, 1990, p. 207). Em contraposição, o sexo torna-se apenas uma tentativa de aproximação estatística ao gênero. A ressalva feita por Eckert (1989) é que, muitas vezes, os dados coletados não possuem uma pesquisa sobre a percepção de gênero na comunidade pesquisada, o que leva a dados de sexo, que, como dito, são categorias evidentes. Isso é razoável. O problema estaria em tratar esses mesmos dados como dados de gênero, o que seria leviano. Sobre isso, Cheshire (2002, p. 432) acrescenta que “operacionalizar a categoria de sexo do falante de uma forma simples e indiferenciada permitiu que estudos quantitativos fossem replicados em uma variedade de comunidades. Além disso, a autora acrescenta que esse contexto também permitiu que se criassem afirmações cada vez mais genéricas e não necessariamente fidedignas. Por conseguinte, conclui que: “o sexo dos falantes não está diretamente relacionado ao comportamento linguístico, mas reflete uma prática social complexa” (Cheshire, 2002, p. 433). Um interessante trabalho do PB que também explora a variável gênero é o de Callou e Marques (1975), em que apontam que as mulheres cariocas realizavam o /s/ não palatal em coda interna o dobro de vezes em relação aos homens.

Para Vida Castro, Ávila Muñoz e Carriscondo Esquivel (2016), o sexo seria uma variável submetida à classe social, funcionando como um subproduto desta, não devendo, pois, ser vista de modo dissociado. “Diferenças de sexo se explicam, então, em termos de diferenças de classe, da mesma maneira, por exemplo, que o estilo se vê como reflexo individual da estratificação” (Vida Castro; Ávila Muñoz; Carriscondo Esquivel, 2016, p. 55). Ao abordar as diferenças de comportamento linguístico ocorridas entre homens e mulheres, tais autores apontam, como possíveis causas, fatores ligados a questões biológicas, a diferenças nas expectativas parentais ou mesmo à educação distinta oferecida a meninos e a meninas. Nesse sentido, observou-se, por exemplo, que, com meninos, as interações parentais estavam mais ligadas à exploração do mundo físico; com as meninas, notou-se a predominância de conversações cotidianas. Outra reflexão levantada pelos autores, nesse contexto, foi que, por exemplo, as brincadeiras com bonecas podiam favorecer uma maior ocorrência de substantivos e um maior grau de interação verbal, enquanto os jogos com objetos inanimados, como carrinhos e bolas, favoreciam ruídos e sons imaginativos, não privilegiando a interação verbal, dada a inanimacidade dos brinquedos. Considerando a complexidade do tema e os diferentes posicionamentos anteriormente expostos, cabe-nos considerar que vamos ao encontro de Coelho *et al.* (2015), que defendem que as variáveis sejam vistas, sempre que possível, interconectadas, e não isoladamente, de modo que se obtenha uma análise mais ampla e menos tendenciosa.

Tem se verificado resultados mais significativos quando se correlaciona essa variável com a faixa etária da população e, se possível, com a história social das diferentes comunidades investigadas, de modo que as transformações culturais e as mudanças comportamentais das faixas mais jovens da população possam ser observadas também (Coelho *et al.*, 2015, p. 44).

Assim, ressaltamos nosso propósito neste trabalho de explorar a marcação da quebra de expectativa nos falares da RMBH e do português moçambicano buscando correlacionar nossas principais variáveis, quais sejam: gênero, idade e escolaridade, para que logremos análises mais satisfatórias.

Nesse sentido, dissertemos, pois, acerca de outra variável social, a idade. À variação favorecida pela faixa etária dos falantes dá-se o nome de variação diageracional (Coelho *et al.*, 2015). Coelho *et al.* (2015) sustentam que se trata de uma variável que deve ser analisada levando-se em conta a relação entre o indivíduo e sua comunidade. Consideremos como exemplo disso os trabalhos de Pagotto (2001) sobre as consoantes oclusivas alveolares antes de /i/ na fala de Florianópolis. Nesse trabalho, o autor encontrou que falantes mais velhos tendem a optar pela forma mais antiga – a não africada –, ao passo que os falantes mais jovens tendem a optar pelas formas mais inovadoras – a africada não palatal e a palatal. A aproximação dos falantes mais velhos à variante mais antiga não é só um aspecto ligado à idade, mas também se refere ao fato de essa variante ser tida como uma marca identitária dos florianopolitanos, o que corrobora a proposição inicial de Coelho *et al.* (2015) sobre a faixa etária não dever ser analisada isoladamente, mas sim levando-se em conta o indivíduo e sua relação com a comunidade a que pertence.

Eckert (1989) afirma tratar-se de uma variável que tem recebido atenção “menos sofisticada” (Eckert, 1989, p. 246). Isso porque, para ela, a idade, que, na verdade, é um *continuum*, geralmente é dividida em partes iguais, sem necessariamente dar-se uma atenção a cada uma dessas partes especificamente e aos estágios da vida significativos de cada uma dessas partes. Não se consideram, por exemplo, quais os marcos sociais de cada um dos períodos ou o que se espera que o falante alcance ou realize em cada um desses estágios, tornando-se, pois, aspectos ignorados, mas que poderiam contribuir para uma melhor compreensão do fator idade em relação ao uso da língua. Vida Castro, Ávila Muñoz e Carricondo Esquivel (2016) afirmam que a maioria dos estudos demonstram, por fim, que as gerações mais jovens apresentam tendência inovadora frente ao conservadorismo associado aos mais velhos. Além disso, destacam que as pesquisas de léxico parecem ser ainda mais sensíveis a essa variável.

Eckert (1997) considera ainda que a variável idade pode indicar a fala de um indivíduo em relação ao tempo em geral, externo, evidenciando, por exemplo, as mudanças históricas pelas quais passam as línguas, e também em relação ao tempo particular, individual, expondo, por exemplo, diferenças linguísticas relativas às gradações etárias. Nesse sentido, segundo a autora, o comportamento linguístico dos indivíduos transforma-se ao longo de suas vidas. Sobre isso, Freitag (2000) recupera a visão clássica de que “a língua de um indivíduo se constitui até cerca de seus quinze anos de idade”, sofrendo alterações posteriores (Freitag, 2000, p. 110). Desse modo, “a cada etapa do ciclo vital, mudanças de ordem biológica e social ocorrem e refletem também na sua língua, é o que Eckert denomina de curso da vida linguística” (Freitag, 2000, p. 111).

Além disso,

a aquisição da língua, a entrada na escola, a aplicação da rede de relações sociais, a entrada e a saída do mercado de trabalho são fatores que se refletem diretamente nas faixas etárias. Logo, a faixa etária não pode ser encarada como um fator simples. Ao contrário, é extremamente complexo, e é preciso muita atenção ao interpretar os resultados de um fenômeno de mudança em função das faixas etárias (Freitag, 2000, p. 111).

Nesse sentido, reforça-se o anteriormente dito sobre o fato de a variável faixa etária ser complexa e interdependente em relação a outras variáveis, pois, inclusive conforme salienta Freitag (2000) e como anteriormente exposto, o fator idade trata-se de um rótulo “que agrupa vários fatores de ordem social e biológica do indivíduo” (Freitag, 2000, p. 111).

Assim, após essa exposição acerca de alguns termos e conceitos pertinentes à pesquisa sociolinguística, vejamos, na seção a seguir, alguns aspectos metodológicos dessa área de estudos.

2.3 Sociolinguística: aspectos metodológicos

Como exposto, a Sociolinguística Variacionista é criteriosa quanto a seus métodos de trabalho. Como bem afirma Wardhaugh (2006 [1986]), a pesquisa dessa área deve ser motivada por questões que possam ser respondidas de forma científica aprovada. Para esse autor, um conjunto qualquer de observações aleatórias sobre como algumas pessoas usam a língua não pode levar

a uma generalização útil. “Não podemos nos contentar com colecionar borboletas, por mais bonitas que sejam as espécies” (Wardhaugh, 2006 [1986], p. 17), criticando os dialetólogos. Sendo assim, os dados linguísticos coletados precisam ter uma finalidade, e essa finalidade deve responder a questões relevantes. Ele conclui seu raciocínio sustentando que, se tais questões não permitem algum tipo de teste empírico, não possuem mais do que um interesse especulativo e vazio em si mesmo. Os pesquisadores da área devem ter, pois, duas preocupações basilares: fazer boas perguntas e encontrar dados que se relacionem coerentemente com esses questionamentos.

Nesse sentido, o autor considera alguns tipos de estudos possíveis, quais sejam: (1) correlacionais; (2) implicacionais; (3) microlinguísticos; (4) macrolinguísticos e (5) generalistas (Wardhaugh, 2006 [1986], p. 17-18). Os primeiros, (1) correlacionais, tentam relacionar duas ou mais variáveis, por exemplo avaliar usos linguísticos de diferentes classes sociais. Nos (2) implicacionais, sugerem-se fórmulas, como “se X, então Y”. Uma possível tentativa, segundo o autor, que pensou sobre a língua inglesa, seria “se alguém diz *tess* para *tests*, também dirá *bes* para *best*?”. Os estudos (3) microlinguísticos, por sua vez, normalmente se concentram em itens linguísticos muito específicos ou em diferenças e usos individuais, como a diferença entre *singing* e *singin*. Os (4) macrolinguísticos, por outro lado, examinam grandes quantidades de dados para estabelecer alguns parâmetros sobre grupos sociais. Por fim, os estudos (5) generalistas busariam estabelecer universais da comunicação humana.

Bell (1976, p. 187-191) estabelece oito princípios que, segundo ele, a pesquisa sociolinguística deveria seguir. São eles: o princípio cumulativo, o da uniformização, o da convergência, o da mudança subordinada, o da mudança de estilo, o da atenção, o vernacular e o da formalidade. Vejamos, detalhadamente, cada um deles.

(1) Princípio cumulativo: segundo esse princípio, o conhecimento adquirido com a pesquisa linguística vai se acumulando. Quanto mais se sabe sobre determinada língua, mais se pode descobrir acerca dela. Inclusive não será uma surpresa se essa busca por conhecimento levar a novas áreas ou disciplinas.

(2) Princípio da uniformização: consoante esse princípio, os processos linguísticos observados hoje são os mesmos ocorridos no passado. Portanto não pode haver uma ruptura brusca entre sincronia e diacronia.

(3) Princípio da convergência: conforme o princípio da convergência, uma vez que se usam dados novos para confirmar velhas descobertas, o valor desses dados é diretamente proporcional aos modos de coleta para obtê-los. Isto é: é natural que surjam novos procedimentos metodológicos, inclusive de outras áreas que não sejam as linguísticas para satisfazer os estudos linguísticos. É o que vemos, por exemplo, no avanço do uso de métodos e técnicas da linguagem de programação contribuindo com a linguística de *corpus*.

(4) Princípio da mudança subordinada: de acordo com esse princípio, falantes de variedades subordinadas (não padrão), como um dialeto, quando questionados sobre próprio falar, tendem a manifestar-se linguisticamente em direção ao padrão ou, ao contrário, distanciando-se dele. Isso permite ao pesquisador observar aspectos valiosos em relação a temas como variedade, normas e mudanças. Esse princípio pôde ser visto, como dito anteriormente, por exemplo, nos estudos de Marthas's Vineyard, segundo os quais, a depender de sua identidade, o falante aproximava-se da variedade continental, considerada padrão, ou afastava-se dela, usando a variedade insular.

(5) Princípio da mudança de estilo: esse princípio pressupõe uma evidência bastante perceptível, a de que os falantes ajustam suas falas segundo as circunstâncias em que se encontram. Desse modo, diz-se que não há falantes de estilo único.

(6) Princípio da atenção: trata-se de outro princípio também bastante firmado nos estudos da pesquisa sociolinguística. Segundo ele, quanto mais atenção os falantes prestam a seu discurso, mais formal este tende a ser.

(7) Princípio vernacular: o princípio vernacular pressupõe que o estilo mais regular em sua estrutura e em sua história com a língua é o vernáculo, no qual há descontração e menos atenção à fala.

(8) Princípio da formalidade: o princípio da formalidade atua em polo oposto ao vernacular. Nesse sentido, quanto mais atenção e consciência se dá à fala, mais difícil será abstrair-se o vernáculo. É no princípio da formalidade que reside o paradoxo do observador, apontado por Labov (1972). A situação do pesquisador linguístico é

paradoxal, posto que, ao mesmo tempo em que ele busca o vernáculo, que tende a ocorrer em situações não controladas de fala, precisa controlar a fala para obter os dados. Resta-lhe, pois, tentar desviar a atenção sobre a língua para que o vernáculo possa emergir. Labov considera que isso pode ser possível levando o entrevistado a emocionar-se, pois, estando emotivo, o falante tende a penetrar em seus próprios sentimentos, esquecendo-se de prestar atenção a seu discurso.

Complementando os princípios supracitados, Labov (2008 [1972]) isola cinco axiomas metodológicos que conduzem ao que vai chamar posteriormente de paradoxo do observador. Vejamos, primeiramente, os cinco fundamentos, para, em seguida, dissertamos sobre tal paradoxo.

O primeiro axioma é a questão da alternância de estilo. Quanto a isso, firma-se a premissa, já anteriormente mencionada, de que não existe falante de estilo único. Evidentemente, o grau de variação de estilo entre um e outro falante é relativo, mas, independentemente de quanto, todo falante exprime alteração de variáveis linguísticas à medida que muda de contexto e de tópico. Essas alterações podem ser percebidas inclusive nas pequenas autocorreções que faz de si mesmo.

O segundo axioma é o da atenção. Existem inúmeros estilos e dimensões estilísticas, o que impossibilita ao pesquisador isolar todos eles. Entretanto Labov (2008 [1972]) verificou que esses inúmeros estilos podem ser dispostos numa macrodimensão, que é medida pelo grau de atenção dado à fala. Nesse sentido, quanto mais emoção e envolvimento o falante tiver em relação ao tema do que está sendo dito, menos atenção ele prestará às variáveis linguísticas utilizadas.

O terceiro axioma é o vernáculo, que é o estilo no qual se dá a mínima atenção à fala. Nele há, pois, pouco monitoramento do discurso. É justamente o vernáculo que nos oferece os dados mais interessantes para a pesquisa sociolinguística.

O quarto axioma chama-se formalidade. Nele observa-se que, por menor que seja o monitoramento dado à fala, ainda assim haverá certo controle e perda de naturalidade. Isso permite supor que, mesmo alcançando o vernáculo, o falante ainda possuirá um comportamento

linguístico mais espontâneo ainda, o que usa numa situação informal com os amigos ou familiares.

O quinto e último axioma refere-se aos bons dados. Quanto a isso, considera-se que, independentemente de outros métodos possíveis para a coleta de dados, como as entrevistas rápidas e anônimas, por exemplo, a maneira indicada de captar bons dados, em quantidade satisfatória, dá-se por meio da entrevista individual e gravada.

Desse modo, após a exploração dos axiomas supracitados, Labov (2008 [1972]) discute o conceito de paradoxo do observador. Esse paradoxo reside no fato de o pesquisador buscar a língua natural, usada quando as pessoas não estão sendo observadas e, portanto, não se preocupam em moldar o seu discurso conforme normas exteriores. Entretanto, só se pode obter essa língua registrando-a por meio de uma observação sistemática, que, inevitavelmente, afetaria a fala. A proposta de Labov (2008 [1972]) para superar o paradoxo do observador, como mencionado, é tentar suavizar os constrangimentos provenientes da observação da fala, desviando a atenção do falante, para que este deixe o vernáculo emergir. Para tanto, pode-se, por exemplo, dar pausas durante as gravações para que o entrevistado pense que não está sendo gravado ou mesmo suscitar nele emoções fortes. Neste caso, pode-se recorrer a perguntas clássicas, como “você já viveu uma situação em que correu sério risco de morrer?” (Labov, 2008 [1972], p. 245).

Ainda sobre a metodologia da pesquisa sociolinguística, Moreno Fernández (1990) considera que três grandes regras norteiam o processo metodológico. São elas: (1) regras de coleta de dados; (2) regras de análise e (3) regras de interpretação. Vejamos separadamente cada uma delas e suas subdivisões.

Como pressupostos das (1) regras de coleta de dados, primeira regra de Moreno Fernández (1990), ele estabelece que, inicialmente, o pesquisador deve abandonar qualquer noção prévia do fenômeno sobre o qual se debruce, para não deixar que suas próprias impressões o induzam a conclusões equivocadas ou mesmo precipitadas. Desse modo, deve esforçar-se para não deixar que suas afinidades pessoais se sobreponham aos interesses científicos, “os fatos devem ser tratados como se fossem contemplados pela primeira vez” (Moreno Fernández, 1990, p. 28). Em segundo lugar, para esse autor, o fenômeno a ser investigado deve ser definido por aspectos exteriores, comuns e constantes. Exteriores no sentido de que sejam objetivos; comuns no

sentido de que sejam aplicáveis a todos os elementos que compõem o fenômeno investigado, e não estejam em função de recortes tendenciosos, e constantes no sentido de que possam ser comparáveis, permitindo uma análise robusta, e não fragilizada. Em terceiro lugar, para o autor, os fatos linguísticos não podem ser confundidos com manifestações individuais. Assim, retomase a fala de Labov (2008 [1972]), quanto ao aspecto individual da linguagem ser observado por meio do contexto social, enquanto o aspecto social é apreendido através da observação do indivíduo. Em quarto e último lugar, quanto à regra da coleta de dados, Moreno Fernández (1990) afirma que se deve observar os fatos linguísticos usando a técnica que melhor se encaixe em cada caso. Assim, procedimentos como as entrevistas rápidas e anônimas e o preenchimento de questionários, por exemplo, podem servir a um *continuum* de possibilidades específicas para cada propósito.

Quanto às (2) regras de análise, segundo grande grupo que direciona a metodologia de pesquisa sociolinguística para Moreno Fernández (1990, p. 33), esse autor considera que tais regras seguem, a princípio, três passos, quais sejam: (a) a codificação das respostas obtidas na fase de coleta, (b) a tabulação desses dados e (c) a aplicação de técnicas estatísticas. Dentro desses passos, o autor ressalta a importância de que sobretudo a fase (c) seja vista como um instrumento a mais de auxílio para a precisão da pesquisa, e não seja vista como um fim em si mesma, ou seja, utilizada meramente porque outros trabalhos já o fizeram.

Quanto ao terceiro e último grupo, (3) regras de interpretação, Moreno Fernández (1990) considera que a interpretação deverá estar em consonância aos objetivos do estudo proposto e à análise dos dados. Dentro desses objetivos, ele destaca quatro principais: (a) o de avanço de algo já conhecido, (b) o de descrição minuciosa, (c) o de apontar frequências em relação a fatores previamente selecionados e o (d) de apontar uma relação causal entre as variáveis observáveis. Ele finaliza sua exposição sobre as regras de interpretação, apontando a necessidade de que os estudos sociolinguísticos sejam sempre postos em continuidade e sejam relacionados a outras pesquisas anteriores. Nesse sentido, para ele, a interpretação deve servir a fins esclarecedores.

A fim de melhor visualizarmos as regras metodológicas da pesquisa sociolinguística, apontadas por Moreno Fernández (1990), observemos o Quadro 8.

**Quadro 8 – Regras do método sociolinguístico segundo Moreno Fernández
(1990, p. 27-38)**

1 Regras de coleta de dados	2 Regras de análise	3 Regras de interpretação
1.1 O pesquisador deve buscar distanciar-se do fenômeno, abandonando noções prévias.	2.1 Fase instrumental	3.1 A interpretação estará em correspondência com a finalidade do estudo e a análise dos dados.
1.2 O pesquisador deve buscar fenômenos exteriores, comuns e constantes.	2.1.1 Codificação dos dados	3.1.1 Avanço no conhecimento de algum fenômeno;
1.2.1 Fenômenos exteriores: fatos objetivos.	2.1.2 Tabulação dos dados	3.1.2 Descrição minuciosa
1.2.2 Fenômenos comuns: fatos que recebem mesmo tratamento analítico devem possuir ao menos um traço comum.	2.1.3 Aplicação de técnicas estatísticas	3.1.3 Determinação de frequências
1.2.3 Fenômenos constantes: fatos comparáveis entre si.	2.1.3.1 Descrição e resumo dos dados	3.1.4 Comprovação de hipóteses
1.3 Fatos sociolinguísticos são diferentes de manifestações individuais.	2.1.3.2 Estimativas de significação e confiabilidade	3.2 Relacionar os resultados obtidos com outros estudos disponíveis
1.4 Cada fato deve receber a técnica mais adequada a suas especificidades.		3.3 Estabelecer conceitos esclarecedores

Fonte: elaborado pela autora.

Após a exposição anterior acerca das questões relativas ao método variacionista, debruçemo-nos sobre dois conceitos capitais: variação e mudança linguística.

2.4 Variação e mudança linguística

Nesta seção, vamos nos concentrar em dois conceitos cruciais para a Sociolinguística: as noções de variação e mudança. Ao se analisarem as variantes, pode-se chegar a duas situações: a de variação, concluindo-se que as variantes apenas coocorrem entre si; e a de mudança em curso, quando uma das variantes tende a suplantá-la outra, e, devido ao fato de ser mais usada, tornar-se preferencial ao falante.

Teorizando a mudança linguística, Weinreich, Labov e Herzog (2006 [1975]) discorrem sobre alguns pressupostos gerais que norteiam o conceito. O primeiro desses pressupostos é que a mudança não pode ser vista como um fenômeno aleatório advindo da variação, que é inerente à fala. Na verdade, para os autores, a mudança começa quando uma alternância particular de

algum subgrupo assume uma direção que lhe dá um caráter de diferenciação ordenada. Nesse contexto, o falante nativo controla tais estruturas diferenciadas e vai fazendo uso delas ou não, conforme as regras que “governam a variação na comunidade de fala” (Weinreich; Labov; Herzog, 2006 [1975], p. 125). Esse processo não é uniforme nem instantâneo e envolve variações e mudanças durante períodos de tempo consideráveis. Além disso, não se pode afirmar que as mudanças estejam confinadas ao núcleo familiar, mas sim que ocorram dentro da comunidade como um todo. Trata-se, pois, de um fenômeno socialmente abrangente: “fatores linguísticos e sociais estão intimamente inter-relacionados no desenvolvimento da mudança linguística” (Weinreich; Labov; Herzog, 2006 [1975], p. 125). Além de socialmente abrangente, é importante salientar o caráter estruturado, ordenado da mudança. Isso implica dizer que uma mudança não impacta o aspecto sistemático da língua, que se mantém estruturada enquanto as alterações vão ocorrendo. Por isso é possível entender por que os falantes continuam se comunicando e se entendendo normalmente, apesar das mudanças ocorridas em sua língua.

Considerando isso, Coelho *et al.* (2015) definem o processo de mudança basicamente da seguinte forma:

é possível que, em seu repertório linguístico, um falante disponha de mais de uma forma para expressar o mesmo significado – temos aí a variação linguística. Ocorre que, dentro do próprio repertório linguístico desse falante, pode acontecer, também, um desfavorecimento gradual da forma original em prol da nova, de modo que a forma mais antiga assuma o estatuto de arcaica ou obsoleta e, aos poucos, deixe de ser usada. Nesse caso estamos diante de um processo de mudança linguística (Coelho *et al.*, 2015, p. 71-72).

Logo, variação e mudança são fenômenos intimamente relacionados. Para Weinreich, Labov e Herzog (2006 [1975]), toda mudança provém de um quadro de variação, embora nem toda variação evolua para uma mudança em progresso. Quando não há a obsolescência de alguma das formas em análise, convencionou-se dizer que se trata de variação estável de longo prazo.

Wardhaugh (2006 [1986], p. 193) compartilha da crença quase geral de que as línguas estão em constantes mudanças e, para ele, se as línguas mudam, os linguistas precisam ver essas alterações em seu andamento e reconhecê-las sistematicamente. Estudos recentes têm abordado a mudança como movimentos linguísticos que se difundem estruturalmente, daí os conceitos de onda e difusão. Nesse sentido, o autor defende que se possa interpretar a variação visível,

mesmo parcialmente, “como uma onda de mudança passando por uma língua”. Além disso, é necessário, ainda, ver “mudanças aparentemente se difundindo através de conjuntos de itens linguísticos semelhantes”. Para tanto, conforme conclui o autor, é preciso ter claro que o que se observa seja uma mudança de fato, e não uma flutuação esporádica.

Para finalizar este capítulo, cabe-nos ressaltar o que disseram Paiva e Duarte (2003, p. 132). Segundo as autoras, só nos é possível tirar conclusões consistentes sobre uma variação linguística em estudo “a partir da comparação cuidadosa e sistemática de amostras de fala representativas de diversos grupos sociais, geográficos e ainda de estilos de fala distintos”. Considerando isso, convém-nos mencionar que o terceiro capítulo deste trabalho, que versa sobre a Sociolinguística Variacionista, nos dá alguns elementos essenciais para que pensemos a variação da quebra de expectativa nos falares da RMBH e de Moçambique. Visto que vamos recorrer a dados de *corpora* representativos das duas variedades, ser-nos-á necessário empregar conceitos essenciais anteriormente discutidos, para que possamos analisar a expressão da adversatividade nas duas variedades propostas. Vejamos, pois, a seguir, no capítulo 3 deste trabalho, os procedimentos metodológicos adotados em nossa pesquisa sobre a marcação da quebra de expectativa no PB e no PM.

3 PROCEDIMENTOS METODOLÓGICOS

Optamos por subdividir este capítulo em algumas seções, objetivando deixá-lo mais claro e didático, embora todas as questões nele abordadas refiram-se, *grosso modo*, a aspectos ligados aos *corpora* e aos dados deles provenientes. Nesse sentido, vejamos, a seguir, nossa primeira seção, em que nos atemos a algumas propriedades gerais dessas questões supracitadas.

3.1 Procedimentos metodológicos: aspectos gerais

Segundo Tarallo (1986), “toda ciência – a linguística em nosso caso particular – tem uma teoria própria, um objeto específico de estudo e um método que lhe é característico” (Tarallo, 1986, p. 17). Como exposto, neste trabalho, o fato linguístico a ser analisado metodologicamente é o fenômeno conhecido como a quebra de expectativa ou adversatividade no PB e no PM. Para tanto, precisamos, antes, sistematizar alguns procedimentos inerentes à pesquisa Sociolinguística, que também nos auxiliarão em nossa tese.

Hudson (1980, p. 144) considera cinco etapas metodológicas essenciais a qualquer estudo sociolinguístico. São elas:

- (a) A seleção dos falantes, as circunstâncias e as variáveis linguísticas;
- (b) A coleta dos textos;
- (c) A identificação das variáveis linguísticas e suas variantes nos textos;
- (d) O processamento dos dados;
- (e) A interpretação dos resultados.

Esse autor salienta que o esperado é que os trabalhos sigam as etapas nessa ordem. Entretanto reconhece que o processo é cíclico e recursivo, na medida em que, por exemplo, nem todos os textos precisam estar definitivamente coletados antes do início do processo ou nem todas as variáveis podem ser identificadas e fechadas de modo a não se alterarem durante o curso da pesquisa. Isso se dá inclusive porque a seleção dos dados envolve hipóteses que podem sofrer adaptações conforme se avance nos estudos do fenômeno. Esse pensamento vai ao encontro do que preveem também Guy e Zilles (2007), segundo os quais:

não se deve esperar concluir a análise definitiva na primeira tentativa. Pelo contrário, como na maioria das pesquisas empíricas, hipóteses serão formuladas, testadas e refinadas, talvez algumas sejam descartadas e outras novas, criadas. Isso pode levar à incorporação de fatores ou grupos adicionais à análise (Guy; Zilles, 2007, p. 39).

Sobre a etapa (a), Hudson (1980, p. 144) considera que a seleção dos falantes, das circunstâncias e das variáveis linguísticas envolve “decisões extremamente importantes, que são, em certa medida, ditadas por hipóteses sobre os resultados esperados”. Em nosso caso, nos *corpora* selecionados para este trabalho¹⁶, encontramos, no PB, 802 ocorrências de *mas*; 72 ocorrências de *só que* e 1 de *porém*. No PM, os números foram: 651 ocorrências de *mas*, 62 de *só que*, 2 de *entretanto* e 1 de *no entanto*. Importante frisar, conforme veremos detalhadamente a seguir, que essas ocorrências referem-se apenas a estruturas de quebra de expectativa, sendo excluídos os itens que não tiveram esse valor.

Quanto à etapa (b) coleta de textos, o autor recomenda que o pesquisador evite deixar que o gravador tome o protagonismo, transformando o que teoricamente deveria ser uma conversa em uma entrevista de rádio, em que se perca a naturalidade do discurso. No que tange à fase (c), Hudson (1980) salienta a necessidade de se atentar ao grau de subjetividade no reconhecimento das variantes, pois, segundo ele, diferentes pesquisadores podem produzir diferentes análises do mesmo texto (Hudson, 1980, p. 145). Além disso, o autor evidencia a necessidade de registro sobre o ambiente linguístico em que se dão as variáveis, uma vez que ele pode influenciar na escolha de uma ou outra variante. Sobre isso, Guy e Zilles (2007, p. 50) afirmam: “cada ocorrência de uma variável vem inserida em um enunciado e em um contexto social, que poderia abarcar um grande número de fatores que influenciam a escolha do falante, feita a partir de uma gama de alternativas”.

Especificamente em nosso caso, optamos por verificar nos *corpora* em análise todos os marcadores adversativos mencionados pelos teóricos citados no capítulo 1 deste trabalho, a fim de não incorrerem no erro de ignorar alguma variante que pode ser produtiva em alguma das variedades em estudo.

¹⁶ Vale lembrar que os resultados obtidos serão mais bem discriminados no capítulo 4 “Resultados e análise dos dados sobre a expressão da quebra de expectativa no PB e no PM”.

No que se refere à fase (d), por sua vez, Hudson (1980) reforça a necessidade de haver uma precisão numérica, recomendando a redução dos números a porcentagens. Especificamente neste trabalho, decidimos ir além do percentual e, utilizando a ferramenta Goldvarb 3.0, tratamos estatisticamente os dados, posto que o *corpus* de PM é numericamente maior do que o do PB, possuindo 222.249 palavras, em contraposição ao *corpus* do PB, que possui 219.012 palavras. Sobre o programa estatístico Goldvarb 3.0, cabe-nos dizer que é um software pertencente ao Varbrul, conjunto de programas computacionais de análise estatística, especificamente estruturado para análise de dados de variação sociolinguística (Guy; Zilles, 2007, p. 105). Nesse sentido, por se tratar de uma ferramenta de análise específica para dados sociolinguísticos, e sendo nosso trabalho, como dito, de base variacionista, cremos ser coerente a utilização desse aplicativo.

Aplicamos, ainda, a normalização da frequência dos dados, por meio da seguinte fórmula:

$$\text{Frequência da palavra} / \text{número de palavras no texto} * 1000$$

Segundo esse autor, normalizar é ajustar as contagens de frequência brutas de textos de diferentes comprimentos para que possam ser comparados com precisão. Conforme mostra a fórmula acima, o número total de palavras em cada texto deve ser levado em consideração na contagem da frequência da norma. Especificamente, a contagem bruta de frequência deve ser dividida pelo número de palavras no texto e depois multiplicada por qualquer base escolhida para a normalização. Em nosso caso, pode-se optar pela base 1000, pelo fato de ser um número amplamente usado nos trabalhos com *corpora* menores, de até cinco mil palavras, como é o nosso caso, e sugerido por Biber (1998). Além disso, matematicamente falando, não faz diferença, por exemplo, dizer que um fenômeno linguístico ocorre com uma frequência de 10 por 100.000 palavras ou 100 por 1.000.000 de palavras, a razão é a mesma. O importante é manter a consistência e adotar a mesma base ao longo de toda a pesquisa. Acreditamos, por fim, que a base em mil, por indicar menores quantidades, pode facilitar nosso entendimento dos números e, conseqüentemente, dos dados.

Para argumentar sobre a seriedade e a necessidade da atenção do pesquisador quanto à normalização, Biber (1998) considera uma situação hipotética em que o linguista estaria pesquisando verbos modais. Como é praticamente impossível encontrar textos exatamente com o mesmo número de palavras, posto que, por exemplo, na entrevista sociolinguística, as

situações são muito variáveis – já que os falantes podem estar tímidos e falar menos ou podem se emocionar e acabar falando mais – é imprescindível garantir dados comparáveis. Para esse autor, encontrar 20 modais num texto de 750 palavras não é o mesmo que encontrar, por exemplo, 20 modais num texto de 1200 palavras. Portanto, a simples comparação das contagens brutas não fornece um relato preciso das frequências relativas dos modais nos dois textos. Dado isso, deve-se recorrer à fórmula da normalização. Vale lembrar, inclusive, que há calculadoras de normalização, como a da Universidade de Lancaster¹⁷, que facilitam o trabalho do pesquisador. Logo, pareceu-nos importante recorrer à fórmula de Biber (1998) para que pudessemos trabalhar com os *corpora* mesmo que apresentassem tamanhos diferentes, daí a importância de aplicarmos a normalização e também os programas estatísticos. Neste caso, como dito anteriormente, optamos pela ferramenta Goldvarb 3.0.

Por fim, em relação à etapa (e) de Hudson (1980), interpretação dos resultados, o autor a considera a fase mais difícil. Para ele, os resultados encontrados devem encaixar-se num quadro teórico sólido: “o sucesso nessa fase depende não só de uma metodologia correta em todas as fases anteriores, mas também de ter um enquadramento teórico geral adequado” (Hudson, 1980, p. 147). Diante disso, cremos ser importante enfatizar, mais uma vez, nosso quadro teórico, dissertado no capítulo 1 deste trabalho, momento em que buscamos elucidar conceitos importantes que perpassam a quebra de expectativa, dentro da perspectiva Funcionalista, e a Sociolinguística Laboviana, fundamentos que nos permitirão observar de modo mais robusto a expressão da adversatividade no PB e no PM.

Naro (2003), também destacando a subjetividade e a responsabilidade do pesquisador em relação aos aspectos metodológicos da pesquisa sociolinguística, considera que

a metodologia da Teoria da Variação constitui uma ferramenta poderosa e segura que pode ser usada para o estudo de qualquer fenômeno variável nos diversos níveis e manifestações linguísticas. As suas limitações são as do próprio linguista, a quem cabe a responsabilidade de descobrir quais são os fatores relevantes, de levantar e codificar os dados empíricos corretamente, e, sobretudo, de interpretar os resultados numéricos dentro de uma visão teórica da língua. O progresso da ciência linguística não está nos números em si, mas no que a análise dos números pode trazer para nosso entendimento das línguas humanas (Naro, 2003, p. 25).

¹⁷ https://www.lancaster.ac.uk/fss/courses/ling/corpus/blue/105_3.htm

Assim, cabe-nos, pois, mais uma vez, destacar que, neste trabalho, visando a dirimir possíveis inadequações interpretativas, nos inspiramos em Biber (1998), que nos adverte para não cairmos no ímpeto de olhar apenas os números brutos encontrados.

Consoante Guy e Zilles (2007, p. 32), “o objetivo final de qualquer estudo quantitativo em pesquisa dialetal não é produzir números, mas identificar e explicar fenômenos linguísticos”. Inclusive, para os autores, os números não respondem a nenhuma das indagações do pesquisador. Na verdade, os números são apenas “estatísticas inferenciais adicionais que podemos usar como indicadores empíricos na nossa busca por respostas” (Guy, Zilles, 2007, p. 42). Se, como dissemos, estamos diante de dados linguísticos desigualmente distribuídos, vale lembrar, por exemplo, a disparidade numérica de nosso *corpora*, em que PB apresenta 219.019 palavras em contraposição ao PM, que possui 222.249 itens, cabe-nos, neste trabalho, conforme inclusive sugerem Guy e Zilles (2007, p. 34), uma análise multivariada. Segundo os autores, essa modalidade de análise dá resultados mais precisos, porque ao mesmo tempo em que calcula o efeito de uma variável independente, consegue controlar a influência das outras variáveis independentes conhecidas. Nesse sentido, cremos estar seguindo os passos orientados pelos autores anteriormente citados. Desse modo, cabe mencionar, primeiramente, a identificação da variável linguística, em nosso caso, a expressão da quebra de expectativa, no PB e no PM. Em seguida, a seleção das variantes encontradas nas duas variedades em questão, quais sejam: *entretanto, mas, no entanto, porém e só que*.

Assim, emerge o conceito de análise de regra variável, que seria uma análise de cunho quantitativo, capaz de lidar justamente com o objetivo da Sociolinguística citado no capítulo 2 deste trabalho, o exame da variação estruturada, regida por regras, no uso da língua, e não o de uma variação supostamente livre e assistemática. Desse modo, as generalizações realizadas tornam-se, por exemplo, mais confiáveis, na medida em que não podem ser refutadas por contraexemplos isolados, mas sim por testes estatísticos robustos, compostos por um conjunto de dados previamente selecionados. Coadunando o que afirmam Guy e Zilles (2007), pretendemos adotar, em nossa metodologia, a análise de regra variável aferindo as ocorrências das variáveis, descrevendo as tendências e a extensão da variabilidade, bem como as “restrições ou fatores que a influenciam, mediante métodos estatísticos” (Guy; Zilles, 2007, p. 50). Além disso, cabe dizer que a análise de regra variável é multivariada, posto que considera que várias forças “simultâneas, interseccionadas e independentes” podem agir na variação. O resultado

desse tipo de análise se dá no “peso” e na “direção” de cada fator atuante. Esses autores definem a análise de regra variável da seguinte maneira:

uma “regra variável” é uma regra de reescrita sensível ao contexto que relaciona um par de variantes como $x - \langle y \rangle$, de modo que, quando a regra se aplica, ocorre “y” e, quando não se aplica, ocorre “x”. Cada uma das dimensões analíticas do contexto que determina se a regra se aplica ou não é chamada de “grupo de fatores” (...). Cada grupo é formado de um conjunto de “fatores” discretos (...). Cada combinação única de fatores descreve uma “célula” (...). Na análise quantitativa, cada fator recebe um “peso (“valor do fator” ou “probabilidade”) – um número entre 0 e 1, que caracteriza o efeito deste fator sobre a regra variável em questão (quanto mais alto o número, maior a chance de a regra se aplicar quando este fator estiver presente no contexto). E, finalmente, um conjunto total de dados caracteriza-se por uma “probabilidade de *input*” (“ p_0 ”), que é, de fato, uma medida global do índice de aplicação da regra. Uma função matemática é usada para combinar os valores dos fatores e a probabilidade de *input*, a fim de produzir os índices esperados de aplicação da regra em cada célula (Guy; Zilles, 2007, p. 51).

Essa concepção vai ao encontro do que Biber (2012 [1988]) chamou de análise multitraço e multidimensional de variação de registro, ou simplesmente análise multidimensional, que é uma metodologia de análise de *corpora* realizada por meio de procedimentos estatísticos. Conforme Sardinha (2004), a vantagem da análise multidimensional é o fato de ela permitir a correlação de diversas variáveis linguísticas a variáveis situacionais, como formalidade, impessoalidade, oralidade, período histórico, estilo etc. (Sardinha, 2004, p. 300).

Vale lembrar que a aplicação efetiva das metodologias de análise variável em nossos dados e os resultados delas obtidos serão detalhados no capítulo 4. Após essa exposição geral, vejamos, a seguir, algumas especificidades relativas às variedades de língua portuguesa estudadas neste trabalho.

3.2 Algumas considerações teóricas sobre o PB e o PM

Antes de tratarmos especificamente dos *corpora* que compõem este trabalho, convém-nos explicitar algumas considerações a respeito das duas variedades do português analisadas neste trabalho.

Sobre o português brasileiro, Perini (2016), por exemplo, considera-o o português falado no Brasil, “a língua que falamos, nós todos, operários, professores, mecânicos, médicos e manicures” (Perini, 2016, p. 31). Esse português, ao contrário da modalidade padrão escrita,

que seria acessada por uma minoria brasileira escolarizada, seria a língua acessada por todos os brasileiros, estudados ou não. Para confirmar essa fala, ele dá exemplos de concordância nominal e verbal, afirmando que estruturas como “aqueles livro” e “vou contar pra todo mundo que o Gui e a Bia *namorava* no jardim” são utilizadas por todos os brasileiros, “inclusive os de instrução universitária” (Perini, 2016, p. 32). Isso justifica inclusive a necessidade de estudo sistematizado dessa língua, que circula no território nacional e é falada por seu povo.

Bagno (2001[1961]) considera “um conjunto de mitos sem fundamento” a visão de que a língua “boa, certa e bonita” seria o português de Portugal. Por isso, aponta como solução pedagógica a denominação da língua falada no Brasil de *brasileiro*, distanciando-a da variedade europeia. Para ele,

estudar o brasileiro é dar voz à língua falada e escrita *aqui*, neste país chamado Brasil, 92 vezes maior que Portugal, habitado por uma população quase 17 vezes mais numerosa. É perceber que todas as línguas mudam, que toda língua é um grande corpo em movimento, em formação e transformação, nunca definitivamente pronto (Bagno, 2001[1961], p. 10).

Vital (2017), por sua vez, chama a língua falada no Brasil de *português do Brasil*, ressaltando que essa denominação pressupõe a existência de outras variedades portuguesas. “Há o português de Portugal, o português de Moçambique, de Timor Leste e outros. Em cada lugar desses, a língua portuguesa é usada de determinada maneira” (Vital, 2017, p. 19). Assim como Bagno (2001[1961]), Vital (2017) também contesta a ideia de uma suposta superioridade do português de Portugal em relação às demais variedades:

é muito natural e esperado que as línguas portuguesas sejam diferentes umas das outras, porque a história, as influências de outras línguas nas várias línguas portuguesas e outros aspectos foram também diferentes. Não podemos dizer, então, que o português de Portugal, por exemplo, é mais certo que o português de Angola: eles são simplesmente diferentes! (Vital, 2017, p. 21).

Nota-se, portanto, que os autores brasileiros apontam como um fator em comum a discrepância que o PB apresenta em relação às outras variedades da língua portuguesa. Nesse sentido, assim como Perini (2016), Vital (2017) e Bagno (2001[1961]) desvinculam o PB do português de Portugal padrão escrito, Timbane (2017), dissertando sobre a variedade moçambicana, afasta-a também do português padrão de Portugal. Para esse autor,

todas as línguas faladas tendem a mudar com o tempo desviando-se constantemente com relação à norma. Sendo assim, a norma não é apenas ou simplesmente um conjunto de formas linguísticas pré-estabelecidas, mas, também, é um agregado de valores socioculturais usados por uma comunidade linguística. Acerca mais especificamente do contexto moçambicano, observa-se que a escola apoia-se no português de Portugal (norma-padrão europeia) para ensinar e avaliar competências em português dos alunos, o que faz com que os alunos não progridam academicamente. É importante deixar claro que a variação não é exclusiva dos falantes não-escolarizados. Ninguém fala ‘norma-padrão’ a todo momento, pois ela é artificial, ou seja, não é língua materna de ninguém (Timbane, 2017, p. 19).

Nesse sentido, para o autor, não se nota a presença do português de Portugal na fala cotidiana moçambicana, mas sim a de uma variedade que se distancia da de Portugal. Trilhando um breve percurso histórico de formação do PM, Timbane (2017) sintetiza que se trata de uma língua que sofreu influências das Línguas Bantu¹⁸, vindas, por sua vez, da família Níger-congo. Espalhando-se pelo território moçambicano, foram criando-se dialetos, que se somaram a línguas fronteiriças e estrangeiras, constituindo uma extensa diversificação linguística no país. Nesse contexto, a imposição do português de Portugal como a modalidade padrão estaria no fato de os colonizadores utilizarem-no como um fator de dominação. Para reforçar essa conjuntura, criavam-se mitos afirmando que as línguas nativas seriam incapazes de exprimir noções científicas e complexas. Inclusive, as Línguas Bantu moçambicanas faladas em Moçambique e em Angola “eram chamadas preconceituosamente por *pretoguês, língua do cão, landim, dialeto, língua dos pretos*, etc.” (Timbane, 2017, p. 22). Por fim, Timbane (2017) afirma que o PM possui características próprias, sobretudo lexicais, semânticas e morfossintáticas, que o distanciariam não só do PE, mas de todas as outras variedades da língua portuguesa.

Quanto à variação lexical, há vocábulos moçambicanos que refletem diretamente a cultura local. É o caso, por exemplo, de *mata-bicho*, chamado de *pequeno almoço* em português de Portugal e *café da manhã* em português brasileiro. O *mata-bicho* moçambicano teria vindo de *ku dlaya nyocana*, que significa matar o bicho em língua xichangana, expressão usada para se referir à primeira refeição do dia (Timbane, 2017, p. 29). Outros casos, também de influência Bantu, seriam *wasso-wasso*, para designar *feitiços de amor* e *tchapo-tchapo*, que, em PE e PB, seria *rápido*. Sobre a variação sintática, Timbane (2017) destaca, por exemplo, alguns casos de concordância redundante, em que se nota a flexão do infinitivo, contrastando com o PE, como vemos em “os alunos propuseram *fazerem* o trabalho em dois dias (PE=fazer)” (Timbane, 2017,

¹⁸ Neste trabalho, adotamos a grafia Bantu, coadunando Timbane (2017).

p. 30 – grifo do autor). Por fim, ilustrando a variação fonológica, o autor cita a inclusão da nasalização depois de vogais, como *convinte*, que, em PB e PE seria *convite* (Timbane, 2017, p. 30).

Observemos, no quadro a seguir, algumas diferenças entre PM e PE apontadas por Timbane (2017), quadro em que também incluímos as particularidades do PB, coadunando a pesquisa comparativa entre PB e PM feita neste trabalho.

Quadro 9 – Alguns casos de variação linguística entre PB, PM e PE, com base em Timbane (2017)¹⁹

Nível da variação	PB	PM	PE	Observações
Lexical	café da manhã	mata-bicho	pequeno almoço	Para Timbane (2017), a variação está intimamente ligada à cultura moçambicana e expressa a integração de várias palavras provenientes das Línguas Bantu moçambicanas.
	casa dos sogros	sograria	casa dos sogros	
	Réveillon	cortar o ano	Réveillon	
	subornar	falar-alto	subornar	
	feitiço de amor	wasso-wasso	feitiço de amor	
	rápido	tchapo-tchapo	rápido	
	mochila ou pasta	pasta	mochila	
	horta	machamba	horta	
	idoso	madala	idoso	
	lanchonete	baraca	lanchonete	
	xilofone	timbila	xilofone	
	chifre	pala-pala	chifre	
	lenço	capulana	lenço	
	carrinho de mão	tchova	carrinho de mão	
	mamãe	mamana	mamãe	
marginal	molwene	marginal		
rapaz	mufana	rapaz		
Sintático	Meu irmão ganhou uma bolsa de estudos.	O meu irmão foi concedido uma bolsa de estudos.	Ao meu irmão foi concedida uma bolsa de estudos.	As interferências sintáticas também provêm da influência das Línguas Bantu moçambicanas. Nota-se, nesse caso, a ausência de preposições, modificando a regência de alguns verbos, como <i>conceder</i> , <i>nascer</i> , <i>afirmar</i> e <i>sair</i> . Além da regência,
	Você também pode ter um filho saudável.	Tu também podes nascer um filho saudável.	Tu também podes dar à luz um filho saudável.	
	O presidente afirmou que não sabia .	O presidente afirmou que não sei .	O presidente afirmou que não sabia .	
	Ele saiu de casa muito cedo.	Ele saiu em casa muito cedo.	Ele saiu de casa muito cedo.	
	Os alunos propuseram	Os alunos propuseram fazerem o	Os alunos propuseram	

¹⁹ No quadro 9, mantivemos a grafia original utilizada em Timbane (2017).

	fazer o trabalho em dois dias.	trabalho em dois dias.	fazer o trabalho em dois dias.	vê-se o não seguimento da concordância padrão europeia, em que se observa a redundância da flexão verbal, como em <i>propuseram fazerem e deviam criarem</i> .
	Os chefes deviam criar melhores condições para todos.	Os chefes deviam criarem melhores condições para todos.	Os chefes deviam criar melhores condições para todos.	
	Jovem universitário, procure o seu lugar.	Jovem universitário, procure o teu lugar.	Jovem universitário, procure o seu lugar.	É comum ocorrer a neutralização das formas de tratamento <i>tu</i> e <i>você</i> .
	Você não tinha nada que falar, porque ele não é seu irmão.	Você não tinha nada que falar, porque ele não é teu irmão.	Você não tinha nada que falar, porque ele não é seu irmão.	
Fonético-fonológico	convite	convinte	convite	Inserção da nasal.
	economia	enkonomiya	economia	
	exagero	enzagero	exagero	
	exame	enzame	exame	
	existe	enzixte	existe	
	caça	kasa	caça	Desvozeamento dos sons consonânticos, /z/-/s/; /d/-/t/.
	Gaza	kasa	Gaza	
dedo	teto	dedo		

Fonte: elaborado pela autora, com base em Timbane (2017).

Um aspecto interessante destacado por Gonçalves (2011) sobre o distanciamento entre o português de Portugal e o português moçambicano é o fato de que “apesar de quatrocentos anos de colonização portuguesa, apenas cerca de 6,5% dos vinte milhões de habitantes de Moçambique têm o português como língua materna; os restantes aprendem primeiro uma língua bantu” (Gonçalves, 2011, p. 202). Endossando Gonçalves, Vieira e Brandão (2025) chamam a atenção para a complexidade envolvendo a questão linguística moçambicana, observando, inclusive, a questão política, pelo fato de o português ser considerado a única língua oficial, apesar de o território compartilhar mais de 20 Línguas Bantu, que, conseqüentemente, não são vistas como oficiais ou sequer nacionais. Ademais, o português não possui o status de língua mais falada no país, sendo esse posto ocupado pela língua emakhuwa, seguida pela língua portuguesa. Esse movimento de preservar o português como única língua oficial, como dito, aproxima-se de questões políticas, na medida em que o uso de um idioma colonial supostamente evitaria conflitos étnicos, no sentido de favorecer uma ou outra língua regional, fazendo da herança colonial uma forma de preservação do território. Por isso o português segue ocupando primazia linguística na nação moçambicana.

Para fechar esse raciocínio, em que consideram as línguas faladas em Moçambique como um *continuum*, sobre o qual transitam os falantes, as autoras reforçam a complexidade implicada nesse processo:

os moçambicanos, embora cientes da importância de falar português, inclusive como meio de ascensão social, parecem resistir a seu domínio, adotando as diferentes normas que se vão constituindo em razão do contato multilinguístico. A tal complexidade somam-se questões relacionadas à busca de identidade dos indivíduos e à atitude deles para com as línguas em contato no meio da sociedade (Vieira; Brandão, 2025, p. 54).

Outro aspecto interessante destacado pelas autoras, que aproxima o argumento delas ao de Timbane, que, por sua vez, como dito, distancia o português moçambicano do português de Portugal, está no fato de elas considerarem esse contexto favorecido por haver pouca exposição ao *input* da variedade europeia e o constante contato com as Línguas Bantu, geralmente línguas maternas aos falantes. Esse fenômeno tem sido chamado de “nativização”, processo no qual o português, língua originalmente colonial, vai se aproximando do contexto sociocultural do novo país formado após a colonização (Vieira; Brandão, 2025, p. 73).

Cabe-nos, pois, neste momento, considerar fatores de similitude e também de discrepância entre o PB e o PM, posto que, tanto os pesquisadores da variedade brasileira quanto os da variedade moçambicana, acima mencionados, não traçam especificamente esse paralelo. Primeiramente, um possível fator de similaridade a ser destacado entre ambas as variedades, conforme apontado no Quadro 9, é o fato de o PB também possuir itens lexicais que refletem as línguas de sua formação histórica. Assim como o PM sofreu influências das Línguas Bantu, o PB sofreu influências, por exemplo, das línguas africanas e indígenas, de sua população originária, como o tupi-guarani, dela originando-se palavras como *carioca*, conforme vemos no exemplo a seguir, de nosso *corpus* de PB.

(60) Olha o Zé *Carioca* ali (Raso; Mello, 2012).

Outros exemplos de PB em nível lexical seriam as palavras *babaca* e *zoar*, ambas de origem africana, conforme vê-se em (61) e (62):

(61) No' / o Galáticos é mesmo / todo mundo é <*babaca*> (Raso; Mello, 2012).

(62) Cê fala que é Futebol Arte / a galera começa a *zoar* (Raso; Mello, 2012).

No nível sintático, por sua vez, também é possível, no PB, a transformação, na fala, de verbos inacusativos em transitivos diretos, como no seguinte exemplo moçambicano, já explicitado no Quadro 9:

(63) Tu também podes *nascer* um filho saudável (Timbane, 2017, p. 30).

No PB, poderíamos produzir frases como (64):

(64) É só você *vacinar* a dose certa (exemplo de própria autoria).

Não encontramos especificamente em nosso *corpus* de PB esse uso de verbos inacusativos como transitivos diretos, mas parece-nos possível tal ocorrência, conforme expusemos em (64)²⁰.

Por outro lado, uma regência de PM influenciada pelas Línguas Bantu, conforme evidenciado no Quadro 9, o uso da preposição *em* como complemento do verbo *sair* não ocorre no PB, que aciona a preposição *de* nesses casos, por exemplo em (65):

- (65) a. Ele saiu *de* casa (PB).
b. Ele saiu *em* casa (PM) (Timbane, 2017, p. 30).

A ausência de concordância verbal apontada por Timbane (2017) é outro fator de similaridade com relação ao PB. Observemos (66), exemplo também presente no Quadro 9:

(66) Muitos já não *respeita* a tradição (Timbane, 2017, p. 30).

²⁰ Chierici (2008) trabalha esses casos de inacusatividade, utilizando exemplos como “os homem chegou tudo”, e também casos de inergatividade, como em “ele não nadou nada”.

Em nosso *corpus* brasileiro, também encontramos casos de concordância verbal em 1ª do singular, quando a norma padrão indicaria 3ª do plural, o que nos mostra que (66) seria um uso possível em PB. Vejamos no exemplo a seguir:

(67) Es *deve* meter o pau (Raso; Mello, 2012).

Por outro lado, um fator de dissonância entre PB e PM, no nível sintático, seria a concordância de infinitivo redundante, que é observada por Timbane (2017) no PM, mas não foi percebida em nosso *corpus* de PB. Esse autor moçambicano cita o exemplo (68), conforme exposto no Quadro 9:

(68) Os alunos *propuseram fazerem* o trabalho em dois dias (Timbane, 2017, p. 30).

Em PB, apenas o primeiro verbo sofreria flexão: “os alunos propuseram *fazer* o trabalho em dois dias”. Vejamos em nosso *corpus* um exemplo desse fenômeno:

(69) Eu acho que es nũ *deviam* mais *participar* (Raso; Mello, 2012).

Em (69), nota-se a flexão apenas do verbo auxiliar *deviam*, ficando o verbo principal *participar* no infinitivo impessoal.

Por fim, no nível fonético-fonológico, conforme apontado no Quadro 9, não se observam em nosso *corpus* de PB a inserção da nasal. ou o desvozeamento dos sons consonânticos, fenômenos descritos por Timbane (2017) relacionados ao PM. Em (70), a seguir, vemos, por exemplo, *arbitragem* e não **arbintragem*.

(70) Negócio de *arbitragem* / também // \$ acho <que a gente> tem / que olhar direito (Raso; Mello, 2012).

Em (71), a seguir, vemos *taça*, e não **tasa*.

(71) A gente podia fazer a *taça* aqui / todo mundo vai adorar (Raso; Mello, 2012).

Nesta seção, nos propusemos a dissertar sobre alguns aspectos pertinentes às duas variedades linguísticas analisadas neste trabalho, quais sejam, PB e PM. Iniciamo-la considerando o que Perini (2016), dentre outros autores, chama especificamente de português brasileiro, que seria a língua acessada pelas mais diversas camadas brasileiras, da classe trabalhadora às elites; dos mais jovens aos mais velhos, em localidades urbanas ou interioranas, escolarizados ou não. Logo, o PB seria a língua espontânea, em contraposição à norma padrão, modalidade mais artificial e ligada a normas e convenções linguísticas. De modo similar a Perini (2016), Timbane (2017) disserta sobre o português moçambicano, distanciando-o da norma, sobretudo da variedade europeia, tida como padrão. Para tanto, esse autor moçambicano aponta fatores que contribuíram para a configuração do PB, como a forte influência das Línguas Bantu. Esse também é outro fator de similitude entre as duas variedades, pois, como vimos, o português brasileiro também foi influenciado por outras línguas, como as indígenas e africanas. Nesta seção, expusemos ainda alguns exemplos que demonstram justamente as variações das duas línguas quanto a aspectos lexicais, sintáticos e fonológicos, fatores estes que ora aproximam as duas variedades, ora as contrastam.

Por fim, após essas considerações teóricas sobre as línguas faladas nas localidades onde os *corpora* foram coletados, apontando fatores de aproximação e também de distanciamento, vejamos, a seguir, algumas especificidades sobre cada *corpora*.

3.3 Os *corpora*

Consoante Moreno Fernández (2017, p. 320), um *corpus* é uma “coleção” de materiais de linguísticos. Percorrendo um histórico da Linguística de *corpus*, esse autor salienta que

não é que antes não se trabalhasse com coleções de materiais; é que o desenvolvimento da informática tem permitido armazenar e consultar, a velocidade assustadora, conjuntos de materiais de grande tamanho, abrindo maiores possibilidades de estudo e aplicação (Moreno Fernández, 2017, p. 320).

Assim, nota-se que a ascensão de ferramentas tecnológicas tem agregado valor a essa área de pesquisa, muitas vezes facilitando o trabalho de coleta do pesquisador e permitindo, também, análises mais acuradas. Rajagopalan (2007) disserta que a Linguística de *Corpus* é um campo sujeito a modificações e inovações constantes, posto que trabalha com a língua, algo em construção e em constante evolução. Para ele, a ideia norteadora por trás desse campo de

conhecimento se resume no fato de que “a natureza e o funcionamento da língua podem e devem partir das observações de dados concretos” (Rajagopalan, 2007, p. 33).

Quanto ao conceito de *corpus*, podemos recorrer às palavras de Vasilévski (2007, p. 48). Segundo essa autora, um *corpus* é “uma densa massa linguística”, que configura um meio de estudo da língua. Para ela, “quaisquer textos, de quaisquer tamanhos” podem compor um *corpus* linguístico, desde que esses textos sejam autênticos. Nesse sentido,

um *corpus* deve ser composto de dados coletados em situações reais de uso da língua para a comunicação, de modo que não pode ter havido qualquer finalidade de pesquisa linguística quando eles foram externalizados pelos usuários da língua. Outros quesitos essenciais de um *corpus* são: que seja eletrônico (o que hoje se tornou óbvio), permita aplicação dos recursos computacionais de armazenamento, busca, seleção e impressão de informações e esteja em formato que o possibilite ser lido pelos programas mais utilizados atualmente (Vasilévski, 2007, p. 50).

Creemos, pois, que nosso propósito se aproxima ao que prevê Vasilévski (2007), posto que estamos lançando mão de dois *corpora* anotados, transcritos e armazenados computacionalmente, quais sejam, C-Oral Brasil I (2012) e Corporaport (2016), passíveis inclusive de serem submetidos a softwares eletrônicos, como propusemos na Introdução deste trabalho. Nesse sentido, pretendemos analisar a expressão da quebra de expectativa utilizando, para o PB, o *corpus* C-Oral Brasil I (Raso; Mello, 2012) e, para o PM, o *corpus* Corporaport (Vieira; Pissurno, 2016). Sobre o primeiro, trata-se de um projeto que visa ao estudo da fala espontânea, sobretudo da diatopia mineira. É composto por textos em forma de diálogos, conversações e monólogos. O C-ORAL-BRASIL tem sua sede no Laboratório de Estudos Empíricos e Experimentais da Linguagem (LEEL) da Faculdade de Letras da Universidade Federal de Minas Gerais, sob a coordenação dos professores Tommaso Raso e Heliana Mello. Seus pesquisadores estão vinculados ao Núcleo de Estudos em Linguagem, Cognição e Cultura (NELC)²¹. Já em relação ao segundo, trata-se de um projeto de *corpora* de variedades do português. Neste trabalho, utilizamos a Amostra Maputo (MAP). As organizadoras desse *corpus* reconhecem a limitação do projeto, posto que a ideia inicial seria coletar a fala apenas de entrevistados que tivessem o português como língua materna, entretanto, dadas as mais variadas razões atinentes às pesquisas de campo, isso não foi possível. Desse modo, a MAP é resultado de amostras de falas de falantes do PM que a tiveram como língua materna ou segunda

²¹ Informações disponíveis no site do projeto: <https://www.c-oral-brasil.org/projeto.php>.

língua (Vieira; Brandão, 2025). Vale lembrar que a escolha por esses *corpora* se deveu ao fato de serem bastante representativos da fala espontânea, tão cara aos estudos sociolinguísticos, e também ao fato de serem *corpora* que seguiram rigorosa metodologia de transcrição e anotação, o que nos aproxima do rigor científico necessário a uma tese.

Como dito, um aspecto importante a ser por nós frisado neste momento seria reconhecer o fato de o português brasileiro ser uma língua materna, em contraposição ao português moçambicano, que se trata de uma segunda língua, em muitos casos. Embora reconheçamos a peculiaridade dos contextos de cada uma das variedades e as especificidades envolvidas, o fato de os dados encontrarem-se estratificados dá-nos segurança e confiabilidade na análise.

Nesse sentido, cabe justificar também nossa opção por diferentes *corpora* neste trabalho. Como iniciamos nossa pesquisa da variedade brasileira utilizando o C-Oral Brasil I (2012), pareceu-nos interessante mantê-lo para esta variedade, posto que o Corporaport (2016) não possui amostras tão vastas da variedade brasileira, embora apresente uma quantidade considerável de dados para a modalidade moçambicana. Assim, em nossa pesquisa, chegamos à conclusão de que as opções mais abrangentes para cada modalidade seriam esses dois *corpora*, uma vez que são extensos, representativos e bem anotados.

Para Sardinha (2004), *corpus* são textos naturais existentes na linguagem, que podem ser usados como base para a pesquisa linguística, mas que, a princípio, não foram criados com esse fim. O termo “natural”, nesse caso, sugere textos autênticos, produzidos por seres humanos, excluindo-se, portanto, produtos advindos de inteligência artificial. Nesse sentido, “embora os textos devam ser naturais (autênticos e independentes do *corpus*), o *corpus* em si é artificial, um objeto criado com fins específicos de pesquisa (Sardinha, 2004, p. 17). Esse autor destaca seis pontos cruciais para o estabelecimento de um *corpus* linguístico, quais sejam: a origem, o propósito, a composição, a formatação, a representatividade e a extensão. Em relação ao primeiro critério, deve-se considerar a autenticidade dos dados, que, como dito, não podem ser, por exemplo, advindos de inteligência artificial ou produzidos por meios tecnológicos. Quanto ao propósito, vale lembrar que um *corpus* tem a finalidade de ser objeto de estudo para a pesquisa linguística, sua finalidade, pois, seria a pesquisa acadêmica. No que se refere à composição, o conteúdo do *corpus* deve ser previamente escolhido. Por exemplo, como dissemos no capítulo 2, deve-se privilegiar narrativas, para buscar uma naturalidade pelo falante, que deverá abstrair-se do conteúdo linguístico em si e demonstrar uma linguagem o mais natural possível. Em relação à

formatação, os dados devem ser passíveis de serem submetidos a ferramentas computacionais. Em relação à representatividade, como o próprio nome já sugere, um *corpus* linguístico deve ser representativo de uma língua ou variedade. E, por fim, quanto à extensão, o *corpus* deve, além de ser representativo, possuir a extensão necessária para que possa efetivamente representar uma comunidade de fala (Sardinha, 2004, p. 18-19).

Sobre o C-Oral Brasil I, trata-se de um *corpus* vinculado à Faculdade de Letras da Universidade Federal de Minas Gerais e tem como objetivo disponibilizar amostras de fala do PB, sobretudo da sintopia mineira. Compõe-se de 139 textos e 208.130 palavras, divididos em diálogos (1/3), conversações (1/3) e monólogos (1/3). O Corporaport, por sua vez, é também um *corpus* acadêmico, vinculado à Faculdade de Letras da Universidade Federal do Rio de Janeiro, que visa a criar um banco de dados para análises linguísticas, compondo-se de amostras de variedades do português de Portugal, moçambicano e brasileiro. Em ambos os *corpora*, os falantes encontram-se estratificados segundo o gênero, a faixa etária e o grau de escolaridade, estratificação esta que pretendemos preservar e adotar em nosso trabalho, para verificar se há influência de fatores extralinguísticos no fenômeno linguístico analisado, tendo em vista as formas para a marcação da quebra de expectativa citadas pelos teóricos mencionados no capítulo 1 deste trabalho.

O C-Oral Brasil I, base para nosso estudo do PB, estratifica suas amostras também em dois gêneros, quais sejam, masculino e feminino. Quanto ao fator idade, a estratificação se dá em quatro faixas. São elas: A: 18 a 25 anos; B: 26 a 40 anos; C: 41 a 60 anos; D: acima de 60 anos. Por fim, quanto ao fator escolaridade, há também quatro faixas: Faixa 1: ausência de escolarização ou escolarização até o nível primário incompleto (não mais de 7 anos de escolarização); Faixa 2: até o ensino superior completo, desde que se exerça uma profissão que não o requeira; Faixa 3: ensino superior, exercendo-se uma atividade que exige esse título; Faixa X: não conhecemos o nível escolar.

O Corporaport, base para nosso estudo do PM, divide suas amostras em três faixas etárias, três níveis de escolaridade e dois gêneros. Quanto às faixas etárias, temos: A: 18 a 35 anos; B: 36 a 55 anos; C: 56 a 75 anos. Quanto aos níveis de escolaridade, há: 1: Ensino Fundamental II (6º ao 9º anos); 2: Ensino Médio (10º ao 12º anos); 3: Ensino Superior. Por fim, quanto aos gêneros, temos: masculino e feminino.

Vejamos, a seguir, o quadro 10, em que ilustramos os fatores extralinguísticos de cada *corpus*.

Quadro 10 – Critérios extralinguísticos do C-Oral Brasil I e do Corporaport

	Corporaport			C-Oral Brasil I			
Faixa etária	A 18-35	B 36-55	C 56-75	A 18-25	B 26-40	C 41-60	D +60
Escolaridade	1 6º-9º	2 9º-12º	3 Ens. Sup.	1 0-9º	2 Ens. Sup. (sem atuação)	3 Ens. Sup. (com atuação)	X Desco- nhecida
Gênero	M / F			M / F			

Fonte: elaborado pela autora.

A fim de aproximarmos os dois *corpora* e minimizarmos divergências metodológicas, desconsideramos, no C-Oral Brasil I, os falantes de nível escolar desconhecido e juntamos numa mesma categoria os falantes que possuem ensino superior, independentemente de trabalharem em sua área de formação ou não, pois a atuação no mercado de trabalho não influencia diretamente o ensino que a pessoa recebeu, posto que a questão profissional é algo posterior ao momento do estudo. Isto é, independentemente de trabalhar na área ou não, o indivíduo graduou-se, recebendo o nível de instrução exigido. No Corporaport, por sua vez, optamos por juntar os falantes de nível fundamental e médio. Desse modo, obtivemos dois grandes níveis de escolaridade em ambos os *corpora*: no primeiro, enquadramos os falantes que possuem escolaridade até o ensino médio; e, no segundo, falantes que possuem ensino superior.

Quanto ao fator idade, optamos por manter os critérios adotados no C-Oral Brasil I, distribuindo os falantes do PM em quatro faixas etárias: 18-25, 26-40, 41-60 e +60. Vejamos, no quadro 11, as decisões quanto à estratificação dos participantes adotada e o resultado final dessas decisões para a nossa análise.

Quadro 11 – Estratificação adotada neste trabalho

Critérios	Subgrupos			
Faixas etárias	A 18-25	B 26-40	C 41-60	D +60
Escolaridades	P escolaridade até o ensino superior		U ensino superior completo	
Gêneros	M masculino		F feminino	

Fonte: elaborado pela autora.

Após essa exposição dos fatores extralinguísticos, como fatores internos, isto é, linguísticos, observamos, a partir dos dados disponíveis, as posições em que os marcadores de quebra de expectativa ocorrem, considerando, portanto, a localização deles no período coordenado (se em início de período, entre as duas orações coordenadas ou em final do período). Portanto, como primeira aspecto linguístico estabelecemos a posição do item conjuncional no período composto por coordenação. Como segundo aspecto, observamos se os marcadores foram antecidos por outros itens conjuncionais adversativos, como se dá em (72).

- (72) Aquilo é fuxico *mas só que* é de uma maneira mais sofisticada (Raso; Mello, 2012).

Considerando, por exemplo, (72), temos como hipótese que *só que*, por ser ainda recente na língua, não se encontra totalmente encaixado no quadro dos marcadores de quebra de expectativa, sendo, portanto, “menos expressivo”, menos carregado de sentido que *mas* dentro do fenômeno, precisando, pois, apoiar-se em outros marcadores para expressar-se. Nesse sentido, teríamos, como segunda variável independente a ser analisada a precedência ou não de outras conjunções ou coordenadores adversativos. Inclusive, para Pezatti e Longhin-Thomazi (2008),

o português dispõe de um conjunto de elementos e sintagmas de classificação fluida – *contudo, todavia, no entanto, entretanto, porém* – que tangenciam as categorias de advérbio e conjunção. Ou seja fazem coordenação mas não atendem a todos os critérios que ajudam a caracterizar uma conjunção coordenativa prototípica (como *e, ou e mas*). Tais partículas de natureza incerta podem, por exemplo, coocorrer com outra conjunção coordenativa (Pezatti; Longhin-Thomazi, 2008, p. 929).

Para ilustrar esse raciocínio da certa dependência dos itens conjuncionais adversativos a outras conjunções, as autoras dão o seguinte exemplo:

- (73) ... a lei é feita para o homem... para proteger o homem... *e no entanto* o homem está sujeito e até certo ponto escravo da lei... (Pezatti; Longhin-Thomazi, 2008, p. 929).

Nesse caso, vemos a locução *no entanto* precedida da conjunção aditiva *e*. Para as autoras, itens de classificação fluida, que transitam entre as categorias de advérbio e conjunção, tenderiam a ser mais dependentes, precisando apoiar-se em itens conjuncionais mais sólidos, que confeririam a esses itens fluidos o estatuto de conjunção. Elas justificam essa dependência pelo fato de os advérbios, que constituíram fontes para as conjunções, passarem pelo processo de gramaticalização:

historiadores da língua sustentam que advérbios e perífrases adverbiais que no português arcaico se justapunham a *mas* com o propósito de ênfase ou reforço do sentido de oposição (formando perífrases como *mas porém*, *mas entretanto*), no curso do tempo e em função da contiguidade sintática, passaram a ter o sentido de oposição mesmo quando usados sozinhos, sem o *mas*, tornando-se assim conjunções, ainda que não prototípicas (Pezatti; Longhin-Thomazi, 2008, p. 929-930).

Após essa breve justificativa quanto aos fatores internos e externos de nossa análise da quebra de expectativa no português moçambicano e no português brasileiro, cabe-nos, pois, neste momento, dissertar sobre o processo de codificação dos dados. Inicialmente, criou-se uma planilha em formato do Excel, na qual foram colocados os contextos de ocorrência de todos os marcadores de quebra de expectativa identificados nos *corpora* brasileiro e africano, quais sejam, como anteriormente dito: *entretanto*, *mas*, *no entanto*, *porém* e *só que*. Posteriormente, foram criados códigos para cada marcador e também para os outros fatores a serem analisados. São eles: gênero, escolaridade, idade, posição do marcador no período composto, precedência ou não de outros marcadores adversativos e origem do *corpora*. No apêndice 1, ilustramos os fatores e os códigos atribuídos.

Após a visualização dos códigos utilizados para o tratamento dos dados estudados neste trabalho, observemos, na seção a seguir, alguns aspectos relativos às conjunções e às locuções conjuntivas adversativas consideradas.

3.4 Algumas especificidades sobre os marcadores de quebra de expectativa

Recuperando os números encontrados na seção 3.1, encontramos, no PB, 802 ocorrências de *mas*; 72 ocorrências de *só que* e 1 de *porém*. No PM, os números foram: 651 ocorrências de *mas*, 62 de *só que*, 2 de *entretanto* e 1 de *no entanto*. Vejamos, na Tabela 1, esses números explicitados.

Tabela 1 – Quantificação dos marcadores de quebra de expectativa encontrados nos corpora de PB e PM

Marcador	Ocorrências em PB	Ocorrências em PM
entretanto	0	2
mas	802	651
no entanto	0	1
porém	1	0
só que	72	62
Total	875	716

Fonte: elaborado pela autora.

A respeito dos conectores de quebra de expectativa, convém ressaltarmos, mais uma vez, que, conforme os objetivos de nosso trabalho, foram desprezados os dados que não correspondessem ao valor adversativo. Consideremos, pois, alguns casos. Primeiramente, vejamos o marcador *antes*. Consoante o dicionário Priberam (2024), o vocábulo *antes* trata-se de um advérbio, que significa, por exemplo, ‘em tempo anterior’. Esse sentido temporal foi, portanto, ignorado em nossa coleta, posto que, como mencionado, nos atemos aos sentidos de quebra de expectativa, que ocorrem em *antes* quando este equivale a *mas*, como no exemplo (74).

(74) O bom educador não proíbe, *antes* (mas) orienta (Só Português, 2007).

Por outro lado, vejamos, a seguir, um exemplo do item *antes* com valor temporal encontrado em nosso *corpus* do PM, dado, conseqüentemente não considerado em nossa análise.

- (75) Se eu tiver a falar frequentemente como falava *antes* tenho que gastar dois mil (Pissurno; Vieira, 2016).

Em (75), o vocábulo *antes* expressa um acontecimento anterior a outro, não havendo, portanto, uma expectativa a ser quebrada. Por isso, optamos por não considerar tais marcadores neste trabalho.

Quanto a *mas*, por sua vez, quando ocorresse em seu sentido circunstancial de intensidade, usado para reforçar ou enfatizar algo, foi também excluído, como em:

- (76) Eu comprei um xampu esses dia... Nossa *mas* ele é uma maravilha (Raso; Mello, 2012).

Além do sentido adverbial, foi excluído o *mas* marcador discursivo, usado sobretudo na função fática da linguagem, para dar sequência ao ato comunicativo, como em:

- (77) <Ô / *mas*> / voltando à questão / falando em e também falando em povo mascarado / esse povo do Galáticos é muito palha (Raso; Mello, 2012).

Em (77), vê-se que o falante recorre ao item *mas* para retomar o assunto anterior. Não se nota, pois, alguma expectativa a ser quebrada, o que justifica sua exclusão dos dados deste trabalho.

Em relação ao item *senão*, por sua vez, foram ignorados os sentidos da conjunção alternativa ou mesmo o significado substantivo, que indica ‘defeito’, ‘imperfeição’. Vejamos cada um desses usos, respectivamente.

- (78) Leve agasalhos, *senão* sentirá frio (Cegalla, 2012, p. 380).

- (79) Não havia um *senão* no texto (Cegalla, 2012, p. 380).

Primeiramente, em (78), vemos a alternância das ideias de levar agasalho ou sentir frio, daí o sentido alternativo. Em (79), vemos *senão* indicando que não havia defeitos no texto em

questão. Logo, devido ao fato de nenhum desses itens sugerir quebra de expectativa, nosso objeto de interesse, como dito, foram excluídos de nossa análise.

Por fim, considerando a perífrase *em todo caso*, excluímos de nossa amostra o valor adverbial intensificador ou modal, como em

(80) *Em todo caso*, vamos à festa (exemplo de própria autoria).

Como o próprio nome sugere, o valor modal de (80) pode ser interpretado pela ideia de irmos à festa *de qualquer modo*, *em todo o caso*, o que não demonstra quebra de expectativa, mas valor circunstancial modal. Dessa forma, posto que não foi encontrado em nenhum dos *corpora* o valor dessa expressão em situação de adversatividade, foi também um conector não considerado neste trabalho. A título de exemplificação, vejamos o uso dessa expressão em valor de quebra de expectativa, retirada do portal eletrônico brasileiro Todo Estudo, como se nota em (81):

(81) Hoje não atendo, *em todo caso* (mas) entre (Nayade, s.d.).

Em síntese, não foram encontrados, em ambos os *corpora*, por fim, os marcadores adversativos citados no capítulo 1 *antes*, *ao passo que*, *contudo*, *e sim*, *em todo caso*, *não obstante*, *senão* e *todavia*. Isso se aproxima do que prevê Perini (2016, p. 447). Para ele, a lista dos coordenadores correntes no PB incluiria apenas os itens *e*, *ou* e *mas*, sendo, pois o *mas* o único representante da quebra de expectativa. Nesse sentido, “há vários outros nas listas tradicionais de ‘conjunções coordenativas’, mas seu comportamento é tão variado que prefiro não os incluir aqui”.

Vejamos, no Quadro 12, os marcadores pesquisados no referencial teórico do capítulo 1 deste trabalho presentes e ausentes em ambos os *corpora*.

Quadro 12 – Marcadores adversativos presentes e ausentes nos *corpora*

Marcador	<i>Corpus PB</i>	<i>Corpus PM</i>
antes	ausente	ausente
ao passo que	ausente	ausente
contudo	ausente	ausente
e sim	ausente	ausente
em todo caso	ausente	ausente
entretanto	ausente	presente
mas	presente	presente
não obstante	ausente	ausente
no entanto	ausente	presente
porém	ausente	ausente
senão	ausente	ausente
só que	presente	presente
todavia	ausente	ausente

Fonte: elaborado pela autora.

Por fim, cabe-nos considerar que, dado o aspecto oral de nossos *corpora*, marcadores repetidos, um traço comum da oralidade, foram contabilizados apenas uma vez. É o caso, por exemplo, de (82).

- (82) ... quem me vê pensa que eu fumo porque eu sou muito acelerado mesmo *mas mas* eu nununca experimentei nada (Vieira; Pissurno, 2016).

Em (82), vemos a repetição da conjunção *mas*, entretanto, apesar de repetida, nota-se que essa repetição se dá por uma retomada da ideia por parte do falante, havendo apenas uma quebra de expectativa, apesar da dupla ocorrência do marcador. Por isso, ele foi contabilizado apenas uma vez.

Após essas considerações sobre os marcadores de quebra de expectativa considerados em nosso trabalho, convém-nos dissertar sobre os dados obtidos e sobre algumas possibilidades de análise. Vejamo-los, pois, no capítulo 4 a seguir.

4 RESULTADOS E ANÁLISE DOS DADOS

Iniciaremos este capítulo pormenorizando as ocorrências de cada marcador de quebra de expectativa pesquisado, em relação à estratificação adotada. Iniciemos pelo *corpus* de PB. A Tabela 2 ilustra os números relativos a cada conjunção ou locução conjuntiva adversativa encontrada neste trabalho.

Tabela 2 – Contabilização dos marcadores adversativos por cada variável no *corpus* de PB

Fatores	Marcadores adversativos			
	Mas	Porém	Só que	Mas só que
A: faixa etária A (18-25 anos)	290	0	27	3
B: faixa etária B (26-40 anos)	332	1	43	5
C: faixa etária C (41-60 anos)	141	0	0	1
D: faixa etária D (+60 anos)	39	0	2	0
P: escolaridade primária (até o ensino médio completo)	108	0	3	1
U: escolaridade universitária (graduação ou pós-graduação completas)	694	1	69	8
M: sexo masculino	338	0	28	5
F: sexo feminino	464	1	44	4
S: marcador adversativo precedido de outros marcadores	0	0	0	9
N: marcador adversativo não precedido de outros marcadores	802	1	72	0
I: marcador em início de período	210	0	12	1
E: marcador em posição entre períodos	592	1	60	8
T: marcador em posição de término de período	0	0	0	0

Fonte: elaborado pela autora.

Vale lembrar que as referências aos códigos utilizados na Tabela 2 encontram-se no apêndice 1 deste trabalho. Quanto à tabela, cabe-nos sumarizar, primeiramente, que os marcadores *entretanto* e *no entanto* não foram encontrados no *corpus* de português brasileiro, figurando, portanto, com um total de 0 ocorrências.

O marcador *mas*, por sua vez, possui 290 ocorrências na faixa etária A, de 18 a 25 anos; 332 ocorrências na idade B, de 26 a 40 anos; 141 ocorrências na idade C, de 41 a 60 anos e 39 ocorrências na idade D, maiores de 60 anos. Esses números, por seu turno, se redistribuem em 108 ocorrências na escolaridade P (primária e média) e em 694 na escolaridade U (universitária). Dando sequência ao raciocínio, percebe-se que 338 ocorrências pertencem ao sexo masculino e 464 ao sexo feminino. A soma entre as ocorrências de todas as idades resulta em 802, que é o total de casos de *mas* identificados no *corpus* brasileiro. O mesmo acontece

com a soma das ocorrências das duas escolaridades e dos dois gêneros. Vemos, ainda na Tabela 2, que esse total, 802 ocorrências, não foi precedido por outros marcadores (N). E, para finalizar, 210 casos iniciam período (I) e 592 localizam-se entre as orações coordenadas (E).

Em relação ao marcador *porém*, vemos que sua única ocorrência se dá na idade B (26-40 anos), na escolaridade U (universitária), no sexo feminino (F), não precedida de outro marcador (N) e na posição entre orações coordenadas (E).

Em referência ao marcador *só que*, temos os seguintes números: 27 ocorrências na escolaridade A (18-25 anos); 43 ocorrências na escolaridade B (26-40 anos) e 2 ocorrências na escolaridade D (+60 anos). Redistribuindo esses números, teríamos 3 de escolaridade primária (P) e 69 de escolaridade universitária (U). Desses, 28 pertencentes ao sexo masculino (M) e 44 ao sexo feminino (F). O total de ocorrências não está precedido de outro marcador (N), 12 estão em posição inicial (I) e 60 em posição entre orações coordenadas (E), totalizando as 72 ocorrências desse marcador em PB.

Por fim, no que se refere ao marcador *mas só que*, que, na verdade, trata-se de *só que* precedido de *mas*, vemos que, de suas 9 ocorrências totais no PB, 3 encontram-se na idade A; 5 na idade B e 1 na C. Também uma ocorrência foi detectada na escolaridade primária (P) e 8 na escolaridade universitária (8). Houve 5 pertencentes ao sexo masculino e 4 pertencentes ao feminino. Por fim, uma iniciou período (I) e 8 ocorreram entre orações (E).

Para melhor ilustrarmos o anteriormente exposto, observemos um exemplo de *mas só que*, que foi o único caso de precedência encontrado neste trabalho, isto é, apenas *só que*, dentre as conjunções e locuções conjuntivas adversativas pesquisadas, aparece antecedido por outro marcador, no caso, a conjunção *mas*.

- (83) Aquilo é fuxico *mas só que* é de uma maneira mais sofisticada (Raso; Mello, 2012).

Convém frisar que vamos discorrer sobre as ocorrências de *mas só que* na seção 4.6. Vejamos, neste momento, na Tabela 3, os números relativos aos marcadores de quebra de expectativa encontrados no *corpus* de PM.

Tabela 3 – Contabilização dos marcadores adversativos por cada variável no *corpus* de PM

Fatores	Marcadores adversativos					
	Entretanto	Mas	No entanto	Porém	Só que	Mas só que
A: faixa etária A (18-25 anos)	0	252	0	0	14	0
B: faixa etária B (26-40 anos)	2	117	1	0	17	0
C: faixa etária C (41-60 anos)	0	252	0	0	30	4
D: faixa etária D (+60 anos)	0	30	0	0	1	0
P: escolaridade primária (até o ensino médio completo)	0	89	0	0	10	0
U: escolaridade universitária (graduação ou pós-graduação completas)	2	562	1	0	52	4
M: sexo masculino	1	316	0	0	30	3
F: sexo feminino	1	335	1	0	32	1
S: marcador adversativo precedido de outros marcadores	0	0	0	0	0	4
N: marcador adversativo não precedido de outros marcadores	2	651	1	0	62	0
I: marcador em início de período	0	0	0	0	2	0
E: marcador em posição entre períodos	2	651	1	0	60	4
T: marcador em posição de término de período	0	0	0	0	0	0

Fonte: elaborado pela autora.

Como mencionado anteriormente, os códigos utilizados na tabela 3 encontram-se disponíveis para consulta no Apêndice 1 deste trabalho. No que diz respeito à Tabela 3, vale observar, inicialmente, que o marcador *porém*, diferentemente do que houve em PB, não ocorre nos dados da variedade moçambicana. Relativamente ao item *entretanto*, por sua vez, houve duas ocorrências, concentradas na faixa etária B, em falantes de nível universitário (U), uma ocorrência em cada sexo (M e F), não precedidas de outro marcador (N) e em posição entre orações coordenadas (E).

No que concerne a *mas*, assim como no PB, foi o marcador de quebra de expectativa mais utilizado, totalizando 651 ocorrências, que se distribuíram da seguinte maneira, primeiramente, em relação ao fator faixa etária: 252 na idade A, de 18 a 25 anos; 117 na B, de 26 a 40 anos; 252 na C, de 41 a 60 anos e 30 na D, maiores de 60 anos. Consoante à escolaridade, houve 89 casos em falantes de nível primário (P) e 562 em falantes de nível universitário (U). Verificam-se, também, 316 casos com falantes do sexo masculino e 335 com falantes do sexo feminino, sendo todos os casos não precedidos de outras conjunções ou locuções conjuntivas (N) e em posição intercalada (E).

Houve ainda um caso da locução *no entanto*, na faixa etária B, em nível universitário (U), sexo feminino (F), não precedida de outro marcador (N), entre orações (E).

Concernente a *só que*, foram identificadas 62 ocorrências. 14 na faixa etária A; 17 na B; 30 na C e 1 na D, em 10 falantes de nível escolar primário (P) e 52 universitários (U), 30 do sexo masculino (M) e 32 do feminino (F); 2 casos iniciando período (I) e 60 em posição intercalada (E).

Assim como no PB, os casos de *só que* precedidos por outro marcador, no caso, *mas*, foram colocados numa categoria à parte. Portanto foram encontradas 4 ocorrências de *mas só que*, todos na faixa etária C, com falantes universitários (U), 3 do sexo masculino (M) e 1 do feminino (F), em posição entre orações coordenadas (E).

Observemos, neste momento, um exemplo dessa ocorrência no PM.

- (84) O preço é aquele razoável *mas só que* lá não tem os medicamentos (Vieira; Pissurno, 2016).

Após essa exposição inicial sobre os números absolutos encontrados relativamente aos marcadores de quebra de expectativa pesquisados e a fim de que possamos melhor analisar esses números, vejamos, na seção a seguir, o que revela a análise estatística dos dados rodados no programa estatístico Goldvarb 3.0.

4.1 Resultados iniciais

Durante a rodagem do programa, foram identificadas algumas incoerências, que estavam impedindo a obtenção de um resultado final. Como dissemos, nossa variável dependente, a marcação da quebra de expectativa no PB e no PM, foi estabelecida, a princípio como eneária, posto que foram encontrados nos *corpora* das duas variedades cinco marcadores de quebra de expectativa, quais sejam: *entretanto*, *mas*, *no entanto*, *porém* e *só que*. Conforme preveem Guy e Zilles (2007),

há certas cautelas que o pesquisador que procura usar uma versão mais-que-binária deve ter. Primeiro, há certas complicações práticas na interpretação dos resultados. Por exemplo, o nível ou o ponto neutro dos pesos relativos,

calculados pelo programa Varbrul, é definido conforme o número de variantes dependentes examinadas: se a análise é ternária, são três as alternativas examinadas; então, é preciso dar a cada alternativa uma probabilidade equivalente de ser escolhida, o que resulta em ter 0,33 como o ponto neutro, equivalente ao valor de 0,50 na análise binária (um valor acima do qual o fator favorece a seleção dessa alternativa, ou abaixo do qual a desfavorece). Além disso, a interpretação do efeito de um fator depende da comparação com duas ou mais alternativas: numa análise ternária, por exemplo, se um fator x mostra um peso de, digamos, 0,20 para a realização de uma variante A, a tentação é de dizer que esse fator desfavorece essa variante, porque o peso é menor do que 0,33. Mas primeiro convém examinar os pesos desse fator para as variantes B e C. Pode ser que uma delas (digamos b) seja superfavorecida nesse contexto, e receba um peso de 0,60, enquanto a outra (C) receba um peso de 0,10. Nesse caso, é, sim, apropriado falar de x desfavorecendo o uso de A em comparação com B. Mas, em comparação com C, ele favorece. Numa análise binária, um fator favorecer uma variante (a aplicação) implica que a alternativa (a não-aplicação) é proporcionalmente desfavorecida, mas, numa análise ternária, pode haver duas alternativas favorecidas pelo mesmo contexto, ou duas desfavorecidas, ou uma favorecida, uma neutra e uma desfavorecida (Guy; Zilles, 2007, p. 142-143).

Nesse sentido, os autores supracitados não desaconselham o trabalho com uma variável eneária, mas reforçam a cautela necessária. Caso optássemos realmente por nossa variável dependente composta por cinco elementos, deveríamos dividir o peso relativo por 5 e analisar esses dados de maneira cruzada para verificar possíveis influências. Esta não seria uma tarefa impossível, mas, considerando que as ocorrências de *entretanto*, *no entanto* e *porém* foram mais baixas, decidimos trabalhar com o Goldvarb em análise binária, considerando *mas* e *só que*, as variantes mais encontrados nos dois *corpora*, e fazer uma análise qualitativa das demais variantes, posto que, numericamente falando, elas não se mostraram expressivas. Em função do baixo número de ocorrências, 13 apenas, também optamos por tratar *mas só que* de modo qualitativo, assim como fizemos com os marcadores supracitados. Começamos, pois, na seção a seguir, a análise de *entretanto*, *no entanto* e *porém*.

4.2 *Entretanto, no entanto e porém nos corpora*

Primeiramente, com relação ao *corpus* brasileiro, o marcador menos frequente detectado foi *porém*, que se deu na fala de uma entrevistada do sexo feminino, na faixa etária de 26 a 40 anos, de escolaridade superior. Vejamos esse caso em (85).

- (85) Como é que chama aquilo do Faustão? “<Arquivo do Faustão” / *porém* num táxi> (Raso; Mello, 2012).

No *corpus* moçambicano, por sua vez, foram encontradas duas ocorrências de *entretanto*, uma em cada sexo, ambas com falantes de idade entre 26 e 40 anos e de escolaridade superior, em posição intermediária, ligando orações coordenadas. Vejamos esses dois casos de *entretanto* nos exemplos a seguir.

- (86) Sentimento que eu não consigo explicar porque que eu gosto *entretanto* há várias coisas que me tiram o sossego (Vieira; Pissurno, 2016).
- (87) São deparadas com a língua materna não português e *entretanto* entre a escola é agredido com a língua portuguesa (Vieira; Pissurno, 2016).

Ainda no *corpus* moçambicano, foi identificada uma ocorrência de *no entanto*, conforme exemplo a seguir.

- (88) D: mas os de português eram moçambicanos?
L: si:m... *no entanto*/ moçambicanos que foram alunos da minha mãe até (Vieira; Pissurno, 2016).

Nessa ocorrência, a falante responde a uma pergunta anteriormente feita, comentando sobre seus professores serem nativos ou não. Como dito, trata-se de uma ocorrência com falante do sexo feminino, na faixa etária 26 a 40 anos, de escolaridade superior.

Interessante notar o fato de que, nos dois *corpora*, todas as ocorrências dos marcadores menos usuais (*entretanto*, *no entanto* e *porém*), que não fossem *mas* e *só que*, terem se concentrado na faixa etária B, que vai de 26 a 40 anos. Freitag (2000, p. 11), como mencionado anteriormente, afirma que “a faixa etária é apenas um rótulo que agrupa vários fatores de ordem social e biológica do indivíduo”. Nesse sentido, o uso de variantes consideradas mais formais por falantes de idade mediana pode se explicar também pela escolaridade deles. Das 4 ocorrências encontradas desses marcadores mais formais, todos os falantes possuem nível de escolarização universitária. Nesses casos, pois, o fato de serem mais escolarizados pode ter dado a eles acesso a variantes mais cultas e formais. Vale lembrar, *entretanto*, que, para confirmarmos essa hipótese, precisaríamos de um *corpus* mais amplo.

Observemos no Quadro 13 a sistematização desses casos de marcadores adversativos menos frequentes.

Quadro 13 – Sistematização dos marcadores menos frequentes em ambos os *corpora*

Marcador	Casos	Corpora	Faixa etária	Sexo
Entretanto	Caso 1	Moçambicano	26-40	Feminino
Entretanto	Caso 2	Moçambicano	26-40	Masculino
No entanto	Caso 1	Moçambicano	26-40	Feminino
Porém	Caso 1	Brasileiro	26-40	Feminino

Fonte: elaborado pela autora.

Após essa análise dos casos menos frequentes, vejamos, na seção a seguir, a pormenorização dos resultados obtidos na análise binária realizada no Goldvarb, com os marcadores *mas* e *só que*, que, como dito, foram os mais numerosos.

4.3 Resultados obtidos após rodagem dos dados de ambos os *corpora*

Conforme mencionado no capítulo anterior, em relação aos fatores linguísticos presentes em nossa análise, havíamos proposto dois critérios, quais sejam: a presença ou não de outros marcadores adversativos próximos ao item analisado e a posição do marcador dentro do período composto. Quanto ao primeiro critério, como já referido, encontramos apenas 13 ocorrências dentro do total de 1604 casos, e o único caso de conjunção ou locução precedendo outra foi *mas* anterior a *só que*. Desse modo, a fim de dirimir inconsistências, optamos por excluir esse grupo de fatores, posto que as 13 ocorrências dentro do total não estavam sendo numericamente significativas. Outra questão a ser considerada é que, ao separarmos as 13 ocorrências precedidas por outro marcador, o grupo 5, que analisa justamente esse fator, ficou com apenas um código, (*n*), que significa não precedido de outra conjunção ou locução conjuntiva. Essa situação gerou um caso de *Singleton Group*, que também impediu a correta rodagem dos dados no programa. Assim, reforçamos que os 13 casos de *mas só que* serão analisados, conforme mencionado, de modo qualitativo, na seção 4.6 deste trabalho.

Outra decisão que julgamos necessária foi a exclusão do código que chamamos de *t*, este atribuído à posição do marcador em término de período, finalizando-o. Em nossas ocorrências, só encontramos os marcadores em posição inicial (*i*) ao período ou majoritariamente em posição intermediária (*e*), conectando as orações coordenadas. Portanto tivemos que excluir o código *t*,

também a fim de dirimir os *KnockOuts* apresentados pelo programa estatístico, pois, conforme Guy e Zilles (2007), haveria um total de 0% para esse valor, o que é justamente um dos motivos para o aparecimento de *KnockOut*.

Expostas, pois, nossas decisões finais e após nova rodagem dos dados recodificados, obtivemos que o grupo mais significativo relativamente ao fenômeno foi o da faixa etária. Numa análise genérica, *mas* foi 91% mais usado que *só que*, que teve 8,4% de uso. Todas as faixas etárias apresentaram um maior uso de *mas* em relação a *só que*. Considerando apenas os usos de *só que*, a faixa B apresentou um leve predomínio, com 11,8%. Observemos esses valores na Tabela 4.

Tabela 4 – Ocorrências de *mas* e *só que* quanto à faixa etária nas duas variedades

Faixa etária	Valores numéricos e percentuais	<i>Mas</i>	<i>Só que</i>	Total numérico <i>mas</i> e <i>só que</i> por faixa	Total percentual <i>mas</i> e <i>só que</i> por faixa																																	
A (18-25)	N	542	41	583	36,7																																	
	%	93	7			B (26-40)	N	449	60	509	32,1	%	88,2	11,8	C (41-60)	N	393	30	423	26,7	%	92,9	7,1	D (+60)	N	69	3	72	4,5	%	95,8	4,2	Total geral	N	1453	134	1587	100
B (26-40)	N	449	60	509	32,1																																	
	%	88,2	11,8			C (41-60)	N	393	30	423	26,7	%	92,9	7,1	D (+60)	N	69	3	72	4,5	%	95,8	4,2	Total geral	N	1453	134	1587	100	%	91,6	8,4						
C (41-60)	N	393	30	423	26,7																																	
	%	92,9	7,1			D (+60)	N	69	3	72	4,5	%	95,8	4,2	Total geral	N	1453	134	1587	100	%	91,6	8,4															
D (+60)	N	69	3	72	4,5																																	
	%	95,8	4,2			Total geral	N	1453	134	1587	100	%	91,6	8,4																								
Total geral	N	1453	134	1587	100																																	
	%	91,6	8,4																																			

Fonte: elaborado pela autora.

Quanto ao nível de escolaridade, *mas* continua sendo absolutamente mais usado que *só que*. No total, houve 91,6% de ocorrências de *mas* contra 8,4% de *só que*. Das 1587 ocorrências, em relação a *mas*, houve 1.256 casos entre falantes de escolaridade superior e 197 casos entre falantes de nível básico e médio. Em relação a *só que*, houve 121 casos entre falantes de nível superior e 13 de nível básico e médio. Observemos tais números na Tabela 5.

Tabela 5 - Ocorrências de *mas* e *só que* quanto à escolaridade nas duas variedades

Nível de escolaridade	Valores numéricos e percentuais	<i>Mas</i>	<i>Só que</i>	Total numérico <i>mas</i> e <i>só que</i> por escolaridade	Total percentual <i>mas</i> e <i>só que</i> por escolaridade
P (primária)	N	197	13	210	13,2
	%	93,8	6,2		
U (universitária)	N	1256	121	1377	86,8
	%	91,2	8,8		
Total geral	N	1453	134	1587	100
	%	91,6	8,4		

Fonte: elaborado pela autora

Com relação ao sexo dos falantes, por sua vez, mulheres usaram o *mas* 91,3% (799 ocorrências) enquanto usaram *só que* 8,7% (76 ocorrências) das vezes. Entre os homens, houve um percentual de 91,9% (654 casos de *mas*) contra 8,1% (58 casos de *só que*). Na Tabela 6, ilustram-se tais números.

Tabela 6 - Ocorrências de *mas* e *só que* quanto ao sexo nas duas variedades

Sexo	Valores numéricos e percentuais	<i>Mas</i>	<i>Só que</i>	Total numérico <i>mas</i> e <i>só que</i> por sexo	Total percentual <i>mas</i> e <i>só que</i> por sexo
Masculino	N	654	58	712	44,9
	%	91,9	8,1		
Feminino	N	799	76	875	55,1
	%	91,3	8,7		
Total geral	N	1453	134	1587	100
	%	91,6	8,4		

Fonte: elaborado pela autora.

Com relação aos *corpora*, houve 651 casos de *mas* e 62 casos de *só que* na amostra moçambicana. Na amostra brasileira, por sua vez, houve 802 incidências de *mas* e 72 de *só que*, mantendo a supremacia da forma simples em contraposição à forma perifrástica dos marcadores mais usados para a quebra de expectativa nas duas variedades. Vejamos tais ocorrências na Tabela 7.

Tabela 7 - Ocorrências de *mas* e *só que* quanto ao *corpus*

<i>Corpus</i>	Valores numéricos e percentuais	<i>Mas</i>	<i>Só que</i>	Total numérico <i>mas</i> e <i>só que</i> por <i>corpus</i>	Total percentual <i>mas</i> e <i>só que</i> por <i>corpus</i>															
Brasileiro	N	802	72	874	55,1															
	%	91,8	8,2			Moçambicano	N	651	62	713	44,9	%	91,3	8,7	Total geral	N	1453	134	1587	100
Moçambicano	N	651	62	713	44,9															
	%	91,3	8,7			Total geral	N	1453	134	1587	100	%	91,6	8,4						
Total geral	N	1453	134	1587	100															
	%	91,6	8,4																	

Fonte: elaborado pela autora

Por fim, em relação à posição dos marcadores dentro do período coordenado, *mas* apareceu 1243 vezes em posição intermediária, conectando as orações coordenadas, e 201 vezes em posição inicial. Por sua vez, *só que* apareceu 120 vezes em posição intermediária, de ligação entre as orações, e 14 vezes em posição inicial, como primeiro elemento do período composto. Observemos tais incidências na Tabela 8. Vale lembrar que esses números serão melhor abordados na seção 4.7 deste trabalho.

Tabela 8 - Ocorrências de *mas* e *só que* quanto à posição no período composto nas duas variedades

Posição no período composto	Valores numéricos e percentuais	<i>Mas</i>	<i>Só que</i>	Total numérico <i>mas</i> e <i>só que</i> por posição	Total percentual <i>mas</i> e <i>só que</i> por posição															
Início de período	N	201	14	224	14,1															
	%	93,8	6,2			Entre orações	N	1243	120	1363	85,9	%	91,2	8,8	Total geral	N	1453	134	1587	100
Entre orações	N	1243	120	1363	85,9															
	%	91,2	8,8			Total geral	N	1453	134	1587	100	%	91,6	8,4						
Total geral	N	1453	134	1587	100															
	%	91,6	8,4																	

Fonte: elaborado pela autora.

As melhores rodadas consideradas pelo programa foram as séries 2 (*step-up*) e 25 (*step-down*). Ambas as rodadas apontaram o grupo 1, relativo à faixa etária, como o mais estatisticamente relevante. Nesse sentido, a faixa B apresentou o maior peso relativo, 0,6. Como a leitura se dá a partir da variante codificada como “5”, ou seja, *só que*, caberia dizer, pois, que o grupo etário B, dos falantes de 26 a 40 anos, foi considerado o mais propenso a utilizar essa locução adversativa. Vejamos, na Tabela 9, os demais pesos relativos.

Tabela 9 – Influência do fator faixa etária no uso da variante inovadora *só que*

Faixa etária	Número de ocorrências	Porcentagem %	Peso relativo
A (18-25)	41	7	0,46
B (26-40)	60	11,8	0,60
C (41-60)	30	7,1	0,46
D (+60)	3	4,2	0,32

Fonte: elaborado pela autora.

Nesse sentido, outro dado que caberia ser esmiuçado é o valor de significância, que aponta justamente o grupo considerado estatisticamente relevante. Como dito, esse grupo foi o 1, relativo à faixa etária dos falantes. Esse foi o único conjunto que manteve o desejável valor abaixo de 0,05 tanto em *stepping up* quanto em *stepping down*, apontando os valores 0,010 e 0,012, respectivamente. Diz-se desejável, porque, valores abaixo do limiar usual de 0,05 sugerem significância, ou seja, confirmam relevância estatística (Tagliamonte, 2006, p. 235). Consoante Tagliamonte (2006), análises que consideram os níveis *step up* e *step-down*, indicando significâncias, são consideradas o padrão ouro (gold standard) na sociolinguística variacionista (Tagliamonte, 2006, p. 226). Por isso, apesar de, na tabela 9, já termos evidenciado os valores dos pesos relativos encontrados, achamos pertinente, a título de complementação, abordarmos também os valores de significância encontrados na rodagem dos dados em ambas as variedades, a fim de deixarmos nossa análise mais completa. Os demais grupos só apresentaram valores abaixo de 0,05 em rodagens que excluía o grupo 1, o que demonstra a presença dele ser significativa. Isso porque, estando presente, foi o único significativo, quando os demais foram irrelevantes. Estando ausente, permitiu que valores anteriormente inexpressivos revelassem-se. O fato é que, se esses valores a princípio inexpressivos tivessem significância, os resultados apresentados por eles deveriam ser abaixo de 0,05 em qualquer rodada, independentemente da presença ou não do grupo faixa etária. Por exemplo, as rodadas 12 e 17 apresentaram valores de significância de 0,012. Na 12, todos os grupos foram incluídos, exceto o 1, e, na 17, foram considerados apenas os grupos 2, 4 e 5. Observemos, na Tabela 10, os valores de significância de cada grupo de fatores separadamente.

Tabela 10 – Grupo de fatores pesquisados e respectivos valores de significância obtidos na rodagem dos dados pelo programa estatístico Goldvarb 3.0

Grupo	Valor de significância
1 faixa etária	0.010
2 grau de escolaridade	0.193
3 sexo	0.700
4 corpus	0.750
5 posição do marcador adversativo	0.190

Fonte: elaborado pela autora.

Na tabela 10, observa-se que o único valor abaixo de 0,05 refere-se ao grupo 1, faixa etária. Os demais grupos, conforme explicitado anteriormente, apresentam valores acima desse número. Convém salientar, ainda, que todos esses dados, gerados automaticamente na rodagem do programa estatístico, encontram-se no Anexo 1 deste trabalho.

Aplicamos o teste qui-quadrado em relação ao grupo de fatores considerado estatisticamente relevante, como dito, faixa etária. Conforme Wicklin (2024), a fórmula para esse teste é:

$$x^2 = 2 (\text{LLcompleto} - \text{LLreduzido})$$

Nessa fórmula, x^2 é o resultado obtido; LLcompleto é o *log likelihood* do modelo completo, valor fornecido pelo programa que inclui o grupo testado, no caso, o grupo faixa etária; e LLreduzido, o valor fornecido pelo programa antes da introdução de qualquer grupo/fator explicativo, ou seja, o modelo nulo. Vale lembrar que ambos os valores são disponibilizados no arquivo denominado *res*, que exhibe os resultados após finalizadas cada rodagem. Esse arquivo completo, como dito, encontra-se no Anexo 1 deste trabalho.

Utilizando os dados fornecidos na rodagem, obtivemos:

$$x^2 = 2 [-453.731 - (-459.392)]$$

$$x^2 = 11,322$$

Portanto, o teste produziu uma estatística de 11,322. Na fórmula acima, subtrai-se o *log likelihood* (com o grupo incluído) do *log likelihood* (com o grupo excluído): (LLcompleto – LLreduzido). Essa subtração busca mostrar qual a importância do grupo considerado

estatisticamente relevante na aplicação da variável em estudo. Por isso, na fórmula, pega-se o valor que inclui o grupo faixa etária subtraindo-o dos demais grupos. Tagliamonte (2006) sugere comparar o valor de χ^2 a uma distribuição qui-quadrado com os graus de liberdade apropriados: “you can determine significance by comparing χ^2 to a chi-square distribution with appropriate degrees of freedom” (Tagliamonte, 2006, p. 148). Nossos graus de liberdade são 3, pois temos 4 faixas etárias (A, B, C, D) subtraídas por uma variável dependente.

Conforme mencionado anteriormente, o valor referencial para determinar a significância estatística de algum fator ou grupo de fatores é 0,05. Comparando esse valor com o χ^2 encontrado na fórmula acima, cujo resultado é 11,322, consoante sugere Tagliamonte (2006), nota-se que se trata de um valor que indica que o modelo completo, que inclui o grupo faixa etária, é estatisticamente significativo. Isso porque, para níveis de significância de 0,05, valores acima de 7,815²² são considerados relevantes para o fenômeno, e, como observado, nosso valor, 11,322 encontra-se acima de 7,815.

A fim de verificarmos se haveria mais algum aspecto estatisticamente relevante ou alguma peculiaridade a ser discutida, decidimos por fazer novas rodadas utilizando o Goldvarb, desta vez analisando separadamente cada *corpora*. Observemos a seguir algumas considerações estabelecidas separadamente em relação a cada variedade. Importante frisar, ainda, que todos os resultados estatísticos obtidos após as rodagens em separado para cada *corpus* encontram-se no Anexo 2 deste trabalho, para a modalidade brasileira, e Anexo 3, para a moçambicana.

4.4 Resultados obtidos após rodagem dos dados em relação a cada variedade

Primeiramente, quanto ao *corpus* brasileiro, foi identificado o *KnockOut* no grupo referente à faixa etária, porque houve uma ocorrência de 100% de *mas* na idade C e zero ocorrências de *só que* nessa mesma faixa etária. Diante disso, mesmo considerando o fato de que o grupo da faixa etária foi o mais significativo na rodada de ambos os *corpora*, eliminamos o grupo C da rodada de amostra brasileira, com o intuito de eliminarmos o erro e verificarmos algum outro aspecto relevante.

²² Segundo a tabela estatística da SCRIBBR (2022), o valor crítico para 3 graus de liberdade e significância $\alpha = 0,05$ é 7,815 (SCRIBBR. *Chi-square distribution table* [PDF]. Disponível em: <https://www.scribbr.com/wp-content/uploads/2022/05/Chi-square-table.pdf>. Acesso em: 04 jun. 2025).

Feito isso, obtivemos que, dentro do *corpus* brasileiro, a variante *mas* continua sendo mais expressiva que a locução *só que* em todos os grupos de fatores, ocorrendo mais de 80% em todos os casos analisados. Vejamos tais percentuais na Tabela 11.

Tabela 11 – Percentuais de *mas* e *só que* em relação aos grupos de fatores no *corpus* brasileiro

Fator	<i>Mas</i> (%)	<i>Só que</i> (%)	Total (%)
Idade A	91,5	8,5	100
Idade B	88,5	11,5	100
Idade C	100	0	100
Idade D	95,1	4,9	100
Escolaridade P	95,9	4,1	100
Escolaridade U	89,5	10,5	100
Sexo F	90,0	10,0	100
Sexo M	90,4	9,6	100
Posição I	93,2	6,8	100
Posição E	89,2	10,8	100

Fonte: elaborado pela autora.

Após essa exposição dos percentuais relativos a cada variante dentro da amostra brasileira, vejamos os resultados obtidos com a rotação dos dados relativos apenas ao português moçambicano. No que se refere a estes, assim como ocorrido no PB, também detectou-se o *knockOut*. Especificamente neste caso, o erro se deu por não haver ocorrências de *mas* em posição de início de período, enquanto que *só que* ocorreu duas vezes nessa posição, concentrando 100% das ocorrências nesse grupo de fatores. Diante desse contexto, optamos por fazer uma nova rotação, excluindo o fator posição do marcador no período composto. Feito isso, obtivemos os seguintes percentuais, mostrados na Tabela 12.

Tabela 12 – Percentuais de *mas* e *só que* em relação aos grupos de fatores no *corpus* moçambicano

Fator	<i>Mas</i> (%)	<i>Só que</i> (%)	Total (%)
Idade A	94,7	5,3	100
Idade B	87,3	12,7	100
Idade C	89,4	10,6	100
Idade D	96,8	3,2	100
Escolaridade P	89,9	10,1	100
Escolaridade U	91,5	8,5	100
Sexo F	91,3	8,7	100
Sexo M	91,3	8,7	100

Fonte: elaborado pela autora.

Assim como no *corpus* brasileiro, *mas* demonstrou-se mais usual que *só que* em todos os grupos de fatores, também com mais de 80% de uso em todos eles. Outro detalhe interessante observado é que, assim como na rodada geral, em que foram considerados ambos os *corpora*, separadamente em relação ao *corpus* moçambicano, o grupo faixa etária também foi considerado significativamente relevante em termos estatísticos, apontando o peso relativo de 0,6 para a faixa etária B. Essa influência pode ser inclusive a responsável por acusar esse mesmo valor de peso relativo na rodada geral.

Atendo-nos, pois, mais especificamente aos pesos relativos para ambos os *corpora*, conforme acusou o programa estatístico, a faixa etária B, como dito, foi considerada a mais produtiva, com um valor de 0,60. Os demais valores de pesos relativos para a faixa etária encontrados foram: A: 0,46; C: 0,46 e D: 0,32. Quanto ao nível escolar, o universitário, por sua vez, foi considerado um pouco mais produtivo, com o peso relativo de 0,51. Por sua vez, o peso relativo para a escolaridade primária foi 0,41. Quanto ao fator sexo, o feminino se mostrou neutro, com o valor de 0,50, em contraposição ao masculino, que apresentou o valor de 0,49. Por fim, quanto à posição do marcador dentro do período composto, obtivemos o peso de 0,51 para a posição intermediária e 0,42 para a posição inicial.

Observando esses valores de pesos relativos, *grosso modo*, tenderíamos a considerar que a faixa etária B, dos falantes de 26 a 40 anos, considerados jovens a medianos, estaria mais propensa ao uso de *só que*, por estar, teoricamente, mais aberta linguisticamente. Os falantes de nível universitário, por sua vez, coadunariam esse aspecto, por ser uma faixa geralmente mais jovem e mediana, também mais aberta a receber termos mais recentes, como *só que*, inclusive pelo fato de ser uma camada que possui mais conhecimento, mais escolaridade. O valor mais alto

entre as mulheres também vai ao encontro desses valores, posto que elas costumam estar mais abertas a variantes inovadoras. E, por fim, a posição intermediária apresentar um valor mais alto de peso relativo também poderia se explicar pelo fato de *só que* ser um marcador adversativo coordenativo, que, em tese, serviria para coordenar orações, ocorrendo mais, portanto, em posição que ligue as sentenças coordenadas e marque a expectativa a ser quebrada entre as duas orações. Entretanto, a fim de não nos limitarmos apenas a esses valores de pesos relativos supracitados e visando a dar mais robustez para a nossa pesquisa, endossamos a opinião de Naro (2003) e partimos para uma análise mais qualitativa dos dados, posto que, conforme esse autor, “o progresso da ciência linguística não está nos números em si, mas no que a análise dos números pode trazer para nosso entendimento das línguas humanas” (Naro, 2003, p. 25). Vejamos, pois, na seção a seguir nossa interpretação dos resultados numéricos anteriormente mencionados.

4.5 Análise qualitativa dos dados

Para Johnstone (2000), o trabalho sociolinguístico é sempre interpretativo. Para ela, interpretar é decidir o que um evento, um padrão linguístico ou um conjunto de números significa. Assim, dados não interpretados são dados vazios e brutos, que não confirmam nem desconfirmam hipóteses. Só a interpretação dá um ponto coerente a uma análise. “Portanto, é sempre necessário interpretar os dados e os resultados da análise, qualquer que seja a abordagem adotada” (Johnstone, 2000, p. 36). Ainda segundo a autora, o contraste entre a pesquisa sociolinguística mais voltada para os aspectos quantitativos ou mais voltada para os aspectos qualitativos reside em se as questões de pesquisa são respondidas por meio de procedimentos relativamente mecânicos ou relativamente não mecânicos. Nesse sentido, destaca-se a palavra “relativamente”. Isso porque as abordagens e os procedimentos são interdependentes. Por exemplo, em sua tese de doutoramento, Mukherjee (1995), citada por Johnstone (2000), decide trabalhar com a identidade étnica e o uso da linguagem por mulheres na comunidade imigrante malaio-bengali. Nesse estudo, ela opta por uma abordagem qualitativa, pois, para ela, mais interessante seria investigar como e por que o bengali e outras línguas eram usadas em vez de preocupar-se com a frequência dessas línguas na comunidade em estudo. Nesse trabalho, por exemplo, entre as mulheres, a autora observou que o bengali era visto como o idioma mais prestigiado, enquanto o inglês malaio era considerado a língua não marcada, que transitava entre mulheres de diferentes faixas etárias, classes sociais e níveis de escolarização. Para chegar a essa conclusão, a autora recorreu à coleta de dados quantitativa, mas, ao longo de sua pesquisa,

foi conversando com os entrevistados a fim de observar as impressões pessoais deles a respeito das línguas que utilizavam. Assim, foi tecendo análises qualitativas para o melhor refinar o entendimento de seus dados.

Nesse sentido, neste trabalho, como dito, também optamos pelo uso das ferramentas quantitativas, como o Goldvarb, mas cremos que seja importante também tecermos uma interpretação desses numerais apontados pelo programa estatístico. Quanto a isso, primeiramente, cabe elucidar o maior uso de *mas* nas duas variedades pesquisadas. Vejamos na Tabela 13 os números referentes às duas variantes, nas duas variedades, em relação a todos os fatores pesquisados.

Tabela 13 – *Mas e só que* no PB e no PM

Fator	<i>Mas</i> PB	<i>Mas</i> PM	<i>Só que</i> PB	<i>Só que</i> PM
Idade A (18-25)	290	252	27	14
Idade B (26-40)	332	117	43	17
Idade C (41-60)	141	252	0	30
Idade D (+60)	39	30	2	1
Escolaridade P (primária)	591	89	3	10
Escolaridade U (universitária)	70	562	69	52
Sexo F (feminino)	265	335	28	32
Sexo M (masculino)	396	316	44	30

Fonte: elaborado pela autora.

Em relação à Tabela 13, primeiramente, no que se refere ao fator faixa etária, cabe destacar a ausência de valores na faixa etária C quanto aos dados do português brasileiro. Como dito, foi necessário excluir essa faixa etária para evitar os *KnockOuts* do programa estatístico, em função de haver zero ocorrências de *só que* na idade em questão. Na faixa A, cujas idades vão de 18 a 25 anos, os números encontrados são relativamente próximos. A diferença de 38 casos de *mas* a mais na variedade brasileira e de 13 casos a mais de *só que* não nos pareceu expressiva a ponto de afirmarmos que essa faixa etária favoreceu algum uso em alguma das variedades.

Entretanto, na idade B, houve uma diferença de 215 casos de *mas* a mais no PB que no PM. Quanto a *só que*, a diferença foi de 26 casos. Trata-se de uma faixa etária considerada jovem e mediana (26 a 40 anos). Além da diferença mais significativa de *mas* entre as duas variedades, chama a atenção o desnível entre *mas* e *só que* em cada variedade separadamente. Isso porque, conforme mostra a Tabela 12, na faixa B, houve 332 casos de *mas* e 43 de *só que* no PB; e 117 ocorrências de *mas* no PB contra 17 de *só que*. As maiores ocorrências de *só que* inclusive se

dão nas faixas etárias A e B, 27 e 43 casos, respectivamente. Isso pode se dever ao fato de que, como dito, tratar-se de uma faixa etária jovial a mediana, que tem mais acesso e mais contato com as variantes mais recentes, como é o caso de *só que*, forma mais inovadora em relação a *mas*.

No PM, a maior ocorrência de *só que* se deu na faixa C, que vai de 41 a 60 anos, com trinta ocorrências. Desses 30 casos, 22 referem-se a falantes de nível universitário e 8 a falantes de nível escolar primário ou médio. Portanto, o maior uso de *só que* entre os universitários pode dever-se ao fato de esses falantes mais escolarizados terem mais contato com outras variantes da língua ou também, justamente por serem mais escolarizados, frequentarem ambientes diversos e terem contato com falantes mais jovens, que fazem mais uso de *só que*.

Esse contexto também poderia explicar a baixa ocorrência de *só que* entre os falantes da faixa etária D, que se situa a partir dos 60 anos, nas duas variedades. Houve duas ocorrências no PB e uma no PM. Esses falantes, justamente por estarem num estágio mais avançado da idade, já estão com a sua linguagem consolidada, não estando tão abertos a receber as mudanças linguísticas encabeçadas pelos jovens.

Em nossa dissertação de mestrado, defendida na Faculdade de Letras da UFMG, em 2017 (SILVA, 2017), e intitulada “A perífrase conjuncional *só que*: gramaticalização e variação linguística”, na qual se analisava o uso de *mas* e *só que* no falar belo-horizontino, notamos que as duas variantes não eram estigmatizadas socialmente, posto que eram livremente usadas por falantes de sexos, escolaridades e idades variadas. O mesmo se percebe neste trabalho, posto que não se observou uma restrição nos usos de *mas* e *só que* em relação aos fatores anteriormente mencionados. Outro aspecto interessante visto em nossa dissertação (SILVA, 2017) é a coocorrência de *mas* e *só que*, não havendo ainda, a princípio, indícios de que a forma inovadora vá ocupar o lugar da forma conservadora na fala da RMBH. O mesmo parece se dar neste trabalho, em que tanto na variedade brasileira quanto na variedade africana, não se notam números expressivos das variantes que justifiquem ou demonstrem um processo de mudança em curso.

Evidenciando mais o fator escolaridade, mantém-se a supremacia do uso de *mas* em relação a *só que*. No português brasileiro, entre os falantes de nível primário, houve 591 usos de *mas* e 69 usos de *só que*. Quanto à escolaridade universitária, há 70 ocorrências de *mas* e 3 de *só que*.

Quanto a *mas*, uma relação inversa ocorre no PM. Isso porque houve 562 incidências de *mas* entre os falantes universitários e 89 entre os falantes de nível escolar primário, o que cremos que se deva ao fato de termos, em nosso *corpus*, mais falantes de nível universitário. Um aspecto interessante desses dados é que, na amostra brasileira, o maior uso de *só que* se deu entre os falantes de nível primário (69), na amostra moçambicana, o maior uso de *só que* se dá entre os falantes de nível universitário (52). Como não se percebe a variante inovadora como estigmatizada, cremos que a maior incidência dela no PB entre o grupo de nível primário escolar pode se dever ao fato de ela estar mais popularizada, ser mais comum entre falantes diversos; no PM, pode se tratar de uma variante que ainda está se difundindo, ocorrendo mais no ambiente universitário, que é mais aberto a mudanças e inovações, e dele se difundir às demais camadas.

Por fim, quanto ao sexo dos falantes, temos números de *mas* relativamente próximos nas duas variedades. No PB, 265 casos entre as mulheres e 396 casos entre os homens; no PM, 335 casos entre as mulheres e 316 casos entre os homens. Logo, não notamos diferença significativa que aponte alguma tendência dessa conjunção em relação ao gênero, dada a proximidade numérica. O mesmo se dá quanto a *só que*, posto que os números também se aproximam relativamente, havendo, no PB, 28 ocorrências no sexo feminino e 44 no masculino; e, no PM, 32 incidências para as mulheres e 30 para os homens.

Em estudo sobre o uso de conectores adversativos na fala natalense, Pedro (2019) havia hipotetizado o maior uso das variantes adversativas mais recentes pelas mulheres, posto que elas costumam liderar os processos de mudança linguística, conforme prevê Labov (2001). Entretanto, assim como em nosso caso, a variável sexo não foi considerada significativa pelo programa estatístico. Nas palavras de Pedro (2019):

os resultados nos permitem notar que as mulheres fizeram um pouco mais de uso do SÓ QUE, forma inovadora, conforme levantado na hipótese. Talvez isso se dê, conforme estudos pioneiros, devido ao fato de que as mulheres tendem a liderar as mudanças linguísticas. Entretanto, embora elas tenham recorrido mais ao SÓ QUE do que eles, considerando que o sexo não foi dado como significativo pelo GOLDVARB X, não é possível aventar explicações sólidas em termos de mudança (Pedro, 2019, p. 82).

Assim, coadunamos as ideias de Pedro (2019), segundo quem ainda é cedo para tecermos afirmações sólidas sobre o processo de encabeçamento de *só que* dentro do panorama das locuções e conjunções adversativas da língua portuguesa.

Feita, pois, essa discussão geral das variáveis em questão, cabe-nos, neste momento, detalhar as treze ocorrências de *mas só que*, em que se nota a junção dos dois marcadores mais frequentes nos *corpora* em estudo e anteriormente submetidos à análise estatística.

4.6 Os casos de *mas só que*

Conforme mencionado anteriormente, foram encontradas 13 ocorrências de *mas só que*. Posto que esse foi um dos critérios de análise pré-estabelecidos, isto é, a precedência de outro marcador dentro da oração adversativa, é interessante notarmos que, dentre todos os marcadores pesquisados neste trabalho, quais sejam: *contudo*, *entretanto*, *mas*, *no entanto*, *porém* e *só que*, houve ocorrência de precedência envolvendo apenas a conjunção *mas*. Observemos, a seguir, os contextos de *mas só que* identificados.

- (89) Ler em casa com mais calma iria ler aí *mas só que* a professora não deu o que está lá no livro (Vieira; Pissurno, 2016).
- (90) A Bíblia diz que há feiticeiros há feitiço *mas só que* nós não podemos (quando nós confiamos em Jeová Deus) (Vieira; Pissurno, 2016).
- (91) O preço é aquele razoável *mas só que* lá não tem os medicamentos (Vieira; Pissurno, 2016).
- (92) Um ano dois anos depois pode se fazer né *mas só que* já se paga com uma parte uma multa (Vieira; Pissurno, 2016).
- (93) Aquilo é fuxico *mas só que* é de uma maneira mais sofisticada (Raso; Mello, 2012).
- (94) Aí eu fui a farmácia *mas só que* eu fui assim com o pé atrás né (Raso; Mello, 2012).
- (95) Um negócio desse tamanho / cem mililitros / enfiar pelo umbigo> não *mas só que* mas nã enfia não você enche o saquinho / lá (Raso; Mello, 2012).

- (96) Os cara eram os cara *mas só que* os cara era assim sabe eles fechavam em casas / com / trinta / quarenta / litros de uísque / cara (Raso; Mello, 2012).
- (97) Fica bem maior *mas só que* a área de serviço ficou muito pequena (Raso; Mello, 2012).
- (98) É um serviço que dá muito dinheiro né *mas só que* ele é pixota demais né (Raso; Mello, 2012).
- (99) Eu tirei meio assim escondido as fotos né *mas só que* obviamente alguns lugares entrei (Raso; Mello, 2012).
- (100) eu mudaria muita coisa né *mas só que* ocê dá a sua opinião e deixa o resto com o cliente que quiser né (Raso; Mello, 2012).
- (101) Eu vou *mas só que* pô / eu tô <desorientado> (Raso; Mello, 2012).

Nos exemplos de (89) a (101), nota-se a presença da quebra de expectativa entre os dois termos ligados por *mas só que*. E, em todas as ocorrências, qualquer um dos marcadores poderia ser retirado, mantendo-se a ideia de adversatividade presente. Por exemplo, em (91), é possível dizermos “o preço é aquele razoável *mas* lá não tem os medicamentos” ou “o preço é aquele razoável *só que* lá não tem os medicamentos”, o que, inclusive, confirma haver entre os marcadores pesquisados nas duas variedades o mesmo valor referencial/representacional, conforme exposto no capítulo 2 deste trabalho.

Dentre as 13 ocorrências de *mas só que* encontradas, 4 foram identificadas no *corpus* moçambicano e 9 referem-se ao *corpus* brasileiro. Das ocorrências do PM, todas encontram-se dentro da faixa etária C, que vai de 41 a 60 anos, e dentro da escolaridade universitária. Dessas 4, três pertencem ao sexo masculino e uma pertence ao feminino.

No PB, das 9 ocorrências identificadas, temos 4 mulheres e 5 homens. Dessas 4 mulheres, todas possuem escolaridade de nível superior; duas pertencem à faixa etária B (26-40 anos) e duas estão na faixa etária C (41-60 anos). Com relação aos falantes do sexo masculino, por sua vez,

4 possuem escolaridade universitária, um encaixando-se na faixa etária A (18-25 anos), dois na B e um na C, enquanto o falante de escolaridade primária encaixa-se na idade B.

Por se tratarem de apenas 13 ocorrências, não conseguimos traçar algum perfil de uso de *mas só que* com base em algum dos fatores extralinguísticos citados, como faixa etária, escolaridade e sexo. Linguisticamente falando, poderíamos coadunar as ideias de Meillet (2014[1921]), segundo quem a expressividade vai comandar o surgimento de novas conjunções. Para esse autor, as conjunções formam uma classe sujeita constantemente a desaparecer e ser renovada. O sujeito falante, tendo a necessidade de fazer-se expressivo e conseguir que seus pensamentos sejam compreendidos por seu interlocutor, mantém um repertório de palavras em sua memória, acessando-as com o objetivo de comunicar-se, o que torna o uso delas praticamente automático. Esse uso frequente das palavras acaba retirando delas o valor expressivo (Meillet, 2014[1921], p. 163). Além disso, para Meillet (2014[1921]), o enfraquecimento expressivo das conjunções não se deve apenas ao desgaste delas pela frequência de uso, mas também a uma redução de volume fonético. Baseados na economia linguística e na redução do esforço, os falantes vão reduzindo as palavras, que, por sua vez, perpetuam-se na língua sofrendo constantes mudanças de pronúncia e diminuição de tamanho. Um exemplo do PB para esse fenômeno poderia ser o processo sofrido, por exemplo, em: “vossa mercê – você – ocê – cê” (Vital, 1996).

Nesse sentido, poder-se-ia pensar que a conjunção *mas*, por ser a mais usual em ambos os *corpora*, estar-se-ia enfraquecendo e se desgastando pelo uso e se apoiando em *só que*, que é o segundo marcador mais usado para a quebra de expectativa, ancorando-se nele para ganhar força significativa. Mas 13 ocorrências são um número insuficiente para tecermos tal afirmação. Esse pode ser, inclusive, objeto de um estudo futuro, que lance mão de um *corpus* mais amplo a fim de investigar mais detalhadamente o fenômeno²³.

Na seção 4.7, buscamos apontar convergências e divergências entre PB e PM.

²³ No Corpus do Português (Davies, 2006), na modalidade Genre/Historical, nota-se um aumento significativo da expressão *só que* no século XX. No XIX, ocorre uma frequência de 122, que aumenta para 862 a partir dos anos de 1900. Na modalidade Now, *só que* apresenta uma frequência de mais de 31.785 no Brasil, enquanto em Moçambique essa frequência é de apenas 76. Esses dados preliminares confirmam a relativamente recente entrada dessa locução conjuntiva no PB e o predomínio dela no PB em comparação ao PM, o que, de fato, nos motiva a ampliar nosso *corpus* a fim de obtermos mais dados a respeito da variação e mudança de *mas* e *só que*. Disponível em: corpusdoportugues.org. Acesso em: 1º jul. 2025.

4.7 Convergências e divergências entre PB e PM

Primeiramente, quanto aos elementos convergentes, podemos ressaltar que o fenômeno da quebra de expectativa se dá de modo bastante similar nas duas variedades. Em ambas, a conjunção *mas* foi mais usual, encabeçando a segunda oração, na posição intermediária dentro do período composto por coordenação, ocupando 91% das ocorrências totais de cada variedade (91,8% para PB e 91,3% para PM).

Observemos, a seguir, um exemplo de PB (102) e outro de PM (103), em que se notam estruturas bastante similares para a adversatividade.

(102) Mauro e Filhos é um time muito legal eu gostaria que eles continuassem *mas* eles não são veteranos (Raso; Mello, 2012).

(103) Nós somos uma família de quatro irmãos *mas* a minha mãe vem de uma família de nove irmãos (Vieira; Pissurno, 2016).

Tanto em (102), exemplo de PB, quanto em (103), exemplo de PM, vemos a conjunção adversativa *mas* iniciando a oração que quebra a expectativa dada na primeira parte do enunciado. Em (102), na primeira informação dada, o falante afirma seu desejo de que o time Mauro e Filhos permanecesse na disputa. Uma possível expectativa gerada é que efetivamente a equipe continuasse na competição. Entretanto, a nova informação dada pela oração adversativa iniciada por *mas*, de que eles não são veteranos, quebra a expectativa da permanência do clube, deixando-o fora da competição. No exemplo (103), de PM, A primeira informação dada é a quantidade de irmãos do falante, isto é, quatro integrantes. Uma possível expectativa gerada é de que, para uma constituição familiar, esse número já seria um número elevado. Entretanto a segunda informação dada, de que a família da mãe do falante possui nove irmãos, quebra essa expectativa de quatro já ser um número grande, apresentando um número maior ainda. Essa similaridade na construção da quebra de expectativa entre as duas variedades foi observada não só com a conjunção *mas*, que, como mencionado, foi a mais frequente em ambas as variedades, mas também foi vista nas demais expressões possíveis para a adversatividade.

Se, conforme mencionado, a conjunção adversativa *mas* ocupou 91% das ocorrências em ambas as variedades, converge entre elas também o fato de a locução conjuntiva adversativa *só que* ocupar o segundo lugar em relação ao uso pelos falantes para a marcação da quebra de expectativa. Observemos, pois, exemplos de uso dessa locução no PB e no PM respectivamente.

(104) É uma moça bonita *só que* a boca dela é muito fina (Raso; Mello, 2012).

(105) Tive alguns amigos brasileiros no whatsapp *só que* depois quando troquei o número perdi todos eles (Vieira; Pissurno, 2016).

Em (104), a segunda informação dada, sobre a moça ter os lábios finos, quebra a expectativa gerada na primeira informação, rompendo o que se fala sobre a beleza dela. Em (105), a segunda informação dada, a respeito da perda de contatos, quebra o anteriormente dito sobre o falante ter amigos brasileiros no aplicativo de mensagens. Assim, nota-se que a configuração da quebra de expectativa tanto em PB quanto em PM se dá de modo análogo, evidenciando um ponto de convergência entre as variedades em questão.

Outro fator convergente entre PB e PM foi que os marcadores de quebra de expectativa *entretanto*, *porém* e *no entanto* foram pouco produtivos. Essas conjunções e locuções conjuntivas são consideradas mais formais. Como os *corpora* desta pesquisa retratam a fala, é natural que esses marcadores tenham se mostrado menos expressivos.

Interessante notar também o comportamento da perífrase conjuncional *só que* em ambas as variedades. Houve 72 ocorrências dela no PB e 62 no PM, totalizando cerca de 8% do quadro da marcação da quebra de expectativa nas duas variedades. Em um trabalho anterior, nossa monografia defendida na Faculdade de Letras da UFMG, em 2013, já havíamos começado a investigar a locução *só que* (SILVA, 2013). Nessa ocasião, selecionamos 30 entrevistas do Censo 1999-2000, com falantes cariocas, coletadas pelo Programa de Estudos sobre o Uso da Língua (PEUL), vinculado à Faculdade de Letras da UFRJ. Dentre os 30 entrevistados, havia cinco idosos. Destes, três não utilizaram a locução *só que*, o que representou 60% dessa faixa etária. Inclusive, a entrevistada mais jovem desse grupo, que na época tinha 61 anos, não usou a perífrase em nenhum momento de sua entrevista. Segundo Labov (2001), é provável que o indivíduo consolide sua linguagem aproximadamente aos 17 anos. Logo, a fala de uma idosa em 1999 reflete a linguagem de mais de 45 anos atrás, isto é, anterior ao ano de 1955. Se a idosa

não usou a perífrase em sua fala, é possível que essa forma linguística tenha surgido quando ela já tivesse atingido os 17 anos, e, conseqüentemente, consolidado sua linguagem. Desse modo, é provável que *só que* tenha se encaixado no PB na segunda metade do século XX, podendo ser considerada uma forma linguística relativamente nova e inovadora em relação a *mas*, que, como visto no capítulo 1, vem do latim *magis*.

Na *Gramática histórica da língua portuguesa*, Said Ali (1971) defende que as conjunções teriam o papel de assinalar a relação lógica entre duas informações. Nesse sentido, diante da necessidade de que houvesse diversos conectores, conforme a língua se desenvolvia e avançava, não foram criados novos “vocábulos especiais” (Said Ali, 1971, p. 220). Ao contrário, o português adaptou advérbios já existentes para exercerem o papel conjuntivo, juntamente com a partícula *que*:

este processo creador de novas conjunções ou locuções conjuncionais revela-se sobremodo fecundo nas combinações de advérbios e dizeres de caráter adverbial com a partícula *que*: *a fim que* (português hodierno *a fim de que*), *sem embargo que*, *contanto que*, *ante(s) que*, *depois que*, etc. Nestas, como em outras locuções conjuncionais, o elemento advérbio nada mais é que um vocábulo deslocado de uma oração para outra. Devia modificar um verbo, mas afasta-se dele, emigra da respectiva oração, atraído por uma partícula, à qual se une, resultando desta liga uma conjunção de nova espécie (Said Ali, 1971, p. 220-222).

Em Silva (2017), abordamos o percurso histórico de *só que* no PB. Considerando o processo descrito por Said Ali (1971), em que advérbios ter-se-iam juntado à palavra *que* para formar itens conjuncionais, postulamos que *só que*, perífrase conjuncional adversativa, tenha sido fruto desse recurso linguístico. Em nossa pesquisa no Corpus do Português (Davies, 2006), encontramos, no português arcaico do século XIV, o advérbio *solamente*, presente em frases como:

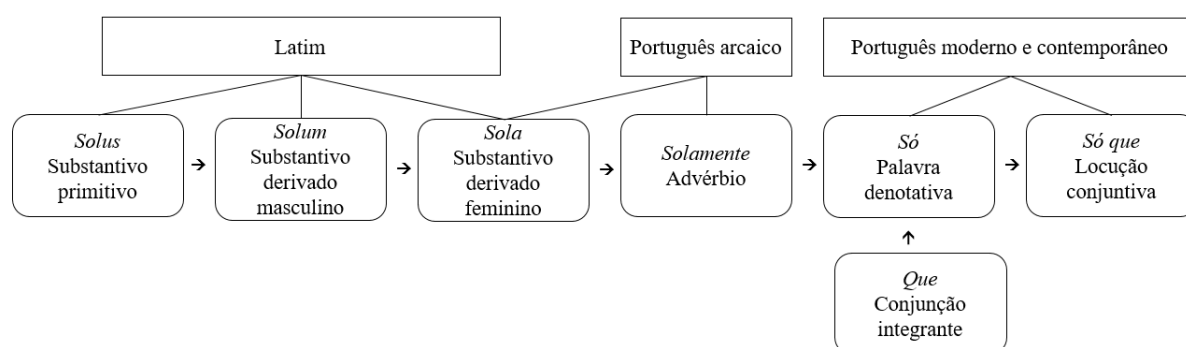
(106) *Tã solamente por fartar ao apetito* (tão somente satisfazer ao apetite – tradução nossa – Davies, 2006).

Nesse período ou em épocas anteriores, não foram encontrados outros vocábulos que compartilhassem a raiz latina de *solus*, palavra primitiva que possuía o feminino *sola* e o neutro *solum*. Como não encontramos *solus* ou *solum* no português, apenas no latim, mas encontramos *sola* no português arcaico, provavelmente essa palavra predominou sobre a neutra e a

masculina, formando, num primeiro momento, *sola mente*, conforme (106), depois tornando-se o advérbio do português arcaico *solamente* (Silva, 2017). *Solamente*, por sua vez, teria passado por um processo de gramaticalização, reduzindo-se a *só*. Seguindo a tendência de produção de itens conjuncionais destacada por Said Ali (1971), o item *só* teria se juntado à conjunção integrante *que* e dado origem à locução *só que*, assim como outros itens conjuncionais da língua portuguesa o fizeram, como *além de que*, *ainda que*, *mesmo que*, *a fim de que*, *contanto que*, *apesar de que* etc. Sobre o processo de gramaticalização sofrido até a formação de *só que*, Longhin-Thomazi (2002, p. 193) afirma que o item *que*, que compõe a perífrase “provavelmente não passa de uma conjunção integrante *que*, com a cristalização gradual da perífrase, perde a transparência e passa a funcionar simplesmente como segundo membro da construção gramaticalizada”.

Sintetizando o percurso, vemos, pois, que o processo se inicia na língua latina, a partir da palavra *solum*. *Solum* avança para o português arcaico e torna-se *sola*. *Sola* se funde a *mente*, formando o advérbio *solamente*. *Solamente* passa por um processo de gramaticalização, que envolve perda de conteúdo fonético, e reduz-se a *só*. *Só*, seguindo a tendência apontada por Said Ali (1971) para a formação de conjunções, liga-se a *que*, formando a locução conjuntiva *só que*²⁴. Vejamos tal percurso na Figura 4.

Figura 4 – Percurso histórico da formação da perífrase só que



Fonte: elaborado pela autora.

²⁴ Em Silva (2017), abordamos o percurso de gramaticalização sofrido por *solamente*, embasados no seguinte processo, elaborado por Heine (1993): dessemantização > alta frequência de uso > perda de valor informacional > erosão.

Após essa explicação sobre a origem de *mas* e *só que* na língua portuguesa, é interessante apontarmos, como fator convergente entre as duas variedades analisadas neste trabalho, a menor incidência da perífrase pela faixa etária D (+60) nas duas variedades, com apenas 4,2% dos casos, contrastando com 7% na faixa A (18-25), 11,8 % na faixa B (26-40) e 7,1% na faixa C (41-60). Não vemos, nesse contexto, ainda, um processo de mudança em curso. Apesar de a faixa etária mediana fazer mais uso dessa locução, houve um equilíbrio entre os sexos, pois homens e mulheres das duas variedades usaram-na em cerca de 8 % dos casos. Outro aspecto de convergência é o comportamento similar da variante *mas* nas duas variedades. Observemos, na tabela a seguir, essa proximidade em termos percentuais.

Tabela 14 – Comparativo percentual das ocorrências de *mas* em PB e PM

Fator	Mas PB (%)	Mas PM(%)
Idade A (18-25)	91,5	94,7
Idade B (26-40)	88,5	87,3
Idade D (+60)	95,1	96,8
Escolaridade P (primária)	95,9	89,9
Escolaridade U (universitária)	89,5	91,5
Sexo F (feminino)	90,0	91,3
Sexo M (masculino)	90,4	91,3

Fonte: elaborado pela autora.

Vale lembrar que a faixa etária C não consta da Tabela 14, pois, como explicado na seção 4.4, não houve ocorrências de *só que* nessa faixa pelos falantes brasileiros. De modo similar, também não figuram na tabela os fatores linguísticos posição entre orações (e) e posição em início de período (i), pois, conforme mencionado na seção 4.3, não foram encontradas ocorrências de *mas* no PM encabeçando o período composto por coordenação, o que inviabilizou a rodagem pelo programa.

Nota-se ainda, na Tabela 14, a proximidade numérica de *mas* nas duas variedades. Na faixa etária A (18-25), o percentual em torno de 90%; na faixa etária B (26-40), o percentual em torno de 80%; na faixa etária D (+60), o percentual de 90%; na escolaridade P (primária), a proximidade aos 90%, e o mesmo ocorrendo entre os dois sexos (masculino e feminino), tanto em PB quanto em PM.

Por sua vez, quanto aos fatores divergentes entre PB e PM, pode-se apontar, primeiramente, um aspecto já mencionado como ausente na Tabela 14. Como exposto, não foram detectadas

ocorrências de *só que* na faixa etária C (41-60) entre os entrevistados brasileiros. Em contrapartida, houve 30 casos dessa perífrase nessa faixa etária pelos falantes moçambicanos. Dessas 30 ocorrências, 8 falantes encontram-se na escolaridade primária e 22 na universitária. Mais uma vez, o grau escolar e conseqüentemente o maior acesso não só a um grau de instrução mais elevado, mas também a um público diverso, que frequenta a universidade, pode explicar esse uso, pois a forma inovadora tenderia a ocorrer mais nesses ambientes. Isso pode indicar também que *só que* não é estigmatizado, posto que é mais frequente no ambiente mais escolarizado. Esse aspecto inclusive é reforçado pelos números das demais faixas etárias entre os falantes de PM. Nesse sentido, na junção de todas as faixas etárias de nível universitário, foram encontradas 121 ocorrências da perífrase conjuncional, contrastando com apenas 13 ocorrências em todas as faixas para o público de escolaridade primária.

Outra questão importante que diverge entre as duas variedades é a possibilidade que o *mas* apresentou no PB de iniciar o período composto por coordenação, com 201 ocorrências, enquanto não foram observados casos como esse no PM. Essa diferença pode ser explicada, em termos, pela modalidade de configuração de cada *corpora*. O C-oral Brasil I (Raso; Mello, 2012), utilizado para as amostras de PB, situa-se como um “*corpus* balanceado, representativo das três principais tipologias da fala – diálogos, monólogos e conversações –, em contextos públicos e privados” (Raso; Mello, 2012, p. 1), de modo a favorecer as interações discursivas. No prefácio à obra, essa questão é explorada por Cresti e Moneglia (2012):

a escolha do C-ORAL-ROM e, com sucesso bem maior, a escolha do C-ORAL-BRASIL tentaram modificar o estado da arte dando espaço à maior variação diafásica possível. Por um lado, renunciou-se a um balanceamento diastrático da amostra, por outro, escolheram-se fixar critérios gerais que aumentassem as probabilidades de ocorrência do maior número possível de situações de interação específicas do uso informal da linguagem: a) variação por estrutura do evento comunicativo (diálogo, monólogo, multidialógico); b) variação por contexto social em que o evento ocorre (vida pública *versus* vida privada e familiar). Cruzando os dois critérios e requerendo a maior variação possível de situações em cada ramificação da amostragem, conseguimos documentar o informal muito além dos bate-papos, com muitas amostras de fala fortemente interativa em situação (enquanto se conserta um carro, enquanto se briga, enquanto se dá aula de direção, enquanto se revelam fotografias, enquanto se tenta entrar em acordo com os amigos sobre o que comprar e como pagar, enquanto se lida com clientes em um bar, no guichê dos impostos, no trabalho etc.) (Cresti; Moneglia, 2012, p. 16).

Portanto, uma vez que a própria configuração do *corpus* estimula a interação entre os falantes, as ocorrências de *mas* iniciando o período demonstram que o entrevistado inicia seu turno de

fala, o que configura a posição inicial, consoante o que abordamos no capítulo 3 deste trabalho, mas esse turno de fala complementa o que anteriormente dito por outro falante. Consideremos o seguinte exemplo:

- (107) *RUT: eu nũ gosto de pensar que tá um pessoal olhando pra mim não // \$
 *TER: <mas quando a Paulinha> casar / cê <vai ter que entrar> // \$ (Raso; Mello, 2012).

Em (107), o *mas* inicia o turno da falante situando-se em posição inicial (i). Entretanto esse marcador quebra uma expectativa dada na fala de outro entrevistado. Desse modo, coadunamos as palavras de Perini (1995), Matos (2003) e Bagno (2012), autores segundo os quais, conforme mencionado no capítulo 1, o *mas* não experimenta mobilidade dentro do período. Nesse caso, ele continua na posição esperada de ligar orações, embora sejam orações ditas por diferentes pessoas. Como o *corpus* de PM se configura dentro da modalidade de entrevistas individuais (Silva, 2025, p. 86), não houve favorecimento à interação entre outros falantes, o que não resultou em marcadores de quebra de expectativa em posição de início frasal.

Nota-se, portanto, que as duas variedades, em relação ao fenômeno da marcação da quebra de expectativa, apresentam convergências e divergências. Neste capítulo, buscamos, primeiramente, pormenorizar as ocorrências de marcadores adversativos encontrados em nosso *corpora*. Nesse sentido, elucidamos as incidências de *entretanto*, *mas*, *no entanto*, *porém* e *só que*. Em seguida, procuramos distribuir tais ocorrências de acordo com os critérios linguísticos e extralinguísticos previamente selecionados em nossa metodologia. Esses critérios foram: faixa etária, escolaridade, sexo, precedência de outro marcador e posição dele dentro do período composto por coordenação. Dando sequência à nossa análise, esmiuçamos esses resultados separadamente com relação a cada variedade pesquisada, quais sejam: português brasileiro e português moçambicano. Posteriormente, submetemos esses dados ao programa estatístico Goldvarb 3.0 e conduzimos algumas considerações numéricas para, em seguida, adotarmos uma análise mais qualitativa desses dados. Logo, demos sequência a este capítulo com algumas considerações sobre as ocorrências de *mas só que*, o único caso de precedência detectado. Por fim, encerramos este capítulo apontando algumas convergências e divergências entre as duas variedades. Vejamos, pois, no capítulo a seguir, nossas considerações finais.

CONSIDERAÇÕES FINAIS

Neste trabalho, vimos que a quebra de expectativa ou adversatividade é um fenômeno linguístico que ocorre quando uma segunda informação é capaz de quebrar a expectativa dada em uma primeira informação, por exemplo, em: “o massacre de Manuel Rubens é uma peça muito bacana, só que infelizmente ela foi proibida pela Censura Federal” (Adaptado de Pezatti; Longhin-Thomazi, 2008, p. 916). Nesse exemplo, a primeira informação dada é o juízo de valor positivo atribuído à peça *O massacre de Manuel Rubens*. Com essa primeira informação, pode-se gerar inúmeras expectativas, que, também, conforme discutimos no capítulo 1 deste trabalho, são subjetivas a cada falante, pois dependem de aspectos sociais e particulares, como a idade, o grau de instrução, o círculo social que esse falante frequenta, suas convicções e visões de mundo etc. (Martelotta, 1998). No exemplo anterior, uma expectativa subjetiva por nós abstraída é que, ao dizer que a peça é “bacana”, pode-se pensar que ela tenha tido uma repercussão positiva e exitosa. Essa possível expectativa do êxito gerado pela peça é quebrada na segunda informação, quando se diz que ela foi proibida pela censura. Nesse caso, a proibição teria impedido o êxito da obra. Vale frisar, mais uma vez, que essas interpretações pessoais sobre as expectativas e as quebras geradas são subjetivas, e o que fazemos neste trabalho é uma tentativa de interpretação, que não necessariamente corresponde ao que o falante originalmente pretendia com sua fala. Como nosso trabalho é de base sociolinguística, embora analisemos e interpretemos as ocorrências de orações adversativas sindéticas no PB e no PM e verifiquemos nelas a presença do rompimento de uma expectativa inicialmente gerada, evidenciando uma oposição ou um contraste, buscamos nos concentrar no fenômeno variável, isto é, na variação existente nas orações que continham *mas*, *só que*, *mas só que*, *porém*, *entretanto* e *no entanto*, não nos atendo, portanto, na explicação de questões discursivas presentes em todas as frases encontradas.

Voltando ao exemplo anterior, temos o marcador de quebra de expectativa *só que*, responsável por encabeçar a segunda informação dada e por quebrar a expectativa anteriormente criada. Propusemo-nos a investigar o fenômeno da quebra de expectativa em duas variedades de língua portuguesa: quais sejam, o português brasileiro (PB), e o português moçambicano (PM), utilizando, para a primeira variedade, o *corpus* C-oral Brasil I (Raso; Mello, 2012), e, para a segunda, o *corpus* Corporaport (Vieira; Pissurno, 2016), ambos de língua falada, modalidade informal. Cientes de que a pesquisa variacionista tem suas limitações, coletamos, em ambos os *corpora*, as estruturas adversativas sindéticas, isto é, aquelas em que o marcador é explícito,

não nos detendo nas estruturas de quebra de expectativa assindéticas, em que há a quebra, mas a conjunção ou locução conjuntiva adversativa não está presente, como em: “a gente perdeu no campo, a gente não perdeu por causa disso não” (Adaptado de Raso; Mello, 2012). Nesse exemplo, é possível colocarmos uma conjunção adversativa: “a gente perdeu no campo, *mas* a gente não perdeu por causa disso não”. Ao tentarmos interpretá-lo, vemos que a derrota em campo poderia ter se dado por questões ligadas à partida em si, quando se declara, na primeira informação, a perda do jogo; mas, segundo o falante, a derrota teria se dado por questões fora de campo, o que quebra a expectativa inicial.

Após selecionarmos as ocorrências de estruturas adversativas nas duas variedades, submetemos tais amostras ao programa estatístico Goldvarb 3.0, estratificando-as segundo a faixa etária, o grau de escolaridade, o sexo dos falantes, a posição do marcador dentro do período composto por coordenação e a precedência ou não de outros marcadores adversativos no período. Os fatores linguísticos pesquisados neste trabalho foram selecionados levando em consideração os estudos sobre as orações adversativas de Neves (2000), Ilari (2008), Pezatti e Longhin-Thomazi (2008), Castilho (2010), Perini (2016) entre outros autores citados no capítulo 1. Já os fatores extralinguísticos foram selecionados tendo como base pesquisadores variacionistas, cujos trabalhos encontram-se consolidados no campo. Nesse sentido, conforme mencionado no capítulo 2, podemos citar obras de Labov (1972), Tarallo (1986), Guy e Zilles (2007), Coelho *et al.* (2015) entre outros.

Após a rodagem dos dados no programa, alguns resultados encontrados foram particularmente interessantes. Primeiramente, com relação aos fatores internos à língua, não foram encontrados, nas duas variedades, marcadores em posição final de período, apenas em posição intermediária, ligando as orações coordenadas, ou em posição de início de período. Esse resultado vai ao encontro do que afirmam Pezatti e Longhin-Thomazi (2008). Segundo as autoras, a construção coordenada compõe-se de dois ou mais membros ligados por mecanismos de conexão. Portanto, se os marcadores adversativos são responsáveis por fazer a conexão entre os membros coordenados, é esperado que apareçam majoritariamente em posição intermediária, conectando os membros. Como vimos neste trabalho, as ocorrências observadas em posição de início de período devem-se geralmente à característica do *corpus* brasileiro, que incentiva a interação entre os diversos falantes que estão sendo entrevistados. Logo, a posição inicial do marcador refere-se ao fato de o falante estar complementando a fala de outra pessoa. Em nossa coleta, ele aparece no início da frase, mas, se observarmos o *corpus*, esse início de frase na verdade está

completando a fala de outro entrevistado. Outro fator interessante foram as ocorrências de *entretanto*, *no entanto* e *porém* em ambas as variedades. Como foram detectadas apenas duas, uma e uma ocorrências respectivamente, optamos por analisá-los mais qualitativamente e os excluimos do programa estatístico. Assim, obtivemos os dois marcadores mais frequentes nas duas variedades, que foram *mas*, considerada a variante conservadora, frente a *só que*, considerada a variante inovadora. Não observamos nesses dois marcadores traços de estigmatização, posto que obtiveram um maior percentual de uso pelos falantes de nível escolar universitário (86,8% – vide Tabela 5), se considerarmos a rodagem que inclui as duas variedades juntamente. Esse resultado, de certo modo, coaduna as questões que traçamos no capítulo 3 a respeito de reconhecermos a divergência de aquisição do PB frente ao PM, posto que a primeira variedade se trata de uma língua materna e a segunda, em muitos casos de nosso *corpus*, é segunda língua para os entrevistados. Como percebemos, pelos dados, que o fenômeno variável em análise não é estigmatizado, as divergências de aquisição suavizam-se.

Em relação ao fator faixa etária, a faixa etária B, dos falantes de 26 a 40 anos, foi a que mais usou *só que*. Inclusive vale lembrar que o grupo faixa etária foi o único considerado estatisticamente relevante. A isso atribuímos o fato de os grupos mais jovens, de 18 a 40 anos (faixas etárias A e B), protagonizarem o uso da variante inovadora, o que poderia indicar um princípio de mudança linguística, embora não o possamos afirmar apenas com esses dados. Nesse sentido, outro aspecto que também sugere um percurso de mudança é o fato de *só que* ser mais utilizado pelas falantes de sexo feminino. Foram 76 ocorrências relativas a mulheres face a 58 ocorrências relativas a homens. Entretanto, precisaríamos ampliar os dados para obtermos conclusões mais robustas.

Ainda apontando tendências de mudança linguística, foram encontradas 13 ocorrências de *mas só que*, nove em PB e quatro em PM. Todas as ocorrências pertencem à faixa etária C, dos falantes de 41 a 60 anos, de escolaridade universitária, quatro de sexo masculino e três de sexo feminino. Como a conjunção *mas*, variante conservadora, foi mais frequente em ambos os *corpora* e em todos os grupos de fatores, obtendo 91,6% de predomínio no grupo faixa etária, 91,2% no grupo grau de escolaridade e 91,6 no grupo sexo, obviamente é mais usual que *só que*, variante inovadora. Para Meillet (2014[1921]), o uso frequente das palavras acaba retirando delas o valor expressivo. Poderíamos, então, hipotetizar que essa maior frequência de *mas* poderia estar desgastando-o, este, por sua vez, estar-se-ia apoiando em *só que*, segunda variante mais usada. Esse possível apoio de *mas* em relação a *só que* apresenta um

comportamento semelhante ao de redobro. O redobro ou reduplicação é um fenômeno linguístico que se caracteriza pela repetição de um ou vários elementos de uma palavra com fins expressivos (Dubois, 1995). Vitral (1999) analisa o redobro na perspectiva da negação em PB. Para ele, a redução da partícula de negação pré-verbal faz surgir itens negativos pós-verbais, como em “Maria *num* fez *não*”. De modo análogo, em nosso trabalho, a possível redução e desgaste de *mas* atrai *só que*, como num processo de redobro, de reduplicação da quebra de expectativa. Inclusive, conforme dissemos no capítulo 3, esse comportamento já foi observado no português arcaico, com *mas porém* e *mas entretanto* (Pezzatti; Longhin-Thomazi, 2008). Não obstante, conforme dissemos anteriormente, precisaríamos ampliar os dados para afirmarmos com mais segurança a possibilidade de redobro de *mas*, acionando *só que*.

Por fim, podemos concluir, com este trabalho, que a marcação da quebra de expectativa se dá de modo bastante similar entre as duas variedades pesquisadas. Tanto em PB quanto em PM, o *mas* foi mais frequente, em posição intermediária, alcançando 91,8% de ocorrências no PB e 91,3% no PM. Em ambas as variedades, também foram mínimas as ocorrências de marcadores considerados mais formais, como *entretanto*, *no entanto* e *porém*. Esse resultado está dentro do esperado, posto que os *corpora* se constituem de amostras de língua falada em situação de comunicação informal. Logo se espera que as variantes consideradas mais formais e restritas à escrita sejam realmente menos frequentes. Para finalizar, destacamos o possível percurso de um princípio de mudança na marcação da quebra de expectativa, sobretudo na variedade brasileira. Isso porque, como exposto, a variante inovadora esteve mais presente entre os falantes mais jovens (vide Tabela 13), alcançando, inclusive, zero ocorrências entre falantes medianos (41 a 60 anos) e apenas duas ocorrências entre falantes idosos (acima de 60 anos). Pode-se observar traços de uma mudança de baixo para cima, posto que, das 72 ocorrências totais, 69 se encontram na faixa mais escolarizada (nível universitário) e apenas 3 encontram-se na faixa dos falantes menos escolarizados, de nível universitário. Além disso, os usos de *mas só que* podem demonstrar um possível enfraquecimento e desgaste da variante mais usual (*mas*), que estaria começando a se apoiar na segunda variante mais usual (*só que*) para ganhar expressividade. Esse possível enfraquecimento e desgaste demonstra um provável rearranjo no quadro da marcação da quebra de expectativa em PB e PM.

Como contribuições deixadas por este trabalho, cabe mencionar a ampliação dos estudos contrastivos entre variedades linguísticas, dado o ineditismo ao compararmos PB e PM quanto à marcação da quebra de expectativa. Outro aspecto contributivo desta pesquisa seria o

incentivo a um maior estudo da variedade moçambicana, pois, segundo Timbane e Berlinck “Moçambique ainda não tem seu próprio dicionário nem gramática que mostra as especificidades do Português falado/escrito pelos moçambicanos.” (Timbane; Berlinck, 2012, p. 209). Assim, uma vez que utilizamos o PM falado e o discutimos, encorajamos novos estudos que também o façam.

Acreditamos, também, que pudemos contribuir para uma maior propagação de conhecimento sobre alguns aspectos linguísticos referentes ao português moçambicano. Além disso, nossa pesquisa estimula a compreensão e o estudo da variante inovadora *só que*, não só nas variedades pesquisadas, mas também em outras variedades da língua portuguesa.

Para futuras pesquisas, reforçamos nosso intuito de analisar um possível processo de mudança linguística envolvendo a marcação da quebra de expectativa, processo em que a variante inovadora poderia ocupar o lugar da conservadora no fenômeno variável. Seria interessante também, em oportunos momentos futuros, ampliarmos os dados a fim de verificarmos mais fortemente a influência do fator linguístico faixa etária na ocorrência do fenômeno. Outro aspecto de interesse poderia ser a verificação de possíveis ocorrências de *mais* marcando a quebra expectativa nas duas variedades. Em nossos *corpora*, não encontramos esse tipo de uso, embora ele possa ocorrer em outras amostras. Nossa hipótese é a de que se trata de uma variação em nível fonológico, não identificável nas transcrições dos *corpora* utilizados neste trabalho. Logo, seria interessante uma pesquisa que se valesse dos áudios, para verificarmos as incidências de *mas* e *mais* como marcadores de quebra de expectativa.

Finalizamos, pois, este trabalho esclarecendo se tratar de um estudo importante para a compreensão do fenômeno chamado quebra de expectativa, mas trata-se, também, de um trabalho não exaustivo, que não tem a pretensão de esgotar o fenômeno, cabendo, inclusive, como dito, a ampliação dos dados para o estabelecimento de novas conclusões ou mesmo o fortalecimento das conclusões já feitas.

REFERÊNCIAS

- AGUILERA, Vanderci de Andrade. **Atlas Linguístico do Brasil. Volume 1**. Londrina: Eduel, 2018.
- ANTES. In: **DICIONÁRIO Priberam**. Disponível em: <https://dicionario.priberam.org/>. Acesso em: 29 mai. 2024.
- ANTUNES, Irlandé. **Gramática contextualizada: limpando ‘o pó das ideias simples’**. São Paulo: Parábola, 2014.
- ASH, Sharon. Social class. In: CHAMBERS, J. K.; TRUDGILL, Peter; SCHILLING-ESTES, Natalie (ed.). **The handbook of language variation and change**. Oxford: Blackwell Publishers, 2002. p. 143–160.
- AULETE, Caldas. **Novíssimo Aulete dicionário contemporâneo da língua portuguesa**. Rio de Janeiro: Lexicon, 2011.
- BAGNO, Marcos. **Português ou brasileiro? Um convite à pesquisa**. São Paulo: Parábola Editorial, 2001 [1961].
- BAGNO, Marcos. **Gramática pedagógica do português brasileiro**. São Paulo: Parábola Editorial, 2012.
- BIBER, Douglas. **Variation across speech and writing**. Cambridge: Cambridge University Press, 2012 [1988]. Disponível em: <https://www.cambridge.org/core/books/variation-across-speech-and-writing/A546CF5ED8F8E62F1432CB2F369CF356>. Acesso em: 1º jun. 2024.
- BELL, Roger T. **Sociolinguistics: goals, approaches and problems**. London: Batsford, 1976.
- BIBER, Douglas. Norming frequency counts. In: BIBER, Douglas; CONRAD, Suzan; REPPEN, Randi. **Corpus linguistics: investigating language structure and use**. Cambridge: Cambridge University Press, 1998. p. 119–131.
- BORIN, Maísa Augusta. **Sociolinguística**. Santa Maria: Universidade Federal de Santa Maria, 2010.
- BURGOON, Michael; DENNING, Vickie Pauls; ROBERTS, Laura. Language expectancy theory. In: DILLARD, James Price; PFAU, Michael (ed.). **The persuasion handbook: developments in theory and practice**. 2. ed. Thousand Oaks: Sage Publications, 2002. p. 117–136.
- CALLOU, Dinah; MARQUES, Maria Helena Duarte. O –s implosivo na linguagem do Rio de Janeiro. **Littera**, Rio de Janeiro, v. 14, p. 9–137, 1975.
- CASTILHO, Ataliba Teixeira de. **Nova gramática do português brasileiro**. São Paulo: Contexto, 2010.
- CASTRO, Matilde Ángeles Vida; MUÑOZ, Antonio Manuel Ávila; ESQUIVEL, Francisco M. Carriscondo. **Manual práctico de sociolingüística**. Madrid: Síntesis, 2016.
- CEGALLA, Domingos Paschoal. **Novíssima gramática da língua portuguesa**. São Paulo: Companhia Editora Nacional, 2005.

- CEGALLA, Domingos Paschoal. **Dicionário de dificuldades da língua portuguesa**. Rio de Janeiro: Lexicon, 2009.
- CEGALLA, Domingos Paschoal. **Dicionário de dificuldades da língua portuguesa**. Rio de Janeiro: Lexicon, 2012.
- CHESHIRE, Jenny. Sex and gender in variationist research. In: CHAMBERS, J. K.; TRUDGILL, Peter; SCHILLING-ESTES, Natalie (ed.). **The handbook of language variation and change**. Oxford: Blackwell Publishers, 2002. p. 223–246.
- CHIERICI, Paloma. Distinção morfossintática entre verbos inergativos e inacusativos. **Portal de periódicos da UFMG**, 2008. Disponível em: <https://periodicos.ufmg.br/index.php/revele/article/view/11233/7971>. Acesso em: 20 ago. 2025.
- COELHO, Izete Lehmkuhl et al. **Para conhecer sociolinguística**. São Paulo: Contexto, 2015.
- CONJUNÇÕES COORDENATIVAS. **Só português**. Virtuoso Tecnologia da Informação, 2007. Disponível em: <https://www.soportugues.com.br/secoes/morf/morf85.php>. Acesso em: 13 set. 2023.
- COULMAS, Florian (ed.). **The handbook of sociolinguistics**. Oxford: Blackwell, 1997.
- CRESTI, Emanuela; MONEGLIA, Massimo. Prefácio. In: RASO, Tommaso; MELLO, Heliana (orgs.). **C-Oral-Brasil I: corpus de referência do português brasileiro falado informal**. Belo Horizonte: Editora UFMG, 2012. p. 13–25.
- CUNHA, Angélica Furtado da. Funcionalismo. In: MARTELOTTA, Mário Eduardo (org.). **Manual de linguística**. São Paulo: Contexto, 2009. p. 157–176.
- DAVIES, Mark. **O Corpus do Português**. Disponível em: <https://www.corpusdoportugues.org/>. Acesso em: 1º jul. 2025.
- DIK, Simon Cornelis. **Functional grammar**. Dordrecht-Holland: Foris Publications, 1978.
- DUARTE, Maria Eugênia L. Avanço no estudo da mudança sintática associando a Teoria da Variação e Mudança e a Teoria de Princípios e Parâmetros. **Cadernos de Estudos Linguísticos**, Campinas, v. 57, Jan./Jun. 2015, p. 85–111. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/cel/article/view/8641473>. Acesso em: 7 mar. 2023.
- DUBOIS, Jean. **Dicionário de linguística**. 5. ed. São Paulo: Cultrix, 1995.
- DUCROT, Oswald; VOGT, Carlos Alberto. De magis à mais: une hypothèse sémantique. **Revue de linguistique romane**, v. 43, p. 317–341, 1979. Disponível em: <https://www.e-periodica.ch/cntmng?pid=rlr-001:1979:43::517>. Acesso em: 7 mar. 2023.
- ECKERT, Penelope. The whole woman: sex and gender differences in variation. In: CACOULLOS, Rena Torres; STANFORD, James N. (ed.). **Language variation and change**, v. 1. Cambridge: Cambridge University Press, 1989. p. 245–267. Disponível em: <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=6c3e385a18d371b03afb63b9afea0d19494207c3>. Acesso em: 2 ago. 2023.
- ECKERT, Penelope. Ages as a sociolinguistic variable. In: COULMAS, Florian (ed.). **The handbook of sociolinguistics**. Oxford: Blackwell, 1997. p. 95–112.

- ETTO, Rodrigo Mazzer; CARLOS, Valeska Gracioso. **Sociolinguística: o papel do social na língua**. São José do Rio Preto: Instituto de Biociências, Letras e Ciências Exatas, 2017. Disponível em: https://anhanguera.edu.br/wp-content/uploads/02_sociolinguistica_o_papel_do_social_na_ling.pdf. Acesso em: 7 mar. 2023.
- FERGUSON, Charles Albert. **Language, structure and language use**. California: Stanford University Press, 1972.
- FLAMENCO GARCÍA, Luis. Las construcciones concesivas y adversativas. In: BOSQUE, Ignacio; DEMONTE, Violeta (dir.). **Gramática descriptiva de la lengua española**, v. 3. Madrid: Espasa, 1999. p. 3855–3878.
- FREITAG, Raquel Meister Ko. Idade: uma variável sociolingüística complexa. **Línguas & Letras**, [S. l.], v. 6, n. 11, p. 105–121, 2000. DOI: 10.5935/rl&l.v6i11.875. Disponível em: <https://saber.unioeste.br/index.php/linguaseletras/article/view/875>. Acesso em: 4 set. 2023.
- GONÇALVES, Perpétua. A gênese do português de Moçambique. **Revista de Estudos Linguísticos da Universidade do Porto**, n. 1, v. 6, p. 201–204, 2011.
- GUY, Gregory Riordan; ZILLES, Ana. **Sociolinguística quantitativa – instrumental de análise**. São Paulo: Parábola Editorial, 2007.
- HALLIDAY, Michael Alexander Kirkwood; HASAN, Ruqaiya. **Cohesion in English**. New York: Longman, 1995 [1976].
- HALLIDAY, Michael Alexander Kirkwood; HASAN, Ruqaiya. **Language, context, and text: aspects of language in a social-semiotic perspective**. Oxford: Oxford University Press, 1989.
- HAUGEN, Einar. Dialect, language, nation. **American Anthropologist**, v. 68, p. 922–935, 1966.
- HEINE, Bernd; CLAUDI, Ulrike; HÜNNEMEYER, Friederike. **Grammaticalization: a conceptual framework**. Chicago: The University of Chicago Press, 1991.
- HUDSON, Richard Anthony. **Sociolinguistics**. Cambridge: Cambridge University Press, 1980.
- HUDSON, Richard Anthony. **Sociolinguistics**. Cambridge: Cambridge University Press, 1996.
- ILARI, Rodolfo. As conjunções. In: ILARI, Rodolfo; NEVES, Maria Helena de Moura (orgs.). **Gramática do português culto falado no Brasil**, v. 2. Campinas: Editora da UNICAMP, 2008. p. 805–860.
- JOHNSTONE, Barbara. **Qualitative methods in sociolinguistics**. New York: Oxford University Press, 2000.
- KEHLER, Andrew. **Coherence, reference and theory of grammar**. Stanford: CSLI Publications, 2002.
- KOCH, Ingedore. A articulação entre orações no texto. **Cadernos de Estudos Linguísticos**, v. 28, p. 9–18, 1995.

- LABOV, William. The linguistic variable as a structural unit. **Washington Linguistics Review**, v. III, 1966. Disponível em: <https://files.eric.ed.gov/fulltext/ED010871.pdf>. Acesso em: 31 jul. 2023. p. 3–22.
- LABOV, William. **Padrões sociolinguísticos**. Tradução de Marcos Bagno, Maria Marta Pereira Scherre, Caroline Rodrigues Cardoso. São Paulo: Parábola Editorial, 2008 [1972].
- LABOV, William. Where does the linguistic variable stop? **Sociolinguistic Working Papers**, Austin, n. 44, 1978. Disponível em: <https://eric.ed.gov/?id=ED157378>. Acesso em: 7 mar. 2023.
- LABOV, William. Building on empirical foundations. In: LEHMANN, Winfred; MALKIEL, Yákov (orgs.). **Perspectives on historical linguistics**. Amsterdam/Philadelphia: John Benjamins, 1982. p. 17–92.
- LABOV, William. The intersection of sex and social class in the course of linguistic change. In: CACOULOS, Rena Torres; STANFORD, James N. (ed.). **Language variation and change**, v. 35, n. 1. Cambridge: Cambridge University Press, 1990. p. 205–254. Disponível em: <https://www.cambridge.org/core/services/aop-cambridge-core/content/view/AAA8227B739187F5D2CBDA51EA212FD8/S0954394500000338a.pdf/the-intersection-of-sex-and-social-class-in-the-course-of-linguistic-change.pdf>. Acesso em: 2 ago. 2023.
- LABOV, William. **Principles of linguistic change II: social factors**. Oxford: Blackwell, 2001.
- LAKOFF, Robin Tolmach. If’s, and’s and but’s about conjunction. In: FILLMORE, Charles; LANGENDOEN, Donald Terence (eds.). **Studies in linguistic semantics**. New York: Holt, Rinehart and Winston, 1971. p. 45–68.
- LAVANDERA, Beatriz Rosario. **Variación y significado**. Buenos Aires: Hachette, 1984.
- MARCUSCHI, Luiz Antônio. **Fenômenos da linguagem: reflexões semânticas e discursivas**. Rio de Janeiro: Lucerna, 2007.
- MARTELOTTA, Mário Eduardo. Gramaticalização e graus de vinculação sintática em cláusulas concessivas e adversativas. **Veredas: Revista de Estudos Linguísticos**, v. 2, n. 2, 1998, p. 37–56. Disponível em: <http://periodicos.ufjf.br/index.php/veredas/article/view/xxxx>. Acesso em: 14 dez. 2023.
- MATOS, Gabriela. Estruturas de coordenação. In: MATEUS, Maria Helena Mira et al. **Gramática da língua portuguesa**, 6. ed. Lisboa: Caminho, 2003. p. 549–592.
- MEILLET, Antoine. **Linguistique historique et linguistique générale**. Paris: Librairie Ancienne Honoré Champion, 2014 [1921].
- MEILLET, Antoine. **The Indo-European dialects**. Tuscaloosa: University of Alabama Press, 2005 [1967].
- MOLLICA, Maria Cecilia; BRAGA, Maria Luiza (orgs.). **Introdução à sociolinguística: o tratamento da variação**. São Paulo: Contexto, 2003.
- MORENO FERNÁNDEZ, Francisco. **Metodología sociolingüística**. Madrid: Gredos Editorial, 1990.
- MORENO FERNÁNDEZ, Francisco. **Principios de sociolinguística y sociología del lenguaje**, 4. ed. Barcelona: Editorial Ariel, 2017.

- MOURA, José de Almeida. **Gramática do português actual**. Lisboa: Lisboa Editora, 2004.
- NARO, Anthony Julius; SCHERRE, Maria Marta Pereira. Variação e mudança linguística: fluxos e contrafluxos na comunidade de fala. **Caderno de Estudos Linguísticos**, Unicamp, v. 20, p. 9–16, Jan./Jun. 1991. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/cel/article/view/8636853>. Acesso em: 7 mar. 2023.
- NARO, Anthony Julius. Modelos quantitativos e tratamento estatístico. In: MOLLICA, Maria Cecília; BRAGA, Maria Luiza (orgs.). **Introdução à sociolinguística: o tratamento da variação**. São Paulo: Contexto, 2003. p. 211–238.
- NAYADE, Priscila. **Conjunção**. Todo Estudo. Disponível em: <https://www.todoestudo.com.br/portugues/conjuncao>. Acesso em: 29 mai. 2024.
- NEVES, Maria Helena Moura. O coordenador interfrasal mas — invariância e variantes. **Alfa**, São Paulo, v. 28, p. 21–42, 1984. Disponível em: <https://periodicos.fclar.unesp.br/alfa/article/view/3665>. Acesso em: 7 mar. 2023.
- NEVES, Maria Helena Moura. **Gramática de usos do português**. São Paulo: UNESP, 2000.
- NEVES, Maria Helena Moura. **A gramática funcional**. São Paulo: Martins Fontes, 2001 [1997].
- PAGOTTO, Emilio Gozze. **Variação e identidade**. Campinas, 2001. Tese (Doutorado em Linguística), Universidade Estadual de Campinas, 311 f.
- PAIVA, Maria da Conceição de; DUARTE, Maria Eugênia Lamoglia. **Mudança linguística em tempo real**. Rio de Janeiro: Contra Capa, 2003.
- PEDRO, Fernando Laerty Ferreira da Silva. **O uso dos conectores adversativos “mas” e “só que” na fala do natalense em entrevistas sociolinguísticas: foco na variação discursiva**. 2019. 111 f. Dissertação (Mestrado em Estudos da Linguagem) – Universidade Federal do Rio Grande do Norte, Natal, 2019.
- PERINI, Mário Alberto. **Gramática descritiva do português**. São Paulo: Ática, 1995.
- PERINI, Mário Alberto. **Gramática do português brasileiro**. São Paulo: Parábola, 2010.
- PERINI, Mário Alberto. **Gramática descritiva do português brasileiro**. Petrópolis: Vozes, 2016.
- PEZATTI, Erotilde Goreti; LONGHIN-THOMAZI, Sanderléia Roberta. As construções coordenadas. In: ILARI, Rodolfo; NEVES, Maria Helena de Moura (orgs.). **Gramática do português culto falado no Brasil**, v. 2. Campinas: Editora da UNICAMP, 2008. p. 865–936.
- RAJAGOPALAN, Kanavillil. A linguística de corpus no tempo e no espaço: visão reflexiva. In: GERBER, Regina Márcia; VASILÉVSKI, Vera (orgs.). **Um percurso para pesquisas com base em corpus**. Florianópolis: Editora UFSC, 2007. p. 101–123.
- RASO, Tommaso; MELLO, Heliana (orgs.). **C-Oral-Brasil I: corpus de referência do português brasileiro falado informal**. Belo Horizonte: Editora UFMG, 2012.
- ROCHA, Ana Paula. Gramaticalização da conjunção mas: reflexões a partir do modelo de Sweetser (1991). **Revista Domínios de Linguagem**, Ano 1, n. 2, 2007. Disponível em: <https://seer.ufu.br/index.php/dominiosdelinguagem/article/view/11435>. Acesso em: 30 mai. 2023.

SAID ALI, Manuel. **Gramática histórica da língua portuguesa**, 7. ed. São Paulo: Melhoramentos, 1971.

SARDINHA, Tony Berber. **Linguística de corpus**. Barueri: Manole, 2004.

SAUSSURE, Ferdinand de; CHELINI, Antonio; PAES, José Paulo; BLIKSTEIN, Izidoro. **Curso de linguística geral**, 28. ed. São Paulo: Cultrix, 2012.

SCHÄFER, Wolf. Finalização em perspectiva: rumo a uma revolução no paradigma social da ciência. **Informação em Ciências Sociais**, v. 18, n. 6, p. 915–943, 1979. Disponível em: <https://doi.org/10.1177/053901847901800607>. Acesso em 24 out. 2025.

SCHILLING-ESTES, Natalie. Investigating stylistic variation. In: CHAMBERS, J. K.; TRUDGILL, Peter; SCHILLING-ESTES, Natalie (ed.). **The handbook of language variation and change**. Oxford: Blackwell Publishers, 2002. p. 316–335.

SCRIBBR. **Chi-square distribution table [PDF]**. Disponível em: <https://www.scribbr.com/wp-content/uploads/2022/05/Chi-square-table.pdf>. Acesso em: 4 jun. 2025.

SILVA, Claudia Roberta Tavares. A importância e os desafios da constituição de corpora de variedades africanas do português: breves relatos de uma experiência de pesquisa. In: ARAUJO, Silvana Silva de Farias; SILVA, Manoel Crispiniano Alves da; ARAUJO, Jean Marcel Oliveira (orgs.). **Lusofonia Afro-Brasileira: questões sócio-históricas e linguísticas**. Campinas: Pontes Editores, 2025. p. 71–102.

SILVA, Edila Vianna da. A pesquisa sociolinguística: a teoria da variação. **Revista da Academia Brasileira de Filologia**, Rio de Janeiro, Ano IX, n. IX, p. 49–57, 2011. Disponível em: <http://www.filologia.org.br/abf/rabf/9/049.pdf>. Acesso em: 7 mar. 2023.

SILVA, Luiza Francisca Ferreira da. **A perífrase conjuncional só que: gramaticalização e variação linguística**. 111 f. Dissertação (Mestrado em Linguística Teórica e Descritiva) – Universidade Federal de Minas Gerais, Belo Horizonte, 2017.

SILVA, Luiza Francisca Ferreira da. **A perífrase conjuncional só que: gramaticalização e variação linguística**. 47 f. Monografia (Bacharelado em Letras – Português, com ênfase em Estudos Linguísticos) – Universidade Federal de Minas Gerais, Belo Horizonte, 2013.

SOTO, Guillermo. Ignacio Bosque y Violeta Demonte (directores): **Gramática descriptiva de la lengua española. Onomázein: Revista de lingüística, filología y traducción de la Pontificia Universidad Católica de Chile**, n. 6, 1999. p. 345–351.

SWEETSER, E. **From etymology to pragmatics**. Cambridge: Cambridge University Press, 1991.

TAGLIAMONTE, Sali. **Analysing sociolinguistic variation**. Cambridge: Cambridge University Press, 2006.

TARALLO, Fernando. **A pesquisa socio-linguística**. São Paulo: Ática, 1986.

TARALLO, Fernando. Por uma sociolinguística românica “paramétrica”: fonologia e sintaxe. **Ensaios de Linguística**, UFMG, v. 13, p. 51–84, 1987. Disponível em: <https://periodicos.ufmg.br/index.php/cltl/article/view/8009>. Acesso em: 7 mar. 2023.

TIMBANE, Alexandre António; BERLINCK, Rosane de Andrade. A norma-padrão europeia e a mudança linguística na escola moçambicana. **Revista Gragoatá**. N. 32. P. 207-226. 2012.

TIMBANE, Alexandre António. A variação linguística do português moçambicano: uma análise sociolinguística da variedade em uso. **Revista Internacional em Língua Portuguesa**, n. 32, 2017, p. 19–38.

TRUDGILL, Peter. **Sociolinguistics: an introduction to language and society**, 4. ed. London: Penguin, [1974] 2000.

TRUDGILL, Peter. Social differentiation. In: CHAMBERS, Jack Kenneth; TRUDGILL, Peter; SCHILLING-ESTES, Natalie (ed.). **The handbook of language variation and change**. Oxford: Blackwell Publishers, 2002. p. 38–62.

VASILÉVSKI, Vera. Aspectos histórico-teóricos da linguística de corpus: surgimento, abandono, levante e uso. In: GERBER, Regina Márcia; VASILÉVSKI, Vera (orgs.). **Um percurso para pesquisas com base em corpus**. Florianópolis: Editora UFSC, 2007. p. 13–45.

VIDA CASTRO, Matilde Ângeles; ÁVILA MUÑOZ, Antonio Manuel; CARRISCONDO ESQUIVEL, Francisco M. **Manual práctico de sociolingüística**. Madrid: Síntesis, 2016.

VIEIRA, Silvia Rodrigues; PISSURNO, Karen Cristina da Silva (orgs.). **Corpus Moçambique-PORT: Maputo**. Rio de Janeiro: Faculdade de Letras/UFRJ, 2016. Disponível em: www.corporaport.letras.ufrj.br. Acesso em: 17 mai. 2023.

VIEIRA, Silvia Rodrigues; BRANDÃO, Silvia Figueiredo. O desafio de analisar variáveis sociolinguísticas frente a um mosaico de etnias e línguas: o caso do português de Moçambique. In: ARAUJO, Silvana Silva de Farias; SILVA, Manoel Crispiniano Alves da; ARAUJO, Jean Marcel Oliveira (orgs.). **Lusofonia Afro-Brasileira: questões sócio-históricas e linguísticas**. Campinas: Pontes Editores, 2025. p. 45–70.

VITRAL, Lorenzo. A forma CÊ e a noção de gramaticalização. **Revista de Estudos da Linguagem**, [S. l.], v. 4, n. 1, p. 115–124, 1996. Disponível em: <https://periodicos.ufmg.br/index.php/relin/article/view/28970>. Acesso em: 4 jan. 2026.

VITRAL, Lorenzo. **Gramática inteligente do português do Brasil**. São Paulo: Contexto, 2017.

VITRAL, Lorenzo. **A negação: teoria da checagem e mudança linguística**. Revista Delta. 15 (1). Disponível em: <https://www.scielo.br/j/delta/a/MxLJmcbB9f769zQ6cjKrWFM/?lang=pt>. Acesso em 30 dez. 2025.

WARDHAUGH, Ronald. **An introduction to sociolinguistics**, 5. ed. Malden: Blackwell Publishing, 2006 [1986].

WEINREICH, Uriel; LABOV, William; HERZOG, Marvin I. **Fundamentos empíricos para uma teoria da mudança linguística**. Tradução de Marcos Bagno. São Paulo: Parábola, 2006 [1975].

WICKLIN, Rick. The likelihood ratio test for linear regression in SAS. **The Do Loop**, 2024. Disponível em: <https://blogs.sas.com/content/iml/2024/03/27/likelihood-ratio-test.html>. Acesso em: 22 mai. 2025.

APÊNDICE

APÊNDICE 1 – Códigos atribuídos às variáveis analisadas neste trabalho

Variável	Marcador	Gênero	Escolaridade	Idade	Posição do marcador no período	Precedência de outros marcadores	Corpus
	1: entretanto	m: masculino	p: até o ensino médio	a: 18-25	i: início do período	s: sim	r: Brasil
	2: mas			b: 26-40	e: entre orações		
	3: no entanto						
	4: porém	f: feminino	u: ensino superior	c: 41-60	t: término do período	n: não	k: Moçambique
	5: só que			d: +60			
	6: mas só que						

Fonte: elaborado pela autora.

ANEXOS

ANEXO 1 – Arquivo *res*, resultados obtidos após rodagem do Goldvarb 3.03 em ambos os *corpora*

• CELL CREATION • 21/08/2024 15:26:49

Name of token file: Untitled.tkn

Name of condition file: Untitled.cnd

(

; Identity recode: All groups included as is.

(1)

(2)

(3)

(4)

(5)

(6)

)

Number of cells: 41

Application value(s): 52

Total no. of factors: 12

Group 5 2 Total %

1 (2) 5 2

c N 30 393 423 26.7

% 7.1 92.9

d N 3 69 72 4.5

% 4.2 95.8

b N 60 449 509 32.1

% 11.8 88.2

a N 41 542 583 36.7

% 7.0 93.0

Total N 134 1453 1587

% 8.4 91.6

2 (3) 5 2

u N 121 1256 1377 86.8

% 8.8 91.2

p N 13 197 210 13.2

% 6.2 93.8

Total N 134 1453 1587

% 8.4 91.6

 3 (4) 5 2
 f N 76 799 875 55.1
 % 8.7 91.3

m N 58 654 712 44.9
 % 8.1 91.9

Total N 134 1453 1587
 % 8.4 91.6

 4 (5) 5 2
 k N 62 651 713 44.9
 % 8.7 91.3

r N 72 802 874 55.1
 % 8.2 91.8

Total N 134 1453 1587
 % 8.4 91.6

 5 (6) 5 2
 e N 120 1243 1363 85.9
 % 8.8 91.2

i N 14 210 224 14.1
 % 6.2 93.8

Total N 134 1453 1587
 % 8.4 91.6

 TOTAL N 134 1453 1587
 % 8.4 91.6

Name of new cell file: Untitled.cel

• BINOMIAL VARBRUL, 1 step • 21/08/2024 15:36:35
 Name of cell file: Untitled quantas vezes cada codigo aparece

Averaging by weighting factors.
 One-level binomial analysis...

Run # 1, 41 cells:
 Convergence at Iteration 4
 Input 0.081

Group Factor Weight App/Total Input&Weight

1: c 0.452 0.07 0.07

d 0.350 0.04 0.05
 b 0.606 0.12 0.12
 a 0.460 0.07 0.07

2: u 0.508 0.09 0.08
 p 0.445 0.06 0.07

3: f 0.508 0.09 0.08
 m 0.490 0.08 0.08

4: k 0.522 0.09 0.09
 r 0.482 0.08 0.08

5: e 0.511 0.09 0.08
 i 0.433 0.06 0.06

Cell Total App'ns Expected Error

dumri 2 0 0.064 0.067
 dumre 4 0 0.174 0.182
 dumke 31 1 1.576 0.222
 dufri 2 0 0.069 0.072
 dufre 4 0 0.187 0.196
 dpmri 5 0 0.126 0.129
 dpmre 12 1 0.409 0.883
 dpfri 5 0 0.135 0.139
 dpfre 7 1 0.256 2.243
 cumri 27 0 1.310 1.377
 cumre 32 0 2.084 2.230
 cumke 92 4 6.973 1.372
 cufri 12 0 0.624 0.658
 cufre 32 0 2.230 2.397
 cufki 1 1 0.061 15.493
 cufke 167 17 13.533 0.967
 cpmri 6 0 0.228 0.237
 cpmre 8 0 0.410 0.432
 cpmki 1 1 0.045 21.471
 cpmke 21 7 1.256 27.932
 cpfri 1 0 0.041 0.043
 cpfri 23 0 1.263 1.337
 bumri 32 5 2.785 1.929
 bumre 111 12 12.791 0.055
 bumke 72 7 9.570 0.796
 bufri 38 4 3.533 0.068
 bufre 152 21 18.671 0.331
 bufke 62 10 8.773 0.200
 bpmri 5 0 0.344 0.370
 bpmre 21 0 1.927 2.122
 bpfri 2 0 0.147 0.159
 bpfre 14 1 1.372 0.112
 aumri 17 0 0.853 0.898

aumre 82 10 5.518 3.904
 aumke 95 8 7.435 0.047
 aufri 68 3 3.654 0.124
 aufre 148 14 10.654 1.133
 aufke 94 4 7.864 2.072
 apmre 2 0 0.106 0.112
 apmke 34 2 2.101 0.005
 apfke 43 0 2.844 3.046
 Total Chi-square = 97.5589
 Chi-square/cell = 2.3795
 Log likelihood = -452.023

• BINOMIAL VARBRUL • 21/08/2024 15:41:19

Name of cell file: Untitled quantas vezes cada codigo aparece

Averaging by weighting factors.

Threshold, step-up/down: 0.050001

Stepping up...

----- Level # 0 -----

Run # 1, 1 cells:

Convergence at Iteration 2

Input 0.084

Log likelihood = -459.392

----- Level # 1 -----

Run # 2, 4 cells:

Convergence at Iteration 5

Input 0.082

Group # 1 -- c: 0.462, d: 0.329, b: 0.601, a: 0.460

Log likelihood = -453.731 Significance = 0.010

Run # 3, 2 cells:

Convergence at Iteration 4

Input 0.084

Group # 2 -- u: 0.512, p: 0.419

Log likelihood = -458.534 Significance = 0.193

Run # 4, 2 cells:

Convergence at Iteration 3

Input 0.084

Group # 3 -- f: 0.508, m: 0.491

Log likelihood = -459.318 Significance = 0.700

Run # 5, 2 cells:

Convergence at Iteration 3

Input 0.084
 Group # 4 -- k: 0.508, r: 0.494
 Log likelihood = -459.339 Significance = 0.750

Run # 6, 2 cells:
 Convergence at Iteration 4
 Input 0.084
 Group # 5 -- e: 0.513, i: 0.421
 Log likelihood = -458.519 Significance = 0.190

Add Group # 1 with factors cdba

----- Level # 2 -----

Run # 7, 8 cells:
 Convergence at Iteration 4
 Input 0.081
 Group # 1 -- c: 0.463, d: 0.343, b: 0.598, a: 0.460
 Group # 2 -- u: 0.508, p: 0.445
 Log likelihood = -453.361 Significance = 0.408

Run # 8, 8 cells:
 Convergence at Iteration 5
 Input 0.081
 Group # 1 -- c: 0.462, d: 0.334, b: 0.601, a: 0.459
 Group # 3 -- f: 0.508, m: 0.490
 Log likelihood = -453.657 Significance = 0.700

Run # 9, 8 cells:
 Convergence at Iteration 5
 Input 0.081
 Group # 1 -- c: 0.451, d: 0.330, b: 0.611, a: 0.460
 Group # 4 -- k: 0.531, r: 0.475
 Log likelihood = -453.012 Significance = 0.236

Run # 10, 8 cells:
 Convergence at Iteration 5
 Input 0.081
 Group # 1 -- c: 0.460, d: 0.333, b: 0.602, a: 0.461
 Group # 5 -- e: 0.513, i: 0.419
 Log likelihood = -452.806 Significance = 0.180

No remaining groups significant

Groups selected while stepping up: 1
 Best stepping up run: #2

 Stepping down...

----- Level # 5 -----

Run # 11, 41 cells:

Convergence at Iteration 4

Input 0.081

Group # 1 -- c: 0.452, d: 0.350, b: 0.606, a: 0.460

Group # 2 -- u: 0.508, p: 0.445

Group # 3 -- f: 0.508, m: 0.490

Group # 4 -- k: 0.522, r: 0.482

Group # 5 -- e: 0.511, i: 0.433

Log likelihood = -452.023

----- Level # 4 -----

Run # 12, 14 cells:

Convergence at Iteration 5

Input 0.083

Group # 2 -- u: 0.513, p: 0.418

Group # 3 -- f: 0.506, m: 0.492

Group # 4 -- k: 0.498, r: 0.502

Group # 5 -- e: 0.514, i: 0.418

Log likelihood = -457.564 Significance = 0.012

Run # 13, 25 cells:

Convergence at Iteration 5

Input 0.081

Group # 1 -- c: 0.451, d: 0.338, b: 0.610, a: 0.459

Group # 3 -- f: 0.509, m: 0.488

Group # 4 -- k: 0.523, r: 0.481

Group # 5 -- e: 0.511, i: 0.435

Log likelihood = -452.386 Significance = 0.412

Run # 14, 23 cells:

Convergence at Iteration 4

Input 0.081

Group # 1 -- c: 0.452, d: 0.345, b: 0.606, a: 0.461

Group # 2 -- u: 0.509, p: 0.444

Group # 4 -- k: 0.522, r: 0.482

Group # 5 -- e: 0.511, i: 0.433

Log likelihood = -452.106 Significance = 0.688

Run # 15, 30 cells:

Convergence at Iteration 4

Input 0.081

Group # 1 -- c: 0.460, d: 0.351, b: 0.599, a: 0.460

Group # 2 -- u: 0.509, p: 0.444

Group # 3 -- f: 0.507, m: 0.491

Group # 5 -- e: 0.514, i: 0.417

Log likelihood = -452.340 Significance = 0.442

Run # 16, 25 cells:
 Convergence at Iteration 5
 Input 0.081
 Group # 1 -- c: 0.450, d: 0.347, b: 0.608, a: 0.460
 Group # 2 -- u: 0.508, p: 0.447
 Group # 3 -- f: 0.508, m: 0.490
 Group # 4 -- k: 0.532, r: 0.474
 Log likelihood = -452.564 Significance = 0.298

Cut Group # 3 with factors fm

----- Level # 3 -----

Run # 17, 8 cells:
 Convergence at Iteration 5
 Input 0.083
 Group # 2 -- u: 0.513, p: 0.417
 Group # 4 -- k: 0.497, r: 0.502
 Group # 5 -- e: 0.514, i: 0.418
 Log likelihood = -457.612 Significance = 0.012

Run # 18, 13 cells:
 Convergence at Iteration 5
 Input 0.081
 Group # 1 -- c: 0.452, d: 0.333, b: 0.609, a: 0.460
 Group # 4 -- k: 0.522, r: 0.482
 Group # 5 -- e: 0.511, i: 0.435
 Log likelihood = -452.496 Significance = 0.395

Run # 19, 15 cells:
 Convergence at Iteration 4
 Input 0.081
 Group # 1 -- c: 0.461, d: 0.347, b: 0.599, a: 0.461
 Group # 2 -- u: 0.509, p: 0.443
 Group # 5 -- e: 0.514, i: 0.418
 Log likelihood = -452.401 Significance = 0.457

Run # 20, 14 cells:
 Convergence at Iteration 5
 Input 0.081
 Group # 1 -- c: 0.451, d: 0.342, b: 0.608, a: 0.461
 Group # 2 -- u: 0.508, p: 0.445
 Group # 4 -- k: 0.531, r: 0.475
 Log likelihood = -452.652 Significance = 0.297

Cut Group # 4 with factors kr

----- Level # 2 -----

Run # 21, 4 cells:

Convergence at Iteration 4
 Input 0.083
 Group # 2 -- u: 0.513, p: 0.417
 Group # 5 -- e: 0.513, i: 0.420
 Log likelihood = -457.619 Significance = 0.016

Run # 22, 8 cells:
 Convergence at Iteration 5
 Input 0.081
 Group # 1 -- c: 0.460, d: 0.333, b: 0.602, a: 0.461
 Group # 5 -- e: 0.513, i: 0.419
 Log likelihood = -452.806 Significance = 0.385

Run # 23, 8 cells:
 Convergence at Iteration 4
 Input 0.081
 Group # 1 -- c: 0.463, d: 0.343, b: 0.598, a: 0.460
 Group # 2 -- u: 0.508, p: 0.445
 Log likelihood = -453.361 Significance = 0.174

Cut Group # 2 with factors up

----- Level # 1 -----

Run # 24, 2 cells:
 Convergence at Iteration 4
 Input 0.084
 Group # 5 -- e: 0.513, i: 0.421
 Log likelihood = -458.519 Significance = 0.010

Run # 25, 4 cells:
 Convergence at Iteration 5
 Input 0.082
 Group # 1 -- c: 0.462, d: 0.329, b: 0.601, a: 0.460
 Log likelihood = -453.731 Significance = 0.180

Cut Group # 5 with factors ei

----- Level # 0 -----

Run # 26, 1 cells:
 Convergence at Iteration 2
 Input 0.084
 Log likelihood = -459.392 Significance = 0.010

All remaining groups significant

Groups eliminated while stepping down: 3 4 2 5
 Best stepping up run: #2
 Best stepping down run: #25

ANEXO 2 – Arquivo *res*, resultados obtidos após rodagem do Goldvarb 3.0 no *corpus* de PB

• CELL CREATION • 19/09/2024 15:21:47

Name of token file: Untitled.tkn

Name of condition file: Untitled.cnd

(

; Identity recode: All groups included as is.

(1)

(2)

(3)

(4)

(5)

)

Number of cells: 21

Application value(s): 25

Total no. of factors: 9

Group 2 5 Total %

1 (2) 2 5
a N 290 27 317 43.2
% 91.5 8.5

b N 332 43 375 51.2
% 88.5 11.5

d N 39 2 41 5.6
% 95.1 4.9

Total N 661 72 733
% 90.2 9.8

2 (3) 2 5
u N 591 69 660 90.0
% 89.5 10.5

p N 70 3 73 10.0
% 95.9 4.1

Total N 661 72 733
% 90.2 9.8

3 (4) 2 5
m N 265 28 293 40.0
% 90.4 9.6

f N 396 44 440 60.0
% 90.0 10.0

Total N 661 72 733
% 90.2 9.8

4 (5) 2 5
e N 497 60 557 76.0
% 89.2 10.8

i N 164 12 176 24.0
% 93.2 6.8

Total N 661 72 733
% 90.2 9.8

TOTAL N 661 72 733
% 90.2 9.8

Name of new cell file: .cel

• BINOMIAL VARBRUL, 1 step • 19/09/2024 15:22:48

Name of cell file: .cel

Averaging by weighting factors.
One-level binomial analysis...

Run # 1, 21 cells:
Convergence at Iteration 7
Input 0.907

Group Factor Weight App/Total Input&Weight

1: a 0.550 0.91 0.92
b 0.454 0.89 0.89
d 0.534 0.95 0.92

2: u 0.474 0.90 0.90
p 0.720 0.96 0.96

3: m 0.507 0.90 0.91
f 0.496 0.90 0.91

4: e 0.472 0.89 0.90
i 0.589 0.93 0.93

Cell Total App'ns Expected Error
dumi 2 2 1.874 0.134

dume 4 4 3.610 0.432
 dufi 2 2 1.869 0.141
 dufe 4 4 3.594 0.452
 dpmi 5 5 4.885 0.118
 dpme 12 11 11.562 0.747
 dpfi 5 5 4.880 0.123
 dpfe 7 6 6.733 2.092
 bumi 32 27 29.282 2.095
 bume 111 99 96.590 0.463
 bufi 38 34 34.638 0.133
 bufe 152 131 131.486 0.013
 bpmi 5 5 4.842 0.163
 bpme 21 21 19.956 1.099
 bpfi 2 2 1.934 0.068
 bpfe 14 13 13.274 0.109
 aumi 17 17 15.992 1.072
 aume 82 72 74.455 0.880
 aufi 68 65 63.795 0.368
 aufe 148 134 133.819 0.003
 apme 2 2 1.931 0.071
 Total Chi-square = 10.7761
 Chi-square/cell = 0.5131
 Log likelihood = -231.152

• BINOMIAL VARBRUL • 19/09/2024 15:23:54

Name of cell file: .cel

Averaging by weighting factors.
 Threshold, step-up/down: 0.050001

Stepping up...

----- Level # 0 -----

Run # 1, 1 cells:
 Convergence at Iteration 2
 Input 0.902
 Log likelihood = -235.417

----- Level # 1 -----

Run # 2, 3 cells:
 Convergence at Iteration 5
 Input 0.904
 Group # 1 -- a: 0.534, b: 0.452, d: 0.675
 Log likelihood = -233.871 Significance = 0.216

Run # 3, 2 cells:
 Convergence at Iteration 5

Input 0.904
 Group # 2 -- u: 0.475, p: 0.711
 Log likelihood = -233.585 Significance = 0.058

Run # 4, 2 cells:
 Convergence at Iteration 3
 Input 0.902
 Group # 3 -- m: 0.507, f: 0.495
 Log likelihood = -235.397 Significance = 0.852

Run # 5, 2 cells:
 Convergence at Iteration 4
 Input 0.903
 Group # 4 -- e: 0.470, i: 0.594
 Log likelihood = -234.147 Significance = 0.116

No remaining groups significant

Groups selected while stepping up: None
 Best stepping up run: #1

Stepping down...

----- Level # 4 -----

Run # 6, 21 cells:
 Convergence at Iteration 7
 Input 0.907
 Group # 1 -- a: 0.550, b: 0.454, d: 0.534
 Group # 2 -- u: 0.474, p: 0.720
 Group # 3 -- m: 0.507, f: 0.496
 Group # 4 -- e: 0.472, i: 0.589
 Log likelihood = -231.152

----- Level # 3 -----

Run # 7, 8 cells:
 Convergence at Iteration 5
 Input 0.906
 Group # 2 -- u: 0.475, p: 0.712
 Group # 3 -- m: 0.501, f: 0.499
 Group # 4 -- e: 0.470, i: 0.595
 Log likelihood = -232.295 Significance = 0.324

Run # 8, 12 cells:
 Convergence at Iteration 4
 Input 0.905
 Group # 1 -- a: 0.533, b: 0.454, d: 0.663
 Group # 3 -- m: 0.513, f: 0.491

Group # 4 -- e: 0.472, i: 0.587
 Log likelihood = -232.760 Significance = 0.078

Run # 9, 11 cells:
 Convergence at Iteration 7
 Input 0.907
 Group # 1 -- a: 0.550, b: 0.454, d: 0.535
 Group # 2 -- u: 0.474, p: 0.721
 Group # 4 -- e: 0.472, i: 0.589
 Log likelihood = -231.168 Significance = 0.867

Run # 10, 11 cells:
 Convergence at Iteration 7
 Input 0.906
 Group # 1 -- a: 0.553, b: 0.450, d: 0.546
 Group # 2 -- u: 0.474, p: 0.717
 Group # 3 -- m: 0.504, f: 0.497
 Log likelihood = -232.277 Significance = 0.143

Cut Group # 3 with factors mf

----- Level # 2 -----

Run # 11, 4 cells:
 Convergence at Iteration 5
 Input 0.906
 Group # 2 -- u: 0.475, p: 0.712
 Group # 4 -- e: 0.470, i: 0.595
 Log likelihood = -232.296 Significance = 0.330

Run # 12, 6 cells:
 Convergence at Iteration 4
 Input 0.905
 Group # 1 -- a: 0.531, b: 0.455, d: 0.666
 Group # 4 -- e: 0.473, i: 0.586
 Log likelihood = -232.820 Significance = 0.074

Run # 13, 6 cells:
 Convergence at Iteration 7
 Input 0.906
 Group # 1 -- a: 0.552, b: 0.451, d: 0.547
 Group # 2 -- u: 0.474, p: 0.718
 Log likelihood = -232.283 Significance = 0.145

Cut Group # 1 with factors abd

----- Level # 1 -----

Run # 14, 2 cells:
 Convergence at Iteration 4

Input 0.903
Group # 4 -- e: 0.470, i: 0.594
Log likelihood = -234.147 Significance = 0.056

Run # 15, 2 cells:
Convergence at Iteration 5
Input 0.904
Group # 2 -- u: 0.475, p: 0.711
Log likelihood = -233.585 Significance = 0.112

Cut Group # 4 with factors ei

----- Level # 0 -----

Run # 16, 1 cells:
Convergence at Iteration 2
Input 0.902
Log likelihood = -235.417 Significance = 0.058

Cut Group # 2 with factors up
Groups eliminated while stepping down: 3 1 4 2

Best stepping up run: #1
Best stepping down run: #16

ANEXO 3 – Arquivo *res*, resultados obtidos após rodagem do Goldvarb 3.0 no *corpus* de PM

1• CELL CREATION • 19/09/2024 17:30:37

Name of token file: Untitled.tkn

Name of condition file: Untitled.cnd

(

; Identity recode: All groups included as is.

(1)

(2)

(3)

(4)

)

Number of cells: 10

Application value(s): 52

Total no. of factors: 8

Group 5 2 Total %

1 (2) 5 2

c N 30 252 282 39.6

% 10.6 89.4

d N 1 30 31 4.3

% 3.2 96.8

b N 17 117 134 18.8

% 12.7 87.3

a N 14 252 266 37.3

% 5.3 94.7

Total N 62 651 713

% 8.7 91.3

2 (3) 5 2

u N 52 562 614 86.1

% 8.5 91.5

p N 10 89 99 13.9

% 10.1 89.9

Total N 62 651 713

% 8.7 91.3

3 (4) 5 2

f N 32 335 367 51.5

% 8.7 91.3

m N 30 316 346 48.5
 % 8.7 91.3

Total N 62 651 713
 % 8.7 91.3

TOTAL N 62 651 713
 % 8.7 91.3

Name of new cell file: .cel

• BINOMIAL VARBRUL, 1 step • 19/09/2024 17:31:28
 Name of cell file: .cel

Averaging by weighting factors.
 One-level binomial analysis...

Run # 1, 10 cells:
 Convergence at Iteration 7
 Input 0.079

Group Factor Weight App/Total Input&Weight

1: c 0.587 0.11 0.11
 d 0.297 0.03 0.04
 b 0.649 0.13 0.14
 a 0.359 0.05 0.05

2: u 0.477 0.08 0.07
 p 0.639 0.10 0.13

3: f 0.499 0.09 0.08
 m 0.501 0.09 0.08

Cell	Total	App'ns	Expected	Error
dum	31	1	0.998	0.000
cum	92	4	9.257	3.320
cuf	168	18	16.813	0.093
cpm	22	8	3.924	5.153
bum	72	7	9.145	0.576
buf	62	10	7.833	0.686
aum	95	8	4.018	4.120
auf	94	4	3.953	0.001
apm	34	2	2.684	0.189
apf	43	0	3.375	3.663

Total Chi-square = 17.8000

Chi-square/cell = 1.7800

Log likelihood = -204.542

• BINOMIAL VARBRUL • 19/09/2024 17:32:49

Name of cell file: .cel

Averaging by weighting factors.

Threshold, step-up/down: 0.050001

Stepping up...

----- Level # 0 -----

Run # 1, 1 cells:

Convergence at Iteration 2

Input 0.087

Log likelihood = -210.648

----- Level # 1 -----

Run # 2, 4 cells:

Convergence at Iteration 5

Input 0.081

Group # 1 -- c: 0.575, d: 0.275, b: 0.623, a: 0.387

Log likelihood = -205.802 Significance = 0.022

Run # 3, 2 cells:

Convergence at Iteration 4

Input 0.087

Group # 2 -- u: 0.493, p: 0.541

Log likelihood = -210.511 Significance = 0.617

Run # 4, 2 cells:

Convergence at Iteration 2

Input 0.087

Group # 3 -- f: 0.500, m: 0.500

Log likelihood = -210.648 Significance = 0.985

Add Group # 1 with factors cdba

----- Level # 2 -----

Run # 5, 6 cells:

Convergence at Iteration 7

Input 0.079

Group # 1 -- c: 0.587, d: 0.298, b: 0.649, a: 0.359

Group # 2 -- u: 0.477, p: 0.639

Log likelihood = -204.542 Significance = 0.118

Run # 6, 7 cells:

Convergence at Iteration 5

Input 0.081
 Group # 1 -- c: 0.576, d: 0.268, b: 0.622, a: 0.387
 Group # 3 -- f: 0.492, m: 0.509
 Log likelihood = -205.769 Significance = 0.797

No remaining groups significant

Groups selected while stepping up: 1
 Best stepping up run: #2

Stepping down...

----- Level # 3 -----

Run # 7, 10 cells:
 Convergence at Iteration 7
 Input 0.079
 Group # 1 -- c: 0.587, d: 0.297, b: 0.649, a: 0.359
 Group # 2 -- u: 0.477, p: 0.639
 Group # 3 -- f: 0.499, m: 0.501
 Log likelihood = -204.542

----- Level # 2 -----

Run # 8, 4 cells:
 Convergence at Iteration 4
 Input 0.087
 Group # 2 -- u: 0.493, p: 0.542
 Group # 3 -- f: 0.502, m: 0.498
 Log likelihood = -210.509 Significance = 0.009

Run # 9, 7 cells:
 Convergence at Iteration 5
 Input 0.081
 Group # 1 -- c: 0.576, d: 0.268, b: 0.622, a: 0.387
 Group # 3 -- f: 0.492, m: 0.509
 Log likelihood = -205.769 Significance = 0.124

Run # 10, 6 cells:
 Convergence at Iteration 7
 Input 0.079
 Group # 1 -- c: 0.587, d: 0.298, b: 0.649, a: 0.359
 Group # 2 -- u: 0.477, p: 0.639
 *** Warning, negative change in likelihood (-0.00002302) replaced by 0.0.
 Log likelihood = -204.542 Significance = 1.000

Cut Group # 3 with factors fm

----- Level # 1 -----

Run # 11, 2 cells:

Convergence at Iteration 4

Input 0.087

Group # 2 -- u: 0.493, p: 0.541

Log likelihood = -210.511 Significance = 0.009

Run # 12, 4 cells:

Convergence at Iteration 5

Input 0.081

Group # 1 -- c: 0.575, d: 0.275, b: 0.623, a: 0.387

Log likelihood = -205.802 Significance = 0.118

Cut Group # 2 with factors up

----- Level # 0 -----

Run # 13, 1 cells:

Convergence at Iteration 2

Input 0.087

Log likelihood = -210.648 Significance = 0.022

All remaining groups significant

Groups eliminated while stepping down: 3 2

Best stepping up run: #2

Best stepping down run: #12