

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

ALESSANDRA EMANUELLE MACIEIRA SILVA

Orações completivas e adverbiais no português falado: uma proposta de análise da interface sintaxe-estrutura informacional em sete variedades

BELO HORIZONTE
Faculdade de Letras da UFMG

2022

ALESSANDRA EMANUELLE MACIEIRA SILVA

Orações completivas e adverbiais no português falado: uma proposta de análise da interface sintaxe-estrutura informacional em sete variedades

Dissertação de Mestrado apresentada como requisito
parcial para conclusão do Mestrado em
Letras/Linguística Teórica e Descritiva na FALE/UFMG
– Linha 1 – Estudos linguísticos baseados em corpora

Orientadora: Prof^ª Dr^ª Giulia Bossaglia

BELO HORIZONTE
Faculdade de Letras da UFMG

2022

S586o Silva, Alessandra Emanuelle Macieira.
Orações completivas e adverbiais no português [manuscrito]: uma proposta de análise da interface sintaxe-estrutura informacional em sete variedades/ Alessandra Emanuelle Macieira. – 2022.
91 f.: il., grafs., tabs., color.

Orientadora: Giulia Bossaglia

Área de concentração: Linguística Teórica e Descritiva.

Linha de pesquisa: Estudos Linguísticos Baseados em Corpora.

Dissertação (mestrado) – Universidade Federal de Minas Gerais, Faculdade de Letras.

Bibliografia: f. 82-91.

1. Língua portuguesa – Português falado – Teses. 2. Atos de fala (Linguística) – Teses. 3. Linguística de Corpus – Teses. 4. Língua portuguesa I. Bossaglia, Giulia. II. Universidade Federal de Minas Gerais. Faculdade de Letras. III. Título.



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

FOLHA DE APROVAÇÃO

Orações completivas e adverbiais no português falado: uma proposta de análise da interface sintaxe-estrutura informacional em sete variedades

ALESSANDRA EMANUELLE MACIEIRA SILVA

Dissertação submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em ESTUDOS LINGUÍSTICOS, como requisito para obtenção do grau de Mestre em ESTUDOS LINGUÍSTICOS, área de concentração LINGUÍSTICA TEÓRICA E DESCRITIVA, linha de pesquisa Estudos Linguísticos Baseados em Corpora.

Aprovada em 23 de setembro de 2022, pela banca constituída pelos membros:

Prof(a). Giulia Bossaglia - Orientadora

UFMG

Prof(a). Janayna Maria da Rocha Carvalho

UFMG

Prof(a). Lucia de Almeida Ferrari

UFMG

Prof(a). Luis Filipe Lima e Silva

Belo Horizonte, 23 de setembro de 2022.



Documento assinado eletronicamente por **Giulia Bossaglia, Professora do Magistério Superior**, em 23/09/2022, às 17:19, 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 **Janayna Maria da Rocha Carvalho, Professora do Magistério Superior**, em 23/09/2022, às 17:32, conforme horário oficial de Brasília, com



Documento assinado eletronicamente por **Lucia de Almeida Ferrari, Professora do Magistério Superior**, em 26/09/2022, às 09:22, 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 **Luis Filipe Lima e Silva, Usuário Externo**, em 26/09/2022, às 17:07, 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 **1710930** e o código CRC **9A4D01CE**.

AGRADECIMENTOS

Agradeço à professora Giulia Bossaglia pela orientação valiosa, pela oportunidade, pela paciência e dedicação infinitas, pelo apoio e incentivo imensuráveis e pelo conhecimento compartilhado. Agradeço pelo exemplo e compromisso, por ter me guiado e ensinado tanto e por tudo o que eu aprendo a cada dia com ela. Muito, muito obrigada.

Agradeço aos meus queridos pais Ana Maria e Marcioni, minha irmã Giovanna e toda a minha família pelo apoio incondicional, por me incentivarem sempre, por serem um espelho para mim, por me inspirarem e por me mostrarem o valor do estudo. Obrigada por estimularem desde a infância o meu aprendizado e a importância do conhecimento. Agradeço infinitamente.

Agradeço à UFMG pelo ensino público, gratuito e de qualidade e a todos os profissionais e colegas que contribuíram para a minha formação. Agradeço à CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) pelo financiamento no início do mestrado e à FUMP pela assistência estudantil. Agradeço a Deus pelo amparo e forças.

Obrigada a amigas pela companhia e a todos que direta ou indiretamente contribuíram para a minha formação. Realmente, muito obrigada.

RESUMO

Neste trabalho propõe-se um olhar sobre subordinação completiva e adverbial na fala espontânea de sete variedades de português que leva em conta a interface da sintaxe com a estrutura informacional. Os dados analisados foram extraídos do corpus *Português Falado – Variedades geográficas e sociais* (GONÇALVES; VELOSO, 2000), que reúne gravações de 10 variedades diatópicas de português (Europa, Brasil, África, Ásia), em sua maioria entrevistas pesquisador-informante de conteúdo livre. As transcrições são alinhadas aos arquivos de áudio e dotadas de anotação de quebras prosódicas, seguindo o modelo do projeto C-ORAL-ROM (CRESTI; MONEGLIA, 2005). Ocorrências de subordinadas completivas e adverbiais foram levantadas para o português falado em Angola, Brasil, Cabo Verde, Guiné-Bissau, Moçambique, Portugal, São Tomé e Príncipe (5 arquivos para cada país; não foram abordadas as variedades asiáticas). Nos dados levantados, após revisão da segmentação prosódica, realizou-se a etiquetagem informacional seguindo o quadro teórico da *Language Into Act Theory* (CRESTI, 2000; MONEGLIA; RASO, 2014). As orações completivas e adverbiais foram classificadas e analisadas de acordo com sua realização dentro de uma mesma ou ao longo de diferentes unidades informacionais (UIs) com relação a suas principais, com foco na análise dos padrões informacionais. Os resultados mostram que as orações estudadas apresentam funções uniformes na fala das demais variedades: as completivas são realizadas, de preferência, em uma única UI junto com suas principais, que tendem a sofrer processos de gramaticalização como marcas de modalidade, ao passo que as adverbiais ocorrem mais frequentemente em UIs (ou até enunciados) diferentes daquelas de suas principais, servindo a funções informacionais diversas (fornecimento de domínios de aplicação da ilocução; realização de ilocuições específicas; parênteses, entre outras). Os resultados confirmam, em novos dados, os achados de estudos similares previamente conduzidos sobre português brasileiro, italiano e inglês americano realizados com base em outros corpora da família C-ORAL (SILVA; BOSSAGLIA, 2019, no prelo; BOSSAGLIA; MELLO; RASO, 2020).

Palavras-chave: Fala espontânea. Orações completivas. Orações adverbiais. Variedades diatópicas do português. Linguística de *corpus*.

ABSTRACT

This work analyzes complement and adverbial clauses in spontaneous speech of seven Portuguese dialects, taking into account the syntax-information structure interface. The data come from the *Spoken Portuguese - Geographic and social varieties* (GONÇALVES; VELOSO, 2000), which gathers recordings of 10 diatopic varieties of Portuguese (between Europe, Brazil, Africa, and Asia), mainly researcher-informant script-free interviews. Their transcripts are aligned to the sound files and provided of prosodic breaks' annotation, following the C-ORAL-ROM project (CRESTI; MONEGLIA, 2005). Occurrences of complement and adverbial clauses were retrieved for the dialects of Angola, Brazil, Cape Verde, Guinea-Bissau, Mozambique, Portugal, and São Tomé and Príncipe (5 recordings for each dialect; Asian dialects were not considered). After revising the prosodic segmentation of the retrieved data, a manual informational tagging was accomplished, following the *Language into Act Theory* framework (CRESTI, 2000; MONEGLIA; RASO, 2014). Complement and adverbial clauses were classified and analyzed according to their fulfillment within the same or throughout different information units (IUs) with respect to their main clauses. The results show that such subordinate clauses display similar functions throughout the dialects: complements are fulfilled, mostly, within the same IU with their main clause, which in turn present a trend of grammaticalization as modality markers, while adverbial clauses occur mostly in different IUs (or even utterances) from their main clauses, accomplishing diverse functions (providing a domain for the application of the illocution; fulfillment of specific illocutions; parenthetical information, among others). These results also confirm, based on new data, what previous studies on spoken Brazilian Portuguese, Italian, and American English found in other spoken corpora, namely those of the C-ORAL family (SILVA; BOSSAGLIA, 2019, no prelo; BOSSAGLIA; MELLO; RASO, 2020).

Key-words: Spontaneous speech. Complement clauses. Adverbial clauses. Diatopic varieties of Portuguese. Corpus linguistics.

LISTA DE FIGURAS

Figura 1: Alinhamento no software EXMARaLDA.	40
Figura 2: Visualização do enunciado no Praat.	48
Figura 3: Processo metodológico final no Praat.....	49

LISTA DE TABELAS

Tabela 1: Assuntos das gravações do corpus Português Falado (GONÇALVES; VELOSO, 2000).....	38
Tabela 2: Quantidade de palavras dos textos do corpus Português Falado.	41
Tabela 3: Assuntos presentes nas gravações do corpus Português Falado (GONÇALVES; VELOSO, 2000).	41
Tabela 4: Comparação entre os símbolos usados nos dois corpora.....	46

LISTA DE GRÁFICOS

Gráfico 1: Tipologia de completivas no corpus.	50
Gráfico 2: Tipos de complementizadores.....	52
Gráfico 3: Linearização e padronização nas completivas do corpus Português Falado.	53
Gráfico 4: Os padrões informacionais.....	57
Gráfico 5: Tipos de orações adverbiais.	64
Gráfico 6: Linearização vs. padronização.	65
Gráfico 7: Padrões informacionais nas adverbiais.	66
Gráfico 8: Quantidade de orações temporais nas variedades do português.	69
Gráfico 9: Quantidade de orações condicionais nas variedades do português.	71
Gráfico 10: Quantidade de orações causais nas variedades do português.....	73

LISTA DE ABREVIACOES E SMBOLOS

[TAG]_r	Discurso Reportado
ALL	Alocutivo
APC	Apêndice de Comentário
APT	Apêndice de Tópico
fam	Familiar
pub	Público
cv	Conversação
dl	Diálogo
mn	Monólogo
CMM	Comentário Múltiplo
CNT	Conativo
COB	Comentário Ligado
COM	Comentário
DCT	Conector Discursivo
EMP	Unidade vazia
EXP	Expressivo
i-[TAG]	Unidade interrompida
IA	Inglês americano
INP	Incipitário
INT	Introdutor Locutivo
IT	Italiano
<i>L-AcT</i>	<i>Language into Act-Theory</i>
PAR	Parentético
PB	Português brasileiro
PHA	Fático
PRL	Lista de Parentéticos
SCA	Unidade de Escansão
TMT	Tomada de Tempo

TOP	Tópico
TPL	Lista de tópicos
UI	Unidade Informacional
(+)	enunciado interrompido
(< >)	sobreposição de fala
([n°])	retracting
(&)	início de palavra interrompida
(&he)	hesitação ou pausa preenchida
(hhh)	ruído paralinguístico
(xxx)	palavra ininteligível
(yyyy)	trecho de enunciado ininteligível
(/)	quebra prosódica não terminal
(//)	quebra prosódica terminal

SUMÁRIO

1.	INTRODUÇÃO	15
2.	REVISÃO TEÓRICA	18
2.1.	CONTEXTUALIZAÇÃO DAS VARIEDADES DIATÓPICAS DO PORTUGUÊS	18
2.1.1.	O PORTUGUÊS NA ÁFRICA.....	18
2.1.2	A SITUAÇÃO LINGUÍSTICA NAS VARIEDADES AFRICANAS DO PORTUGUÊS	18
2.1.2.1.	ANGOLA.....	19
2.1.2.2.	CABO VERDE.....	19
2.1.2.3.	GUINÉ-BISSAU.....	19
2.1.2.5.	SÃO TOMÉ E PRÍNCIPE.....	21
2.2.	CONSIDERAÇÕES SOBRE A SINTAXE NA DIAMESIA FALADA	21
2.3.	A <i>LANGUAGE INTO ACT-THEORY</i>	25
2.3.1.	AS UNIDADES INFORMACIONAIS.....	28
2.6.1.1	O COMENTÁRIO (COM)	29
2.6.1.2	O TÓPICO (TOP).....	30
2.6.1.3	O APÊNDICE DE TÓPICO (APT) E APÊNDICE DE COMENTÁRIO (APC).....	31
2.6.1.4	O PARENTÉTICO (PAR)	31
2.6.1.5	O INTRODUTOR LOCUTIVO (INT).....	32
2.6.2	OS LIMITES DO ISOMORFISMO	33
2.6.2.1	A ESCANSÃO (SCA)	33
2.6.2.2	OS COMENTÁRIOS MÚLTIPLOS (CMMS)	34
2.6.2.3	OS COMENTÁRIOS LIGADOS (COBs)	35
2.6.3	A SINTAXE NA FALA: UMA ABORDAGEM DA L-ACT	36
3.	METODOLOGIA	38
3.1.	O CORPUS PORTUGUÊS FALADO	38
3.2.	TRATAMENTOS DOS DADOS.....	45
4.	ANÁLISE E DISCUSSÃO DOS DADOS	50
4.1	AS ORAÇÕES COMPLETIVAS	50
4.1.1	TIPOLOGIA E COMPLEMENTIZADORES	50
4.1.2	LINEARIZAÇÃO VS. PADRONIZAÇÃO.....	52
4.1.3	OS PADRÕES INFORMACIONAIS	56
4.1.3.1	TÓPICO-COMENTÁRIO (TOP/COM).....	56
4.1.3.2	COMBINAÇÕES DE COMENTÁRIOS.....	60
4.1.3.3	COMPLETIVAS ENTRE ENUNCIADOS	62
4.2	AS ORAÇÕES ADVERBIAIS	64

4.2.1 TIPOLOGIA DE SUBORDINAÇÃO.....	64
4.2.2.1 AS ORAÇÕES TEMPORAIS	68
4.2.2.2 AS ORAÇÕES CONDICIONAIS.....	71
5. CONSIDERAÇÕES FINAIS	82
6. REFERÊNCIAS BIBLIOGRÁFICAS.....	84

1. INTRODUÇÃO

O trabalho aqui apresentado propõe o estudo da subordinação na fala espontânea de variedades diatópicas do português, com base no corpus Português Falado – Variedades Geográficas e Sociais (GONÇALVES; VELOSO, 2000, disponível em: <http://www.clul.ulisboa.pt/recurso/portugues-falado-variedades-geograficas-e-sociais> e, na versão mais atualizada, em <http://catalog.elra.info/en-us/repository/browse/ELRA-S0345/>). Os corpora orais, desde que respeitem critérios metodológicos bem definidos e apresentem especificações adequadas, constituem ferramentas fundamentais para o estudo da fala que, durante muito tempo, foi estudada apenas a partir de textos transcritos, ou seja, da escrita. Desse modo, os estudos sobre a diamesia falada são mais recentes e consistem em uma área a ser ainda muito explorada, e é importante aproveitarmos os recursos finalmente disponíveis e que são projetados para essa finalidade de estudo. Logo, a pesquisa desenvolvida neste trabalho se faz necessária para contribuir para os estudos sobre português falado, sobretudo tendo em vista a amplitude da comparação entre seus dialetos, a partir de uma ferramenta adequada e disponível.

O objetivo central deste estudo é analisar as estruturas de subordinação completiva e adverbial na fala de sete variedades diatópicas do português, com base em corpus, de modo a aprofundar e ampliar o espectro dos estudos sobre sintaxe da fala espontânea do português - o que até o presente momento não foi feito entre as pesquisas sobre sintaxe com base neste corpus.

Em nível mais específico, tem-se como objetivo, por meio do mapeamento dessas orações subordinadas, descrever sua realização com base nos tipos de subordinadores e de subordinadas. Dentro desse contexto, pretende-se, também, estudar os padrões informacionais em que as orações em estudo são realizadas, observando a proporção de estruturas que são realizadas sem padronização e com padronização, seguindo o quadro proposto pela *Language into Act-Theory* (L-Act, CRESTI, 2000; MONEGLIA; RASO, 2014). A análise informacional a partir dessa teoria possibilita investigar como os itens lexicais são distribuídos ao longo dos padrões informacionais, observando a relação entre a estrutura informacional e as tipologias de subordinação. Por fim, objetiva-se, ainda, realizar a comparação com estudos sobre o português brasileiro, italiano e inglês americano previamente conduzidos com base nos

corpora da família C-ORAL, a fim de verificar se os resultados se alinham àqueles do presente trabalho.

A presente pesquisa não apenas pretende analisar como ocorre a distribuição dos índices lexicais de subordinação na fala e se existem ocorrências que desafiam construções sintáticas canônicas, como também explora quais funções pragmático-informacionais tais construções sintáticas desempenham na fala em variedades diatópicas distintas do português.

Ao longo dos estudos linguísticos, algumas variedades foram mais estudadas do que outras, visto que poucos são os trabalhos sobre o português na Guiné-Bissau e em Cabo Verde enquanto existem muitas pesquisas sobre as variedades em Angola, Portugal e Brasil, por exemplo. Portanto, ainda são escassas as análises sobre sintaxe nas variedades do português com base em corpus com dados autênticos de fala e com abrangência comparativa.

Nesse sentido, o presente estudo, apesar de exploratório e de caráter qualitativo, se faz relevante uma vez que analisa dados de um corpus de fala pouco explorado, mas que se presta a ser uma ferramenta adequada para o estudo do português falado, pois apresenta as especificações necessárias (SILVA, 2019; SILVA; BOSSAGLIA, 2019).

A análise da subordinação completiva e adverbial na fala foi foco de alguns estudos de natureza comparativa ou não, tanto recentes (THOMPSON, 2002; HOPPER; THOMPSON, 2008; BOSSAGLIA, 2014, 2015; SILVA; BOSSAGLIA, 2019; SILVA, 2019) como também anteriores (CHAFE, 1984; MILLER; WEINERT, 1998; COUPER-KUHLEN; 1996). Porém, os estudos sobre a fala ficaram restritos ao inglês americano (SILVA, 2020) e francês (DEBAISIEUX, 2013), italiano (SCARANO, 2002), português brasileiro (SILVA, 2016; 2020) e hebraico (INBAR, 2016) e comparações entre variedades africanas, europeias e brasileiras do português (NASCIMENTO et al 2008; CHIMUKU, 2019; GONÇALVES, 2015; RABÊLO, 2016). Desse modo, algumas variedades foram mais estudadas que outras e, além disso, os estudos não compararam estruturas de subordinação em muitas variedades ao mesmo tempo com base em corpus de fala. Portanto, o presente trabalho se faz importante por propor uma comparação abrangente, isto é, entre muitas variedades diatópicas do português falado; aprofundando o olhar para as construções sintáticas, o que ainda foi pouco explorado nas pesquisas com base no corpus. Além disso, há a possibilidade de ampliar o espectro de comparação para as variedades asiáticas e as ilhas, como Açores e Madeira, podendo, também, servir de inspiração para projetos futuros, sobretudo de cunho sintático, com base nas seções pouco exploradas do corpus.

Em relação à análise, as orações foram organizadas quanto à tipologia (completiva e adverbial), tipo de realização informacional e, posteriormente, os dados foram analisados com

foco nos padrões informacionais. Além disso, o corpus das variedades diatópicas possui arquitetura diferente dos corpora do projeto C-ORAL-ROM (CRESTI; MONEGLIA, 2005) e C-ORAL-BRASIL (RASO; MELLO, 2012). Durante o processo metodológico, foi feita a revisão da segmentação original, etapa essencial para qualquer estudo sobre fala, e a etiquetagem informacional nos dados de um corpus ainda pouco utilizado.

Por fim, o trabalho é estruturado em seis seções. A seção de revisão teórica será dividida em três subseções: a primeira apresenta um breve contexto histórico das variedades diatópicas do português examinadas; na segunda, são mostrados os principais estudos linguísticos sobre a diamesia falada no português; na terceira, são abordados os conceitos fundamentais da teoria que serve de base para esse trabalho, a *Language into Act-Theory* (L-AcT, CRESTI, 2000; 2014).

A terceira seção descreve a metodologia utilizada e os softwares necessários para a busca pelas orações subordinadas. A discussão e a análise dos dados são apresentadas na sequência, focando na descrição da articulação informacional das orações completivas e adverbiais e, em seguida, são discutidos os resultados da validação estatística pelo teste estatístico. As conclusões e observações sobre a sintaxe da fala no que diz respeito às relações entre principal e subordinada nas variedades do português são expostas na última seção sobre os dados.

2. REVISÃO TEÓRICA

2.1. CONTEXTUALIZAÇÃO DAS VARIEDADES DIATÓPICAS DO PORTUGUÊS

2.1.1. O PORTUGUÊS NA ÁFRICA

Sob a ótica linguística, as consequências da colonização portuguesa na África foram muitas, como o surgimento de línguas crioulas a partir do contato entre portugueses e africanos, emergências de novas variedades do português e a integração do léxico de origem portuguesa com as línguas africanas que o português teve contato (HAGEMEIJER, 2016).

Atualmente, restam algumas heranças da colonização. Como Hagemejjer aponta, até à independência dos países em 1975, o número de falantes de português como segunda língua era baixo. Porém, na atualidade, já é possível distinguir onde a língua portuguesa é falada como primeira ou segunda língua. Segundo o Ethnologue (EBERHARD, D. M., GARY F. S.; CHARLES, 2022), o português é a nona língua mais falada do mundo, somando 257,7 milhões de falantes. Segundo dados oriundos de censos, em 2009 o português era falado por 27,1% da população e Guiné-Bissau, e em 2014, por 71,15% em Angola. Já em Moçambique, de acordo com dados do governo, em 2017 esse número representava 47,4% da população e em São Tomé e Príncipe, no último censo em 2012, 98,4% dos santomenses falavam português. Em Portugal e no Brasil, os dados de Camões - Instituto da Cooperação e da Língua de 2013 indicam que eram 10,6 e 194,9 milhões, respectivamente.

2.1.2 A SITUAÇÃO LINGUÍSTICA NAS VARIEDADES AFRICANAS DO PORTUGUÊS

Entretanto, o cenário da língua portuguesa em cada país revela algumas consequências para outras línguas nacionais¹. Em São Tomé e Príncipe, por exemplo, o contexto atual aponta para uma diminuição da porcentagem de falantes do forro, uma língua crioula de base lexificadora portuguesa. Em Moçambique, Gonçalves (2012) destaca que em relação à população mais escolarizada, o português passa a ser a língua dominante de moçambicanos que tinham uma língua bantu como primeira língua na infância. Diante desse cenário, o governo local busca alternativas para promover as línguas nacionais incentivando o projeto criado em 1993 para escolarização bilíngue. Considerando esse cenário, nas próximas subseções iremos nos ater apenas ao contexto histórico das variedades africanas, justamente

¹ Fonseca (2012) apresenta uma sólida contextualização sobre a situação linguística das línguas nacionais na África, especificamente em Angola.

por serem menos estudadas e receberem menos atenção no que tange aos estudos linguísticos sobre a fala espontânea.

2.1.2.1. ANGOLA

Em Angola, até meados do século XVIII, a elite administrativa afro-portuguesa tinha o português como segunda língua. Durante o regime de Salazar (1928-1974), a presença portuguesa em Angola aumentou e houve uma exigência para que os colonos fossem fluentes em português para participarem das decisões públicas. Entretanto, as novas regras não surtiram efeito, pois o número de colonos na década de 70 era 5% (BENDER, 2004, p. 71) e após a independência política esse número foi bastante reduzido.

Nesse cenário, houve uma política de assimilação da língua e cultura portuguesa pelos nativos, no sentido de conduzir os colonos a aprenderem tanto a diamesia escrita quanto a falada. A presença portuguesa em Angola foi bastante significativa, o que pode ter contribuído para o português adquirir status de língua franca. Após a independência, com a migração dos povos que falavam diferentes línguas bantu para a capital, Luanda conferiu ao português o status de língua franca. Atualmente, o português é a única língua oficial de Angola e tem papel unificador, mas no país coexistem mais de vinte línguas bantu com outras línguas europeias².

2.1.2.2. CABO VERDE

A situação linguística em Cabo Verde é diferente de Angola, pois a língua materna é o crioulo cabo-verdiano, instrumento de comunicação e interação diária entre a população. O uso do português se caracteriza a partir de duas formas: os falantes escolarizados usam na fala e na escrita uma forma vernacular e mais rebuscada; já as camadas com menor grau de instrução utilizam um português rudimentar.

Nesse contexto, em situações comunicativas mais formais, como transmissões pela televisão, o português é utilizado. Porém, em interações informais, como conversas em amigos, utiliza-se o crioulo. Portanto, em Cabo Verde, o cabo-verdiano desempenha o papel de língua de unidade nacional.

2.1.2.3. GUINÉ-BISSAU

Historicamente, a presença portuguesa em Guiné-Bissau não foi tão significativa, por isso, atualmente, um número reduzido de guineenses fala e escreve o português. Como

² Para reflexões e contextualizações mais aprofundadas, ver Ndombele e Timbane (2020).

apontam Peixoto e Carioca (2013), o português assume status de língua oficial para fins diplomáticos, de ordem social e política; como a divulgação em programas de rádio e propagandas eleitorais.

A diversidade linguística existente no território também contribuiu para oficializar o português, pois não havia como selecionar, entre as línguas nativas, apenas uma como a língua oficial de Guiné-Bissau. Além disso, como aponta Ferreira (1988), fatores sociais e políticos também foram considerados nessa decisão, visto que, mesmo que uma língua de alguma das etnias fosse escolhida, seria difícil uniformizar e estruturar o ensino nas escolas; considerando-se, ainda, que o país havia enfrentado um conflito armado e estava sob péssimas condições econômicas.

Atualmente, as vinte línguas étnicas são bastante utilizadas em situações comunicativas informais entre familiares e amigos, entre elas a que possui maior número de falantes é o crioulo guineense (80%), que surgiu do contato entre a população de Guiné-Bissau e os portugueses e tem como base o léxico do português. Por outro lado, a língua portuguesa é usada em contextos formais pelo governo e para fins administrativos, e possui um número reduzido de falantes, cerca de 13%, segundo apontam Couto e Embaló (2010).

2.1.2.4. MOÇAMBIQUE

A ocupação portuguesa em Moçambique no período colonial foi mais reduzida, o que pode estar relacionado à baixa quantidade de falantes de português, embora seja a língua oficial do território. Nos anos 90, o governo desenvolveu um programa de ensino bilíngue para incentivar o aprendizado da língua portuguesa em paralelo ao das línguas nacionais.

Nesse contexto, o português coexiste com cerca de 20 línguas bantu³ do território, além de existirem cidadãos falantes de inglês, hindi e árabe (CHIMBUTANE, 2018). Do ponto de vista de veiculação da língua, a forma culta do português é disseminada pela imprensa e jornais, já as línguas nacionais são usadas em contextos informais. De acordo com o Censo de 2017, as línguas bantu são mais faladas pela população do que o português. O mesmo relatório indicou que 47,4% dos cidadãos sabem falar português, ou seja, embora seja limitada a contextos midiáticos e institucionais, a língua portuguesa é empregada por moçambicanos ou como língua materna, ou como segunda língua, conforme apontam os dados.

³ Para mais detalhes sobre o grupo das línguas bantu, veja Moisés; Cande; Jesus (2012) e Chimbutane (2012).

2.1.2.5. SÃO TOMÉ E PRÍNCIPE

Com a abolição da escravatura, os portugueses enviaram às ilhas de São Tomé e Príncipe mão de obra vinda de Cabo Verde, Angola e Moçambique e, nesse contexto, as línguas crioulas deixaram de ser o principal meio de comunicação devido à disseminação da língua portuguesa como L2 e língua franca pelos migrantes chegados das outras regiões.

Segundo Gonçalves e Hagemeyer (2015), assim como em Angola e Moçambique, a hegemonia do português se relaciona ao contexto do período de pós-independência. Atualmente, o português tem um papel dominante no país, mas o crioulo santomense ainda é usado por grande parte da população. De acordo com o último Censo (2012), 170.309 pessoas falavam português e 62.889 usavam o santomense. Balduino e Bandeira (2022) comparam os dados do Censo anterior (2001) e destacam que, embora ainda seja disseminado, o crioulo sofreu uma redução de 8.4% do número de falantes. Nesse cenário, há uma queda no uso de línguas autóctones a favor do emprego de português como L1, Agostinho (2015) destaca que isso gera uma resistência em incentivar o uso das línguas autóctones pelas crianças. Nesse contexto, a língua portuguesa e as línguas crioulas coexistem em um cenário social e ecolinguístico complexo, onde especificamente a população mais jovem limita o uso das línguas crioulas e prefere manter a comunicação em português. Como aponta Gonçalves (2010), a língua portuguesa é usada nos centros urbanos e sofre mudanças advindas do contato linguístico.

Apesar de os estudos, em sua maioria, se voltarem para observações linguísticas em Angola, Brasil e Portugal, há algumas pesquisas que se propuseram a investigar alguns padrões recorrentes no português de outras variedades. No entanto, além de serem escassos, muitos estudos são desenvolvidos a partir de corpora escritos, e as pesquisas voltadas para a diamesia falada se concentram, principalmente, em compreender fatores relacionados ao ensino (MARTINS; GOMES, 2016; LOPES; SOARES, 2017) e aspectos fonético-fonológicos (SANTOS; 2015; DOS SANTOS; SVARTMAN, 2014). Por outro lado, muitas pesquisas se debruçaram em observar as construções sintáticas na fala do português em Portugal e no Brasil. Além disso, às vezes são desenvolvidos mais estudos sobre os crioulos do que sobre as variedades do português, em decorrência também dos diferentes status que o português tem em cada país.

2.2. CONSIDERAÇÕES SOBRE A SINTAXE NA DIAMESIA FALADA

Segundo Raso (2012, p. 58), a fala espontânea é entendida como a fala que é planejada ao mesmo tempo em que é executada, sem planejamento “prévio”. Essa diamesia foi foco de estudos realizados ao longo dos anos, que culminaram na análise de que as estruturas da fala certamente se diferenciam daquelas da escrita. Lyons (1968), seguindo os preceitos de Bloomfield (1935), ressalta que fala e escrita são diamesias que se diferem e são independentes em relação ao vocabulário e gramática. Seguindo os preceitos de Lyons (1968), Miller e Weinert (1998) pontuam que a diamesia falada apresenta além do vocabulário e gramática, morfologia, distribuição de itens lexicais e sintaxe própria.

As diamesias falada e escrita se diferem não apenas em relação a questões estruturais, mas também apresentam distinções na organização da informação devido a características próprias da fala. Miller e Weinert (1998) ressaltam que não é possível “apagar” aquilo que foi dito no momento em que se fala. Assim, falante e ouvinte estão sujeitos a limitações de pouco tempo (de memória), isto é, de “pedaços” de informações que conseguem armazenar em um curto período de tempo. Pawley e Syder (1983) denominam esse conceito de “*one clause at time*”, isto é, os falantes e ouvintes registram por um curto período as informações projetadas na fala. Desse modo, a organização da informação na fala caracteriza a sintaxe da fala como fragmentada, pois a informação é distribuída ao longo da estrutura e cada parte da construção recebe um conteúdo (Miller, Weinert, 1998). Assim, na fala espontânea, o falante “acompanha” seu pensamento enquanto fala e, geralmente, as estruturas realizadas na fala são menos complexas que na escrita, dado que a informação é distribuída em poucos itens lexicais e a extensão do vocabulário na fala não guiada tende a ser a menos complexo. É possível reconhecer que essa característica da fala está relacionada ao fato de que, tendencialmente, as construções realizadas são orientadas ao contexto comunicativo e à interação com o interlocutor, portanto, não é necessário que haja articulação, como a da escrita para que o ouvinte seja constantemente situado no tempo, visto que a marcação no tempo é feita frequentemente por marcações nos itens lexicais (MILLER; WEINERT, 1998).

De acordo com Miller e Weinert (1998), a fala é acompanhada, ainda, por gestos, interações faciais, contato visual e postura corporal. Os elementos prosódicos, como *pitch*, ritmo, *voice quality* e amplitude também são próprios dessa diamesia. Dentre os componentes prosódicos da fala, a entonação é responsável por sinalizar a relação entre as construções (sentenças, orações ou frases). Assim, é a entonação que veicula uma marcante diferença entre fala e escrita. Segundo Ladd (1996), a entonação utiliza componentes fonéticos além das unidades sonoras individuais (sílabas, acento, etc) para sinalizar diferentes significados pragmáticos.

As pesquisas desenvolvidas ao longo dos anos mostraram relevantes diferenças entre a diamesia falada e escrita. Kroll (1987) mostrou que há maior ocorrência de subordinação em registros escritos (35% dos dados) do que em falados (14%). Poole e Field (1976) também observaram particularidades nas estruturas realizadas na fala, como a maior quantidade de subordinadores do que em dados escritos. Apesar de apresentarem observações comparativas, Beaman (1984) pontua que as conclusões as quais os autores chegaram pode ser fruto de um equívoco de interpretação entre modalidade, que se divide em *falada* e *escrita*, e registro, dividido em *formal* e *informal*, ou seja, de acordo com o autor, fala e escrita apresentam características distintas de formalidade e tempo de planejamento.

Do ponto de vista informacional, Schulz (1973) e Fernandez (1994) observaram que as relações entre as orações podem ser indicadas por elementos gramaticais ou pela entonação. Como observado no francês⁴, esse elemento sinaliza que há integração entre as estruturas, mas que não necessariamente exista uma relação de coordenação ou subordinação:

(1) (a) T'auras pas de dessert t'es pas venu avec nous

"Você não terá nenhuma sobremesa (porque) você não veio conosco"

(b) Il n'a pas plu le linge est sec

"Não choveu (já que) a roupa está seca"

Os exemplos apresentados por Fernandez (1994, pg. 95-96) mostram que o padrão prosódico sinaliza que *t'es pas venu avec nous* e *le linge est sec* estão unidos a suas respectivas primeiras orações. Logo, a prosódia, veiculada na língua falada, indica essa ligação entre as duas cláusulas e não apenas pelo índice lexical. O autor, ainda, destaca que, em alguns casos, a repetição e a inversão de ordem das palavras funcionam como uma forma de persuasão, para incentivar o interlocutor a fazer ou pensar algo. As observações mostram, portanto, algumas estratégias possíveis na diamesia falada que não possuem recursos para serem aplicadas na escrita.

Sornicola (1981) e Enkvist (1982) fazem observações interessantes a respeito da integração sintática entre os elementos na fala. De acordo com os estudos, em alguns casos, por motivações prosódicas ou relacionadas ao contexto, torna-se complicado reconhecer a dependência sintática entre alguns termos:

(2) A whose idea// was it

BI X's

C uhuh

⁴ Não é possível ter acesso aos áudios da pesquisa do autor.

B2 they got one of the teachers that we always play jokes
 on/one of the young
 women/ they got her to write it

O exemplo apresenta uma transcrição de conversações espontâneas do inglês americano e, apesar de não termos acesso ao áudio, as pesquisas reportaram uma fragmentação sintática. O sintagma *one of the young women* é inserido entre duas orações (*they got one of the teachers that we always play jokes* e *they got her to write it*) e parece interrompê-las. Entretanto, é possível interpretar isoladamente o fragmento dividido da oração *they got one of the teachers that we always play jokes, they got her to write it*, de modo que as duas estruturas são independentes sintaticamente, apesar de parecer haver uma relação no nível pragmático. Desse modo, pode-se interpretar uma relação no plano pragmático entre o sintagma e a primeira oração, mas também é possível reconhecer que esse elemento, sintaticamente, está "solto" dentro da construção sintática. Nesse sentido, novamente, ressalta-se a importância da prosódia para reconhecer as interpretações possíveis dos enunciados e as relações informacionais entre os itens lexicais.

Miller e Weinert (1998) pontuam algumas diferenças entre fala e escrita no inglês americano a partir da análise das WH-clauses. Os autores comparam dois exemplos que podem gerar interpretações equivocadas sobre a forma escrita e a falada:

- (3) what you're going to do—you're going to go up past the allotments
 (4) What you are going to do is go up past the allotments

A princípio, pode-se imaginar que as ocorrências são contrastivas e que (3) é a representação na forma escrita, de (4). Com base nessa perspectiva, a construção em (3) tem a segunda oração principal com aspecto progressivo e SN sujeito explícito, já a construção escrita em (4) não tem. Contudo, a diamesia falada não deve ser analisada a partir de parâmetros comparativos com a escrita, pois isso seria minimizar as características próprias da fala de modo a encaixá-las em análises escritas. Miller e Weinert (1998) destacam que a oração-WH codifica instruções para o interlocutor "eliminar" as informações antigas e inserir novas informações, como uma reformulação. Além disso, em (3) *what you are going to do* aponta para um conteúdo que o falante especifica na segunda oração *you're going to go up past the allotments*, em uma espécie de pausa.

2.3. A LANGUAGE INTO ACT-THEORY

O presente trabalho tem como base teórica a *Language into Act-Theory* (L-AcT, CRESTI, 2000; MONEGLIA; RASO, 2014). A L-AcT é uma teoria indutiva, desenvolvida a partir de verificação empírica nos *corpora* compilado pelo LABLITA (*Laboratorio di Linguistica del Dipartimento dell'Università di Firenze*, MONEGLIA, 2005). Além disso, é uma teoria *corpus-driven* e voltada para a análise pragmática e prosódica da fala espontânea, reconhecida como a fala que é planejada ao mesmo tempo em que é executada, isto é, não é uma fala guiada (NENCIONI, 1983 *apud* CRESTI, 2000). O estudo da fala, durante muito tempo, se ateuve à transcrição de textos escritos e, portanto, uma categoria adequada à análise dessa diamesia. Contudo, devido ao fato de o estudo da fala exigir aparatos tecnológicos adequados, as pesquisas ficaram enviesadas e restritas à escrita. Com o avanço dos equipamentos de gravação e dispositivos para análise da fala no século XXI, as características da diamesia falada tendem a ser preservadas nas transcrições de textos de fala por meio do uso de convenções e critérios específicos para tal diamesia, a fim de que a fala não seja estudada a partir de meios convencionados à escrita.

A partir disso, o estudo da fala começou a ser feito através de critérios próprios e por meio de aparatos tecnológicos adequados. Essa metodologia é adotada pela L-AcT, que individualiza a unidade de referência da fala no enunciado, considerado uma unidade com autonomia prosódica e pragmática; e não em uma unidade sintática, como a oração (CRESTI, 2000). A L-AcT é considerada uma extensão da Teoria dos Atos de Fala de Austin (1962), na medida em que considera que o enunciado corresponde a um ato de fala locutivo (material linguístico transmitido) que veicula uma força ilocucionária. Os atos de fala são compostos por três atos simultâneos (AUSTIN, 1962). O ato locutivo é material linguístico transmitido, o ato ilocutivo consiste na ação realizada por meio do ato locutivo (perguntar, ordenar, convidar, ameaçar etc) e o ato perlocutivo corresponde aos efeitos que o ato ilocutivo provoca no interlocutor (por exemplo, ao receber uma ordem o interlocutor a cumpre ou não). Na visão de Cresti (2000), porém, a ilocução corresponde à atitude do falante em relação ao ato de fala, transformada em comportamento social convencionalizado.

Ao produzir um ato de fala, o falante tem a intenção de realizar um objetivo comunicativo, como pedir e aconselhar. As ilocuções produzidas pelo falante podem ser distinguidas por meio da prosódia, ou seja, através do elemento prosódico é possível distinguir funções comunicativas distintas. Além disso, a prosódia é o componente que possibilita diferenciar um registro falado e escrito e responsável por veicular a autonomia

pragmática do enunciado. Segundo a L-AcT, a prosódia também segmenta o fluxo da fala em enunciados e em unidades internas a ele. A segmentação obedece a critérios prosódicos e não necessariamente sintáticos, que refletem a autonomia prosódica e pragmática do enunciado:

(1) bpubcv02⁵

*OSV: hum hum //

(2) bfammn04

*REG: psiu //

(3) bfamdl02

*BAL: Nossa //

(4) bfamdl01

*REN: acho que não //

*REN: eu acho que no saco de lixo nós já passamo //

(5) bfamdl03

*LUZ: porque eu acho que no mesmo concurso /

Após escutar os enunciados, pode-se perceber que as ilocuções acima veiculadas pela negação (*hum hum*), interjeição (*psiu*) e exclamação (*Nossa*) são autônomas do ponto de vista prosódico e pragmático mesmo que não apresentem conteúdo sintático e semântico e, portanto, são considerados enunciados. Nesse caso, as ilocuções são reconhecidas como prosódica e pragmaticamente completas, de modo que é possível reconhecer a intenção comunicativa do falante ao proferir os enunciados (possivelmente, concordar, chamar e expressar surpresa). Contudo, os exemplos (4) e (5), apesar de apresentarem preenchimento sintático, são contrastantes, pois ao contrário de (4), a estrutura em (5) apresenta perfil prosódico não conclusivo, de modo que se espera algo a seguir. Desse modo, como mostrado em (1-3) não é necessário que haja autonomia sintática para haver autonomia pragmática,

⁵ Com base nos critérios de compilação dos corpora do projeto C-ORAL, as barras são usadas para marcar um enunciado como concluído (“//”) ou não concluído (“/”). Para maior detalhamento dos critérios de transcrição, conferir Raso e Mello (2012). Considerando-se a relevância da prosódia, os áudios referentes aos enunciados são disponibilizados neste trabalho no link: <https://drive.google.com/drive/folders/1JuiOQ2tQtfoh7ED3dTga27aESgtjMa4V?usp=sharing>. Os exemplos aqui presentes foram retirados do corpus Português Falado (GONÇALVES; VELOSO, 2020) e C-ORAL BRASIL (RASO; MELLO, 2012).

assim como não é preciso autonomia pragmática para que exista autonomia sintática (como em (5). Portanto, ao contrário de (5), os exemplos anteriores evidenciaram que, mesmo sem o contexto ampliado, é possível interpretá-los como ato de fala e, desse modo, reconhecê-los como unidades com autonomia prosódica e pragmática, isto é, um enunciado:

(6) bfamd103

*LUZ: porque eu acho que no mesmo concurso / cê nã pode fazer duas //

O contexto ampliado de (5) mostra que a construção não era interpretada como ato de fala e não constituía um enunciado, visto que *porque eu acho que no mesmo concurso* constitui uma parte não autônoma do enunciado cuja autonomia pragmática é veiculada por *cê nã pode fazer duas*, que carrega a proeminência acústica (núcleo prosódico) da ilocução. Logo, apenas em (6) o ato de fala é concluído e, assim, a estrutura é considerada um enunciado.

Como já apresentado, a prosódia é responsável por segmentar o fluxo da fala em enunciados e unidades internas a ele. A partir de percepções de variações prosódicas, segmenta-se o enunciado em quebras prosódicas que podem ser: conclusivas (“//” na transcrição), que segmentam as fronteiras entre enunciados; e não conclusivas (“/” na transcrição) que marcam unidades menores dentro do enunciado (as unidades tonais). Para critério de marcação das quebras prosódicas, são considerados elementos como velocidade de fala, mudanças de intensidade, variações na curva de f_0 , etc (RASO, 2013, p. 30). Com base na percepção das variações prosódicas, os enunciados podem ser divididos em (i) *simples*, formados por apenas uma unidade tonal, como em (7); e (i) *complexos*, quando há mais de uma unidade tonal interna a ele, como em (8):

(7) eu não vou dizer que eu consigo directamente dirigir // (ST)⁶

(8) eu vou dizer / como é que se faz açucarinha // (ST)

Nota-se que em (7) o enunciado apresenta apenas uma quebra prosódica, sendo essa, terminal e, portanto, é um enunciado simples. Já em (8), o enunciado contém unidades tonais internas a ele, ou seja, quebras não terminais e, por isso, é um enunciado complexo. A partir da segmentação prosódica, é possível ter diferentes interpretações sobre o mesmo enunciado:

(9) mas / felizmente / que o que estou a ver que há [/1] há pessoas aqui na juventude que estão interessados / mesmo pela raiz / da cultura cabo-verdiana //⁷(CV)

⁶ As abreviaturas indicam os países: ST (São Tomé e Príncipe), M (Moçambique), CV (Cabo Verde), A (Angola), GB (Guiné-Bissau), PT (Portugal) e BR (Brasil).

Após o acesso ao áudio, é possível interpretar que a construção *estou a ver que há pessoas aqui na juventude* pode ser uma completiva ([o que] estou a ver que há pessoas aqui na juventude) ou pseudo-clivada (o que estou a ver [é] que há pessoas aqui na juventude). Portanto, nota-se que as diferenças no reconhecimento das estruturas são veiculadas pela prosódia, a partir da segmentação do enunciado.

Com base na L-AcT, as quebras prosódicas delimitam unidades tonais que, a depender de determinados perfis prosódicos, funções e distribuições, veiculam as unidades informacionais. A dimensão prosódica, como já mencionado, é veiculada pelas percepções de variações prosódicas (terminal ou não terminal), ou seja, como unidades tonais. No nível pragmático, por sua vez, as unidades tonais recebem marcações com base em determinados perfis prosódicos, funções e distribuições próprias dentro do enunciado. Nesse sentido, cada unidade tonal veicula um valor informacional que é carregado pelas unidades informacionais, que possuem natureza pragmático-informacional. Segundo o modelo IPO (*Institute for Perception Research*), apesar de os enunciados possuírem muitos movimentos de *pitch* (HART; COLLIER; COHEN, 1990), apenas os movimentos produzidos de forma intencional (ainda que inconsciente) são percebidos pelos falantes.

2.3.1. AS UNIDADES INFORMACIONAIS

Com base na L-AcT, as unidades informacionais são divididas em textuais e dialógicas. As unidades textuais compõem o texto do enunciado e se dirigem a ele, ou seja, fornecem informações sobre como interpretá-lo. As unidades dialógicas, por sua vez, são dirigidas ao interlocutor a fim de regular a interação e não compõem o texto do enunciado (correspondem, em geral, aos marcadores discursivos). Dentro desse cenário, os enunciados simples possuem apenas uma unidade informacional, sendo necessariamente o Comentário, que é a unidade que carrega ilocução e fornece autonomia pragmática ao enunciado. Por outro lado, os enunciados complexos apresentam mais de uma unidade informacional, além do Comentário; como mostram os exemplos a seguir:

(10) não sei se as condições actuais serão as mesmas //=**COM**= (ST)

(11) mas se não pode /=**TOP**= pode fazer qualquer coisa //=**COM**= (ST)

⁷ A barra entre colchetes indica o *retracting*, ou seja, quando o falante reformula o conteúdo locutivo. O número entre os colchetes marca a quantidade de palavras reformuladas.

O exemplo em (10) apresenta um enunciado formado por apenas a unidade de Comentário e, portanto, não constitui um padrão informacional. O enunciado complexo em (11), por sua vez, apresenta um padrão recorrente na fala, formado pelas unidades de Tópico e Comentário. Nessa construção, a primeira unidade tonal, que apresenta o perfil prosódico do Tópico (*mas se não pode*), identifica um domínio semântico (no caso, a condicionalidade) e fornece informação de *background* para a ilocução que está no Comentário. Existe, portanto, uma relação de *aboutness* pragmática entre as duas unidades, de modo que o Tópico referencia semanticamente a ilocução presente no Comentário.

Como já apresentado, segundo o modelo IPO, os falantes percebem os movimentos de *pitch* voluntários, denominados perfis prosódicos, que servem para a interpretação do enunciado. Nesse sentido, a prosódia é responsável por intermediar a relação entre unidade tonal e unidade informacional de modo que os perfis prosódicos, segundo a L-AcT atribuem valores informacionais às unidades tonais. Assim, de acordo com a Teoria da Padronização da Informação (CRESTI, 1994; MONEGLIA; CRESTI, 2006; CRESTI; MONEGLIA, 2010; MELLO; PANUNZI; RASO, 2011), as unidades tonais carregam valores informacionais com base nos perfis prosódicos, além da posição em relação ao COM (distribuição no enunciado) e funções próprias.

Os perfis prosódicos são de três tipos: *root* (raiz), *prefix* (prefixo) e *suffix* (sufixo). O perfil *root* define o tipo de ilocução e corresponde ao Comentário, o perfil *prefix* é típico do TOP e ocorre antes de perfis tipo *root* (no PB foram individualizados quatro tipos de perfis de Tópico, cf. Mittmann, 2012, Cavalcante, 2015); e o perfil *suffix* é característico das unidades de Apêndice de Comentário e segue as unidades de tipo *root*. Contudo, apenas o perfil de tipo *prefix* pode ser recursivo, isto é, pode permitir que uma unidade de tipo *prefix*, como o TOP, seja acrescentada a outra unidade de *prefix*. Logo, um TOP pode ser acrescentado a outro TOP, mas o mesmo não é possível para o Comentário e Apêndice de Comentário, por exemplo.

É importante salientar que neste trabalho apenas as unidades textuais foram consideradas na análise, pois são aquelas que compõem o texto do enunciado, visto que as dialógicas são usadas pelo falante para regular a interação com o interlocutor.

2.6.1.1 O COMENTÁRIO (COM)

O Comentário é a unidade necessária da fala e veicula a força ilocucionária do enunciado, isto é, carrega o núcleo prosódico da ilocução. A unidade de Comentário carrega o

foco funcional⁸, isto é, o núcleo prosódico que marca o valor funcional da ilocução no COM. Além disso, devido ao perfil prosódico de tipo *root*, sua distribuição é livre dentro do enunciado, ou seja, as unidades informacionais se distribuem a partir do COM de modo que esta unidade serve como orientação para as outras UIs se ordenarem ao redor do Comentário⁹:

(12) como é que eu vou saber //COM=

(13) de festa //COM=

(14) uma prova //COM=

(15) psiu //COM=

Os exemplos mostram diferentes ilocuições realizadas no Comentário: em (12) um enunciado sintaticamente completo (frase interrogativa), em (13) um sintagma preposicional, em (14) um sintagma nominal e em (15) uma interjeição. Portanto, tanto em construções sintaticamente completas quanto naquelas em que o falante utiliza apenas interjeições, advérbios ou ruídos paralinguísticos, por exemplo; é possível reconhecer os objetivos comunicativos do falante. Além disso, diferentes ilocuições no COM podem ser realizadas pelo mesmo conteúdo locutivo: por exemplo, a interjeição *ahn* pode veicular uma pergunta ou expressar surpresa.

2.6.1.2 O TÓPICO (TOP)

O Tópico (SIGNORINI, 2003; 2005) estabelece o domínio semântico para ilocução carregada pelo Comentário, ou seja, identifica, no enunciado, o âmbito semântico ao qual a ilocução no COM remete. O perfil prosódico do TOP é do tipo *prefix* e, portanto, ocorre antes do COM e permite a estratégia de recursividade, ou seja, um TOP pode ser adicionado após outro TOP formando, assim, uma Lista de Tópicos (TPL). O Tópico apresenta três formas prosódicas, que foram individualizadas no italiano (CRESTI, 2000; FIRENZUOLI; SIGNORINI, 2003), inglês americano (CAVALCANTE, 2020) e português brasileiro (MITTMANN, 2012). No Tópico podem ocorrer diferentes construções:

(16) quando eu cheguei aqui /=TOP= todas as minhas calças tinham ficado lá hhh //COM=

⁸ Na maior parte das abordagens na literatura, o foco funcional está relacionado à Focalização (veja SILVA, 2013), o que não corresponde à definição adotada pela L-AcT.

⁹ Para fins explicativos, alguns exemplos foram retirados do corpus C-ORAL BRASIL (RASO; MELLO, 2012), visto que apenas as ocorrências de subordinação foram etiquetadas no corpus das variedades do português.

(17) desinfetante /=TOP= não //=COM=

(18) e depois da instrução /=TPL= organização /=TPL= o efectivo da banda /=TOP= quando aumentar /=TOP= nós temos muitos programas /=COB= que queremos levar a cabo //=COM= (costureira)

(19) departamento /=TOP= Artes Plásticas //=COM=

O exemplo (16) contrasta com (17) e mostra que não é necessário que haja composicionalidade sintática entre Tópico e Comentário, pois o primeiro apresenta uma estrutura subordinada completa e o segundo, um exemplo em que não há dependência sintática entre as UIs. Além disso, outras construções podem ocorrer em TOP, como o sintagma nominal em (17) e o exemplo de recursividade do Tópico em (18), no qual a unidade forma uma padronização melódica e constituem um único âmbito de aplicação da ilocução no Comentário. O enunciado em (19) apresenta a relação de *aboutness* pragmática que ocorre entre Tópico e Comentário. Nesta relação, *departamento* que está no TOP fornece informação de *background* para ilocução no Comentário, mesmo que não haja mais itens lexicais que especifiquem sintaticamente a relação entre as UIs e a ilocução veiculada.

2.6.1.3 O APÊNDICE DE TÓPICO (APT) E APÊNDICE DE COMENTÁRIO (APC)

As unidades de Apêndice (TUCCI, 2006; RASO; ULISSES, 2008; ULISSES, 2008) possuem perfil prosódico tipo *suffix* e são realizados imediatamente depois da unidade a qual integram. Do ponto de vista funcional, tais unidades integram textualmente, o Tópico, no caso do Apêndice de Tópico e o Comentário, no caso do Apêndice de Comentário:

(20) porque eu /=TOP= também se fosse pela mãe /=APT= nã levaria não //=COM=

(21) e sem material /=COM= também /=APC= né //=PHA=

(22) uma /=SCA¹⁰= tora /=COM= da madeira //=APC=

As unidades de Apêndice carregam oração condicional em (20), advérbio em (21), e sintagma preposicional em (22). Nota-se como não há restrição morfossintática, visto que o enunciado é autônomo pragmática e prosodicamente.

2.6.1.4 O PARENTÉTICO (PAR)

¹⁰ Existem algumas unidades que não possuem valor informacional, como a SCA (Escansão), que ocorre quando uma unidade informacional textual é realizada ao longo de mais de uma unidade tonal.

A unidade de Paréntetico (TUCCI 2004; 2009; SANTOS; BOSSAGLIA, 2018), possui função metalinguística, pois fornece instruções sobre como o enunciado ou parte dele deve ser interpretado (RASO; MELLO, 2012). Prosodicamente, a unidade, em geral, apresenta frequência fundamental (f_0) mais baixa e velocidade mais alta que o restante do enunciado, além de perfil nivelado. Do ponto de vista distribucional, o Parentético pode ocorrer em qualquer posição dentro do enunciado, desde que não seja no início:

(23) deixa ferver /=COB= mas não pode mexer /=COB= porque tem tempero /=PAR= para não cheirar /=COB= tempero //COM= (ST)

(24) &he /=TMT= Filosofia /=COB= se não contarmos com a opção seminário /=i-PAR= portanto /=AUX= que tem /=i-PAR= como sabe /=PAR(2)= Filosofia e Teologia na sua formação /=PAR= &he /=TMT= nunca houve //COM= (A)

(25) pôs na liquidação por cinquenta-e-nove /=COM= quando tava na liquidação //PAR=

O exemplo em (23) apresenta o PAR com função explicativa, em que o falante justifica o que foi dito por ele, isto é, o motivo pelo qual não se pode mexer alimento enquanto ocorre o processo de cozimento. A ocorrência em (24) mostra como o Parentético pode interromper uma unidade, no caso ocorre entre unidades de Comentários Ligados (COBs). Em (25), o Parentético especifica um domínio temporal de um evento. De acordo com os critérios de Santos e Bossaglia (2018)¹¹, em (23) o PAR é realizado em posição média, em (24) em posição encaixada, e em (25), em posição final.

2.6.1.5 O INTRODUTOR LOCUTIVO (INT)

O Introdutor Locutivo (GIANI, 2003; 2004; ROCHA, 2011; ROCHA, RASO, 2011) tem como função sinalizar que o conteúdo locutivo que segue pertence a um nível hierárquico diferente daquele da enunciação e, portanto, deve ser interpretado em planos pragmáticos diferentes. A unidade possui perfil descendente e apresenta claramente um contraste entre a f_0 do INT e da unidade seguinte, de modo que o Introdutor Locutivo possui uma f_0 muito baixa e velocidade de elocução mais rápida que o restante do enunciado.

(26) hoje /=TOP= acho que /=INT= se /1=SCA= muitas coisas que eu fiz /=TOP= talvez não fizesse /=COB= se voltasse ao ensino //COM= (ST)

¹¹ A descrição detalhada da unidade de Parentético encontra-se em Santos e Bossaglia (2018).

(27) é preciso que /=INT= toda a gente /=COB= a todos os níveis /=COB= cada um a seu nível /=PAR= efectivamente /=PAR= saiba aquilo que deve fazer /=COB= e /=DCT= de que é que dispõe para fazê-lo //COM= (ST)

Do ponto de vista distribucional, o Introdutor Locutivo é posicionado antes da unidade a qual introduz. A unidade está introduzindo um Tópico com exemplificação em (26) e Comentários Ligados em (27). O exemplo abaixo exemplifica uma relação pragmática entre a unidade informacional e o conteúdo do enunciado:

(28) mas primeiro cê nũ fez assim /=COB= tipo +=INT= fez //COM=

O Introdutor Locutivo está em uma unidade interrompida, ou seja, o falante não dá sequência ao conteúdo locutivo. Porém, ao escutar o enunciado, é possível reconhecer que o falante usa um gesto para exemplificar a ação introduzida pelo INT¹². Portanto, o exemplo ilustra a multimodalidade da fala e como os fatores pragmáticos, e não linguísticos, veiculam a interpretação do enunciado.

2.6.2 OS LIMITES DO ISOMORFISMO

De acordo com a L-AcT, é possível reconhecer um isomorfismo entre sequência terminada (o que está entre duas quebras terminais, “//”) e ilocução (intenção comunicativa do falante), ou seja, cada sequência terminada veicula uma ilocução. Do mesmo modo, há uma correspondência biunívoca entre unidade tonal e unidade informacional, de modo que cada unidade tonal veicula uma unidade informacional. Entretanto, existem três unidades nas quais tais correspondências não são observadas: unidade de escansão (SCA), Comentários Múltiplos (CMMs) e Comentários Ligados (COBs).

2.6.2.1 A ESCANSÃO (SCA)

Na unidade de Escansão, o isomorfismo entre unidade tonal e unidade informacional não é respeitado, pois constitui casos em que unidades informacionais textuais são realizadas em mais de uma unidade tonal. Segundo Raso (2012), os motivos pelos quais isso ocorre podem ser variados: a) grande dimensão silábica do conteúdo locutivo, de modo que fisiologicamente não é possível realizá-lo em uma única unidade tonal; b) questões

¹² Uma descrição mais completa do Introdutor Locutivo foi feita por Rocha (2011) e Rocha e Raso (2011).

relacionadas à escassa perícia na fala, como ocorre com falantes muito jovens ou de baixa diastratia; c) razões de ênfase; ou d) por algum motivo de hesitação:

(29) acho que era muito /=SCA= rígida //COM= (ST)

(30) e comiam bem /=COB= porque /=SCA= este prato leva muito //COM= (CV)

(31) e se era para estar a [/1]=SCA= a dividir aos [/1]=SCA= aos pratinhos +=TOP= e essa gamela /=TOP= era colocada ao centro da [/1]=SCA da mesa //COM= (CV)

A unidade de Escansão não possui valor funcional e o perfil prosódico próprio da unidade informacional escansionada é realizado na última unidade tonal, de modo que as unidades tonais anteriores a ela carregam perfil prosódico neutro. Portanto, as unidades de Escansão são distribuídas em unidades tonais diferentes de uma mesma unidade informacional e não interferem na composicionalidade sintática dentro da mesma UI, de modo que é possível reconhecer relações sintáticas quando as UIs ocorrem escansionadas. Aliás, é válido mencionar que foi observado que nas variedades do português analisadas, justamente por se tratar de textos monológicos, os falantes escansionam a fala principalmente quando relatam determinados assuntos em longos trechos.

2.6.2.2 OS COMENTÁRIOS MÚLTIPLOS (CMMs)

No caso dos Comentários Múltiplos, a relação entre sequência terminada e ilocução não é observada, pois em uma única sequência terminada é realizada uma sequência de Comentários em um padrão retórico convencionalizado. Desse modo, as ilocuções ocorrem em um mesmo padrão prosódico e cada ilocução possui sua própria força ilocucionária. Quando há ocorrência desse padrão, mesmo que as ilocuções pertençam a classes diferentes, gera-se uma interpretação holística no interlocutor. Assim, os CMMs possuem autonomia isoladamente, porém o padrão ilocucionário é interpretado como um todo e não como uma soma das ilocuções de cada Comentário Múltiplo.

(32) e /=SCA= &v [/1]=EMP= nota-se /=CMM= de avião /=CMM= quem chega /=CMM= não é //AUX= (CV)

(33) agora /=COB= se [/1]=EMP= o professor não pode exigir nada /=CMM= porque o professor não me paga //CMM= (GB)

Ainda não foram individualizados todos os tipos de CMMs, porém sabe-se que as unidades podem estabelecer entre si relações de: padrão de lista, pedido de confirmação, comparação, chamamento funcional, clímax, correção, persuasão, relações de coordenação (alternativa e aditiva), dentre outros. Além disso, a relação entre CMMs pode ter valores lógicos, como temporais, causais e hipotéticos. Do ponto de vista prosódico, o perfil das unidades é de tipo *root*.

2.6.2.3 OS COMENTÁRIOS LIGADOS (COBS)

Nas estrofes, a correspondência biunívoca entre sequência terminada e ilocução também não é obedecida. As estrofes são formadas por sequências de Comentários, os Comentários Ligados, unidos entre si por um sinal prosódico de continuidade. O elemento prosódico, nas estrofes, indica que, embora cada ilocução possa ser interpretada autonomamente, a sequência não deve ser considerada concluída, pois cada Comentário Ligado deve ser interpretado um a um até o último COB da estrofe. Portanto, as unidades estão ligadas entre si por andamento processual.

As estrofes, ou *stanzas* (CRESTI, 2010), são compostas por uma sequência de Comentários, os Comentários Ligados, unidos uns aos outros por um sinal prosódico de continuidade e ligados entre si por um andamento processual. Isso significa que cada Comentário Ligado deve ser interpretado por vez até se chegar ao último COB. As estrofes são típicas de textos monológicos. À proporção que cada Comentário Ligado é realizado, a força ilocucionária enfraquece e as ilocuições tendem a ser homogêneas, isto é, da mesma classe. O corpus aqui utilizado se caracteriza pela predominância de estrofes, porque nesse tipo de texto a interação entre os interlocutores é baixa e o falante não precisa fazer esforços para deter o turno de fala:

(34) se haverá /=SCA= licença de parto /=SCA= para aquelas que têm cesariana /=COB=
 porque /=DCT= o sofrimento /=TOP= não é o mesmo /=COB= logicamente /=PAR=para
 curar a ferida /=COB= as três carnes /=COB= assim digamos /=PAR= as três carnes que eu
 sofri de corte /=COB= para eu poder tirar a minha criança //COM= (GB)

(35) isso também acontece em Cabo Verde /=COB= porque /=DCT= apesar de [/1]=SCA= de
 aqui &a [/1]=SCA= as raparigas têm muito mais que fazer /=COB= como sabe /=PAR= vão
 apanhar água /=COB= sobretudo se for numa zona rural /=PAR= têm que apanhar lenha

/=COB= &he /=TMT= ajudam os pais com os /=SCA= bebês /=COB= os irmãos mais pequeninos /=COB= às vezes há &m /=SCA= pais que saem de manhã e só regressam à tarde /=COB= são as meninas é que tomam conta dessas [/1]=SCA= &de /1=EMP= dessas crianças /=COB= dos irmãos mais pequeninos /=COB= e /=DCT= mesmo assim /=SCA= com toda essas tarefas /=TOP= às vezes os alunos [/2]=EMP= as alunas /=TOP= ou /=SCA= as raparigas /=TOP= têm muito mais /=i-COB= &he /=TMT= notas /=COB= ou estudam muito mais /=COB= esforçam-se muito mais //COM= (CV)

Os monólogos apresentam narrações de acontecimentos e fatos pelo falante. Em (34), há um relato sobre a experiência em uma banda, com a organização informacional distribuída em uma estrutura complexa. Em (35), há a narração da rotina das mulheres em Cabo Verde, na qual o falante constrói uma narrativa sobre esse acontecimento, através de um andamento processual, característico dos COBs.

Prosodicamente, assim como os CMMs, os COBs são unidades de tipo *root*. Entretanto, diferentemente dos Comentários Múltiplos, os Comentários Ligados não estão em relação padronizada e não geram uma interpretação holística, ao contrário, os COBs estão justapostos e ligados entre si por andamento processual e o padrão ilocucionário é interpretado com uma soma dos valores ilocucionários de cada unidade.

2.6.3 A SINTAXE NA FALA: UMA ABORDAGEM DA L-Act

De acordo com a *Language into Act-Theory* (L-Act; CRESTI, 2000; 2014; MONEGLIA; RASO, 2014), a fala deve ser analisada com base no nível pragmático-informacional, de modo que as relações sintáticas entre as unidades informacionais são, primeiramente, orientadas pelo nível pragmático. Nesse sentido, as UIs têm funções pragmáticas e comunicativas e, sobretudo, estabelecem entre si relações de tipo informacional, e não necessariamente sintático.

Segundo a L-Act, a sintaxe de fala é fragmentada e não deve ser interpretada com base em dependências sintáticas tradicionais, mas como combinação de fragmentos sintáticos, pois a “a sintaxe do enunciado não corresponde a uma configuração unitária hierárquica, mas à combinação de orações, sintagmas ou fragmentos sintáticos locais” (CRESTI, 2014, p. 368). As relações sintático-semânticas ocorrem dentro de cada unidade informacional, concebida como ilha sintático-semântica. Assim, não haveria composicionalidade sintático-semântica entre as ilhas, de acordo com Cresti (2014) a articulação sintática ocorre localmente dentro

das UIs, que possuem funções pragmáticas veiculadas prosodicamente. A sintaxe da fala é dividida em dois tipos, sintaxe linearizada e padronizada:

(36) eu não vou dizer que eu consigo directamente dirigir //COM= (ST)

(37) eu acho que se nós estivéssemos perto dela /TOP= talvez nós nem teríamos dado apoio //COM= (M)

Na sintaxe linearizada em (36), as estruturas de subordinação e coordenação são realizadas dentro da mesma unidade informacional. No exemplo, a oração principal *eu não vou dizer*, o complementizador *que* e o predicado da completiva *eu consigo directamente dirigir* ocorrem na mesma UI, o Comentário. Já na sintaxe padronizada em (37), as estruturas de subordinação e coordenação são distribuídas ao longo de mais de uma UI e possuem uma relação primeiramente no nível informacional. No caso, a oração condicional *se nós estivéssemos perto dela* estão no Tópico e a matriz *talvez nós nem teríamos dado apoio*, no Comentário. Segundo a L-AcT, a ocorrência das construções sintáticas para além das ilhas veicula funções pragmático-informacionais distintas, e as relações sintáticas entre as UIs, como uma oração condicional em Tópico e a principal no Comentário, não entra em conflito com a relação informacional entre as UIs (cf. SILVA, 2020).

Nessa perspectiva, para estudar a sintaxe da fala de acordo com esse aporte teórico, é preciso entender que a partir da combinação das ilhas, que estão relacionadas ao padrão informacional do enunciado. Contudo, apenas as unidades textuais são consideradas para fins da combinação nas UIs, pois são aquelas que compõem o texto do enunciado, já as dialógicas são orientadas ao interlocutor e, prosodicamente, a duração curta não permite comportar um material fonológico maior, como ocorre em construções mais complexas. Em suma, nesse arcabouço teórico, as relações entre as unidades informacionais são veiculadas pela prosódia e de tipo pragmático-informacional. Porém, considerando que o corpus utilizado nessa pesquisa possui textos predominantemente monológicos, a sintaxe se mostra menos fragmentada que aquela da fala espontânea acional.

3. METODOLOGIA

3.1. O CORPUS PORTUGUÊS FALADO

Os dados da pesquisa foram extraídos do corpus *Português Falado – Variedades Geográficas e Sociais* (GONÇALVES; VELOSO, 2000), retirado da seção oral do *Corpus de Referência do Português Contemporâneo* (NASCIMENTO, 2000; GÉNÉREUX; HENDRICKX; MENDES, 2010), compilado pela equipe do *CLUL - Centro de Linguística da Universidade de Lisboa*. O recurso está disponível em: <http://www.clul.ulisboa.pt/recurso/portugues-falado-variedades-geograficas-e-sociais> e, na versão mais atualizada, em <http://catalog.elra.info/en-us/repository/browse/ELRA-S0345/>. Apesar de ter sido compilado com o objetivo específico de auxiliar o ensino do português como língua estrangeira, o corpus é um importante recurso também para o estudo do português falado em geral por ter características adequadas ao estudo dessa diáspora, colaborando para pesquisas sobre as variedades diatópicas do português.

O recurso compreende interações registradas entre os anos 1970 a 2001, sendo que a maioria remete aos anos 90, em países onde o português é a língua oficial, sendo estes: Portugal (30 textos), Brasil (20), Angola, Cabo Verde, Guiné-Bissau, Moçambique, São Tomé e Príncipe, Macau (5 cada), Goa e Timor-Leste (3 cada) totalizando 128.542 palavras. As 86 gravações que compõem o corpus são em formato *.wav* e as interações correspondem a entrevistas informais, entre pesquisadores e informantes, e (só em pequena parte) formais, como entrevistas em programas de rádio. Sem considerar as variedades asiáticas, que não foram abordadas nesta pesquisa, os textos apresentam, em sua totalidade, falantes com idades de 20 a 80 anos e uma leve desproporção entre vozes masculinas e femininas, sendo 43 vozes masculinas e 36 femininas. O recurso possui variação com relação aos temas das interações comunicativas, como mostra a tabela abaixo:

Tabela 1: Assuntos das gravações do corpus Português Falado (GONÇALVES; VELOSO, 2000).

Variedade	Arquivo	Assunto da gravação
Angola	A guerra e o ambiente	Guerra e fluxo migratório
Angola	Meninos de rua	Crianças de rua
Angola	O jovem Gaspar	Um dos informantes comenta sobre a sua infância
Angola	Um conto tradicional	Um conto tradicional na região que envolve lições sobre prudência e justiça

Brasil	Economia e sociedade	Sistema político
Brasil	Festa de estudante	Festas escolares
Brasil	O mundo do direito	Direitos civis, leis e costumes
Brasil	Surpresas da fotografia	Experiências com a fotografia
Brasil	Muito iguais e muito diferentes	As experiências de irmãs gêmeas
Cabo Verde	Raparigas de Cabo Verde	A vida das mulheres em Cabo Verde
Cabo Verde	As mornas	O significado de "mornas" no léxico de Cabo Verde
Cabo Verde	Ilha do Fogo	Os aspectos geográficos da Ilha do Fogo em Cabo Verde
Cabo Verde	Colher de Panela	Um ritual de nascimento chamado "Colher de Panela"
Cabo Verde	Coleccionismo	Coleções de borboletas
Guiné-Bissau	A juventude guineense	A juventude em Guiné-Bissau
Guiné-Bissau	A mulher africana	A vida e os desafios das mulheres africanas
Guiné-Bissau	A Sida	A situação da Aids no país
Guiné-Bissau	Democracia	Como é a democracia em Guiné-Bissau
Guiné-Bissau	O aborto clandestino	As complicações do aborto clandestino no país
Moçambique	A chuva	Relato de situação preocupante vivida em uma manhã chuvosa
Moçambique	Cantar e pintar ¹³	Música e pintura
Moçambique	Maternidade ¹⁴	Projeto de lei sobre licença maternidade
Moçambique	Meninice na Machamba	A infância na Machamba, uma região em Moçambique
Moçambique	Sentimento e desporto	Prática de desporto em Moçambique
São Tomé e Príncipe	A banda	A rotina de um grupo musical em Moçambique
São Tomé e Príncipe	Costureira de sucesso	Uma costureira bem sucedida conta sobre a sua vida
São Tomé e Príncipe	Sabores	A informante ensina uma receita famosa

¹³ Durante alguns minutos da gravação, o informante canta uma música.

¹⁴ A gravação é de uma participação em um programa de entrevista.

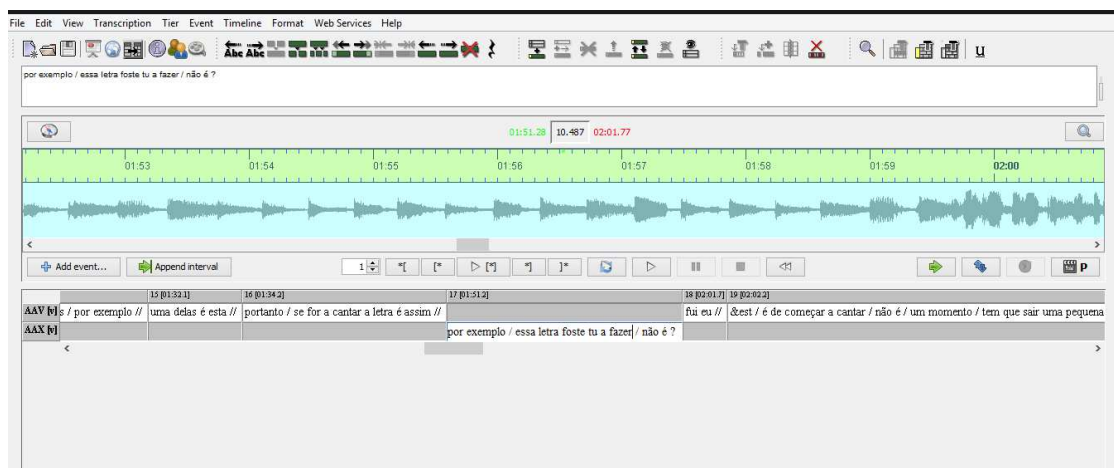
São Tomé e Príncipe	Ser professor	Os desafios e aprendizados da docência no país
Portugal	As grandes cidades	O informante comenta sobre as principais cidades de Portugal
Portugal	O quotidiano e a lei ¹⁵	Conselhos jurídicos
Portugal	Saber vender	Como saber vender em uma loja de presentes
Portugal	Vida de estudante	A vida dos estudantes em Portugal

Fonte: adaptado de Gonçalves e Veloso (2000)

Por outro lado, o corpus apresenta não apresenta variação do ponto de vista da tipologia interacional, já que todos os textos são de natureza monológica porque apresentam situações comunicativas nas quais o informante fala sem que o pesquisador interaja muito.

A transcrição apresenta anotação das quebras prosódicas terminais e não terminais, e é alinhada ao arquivo de som por meio do programa *Partitur Editor* do software EXMARaLDA (SCHMIDT; WÖRNER, 2009, http://www.exmaralda.org/en_downloads.html), sendo possível acessar o áudio e os textos simultaneamente, como mostra a figura abaixo:

Figura 1: Alinhamento no software EXMARaLDA.



A disponibilidade dos áudios alinhados aos textos das transcrições é uma característica altamente desejável em um recurso destinado ao estudo da fala e, infelizmente, pouco comum. Portanto, o corpus é uma ferramenta muito importante para estudar o português falado. No caso do recurso aqui utilizado, os textos monológicos são uma tipologia interacional pouco acional, portanto há menor interação entre o falante e o interlocutor. Nesse sentido, as

¹⁵ Essa gravação também é de uma entrevista em um programa de rádio.

estruturas sintáticas encontradas nos textos do corpus se assemelham àquelas realizadas na diamesia escrita.

Outra limitação encontrada foi o fato de o corpus apresentar pouco equilíbrio quanto ao número de palavras e, para contornar essa desproporção, o balanceamos com base no número de palavras. As seções do corpus das variedades africanas contêm cinco arquivos, enquanto as do Brasil possuem 20 e as de Portugal, 30. Para balancear essa desproporção, para estas seções com mais de cinco textos foi adotado um limite de 5500 a 6000 palavras, para que sejam semelhantes ao número de palavras das variedades africanas. Em alguns casos, foi necessário fazer um corte no final dos textos para respeitar essa definição. A tabela abaixo apresenta o número de palavras no corpus:

Tabela 2: Quantidade de palavras dos textos do corpus Português Falado.

Seção do corpus	Número de palavras
Brasil	5951 palavras
Portugal	5745 palavras
Angola	5994 palavras
Cabo Verde	5522 palavras
Guiné-Bissau	3329 palavras
Moçambique	4681 palavras
São Tomé e Príncipe	4443 palavras
Total	35665 palavras

Para a análise, apenas as palavras dos informantes foram contabilizadas, dado que os entrevistadores são todos de origem portuguesa. A tabela abaixo apresenta os metadados disponíveis para o corpus:

Tabela 3: Assuntos presentes nas gravações do corpus Português Falado (GONÇALVES; VELOSO, 2000).

Variedade	Arquivo	Ano	Tipo	Lugar	Sexo/ idade	Escolaridade	Profissão
Angola	A guerra e o ambiente	1997	monólogo-espontâneo	Luanda	M/40 anos	curso superior	Professor universitário

Angola	Meninos de rua	1997	monólogo-espontâneo	Luanda	M/27 anos	sem info.	Educador do centro infantil
Angola	O Ensino em Angola	1997	monólogo-espontâneo	Luanda	F/sem info.	curso superior	professora
Angola	O jovem Gaspar	1997	diálogo-espontâneo	Luanda	M/22 anos	frequência de um curso superior	empregado de mesa
Angola	Um conto tradicional	1997	monólogo-espontâneo	Luanda	M/sem info.	frequência de um curso superior	Estudante (usa um discurso muito formal)
Brasil	Economia e sociedade	1987	Diálogo-espontâneo	São Paulo	M/22 anos	frequência de um curso superior	estudante
Brasil	Festa de estudantes	1993	Diálogo-espontâneo	Araraquara	F/26 anos	Sem info.	estudante
Brasil	O mundo do direito	1995	Diálogo-espontâneo	Araraquara	F/20 anos	frequência de um curso superior	estudante
Brasil	Surpresas da fotografia	1993	Diálogo-espontâneo	São Paulo	F/23 anos	frequência de um curso superior	estudante
Brasil	Muito iguais e muito diferentes	1995	Diálogo-espontâneo	Araraquara	F/22 anos	frequência de um curso superior	estudante
Cabo Verde	Raparigas de Cabo Verde	1995	Diálogo-espontâneo	Praia (Santiago)	F/41 anos	frequência de um curso superior	Inspetora do magistério primário
Cabo Verde	As mornas	1995	Diálogo-espontâneo	São Vicente	M/54 anos	Ensino básico	marinheiro
Cabo Verde	Coleccionismo	1995	Diálogo-espontâneo	São Vicente	M/48 anos	Ensino médio	Tipógrafo
Cabo Verde	Colher de panela	1995	Diálogo-espontâneo	São Vicente	F/55 anos	Ensino médio	comerciante

Cabo Verde	Ilhas do fogo	1995	Diálogo-espontâneo	São Vicente	M/25 anos	frequência de um curso superior	estudante
Guiné-Bissau	A juventude guineense	1995	Diálogo-espontâneo	Bissau	F/19 anos	Ensino médio em curso	estudante
Guiné-Bissau	A mulher africana	1995	Diálogo-espontâneo	Bissau	F/45 anos	Curso superior	jornalista
Guiné-Bissau	A Sida	1995	monólogo-espontâneo	Bissau	F/17 anos	Ensino médio em curso	estudante
Guiné-Bissau	Democracia	1995	monólogo-espontâneo	Bissau	M/18 anos	Ensino médio em curso	estudante
Guiné-Bissau	O aborto clandestino	1995	monólogo-espontâneo	Bissau	F/16 anos	Ensino médio em curso	estudante
Moçambique	A chuva	1986	monólogo-espontâneo	sem info.	M/32 anos	sem info.	professor
Moçambique	Cantar e pintar	1983	monólogo-espontâneo	Maputo	M/sem info.	sem info.	sem info.
Moçambique	Maternidade	1997	Não-natural/planejada	Maputo	1M/1F - sem info.	Sem info.	Sem info.
Moçambique	Meninice na Machamba	1986	Monólogo-espontâneo	Maputo	M/21 anos	frequência de um curso superior	Sem info.
Moçambique	Sentimento e desporto	1997	Não natural/planejada	Sem info.	2F/25 e 26 anos	frequência de um curso superior	Professora e estudante
Portugal	As grandes cidades	1995	Diálogo-espontâneo	Porto	M/42 anos	Ensino médio	Secretário
Portugal	Bom senso e bom rosto	1996	Diálogo-espontâneo	Lisboa	F/55 anos	Ensino médio	vendedora
Portugal	O quotidiano e a lei	1989	planejada	Porto	M/25 anos	Curso superior	advogado

Portugal	Saber vender	1995	Diálogo/espontâneo	Lisboa	F/46 anos	Ensino médio	vendedora
Portugal	Vida de estudante	1995	Diálogo/espontâneo	Porto	F/47 anos	Ensino superior completo	farmacêutica
São Tomé e Príncipe	A banda	1996	Planejada-não espontânea	São Tomé	M/sem info.	Sem info.	Sem info.
São Tomé e Príncipe	Costureira de sucesso	1996	Diálogo-espontâneo	São Tomé	F/35 anos	Ensino médio	costureira
São Tomé e Príncipe	Pesca arriscada	1996	Monólogo-espontâneo	Praia Melão	M/41 anos	Sem info.	pescador
São Tomé e Príncipe	Sabores	1997	Diálogo-espontâneo	São Tomé	F/20 anos	4 anos de escolaridade	Empregada doméstica
São Tomé e Príncipe	Ser professor	1996	Diálogo-espontâneo	São Tomé	F/39 anos	Curso superior	Professora de ensino secundário

Fonte: Gonçalves e Veloso (2000).

Apesar das limitações, tentou-se alcançar um melhor balanceamento no que diz respeito a algumas variáveis. Em primeiro lugar, para as variedades com mais de cinco textos, os arquivos foram escolhidos buscando obter homogeneidade de faixa etária para cada país. Para as variedades que possuem apenas cinco gravações, não foi possível realizar a escolha da idade dos informantes. O recurso também apresenta, para Brasil e Portugal, registros feitos em diferentes diatópias e, para atingir uma consistência, foram selecionados apenas textos dos principais centros econômicos dos países, visto que a informação da origem do informante não é disponibilizada no corpus. Além disso, a maioria dos textos é dos anos 90 e, por esse motivo, foram selecionados mais textos dessa época.

Desse modo, a seleção dos textos foi feita com base nos seguintes critérios: número de palavras, idade dos falantes e o local da gravação. É importante ressaltar que, apesar de buscarmos o balanceamento de forma mais adequada e criteriosa possível, este foi feito de modo a permitir um maior grau de comparabilidade entre as seções, o que é valioso devido à abrangência da comparação dialetal.

O corpus é um recurso relevante para análise da fala das variedades do português, pois apresenta dados orais autênticos de variedades ainda pouco estudadas, além de transcrições e áudios disponibilizados gratuitamente e com alinhamento texto-som. Além disso, as

dimensões do corpus e a comparabilidade entre suas seções foram julgadas adequadas para esta pesquisa de mestrado.

3.2. TRATAMENTOS DOS DADOS

Após a seleção dos textos, os arquivos de alinhamento foram exportados no formato TextGrid do Praat, utilizando uma função específica do programa *Partitur Editor* do EXMARALda, pois a ferramenta possibilita a visualização do espectro e da curva da frequência fundamental, importantes para identificar as unidades informacionais durante o processo de etiquetagem. Cada arquivo de alinhamento foi analisado no Praat junto com a busca pelas orações completivas e adverbiais.

O plano metodológico inicial para o tratamento dos dados previa um passo a passo semelhante ao adotado para a compilação do corpus C-ORAL BRASIL (RASO; MELLO, 2012): uma revisão da segmentação prosódica de todo o corpus seria feita por uma equipe de quatro pesquisadoras treinadas (a autora desta dissertação, a coordenadora do projeto de estudo da fala de variedades diatópicas do português em que esta pesquisa de mestrado se insere, e mais duas estudantes de graduação). Pretendia-se fazer, antes da revisão, um teste Kappa (FLEISS, 1971) para verificar o coeficiente de concordância entre as pesquisadoras na delimitação de enunciados e de unidades tonais e, portanto, a confiabilidade da segmentação. Após essa etapa de revisão da segmentação, seria realizada a etiquetagem informacional de enunciados e estrofes contendo as orações em estudo.

Contudo, devido ao período de pandemia e às limitações advindas, foi necessário reformular a metodologia adotada. Assim, seguiu-se com o treinamento remoto de etiquetagem junto à orientadora, mas apenas os enunciados contendo as orações subordinadas foram etiquetados, utilizando como aporte teórico a *Language into Act-Theory* (CRESTI; MONEGLIA, 2005, cf. seção 2.4). Na etapa final, as construções subordinadas foram classificadas em objetivas, interrogativas indiretas e subjetivas para as completivas; e as adverbiais de acordo com o valor semântico da conjunção (tempo, causa, condição, etc.).

O corpus Português Falado segue o mesmo modelo de referência e segmentação do corpus C-ORAL ROM. Porém, por já termos usado corpora que seguem os critérios mais atualizados do C-ORAL BRASIL em trabalhos anteriores, optamos por seguir essa adaptação no recurso aqui utilizado. Para tanto, em primeiro lugar, foi feita uma revisão dos critérios de transcrição, adaptando-se os símbolos que diferem daqueles usados no corpus que serviu como referência: marcações adotadas para reformulações de conteúdo locutivo - *retractings*

([///]), enunciados com interrogações (?) e pausas preenchidas (&eh, &ah). A tabela abaixo compara os critérios de etiquetagem dos dois corpora:

Tabela 4: Comparação entre os símbolos usados nos dois corpora.

Símbolo	Corpus Português Falado	Corpus C-ORAL BRASIL
//	quebra prosódica terminal (exceto para interrogação e interrupção)	quebra prosódica terminal (usada também para interrogação)
?	usado para enunciados interrogativos	o enunciados interrogativos são transcritos com o símbolo //
/	quebra prosódica não terminal, usada quando o enunciado não foi concluído	quebra prosódica não terminal, usada quando o enunciado não foi concluído
...	indica enunciado suspensivo	não é usado no corpus
&ah, &eh, &hum	hesitação fonética	é usado apenas &he para transcrever hesitação
+	enunciado interrompido	enunciado interrompido
<>	no corpus, a sobreposição é marcada apenas no arquivo de alinhamento	sobreposição entre falantes
xxx	quando uma palavra não é inteligível	quando uma palavra não é inteligível
yyyy	quando uma sequência de palavras não é inteligível	quando uma sequência de palavras não é inteligível
&	usado para palavras incompletas	usado para palavras incompletas
[/]	repetição de uma única palavra	repetição de uma única palavra
<text> [/]	repetição de uma sequência de palavras	o mesmo símbolo [/] é usado
[//]	repetição com mudança de uma palavra	o mesmo símbolo [/] é usado
<text> [//]	repetição com mudança de uma sequência de palavras	o mesmo símbolo [/] é usado
[///]	reformulação total do enunciado	o mesmo símbolo [/] é usado
hhh	ruído paralinguístico	ruído paralinguístico
“ “	discurso direto	o símbolo usado é _r

Segue abaixo a demonstração, em um texto da seção angolana, de como foi feita a revisão da segmentação prosódica e a adaptação da transcrição:

(1) segmentação original

primeiro / porque < é uma zona semi > [//] são &á / zonas semi-áridas / não é //

(2) após a revisão e segmentação

primeiro / porque é uma zona semi [/4] são &á [/1] zonas semi-áridas / não é //

No exemplo (1), a transcrição original contém símbolos que não são usados segundo os critérios do corpus C-ORAL-BRASIL, como [//] para indicar cancelamento das palavras, o uso de <> para indicar palavras repetidas pelo falante e ausência da marcação de *retracting*. Na adaptação em (2), a correção foi feita indicando, respectivamente, o número de palavras canceladas, retirando a marcação <> e colocando a marcação de retracting em &á. Além disso, após ouvir os enunciados foi acrescentada a etiquetagem com as devidas unidades informacionais reconhecidas.

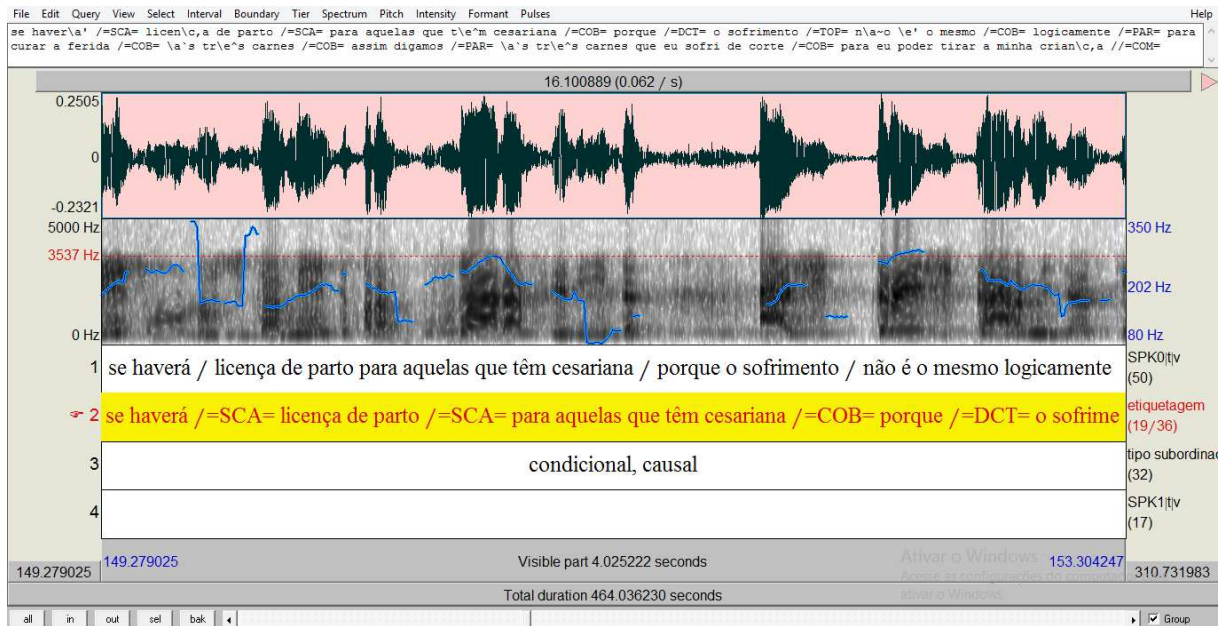
Em relação à etiquetagem, é importante ressaltar que, em alguns casos, foi difícil realizá-la devido a questões como a velocidade de fala, entonações menos familiares às etiquetadoras, fala muito escandida e até a ilegibilidade do espectro:

(3) se haverá /=SCA= licença de parto /=SCA= para aquelas que têm cesariana /=COB= porque /=DCT= o sofrimento /=TOP= não é o mesmo /=COB= logicamente /=PAR= para curar a ferida /=COB= as três carnes /=COB= assim digamos /=PAR= às três carnes que eu sofri de corte /=COB= para eu poder tirar a minha criança //COM= (PM)

(4) como /=DCT_d= uma pessoa assim de dezasseis anos /=PAR= se os pais acharem que /=SCA= ela é muito nova /=SCA= para dar luz àquele bebê /=TOP= então vão a uma clínica /=COB= legalmente ou hospital /=PAR= fazer o aborto //COM= (PGB)

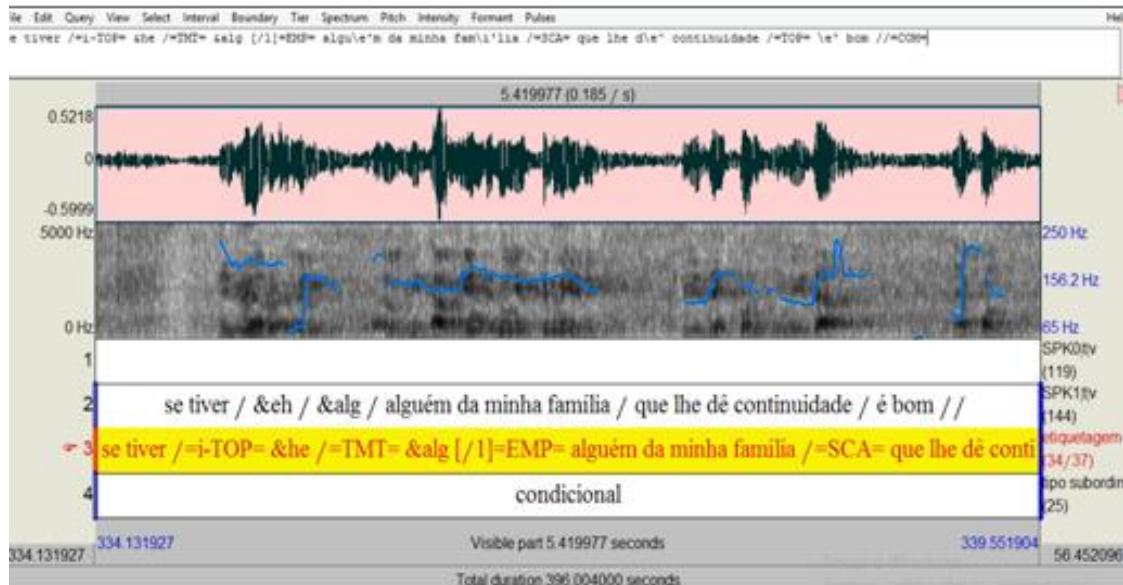
Apresentamos também o espectro do exemplo (3);

Figura 2: Visualização do enunciado no Praat.



Em (3), a etiquetagem gerou dúvidas tanto pela oitiva quanto pelo espectro. Do mesmo modo, houve casos como em (4), que, por razões prosódicas, a etiquetagem foi particularmente difícil e optamos por incluir uma marcação (TAG_d) para sinalizar a pouca confiabilidade da etiquetagem. Assim, em alguns casos etiquetamos como DCT e em outros, como AUX. A etiquetagem foi feita apenas nas sequências terminadas (i.e. enunciados e estrofes) contendo as estruturas de subordinação em estudo, devido à impossibilidade de fazê-la no corpus inteiro durante o tempo de realização da pesquisa. Após a etiquetagem, as ocorrências foram divididas conforme a articulação informacional em linearizadas (subordinação dentro da mesma unidade informacional) e padronizadas (ao longo de mais de uma unidade informacional). Em seguida, os dados foram organizados e analisados com foco nos padrões informacionais nos quais as orações ocorrem. A Figura 3 abaixo mostra o processo final das etapas metodológicas:

Figura 3: Processo metodológico final no Praat.



A Figura 3 apresenta que foi adicionada uma camada no arquivo para revisar a segmentação e outra para acrescentar a etiquetagem informacional, buscando adequar possíveis erros na marcação de quebras prosódicas, enunciados interrompidos e *retractings* dos falantes.

Por fim, as ocorrências foram comparadas entre as variedades do português com atenção à tipologia de padrões informacionais e à proporção de padronização e linearização das orações completivas e adverbiais. Foi realizado um teste estatístico a fim de validar os resultados obtidos, e um modelo linear, por meio do programa R, para verificar como os dados estão distribuídos e validar a correlação estatística.

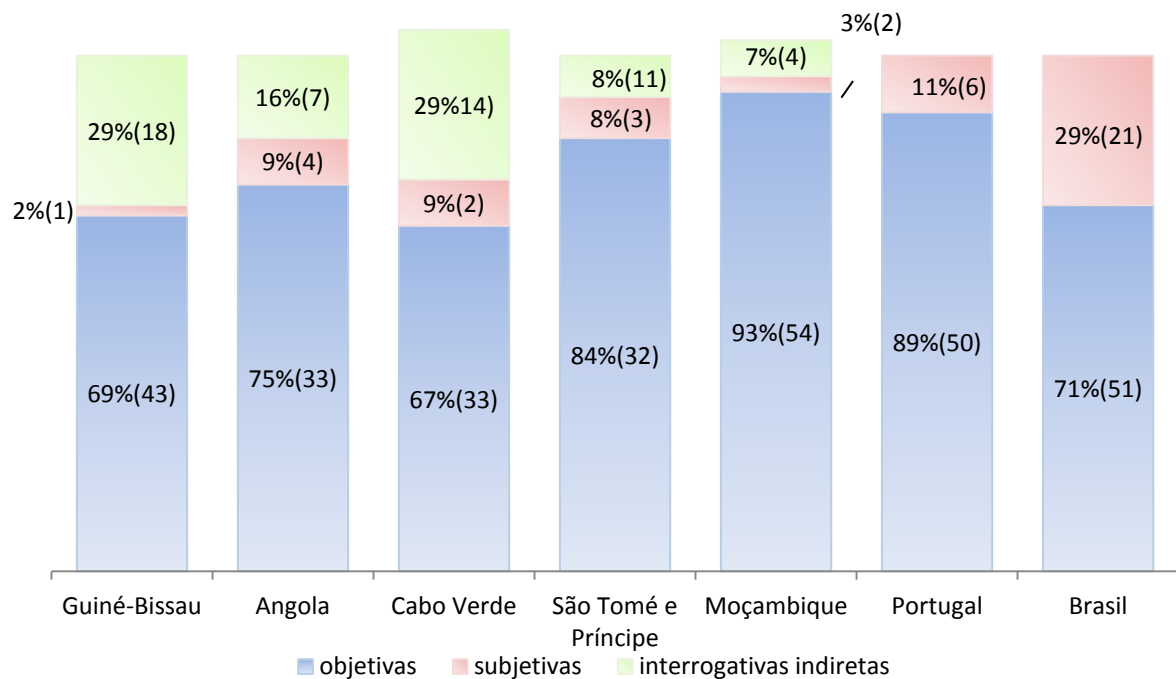
4. ANÁLISE E DISCUSSÃO DOS DADOS

4.1 AS ORAÇÕES COMPLETIVAS

4.1.1 TIPOLOGIA E COMPLEMENTIZADORES

As orações completivas apresentaram organização semelhante na tipologia nas variedades, como mostra o gráfico abaixo:

Gráfico 1: Tipologia de completivas no corpus.



O resultado apontou um total de 296 objetivas, 81 interrogativas indiretas e 12 subjetivas. A seguir, alguns exemplos de ocorrências:

(1) eu acho que ela até teria conseguido a medalha de prata //COM= (M)

(2) e /=SCA= &v [/1]=EMP= nota-se de avião /=COM= **quem chega** /=APC= não é //AUX= (padrão COM/APC) (A)

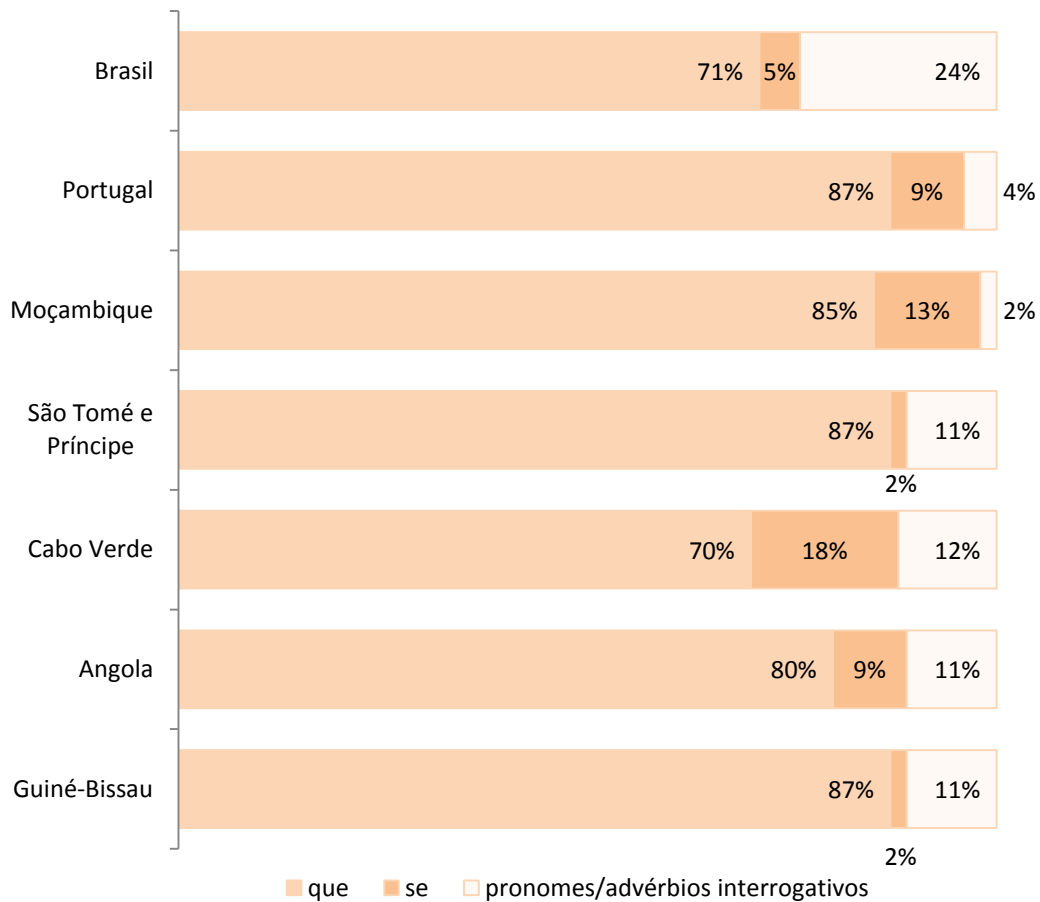
(3) eu não sei se é verdade /=CMM= **ou se é mentira** //CMM= (padrão CMM/CMM) (CV)

(4) ela falou assim que o escritório dela é carpetado /=COB= é lindinho /=COB= e que vai aqueles peão /=COB= de chinelo /=COB= e tudo sujo de barro /=COB= sabe //COM= &he

/=TMT= e **que eles vão lá assim** /=COB= &e [/1]=EMP= &o [/1]=EMP= as &pe [/2]=EMP= eles /=TOP= têm muito medo do advogado //COM= (BR)

O exemplo em (1) apresenta uma completiva objetiva linearizada, na qual a principal, o complementizador e a subordinada ocorrem na mesma unidade informacional. Diferentemente, em (2) a linearização também ocorre, porém a principal *nota-se de avião* está Comentário e a subjetiva *quem chega* é realizada no Apêndice de Comentário. Nesse caso, a estrutura é linearizada, pois o APC tem como função completar e integrar a unidade de COM. Portanto, trata-se do mesmo nível informacional. Já em (3), a interrogativa indireta é distribuída entre dois Comentários Múltiplos e está em relação de coordenação alternativa. Há, ainda, em (4), outra distribuição das orações em outro nível, para além das unidades informacionais. Nessa configuração, a matriz está em um enunciado e a subordinada, em outro.

Em relação aos complementizadores, os mais frequentes são *que* (297 ocorrências), advérbios e pronomes interrogativos, como *quem, o que, como*, etc (43 ocorrências); e *se* (35 ocorrências):

Gráfico 2: Tipos de complementizadores.

O complementizador *que* é o mais frequente em todas as variedades, *se* e os pronomes interrogativos alternam como a segunda maior ocorrência, o que está relacionado à estrutura sintática das orações subordinadas em termos de itens lexicais. Vejamos algumas ocorrências linearizadas das tipologias de subordinação:

(5) isso acho que não tem razão de ser //COM= (PT)

(6) às vezes nem as pessoas &sa [1]=SCA= sabem o que é que é a cultura //COM= (GB)

(7) conhecer como é que eles vão entrar nesse meio //COM= (GB)

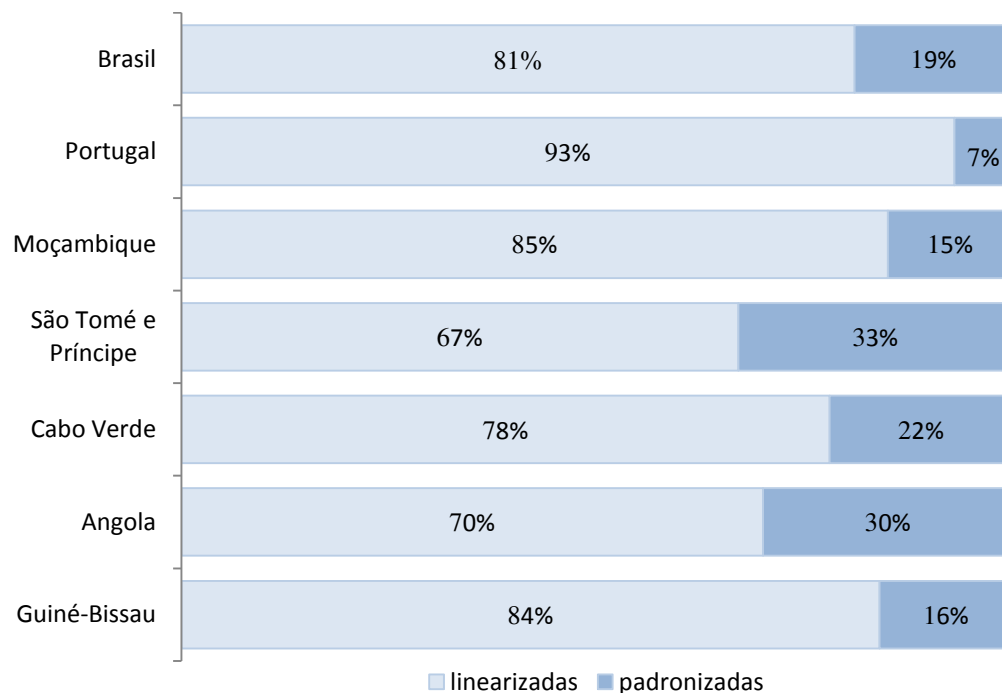
Nos três exemplos, as orações subordinadas são realizadas na mesma configuração: a principal e a completiva ocorrem na mesma unidade prosódica.

4.1.2 LINEARIZAÇÃO VS. PADRONIZAÇÃO

Orações completivas foram encontradas em 389 enunciados/estrofes, sendo 81% linearizadas (313 enunciados) e 19% padronizadas (76). Em sua maioria, portanto, as

completivas ocorrem junto de sua principal dentro da mesma unidade informacional, sendo preservada a relação de dependência sintática. Nesse corpus, a padronização é motivada pela tipologia interacional uma vez que nos monólogos apenas um falante detém o turno de fala. Assim, devido a essa característica própria do monólogo, as estruturas sintáticas apresentadas são mais complexas, porém os padrões informacionais são menos diferenciados por se tratar de longos turnos de fala construídos com unidades informacionais características desse tipo de texto. Além disso, por se tratar monólogos, os textos apresentam pouca acionalidade e há pouca, ou nenhuma, interação entre o falante e o interlocutor, o que torna a padronização menos significativa já que é motivada pela tipologia interacional. Desse modo, os monólogos do corpus apresentam uma fala espontânea menos acional. O gráfico abaixo apresenta a ocorrência da articulação informacional dos dados:

Gráfico 3: Linearização e padronização nas completivas do corpus Português Falado.



Consideramos também como padronização, além dos casos de não linearização, a ocorrência de completivas entre enunciados, isto é, que ultrapassam as fronteiras prosódicas. É interessante observar que estudos anteriores (BOSSAGLIA, 2014; SILVA; BOSSAGLIA, 2019) observaram a mesma preferência das completivas pela configuração linearizada, uma vez que no italiano, as linearizadas somaram 67% das ocorrências e no inglês americano, 90%.

Os dados mostram que, em algumas variedades, a taxa de padronização é maior, o que pode estar relacionado com a tipologia interacional monológica, que admite estruturas textualmente e, por conseguinte, sintaticamente mais complexas¹⁶. É interessante observar que, cotejando com dados anteriores (SILVA, 2019), a articulação informacional é similar em dados do português brasileiro retirados de um corpus diferente, o C-ORAL BRASIL (RASO; MELLO, 2012), porém, com menor proporção de padronizadas, sendo 34 (16%) ocorrências de 211 enunciados. Como padronização, foi considerada também a ocorrência de completivas entre enunciados, isto é, que ultrapassam as fronteiras prosódicas. Apesar de não ser o propósito deste trabalho, seria interessante fazer algumas comparações com outras línguas. No italiano, dos 165 enunciados, 54 (33%) são padronizadas e no inglês americano, são apenas 19 (10%) de 197 enunciados. É possível notar que mesmo em línguas diferentes, é mantida a tendência de as completivas ocorrerem, em sua maioria, na mesma unidade informacional.

No caso das variedades aqui examinadas, a linearização ocorre na maioria dos dados, o que pode estar relacionado ao fato de que as completivas estão mais integradas à oração principal funcionando como argumento. Desse modo, estabelecem uma relação de dependência com a principal (GIVÓN, 1980; NOONAN, 1985; DIK, 1997; HENGEVELD; MACKENZIE, 2008) e tendem a ocorrer dentro da mesma unidade informacional da matriz. Em seguida, fornecem-se exemplos de padronizadas:

(8) bem /=AUX= &he /=TMT= dizem /=COB= **que** /=SCA= **nós todos temos um pouco de** /=SCA= **coleccionistas** //COM= (padrão COB/COM)¹⁷ (CV)

(9) mas de qualquer forma /=TOP= convém também recordar /=COB= **que há certos cursos que nunca tivemos aqui no período colonial** //COM= (padrão COB/COM) (A)

(10) o que não quer dizer que alguns angolanos /=TOP= das mais variadas origens /=APT= **não tenham também apanhado já esses estudos universitários** //COM= (padrão TOP/APT/COM) (A)

(11) ah /=AUX= também servia para quando /=i-COB= enfim /=AUX= &he /=TMT= se matava porcos /=COB= salgar /=COB= e era muito bom /=COB= porque /=AUX= esse

¹⁶ Têm sido levantadas hipóteses (BOSSAGLIA, 2016), ainda não suficientemente exploradas, que relacionam a taxa de padronização a características rítmicas das línguas. Nesse sentido, haveria isocronismo entre as sílabas tônicas, o que leva a uma maior quantidade de material fonológico na unidade tonal. Sob esse ponto de vista, línguas de ritmo mais acentual, como o português europeu, teria maior taxa de padronização do que o português brasileiro - o que está alinhado com a ideia de Bossaglia (2016).

¹⁷ Ao lado dos exemplos, fornecemos uma descrição do padrão apresentado no enunciado.

líquido /=TOP= &sa [/1]=EMP= essa água /=SCA= salgada que /=SCA= ficava /=PAR= era bom que ficasse aí na +=COB= (CV)

Os exemplos em (8) e (9) são semelhantes, visto que em ambos o complementizador e o predicado da completiva estão em unidades informacionais diferentes, entre COB e COM. Já em (10), a principal *o que não quer dizer* e o complementizador ocorrem em um TOP complexo (ou seja, integrado por um Apêndice de Tópico) junto com o tema, que é o sintagma nominal *alguns angolanos*, e o predicado da subordinada é realizado no COM, com o Apêndice de Tópico completando o TOP, um padrão com apenas uma ocorrência nos dados. Os casos como em (11), isto é, ocorrências de enunciados interrompidos, também foram contabilizadas, mas apenas nos casos em que, apesar da interrupção, houvesse suficiente material linguístico para analisar as estruturas de subordinação em estudo.

Os casos em que as completivas são realizadas sem complementizador também foram contabilizados dentro da categoria *que*, pois corresponde a uma possibilidade de que o item lexical fosse utilizado. Nesse caso, os resultados mostraram que em algumas variedades do português é comum a omissão, ao contrário do que os dados do português brasileiro do corpus C-ORAL BRASIL (RASO; MELLO, 2012) apontaram sobre a mesma língua. Os exemplos abaixo mostram algumas ocorrências não canônicas apresentadas nos dados:

(12) e dali /=i-COB= prontos /=AUX= vimos pelas horas /=SCA= **era um pouco tarde** /=COB= evacuamos a [/2]=EMP= levamos até `a Quinta Avenida //COM= (padrão COB/COM) (A)

(13) não há nenhum filho que /=SCA= pode dizer nos vai /=SCA= **continuar** /=SCA= **o meu trabalho** //COM= (CV)

(14) os rapazes não [/1]=SCA= não fazem mais nada //COM=

eu acho que não /=COB= a não ser <no interior> /=COB= no interior /=PAR= na zona rural /=PAR= têm [/1]=EMP= já /=SCA= vão dar de beber às cabras /=COB= vão atirar palha para os animais /=COB= fazem mais uma coisinha //COM= (CV)

(15) **claro que ao retirar o coberto vegetal** /=TOP= **eu estou imediatamente** /=SCA= **a provocar** /=SCA= **a situação do surgimento de ravinas** //COM= (padrão TOP/COM) (A)

Os exemplos (12) e (13) mostram ocorrências de completivas linearizadas sem complementizador realizadas na unidade de Comentário. Já em (14), há um exemplo de uma

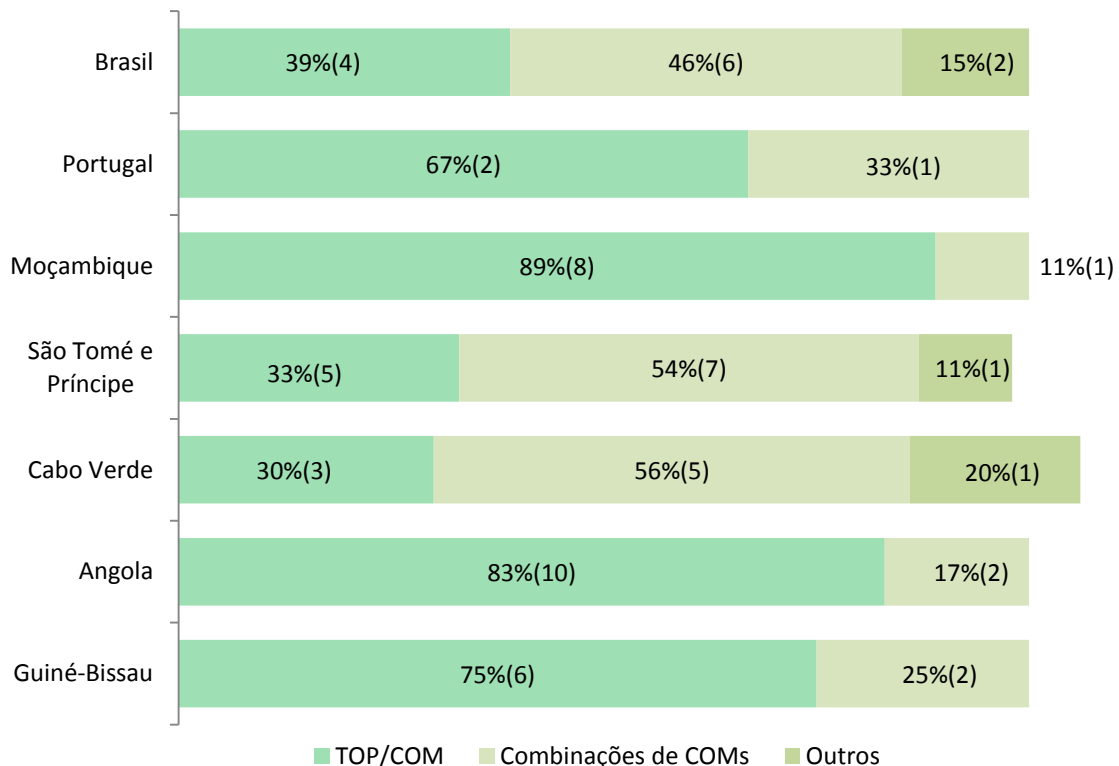
completiva que não apresenta predicado e é uma resposta à pergunta da outra informante. A ocorrência em (15) apresenta a gramaticalização da principal “(é) claro que” com o complementizador no Tópico, e a circunstancial *ao retirar o coberto vegetal* encaixada como tema em TOP, mais o predicado da completiva no COM. Casos como esse, evidenciam que no padrão TOP/COM as principais se reduzem a marcadores de modalidade epistêmica, enquanto as completivas exercem a função pragmática e informacionalmente mais relevante, aquela de veicular a ilocução. O caso extremo disso são construções como as de "claro que", bastante frequentes nos dados, em que a principal se gramaticalizou, perdendo a cópula. Foram poucos (7 ocorrências) os exemplos não canônicos, porém é igualmente interessante observá-los uma vez que fogem à ocorrência comum dessa tipologia. Além disso, os resultados da ausência do complementizador foram encontrados apenas para as completivas objetivas introduzidas por *que*.

4.1.3 OS PADRÕES INFORMACIONAIS

4.1.3.1 TÓPICO-COMENTÁRIO (TOP/COM)

Embora sejam poucas, as ocorrências de orações padronizadas evidenciam a ocorrência de padrões consistentes em todas as variedades:

Gráfico 4: Os padrões informacionais.



O padrão informacional mais comum é o TOP/COM e suas variantes textualmente mais complexas¹⁸, seguido das combinações de unidades ilocucionárias e outras ocorrências¹⁹. Segundo Mittmann (2012), o Tópico-Comentário é o principal padrão informacional da fala e os dados aqui analisados seguem essa observação. Nessa configuração, as unidades estão em relação de *aboutness* pragmática, na qual o TOP fornece um domínio semântico para a ilocução que está no COM:

(1) porque /=DCT= eu /=TOP= no [/1]=EMP= no meu ponto de vista /=TOP= **eu acho que a democracia na Guiné-Bissau** /=TOP= não devia ser uma democracia importada //COM= (padrão TOP/COM) (GB)

(2) e /=DCT= **acho que** /=SCA= neste momento /=TOP= **as pessoas estão** /=SCA= **fazendo um bocado só de rotina** /=COB= sem saber efectivamente o que é que deve fazer /=COB= o que é que pode fazer /=COB= o que é que /=SCA= não pode fazer /=COB= e [/1]=EMP= e

¹⁸ Algumas variantes complexas de TOP/COM foram observadas nos dados, porém com poucas ocorrências, sendo estas: COB/INT/TOP/COM, INT/TOP/COM, INT/COB, TOP/PAR,

¹⁹ Outros padrões, em menor número, foram encontrados, como: INT/COB, COM/APC e PAR/COM.

que /=SCA= contornos dar para /=SCA= poder resolver minimamente a situação //COM= (padrão TOP/COM) (ST)

(3) eu [/1]=EMP= eu [/1]=EMP= eu acho que hoje em dia /=TOP= com a proliferação de doutores que a gente tem /=TOP= &he /=TMT= **tende a projectar para o passado** esses nossos [/2]=SCA= **essas nossas maneiras de observar** //COM= (padrão TOP/COM) (A)

(4) aconteceu que essa família /=TOP= **no seio desta família** /=APT= **não havia** /=SCA= **ninguém** /=SCA= **que possuía** /=SCA= **a pulseira recomendada pelo** /=SCA= **médico tradicional** //COM= (padrão TOP/COM) (A)

(5) eu acho que /=SCA= **mesmo a nível de moda** /=TOP= **e de decoração** /=TOP= **as pessoas estão a voltar um bocado aos anos cinquenta** //COM= (padrão TOP/COM) (PT)

Os exemplos (1-5) mostram ocorrências do padrão *Theme in Topic* (BOSSAGLIA; MELLO, 2016; CRESTI, 2014), uma construção usada pelo falante para focalizar um tema (conteúdo) específico no TOP. Nessa distribuição, o núcleo da regência sintática é reduzido à marca de modalidade epistêmica, parte da completiva é tematizada/focalizada no Tópico, e o resto da completiva veicula a ilocução. Em (1), o tema é o sintagma preposicional *na Guiné-Bissau* assim como *neste momento* em (2), *hoje em dia* em (3), *essa família* em (4) e *a nível de moda* em (5). Nesses casos, a principal ocorre no TOP junto com o complementizador e o predicado é realizado no COM e todas as ocorrências deste padrão foram realizadas nesta construção. Em outros casos, as ocorrências apontam para uma função comunicativa da completiva:

(6) sim /=COB= mas eu acho que noutros países /=TOP= por mais que os jogadores percam /=TOP= **mas são recebidos com um pouco** de [/1]=SCA= **de carinho** /=COM= não é //AUX= (padrão TOP/COM) (M)

(7) bom /=AUX= eu creio que os jovens agora /=TOP= **têm mais** /=SCA= **acesso à** /=SCA= **informação** /=COB= e têm mais acesso /=SCA= a poderem /=i-COB= &he /=TMT= que [/1]=EMP= &he /=TMT= como é que se diz /=PAR= &he /=TMT= se cultivarem /=COB= a eles próprios /=APC= sobre /=SCA= tal //COM= (padrão TOP/COM) (GB)

Como já observado por Thompson (2002) e Silva e Bossaglia (2019) nos dados do inglês americano, algumas construções das orações completivas assumem uma função modalizadora. Nesse sentido, a matriz é "esvaziada" da sua função, podendo ser retirada sem que o enunciado perca o sentido. Nessas estruturas, o falante assume um posicionamento incerto sobre aquilo que diz e, para isso, usa uma marca de modalidade epistêmica. Em (6), o falante não tem certeza se em outros países os jogadores são mais bem recebidos e em (7) prefere não afirmar certamente que os jovens têm mais acesso à educação. Portanto, nos exemplos é possível notar que se a oração principal fosse retirada não afetaria a comunicação da mensagem. Assim, *acho que* e *creio que* funcionam como modalizadores epistêmicos e todas as construções ocorrem em primeira pessoa do singular. Diferentemente de *dizem que* que, como vimos, também associa o posicionamento do falante em relação ao conteúdo locutivo e de certa forma se relaciona à modalidade epistêmica, porém em experiências evidenciais. Portanto, a principal funciona como modalizadora e, em geral, é reduzida foneticamente, se "afastando" de certa forma da sua função gramatical. Vejamos outros dados:

(8) é quem não sabe fazer izaquente /=SCA= de azeite /=TOP= então dizem que aqui em São Tomé /=TOP= dizem que izaquente de açúcar /=TOP= é izaquente de frigida //COM= (padrões TOP/TOP e TOP/COM) (ST)

No exemplo, os dois Tópicos definem de maneira cada vez mais específica o domínio da mesma ilocução no Comentário. Nesse sentido, os dois temas (*izaquene de açúcar* e *izaquente de frigida*) nos dois TOPs poderiam ser considerados como partes da mesma completiva (*dizem que aqui em São Tomé, izaquente de açúcar é izaquente de frigida*). Além disso, o verbo *dizer* é usado como marca de evidencialidade, utilizada estrategicamente pelo falante para indicar ao interlocutor que não quer/pode identificar a fonte da informação e/ou que aquela informação não partiu dele (GALVÃO, 2001)²⁰. Sintaticamente, apesar de ser um verbo *dicendi*, que prototipicamente introduz orações completivas, no exemplo há um esvaziamento semântico, de modo que corresponde a “comenta-se que/estão falando que” no uso coloquial. Logo, a construção usada pelo falante não está no nível sintático, e sim semântico. Essa conclusão corrobora, novamente, o fato de as orações principais das completivas serem “gramaticalizadas” na fala.

²⁰ Galvão (2001) desenvolve um extenso trabalho descrito sobre a evidencialidade no uso da expressão *dizer que*.

4.1.3.2 COMBINAÇÕES DE COMENTÁRIOS

Em alguns casos, na distribuição dos itens lexicais, uma única oração principal pode introduzir mais de uma oração completiva, em estruturas de coordenação umas com as outras:

(9) eu não sei /=SCA= **se é por falta de finanças** /=COB= <ou é> [/2]=EMP= <ou é> [/2]=EMP= **se é por falta de** /=SCA= **cultura no seu eu** /=COB= ou se é hhh /3=SCA= **ou não sei** &p [/1]=SCA= **o que é que é esse fenômeno** //COM= (padrão COB/COM) (CV)

(10) eu não sei **se é verdade** /=CMM= **ou se é mentira** //CMM= (padrão CMM/CMM) (CV)

Em (9), há uma oração interrogativa linearizada *eu não sei se é por falta de finanças* que é a principal de outras duas subordinadas: a primeira *se é por falta de cultura no seu eu* no COB e a segunda *ou não sei o que é que é esse fenômeno* no COM que, por sua vez, é uma coordenada alternativa da oração anterior. Em (10), há uma construção semelhante: a primeira oração *eu não sei se é verdade* está em linearização e o verbo *saber* introduz outra completiva (*ou se é mentira*) em padronização que também está em relação alternativa com a primeira apódose entre CMMs. Além da análise sintática, a prosódia também é importante para reconhecer os níveis de dependência:

(11) e /=DCT= também o [/1]=SCA= os pais /=SCA= nas zonas piscatórias /=TOP= as mulheres /=i-COB= mesmo os rapazes /=PAR= também não iam /2=SCA= as raparigas não iam muito /=SCA= para [/1]=EMP= para a escola /=COB= porque diziam que /=INT= enfim /=AUX= &he /=TMT= **somos pescadores** //COM_r= sem escola /=CMM_r= com escola //CMM_r= que seria &nor [/2]=EMP= portanto era normal /=COB= **que não valia a pena** //COM=

No grupo de três sequências terminadas em (11), a oração *diziam que* realizada na unidade de Introdutor Locutivo da primeira estrofe introduz, apesar da presença do complementizador *que*, o discurso direto *somos pescadores* no COM final. De fato, não só escutando o áudio, fica evidente que *somos pescadores* não é uma oração objetiva, mas o tempo presente do indicativo sinaliza também que o sujeito desse predicado não é correferencial ao sujeito da principal. Por outro lado, *diziam que* poderia também ser analisada como principal da completiva *que não valia a pena*, que está no último COM da última

estrofe. Nesse caso, não se trata de um discurso direto e a *consecutio temporum* também aponta para isso. Desse modo, é o falante quem conta o que foi dito por outra pessoa (*diziam que não valia a pena*), mantendo, assim, a relação sintático-semântica entre a matriz e a subordinada. Em outros casos, a prosódia ajuda a reconhecer determinadas tipologias de subordinação:

(12) depois / no / dia seguinte / fizemos uma palestra / na Vila da Igreja / Mosteiros / sobre o vulcanismo // e logo / às / doze horas / aconteceu a erupção // e /=DCT= a gente /=SCA= acharam que nós já sabíamos que ia acontecer /=SCA= a erupção xxx // =COM= até inclusive um [/1]=SCA= um indivíduo **perguntou-me** /=COB= **quando** é que vai haver uma actividade vulcânica na ilha do Fogo // =COM= eu disse /=INT= o vulcão da ilha do Fogo /=TOP_r= pelo que eu sei /=PAR_r= é activo // =COM_r= (padrão COB/COM) (CV)

O exemplo apresenta duas possíveis interpretações dependendo da perspectiva de análise. No Comentário, é realizado um discurso direto introduzido por *quando é que vai haver uma actividade vulcânica na Ilha do Fogo*, porém com prosódia de discurso indireto. No entanto, apesar da entonação, a *consecutio temporum* indica que o discurso é direto, visto que do ponto de vista morfossintático, a construção seria *perguntou-me quando haveria uma actividade vulcânica*.

Nos resultados encontrados, o padrão de combinação de Comentários também foi recorrente:

(13) e eu achava que havia um certo vazio /=COB= que as pessoas que [/4]=EMP= **que era assim um bocado falso** /=COB= coisa muito plástica // =COM= (padrão COB/COB) (PT)

(14) para /=SCA= dizer o que é que quer dizer a sida /=COB= **o que é que a sida faz** /=COB= a sida é uma doença que ainda não tem cura // =COM= (padrão COB/COB) (GB)

(15) mas de qualquer forma /=TOP= convém também recordar /=COB= **que há certos cursos que nunca tivemos aqui no período colonial** // =COM= (padrão COB/COM) (A)

Em (13), a oração principal *e eu achava* é realizada no COB e o complementizador mais a completiva *que era assim um bocado falso* em outro COB; e o mesmo acontece em (14). Já em (15), a matriz *convém também recordar* ocorre no COB e a subordinada *que há certos cursos que nunca tivemos aqui no período colonial* está no COM. Apesar de

predominarem os padrões TOP/COM e as combinações de COMs, outras distribuições foram encontradas em raras ocorrências, como o exemplo abaixo:

(16) é preciso que /=INT= **toda a gente** /=TOP= **a todos os níveis** /=TOP= cada um a seu nível /=PAR= efectivamente /=PAR= **saiba aquilo que deve fazer** /=COB= e /=DCT= de que é que dispõe para fazê-lo //COM= (padrão TOP/COM) (ST)

No exemplo (16), a completiva subjetiva *é preciso*, mais o complementizador *que* estão no Introdutor Locutivo, e os temas *toda a gente*, *a todos os níveis* estão no Tópico, e a principal é realizada no Comentário Ligado.

4.1.3.3 COMPLETIVAS ENTRE ENUNCIADOS

Em alguns casos, as completivas mantêm uma relação com a matriz para além do nível prosódico:

(17) quer dizer que não só foi uma universidade tardia /=COB= como limitada /=COB= e /=DCT= com muitas reservas desde [1]=SCA= desde o início //COM= e que &l [3]=EMP= **e que nunca chegou a tomar** /=SCA= **desenvolvimento pleno antes da independência** /=COM= não \e' //AUX= (padrão COB/COM) (A)

(18) eu acho que sim //COM= **que com o tempo que o homem vai** /=SCA= **criando hábitos e vai mudando** //COM= (PT)

As completivas podem ultrapassar as fronteiras dos enunciados e serem realizadas para além do escopo de dependência sintática tradicional. Em (17), o verbo *dizer* introduz duas subordinadas que estão coordenadas entre si (*quer dizer que não só foi uma universidade tardia como limitada e com muitas reservas desde o início e que nunca chegou a tomar desenvolvimento pleno*). Desse modo, as estruturas mantêm, ao mesmo tempo, uma relação de coordenação, sinalizada pela conjunção aditiva *e*, e subordinação, veiculada pela dependência entre o verbo *dizer* no primeiro enunciado e o predicado no segundo. De modo semelhante, *eu acho* em (18), por um lado é realizado junto com o complemento *que sim* no primeiro enunciado, por outro parece introduzir as duas completivas coordenadas (*que com o tempo que o homem vai criando hábitos e vai mudando*) do enunciado seguinte. Embora seja possível reconhecer uma relação sintática entre as estruturas, observamos que as completivas

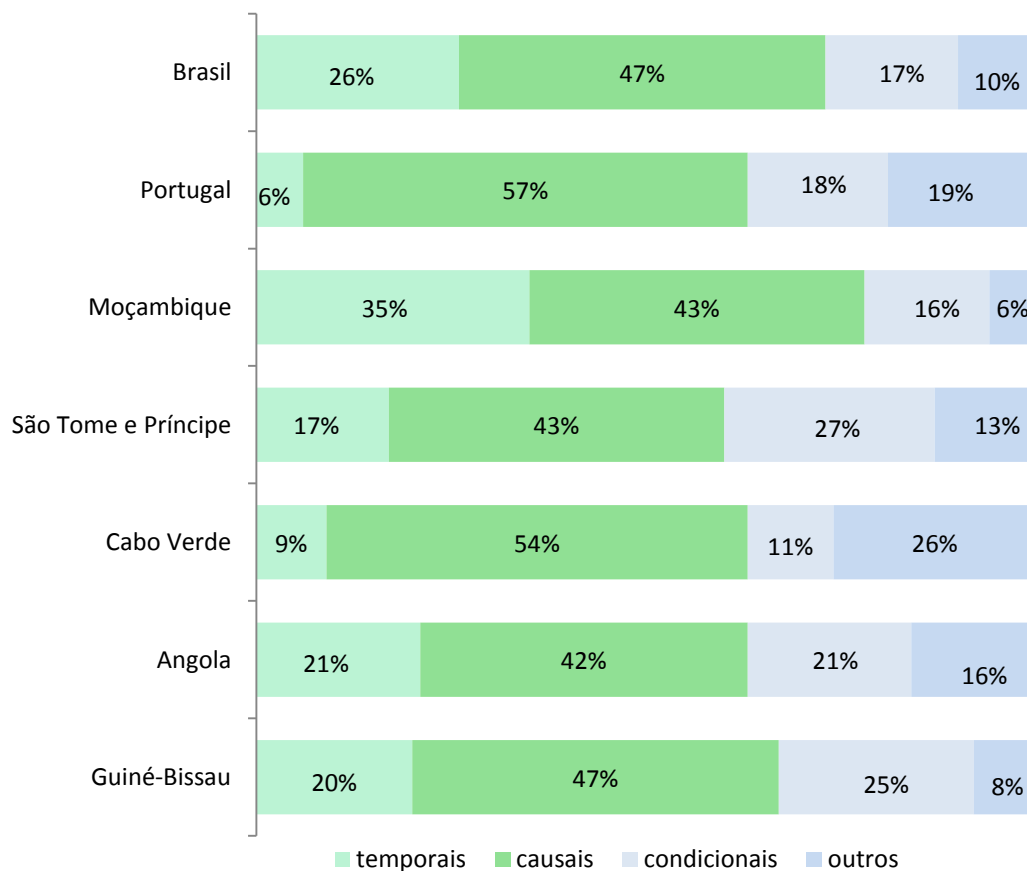
nessa distribuição têm valor anafórico, uma vez que retomam, em dependência, algum termo que foi apresentado no enunciado anterior. Dessa forma, no caso das completivas, as orações entre enunciados possuem uma função mais anafórica do que necessariamente pragmática, visto que desempenham o papel de acrescentar informações no processo de construção do discurso monológico, reaproveitando alguma estrutura no co-texto anterior que funciona como principal.

4.2 AS ORAÇÕES ADVERBIAIS

4.2.1 TIPOLOGIA DE SUBORDINAÇÃO

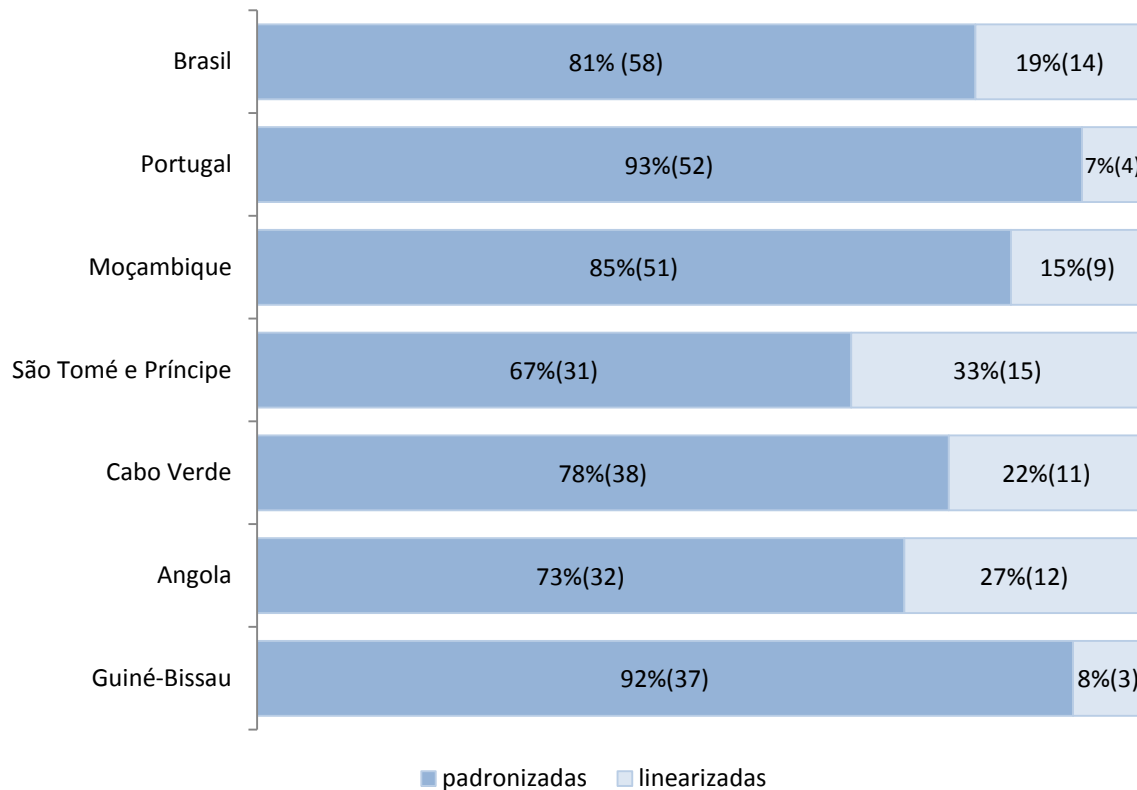
As orações adverbiais somam 439 ocorrências no corpus, distribuídas entre as variedades segundo mostra o gráfico:

Gráfico 5: Tipos de orações adverbiais.



Assim como no IA e no IT, a maioria das orações são temporais (91 enunciados, 21%), causais (205, 47%) e condicionais (84, 19%)²¹. A categoria *outros* inclui ocorrências menos frequentes nos dados, como as orações proporcionais, finais, comparativas e concessivas. Diferentemente das completivas, as construções adverbiais são realizadas, em sua maioria, em usos não linearizados:

²¹ Para as temporais, são oito enunciados para a Guiné-Bissau, 15 para Angola, quatro para Cabo Verde, 12 para São Tomé e Príncipe, 27 para Moçambique, três para Portugal e 22 para o Brasil. Em relação às causais são: 19, 30, 25, 30, 33, 29 e 39 enunciados, respectivamente. Já para as condicionais são 10, 15, 5, 18, 12, 9 e 14, nessa ordem para cada país.

Gráfico 6: Linearização vs. padronização nas adverbiais.

Como mencionamos, a integração²² sintático-semântica entre a principal e a subordinada pode influenciar a distribuição das estruturas. As adverbiais funcionam como adjunto, logo possuem, tendencialmente, maior autonomia semântica com relação à principal, o que não acontece com as completivas, que são mais dependentes semanticamente da matriz. Como observa Van Valin (1984; cf. também Lehmann (1988) sobre Hierarquia de subordinação, Givón (2001), e Cristóforo (2003) para o conceito de Hierarquia de *Deranking*) quanto maior a integração semântica entre os eventos, mais eles tendem a ser codificados por meio de uma relação de dependência sintática (LEHMANN, 1988, GIVÓN, 2001; FOLEY; VAN VALIN, 1984). Para avaliar o grau de integração sintático-semântica, Lehmann (1988) usa parâmetros como a presença/ausência de uma marca explícita de subordinação e qual o caráter oracional da subordinada. De certa forma, a integração semântica entre as orações também se reflete na maneira como elas são realizadas do ponto de vista do contorno entonacional (DECAT, 1993, 1999a, 1999b; NEVES, 2001; LIMA, 2002), uma vez que quando a subordinação é realizada em uma única unidade prosódica, o grau de integração é maior do que se houver distribuição ao longo de mais unidades entonacionais/informacionais.

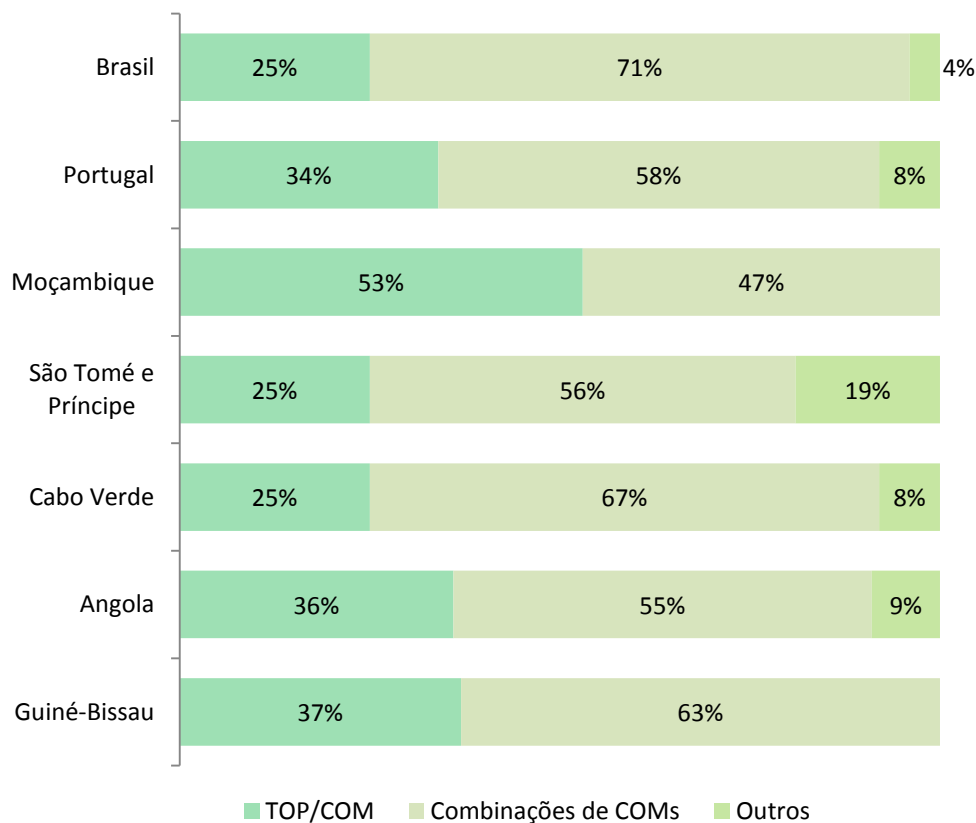
²² A integração semântica entre principal e subordinada na Hierarquia de Subordinação por Lehmann (1988) e Cristóforo (2003) na Hierarquia de *Deranking* na Subordinação.

Do ponto de vista da L-AcT, as completivas tendem a serem realizadas em uma única unidade informacional e as adverbiais manifestam comportamento oposto.

4.2.2 PADRÕES INFORMACIONAIS

Os padrões informacionais mais frequentes foram TOP/COM, e suas variantes complexas, e combinações de COMs. Além disso, foram considerados outros padrões categorizados como *outros*, os quais correspondem a ocorrências raras, como PAR/COM, COB/PAR e TOP/INT:

Gráfico 7: Padrões informacionais nas adverbiais.



Diferentemente das completivas, as orações adverbiais podem ocorrer de forma deslocada, antes ou depois da matriz, e essa organização se sobrepõe às funções comunicativas das orações. Os exemplos abaixo mostram algumas ocorrências no corpus:

(1) não /=AUX= assim /=AUX= eu achei legal isso /=COB= **porque** /=DCT= **cada um cê sabe que** /=SCA= cria um vínculo assim de amizade /=COB= assim /=AUX= um grupinho //COM= (padrão COB/COB) (B)

(2) então /=DCT= há dois aspectos importantes /=COB= que é /=INT= primeiro /=COB_s= fazer a &ed [2]=SCA= a campanha de educação ambiental /=COB= alertar as populações /=COB= para que a maioria tenha consciência /=SCA= do que está a fazer e do que deve ser feito /=COB= como pode ser feito /=COB= segundo /=COB_s= ao mesmo tempo /=PAR= fazer publicação de leis /=COB= **para que logo a seguir** /=SCA= **seja possível a aplicação de acções coercivas** /=COB= obrigar as pessoas a /=COB= &outr /1=EMP= há muitas maneiras de acção coerciva /=PAR= como disse /=PAR= e depois /=DCT= continuar com a investigação científica /=COB= para cada vez mais /=SCA= aumentar o nosso grau de conhecimentos //COM= (padrão COB/COB) (A)

Em (1), a causal é distribuída entre COB e COB e a final em (2) ocorre em uma estrofe, sequência terminada típica de textos monológicos e muito frequente no corpus. Em relação ao deslocamento, Chafe (1984), sobre as adverbiais causais, divide as ocorrências em *preposed*, quando ocorrem antes da principal, e *postposed*, quando ocorrem depois. O autor também categoriza os dados de acordo com a distribuição em unidades prosódicas, sendo *bound* (que corresponderia à linearização no quadro da L-Act) quando ocorrem na mesma unidade e *free* (padronização), em unidades diferentes. Diessel (2005) também estudou a relação entre a posição das adverbiais e suas funções comunicativas nas adverbiais temporais no inglês e observou resultados parecidos ao de Chafe (1984).

No presente trabalho, observamos que algumas subordinadas tendem a ocorrer prepostas em TOP, como as temporais, condicionais e causais introduzidas por *como*, servindo como *guidepost*. Já as causais iniciadas por *porque* são sempre pospostas e, conseqüentemente, não ocorrem em Tópico:

(3) não sei /=COB= quer dizer /=PAR= **como eu estava a treinar sozinho** /=TOP= para mim aquilo era um pouco difícil //COM= (padrão TOP/COM) (M)

(4) depois /=DCT= **quando o meu pai disse que** /=i-TOP= ontem /=PAR= **por exemplo** /=PAR= **no dia anterior** /=SCA= **sentiu-se tremor de terra** /=TOP= logo levantei /=COB= porque relacionei as duas coisas //COM= (padrão TOP/COM) (CV)

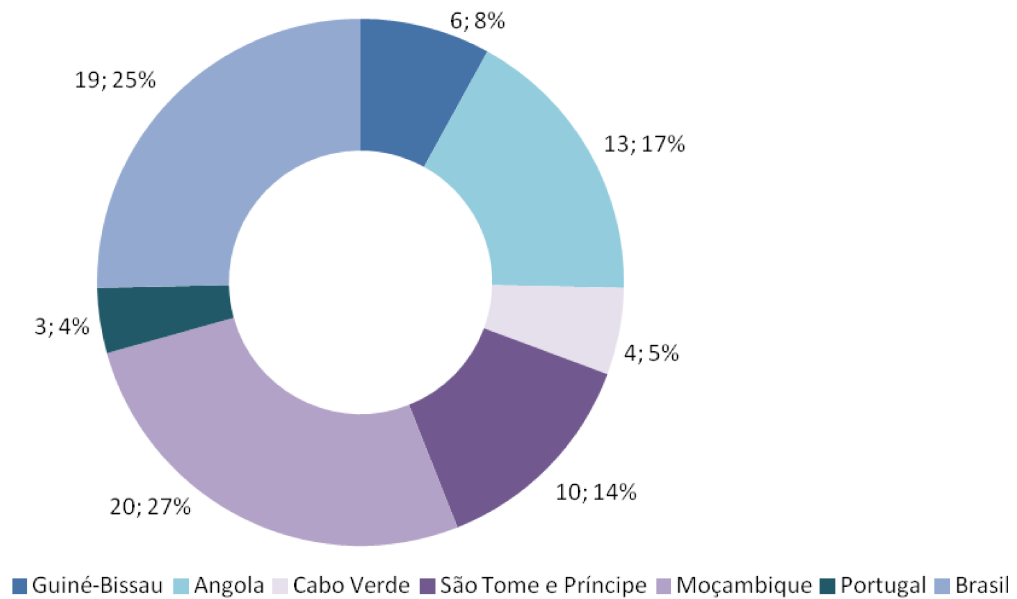
Nas ocorrências aqui analisadas, as orações causais prepostas ocorrem, geralmente, com a conjunção *como* em TOP/COM e não ocorrem em combinações de COMs, sendo esta a configuração mais frequente nas estruturas introduzidas por *porque*. Segundo Chafe (1985),

as adverbiais causais funcionam como um *guidepost* de modo que a informação conhecida está na principal e o conteúdo desconhecido se encontra na adverbial e quando ocorrem em unidades prosódicas diferentes pospostas, funcionam como *afterthought*. Neste trabalho, verificamos que, apesar de outras abordagens considerarem que o Tópico (tema) veicula informação já conhecida, observamos que as causais introduzidas por *como* e *que* sempre são realizadas no TOP, apresentam informações novas. O autor ressalta, ainda, que essas estruturas veiculam funções comunicativas diferentes dependendo da posição da adverbial. Sob esse ponto de vista, em (3) a causal ocorre preposta no Tópico e, portanto, fornece uma orientação pragmática para o ouvinte sobre o conteúdo que vem a seguir. Já em (4) a subordinada é posposta à matriz, acrescentando novas informações após um turno de fala já concluído. Certamente, as orações subordinadas adquirem novas funções comunicativas na fala espontânea, porém o corpus utilizado neste trabalho possui textos monológicos, com raras interações entre o informante e o interlocutor.

4.2.2.1 AS ORAÇÕES TEMPORAIS

De acordo com Said Ali (1969, p. 140), as orações temporais funcionam, com relação ao verbo da principal, de forma análoga a advérbios expressando circunstâncias de tempo. Elas são iniciadas por algumas conjunções prototípicas, como: *quando*, *cada vez que*, *depois que enquanto* e *no momento em que*. No corpus, foram encontradas 75 orações temporais distribuídas, principalmente, em TOP/COM (65), com poucas ocorrências de PAR/COM (1) e 9 combinações de Comentários:

Gráfico 8: Quantidade de orações temporais nas variedades do português.



(5) isto /=TOP= aconteceu porque eu gosto /=SCA= de ver chover /=COB= e mesmo agora /=SCA= que sou adulto /=TOP= **quando começa a chuviscar** /=TOP= e à noite /=PAR= de dia /=PAR= prefiro ficar na janela /=COB= ou saio mesmo para ver /=SCA= as águas a cair //COM= (padrão TOP/COB) (M)

(6) eu /=TOP= se [/1]=EMP= **quando partir** /=TOP= pronto /=COB= adeus borboletas /=CMM= e adeus colecção //CMM= (padrão TOP/CMM/CMM) (CV)

(7) mas **enquanto isso não acontecer** /=TOP= acho que /=SCA= estaremos um bocado a lutar contra a maré //COM= (padrão TOP/COM) (ST)

São raras as ocorrências de temporais em padrões diferentes de Tópico-Comentário, como em (5), em que a oração principal está no COB e a temporal, no TOP. Os demais enunciados também correspondem ao padrão Tópico-Comentário, função prototípica temporal na fala: a ocorrência não canônica em (6), com o advérbio mais frequente *quando* e em (7), com a ocorrência isolada de *enquanto isso*. Assim como em trabalhos anteriores (SILVA; BOSSAGLIA; 2019; BOSSAGLIA, 2015), a preferência por essa distribuição é esperada, uma vez que os Tópicos têm como função estabelecer o domínio semântico da aplicação da ilocução no Comentário e as orações temporais especificam um domínio temporal de determinado evento. Nesse sentido, as funções desempenhadas pelo Tópico se relacionam

semanticamente à função semântica das orações temporais. Além disso, as adverbiais também codificam valores semânticos diferentes daqueles do item lexical:

(8) normalmente **quando a pessoa** /=i-TOP= &he /=TMT= **ama a sua profissão** /=TOP= <tem &probl> [/2]=EMP= &he /=TMT= tem problemas &n [/1]=EMP= mas não são assim tão graves //=-COM= (padrão TOP/COM) (ST)

(9) porque /=DCT= bom /=AUX= em princípio /=AUX= **quando não se tem nada** /=TOP= não se consegue nada de nada //=-COM= (padrão TOP/COM) (ST)

Em (8), o item lexical *quando* introduz, prototipicamente, orações temporais. Porém, existe uma aproximação semântica com o valor condicional, de modo que oração pode ser interpretada como condicional ou temporal. O mesmo é verificado em (9), onde a subordinada iniciada por *quando* também possibilita, semanticamente, duas interpretações possíveis semanticamente: temporais ou condicionais (*quando/se não se tem nada, não se consegue nada de nada*). Além disso, algumas ocorrências mostraram fenômenos interessantes, como apresentado no exemplo abaixo:

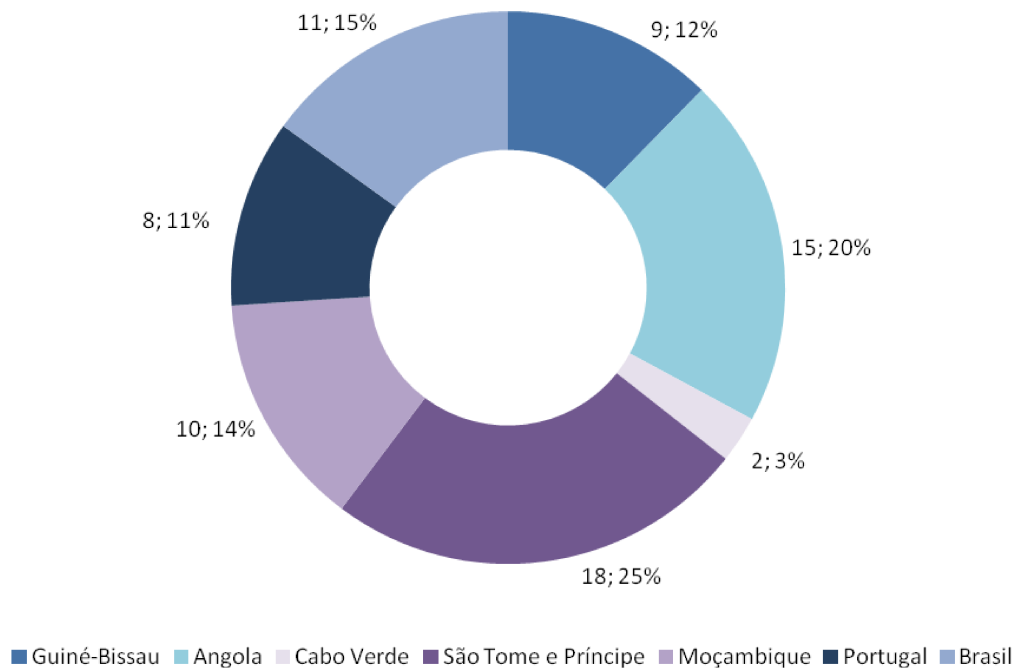
(10) há também **quando** /=SCA= **uma** /=SCA= **pessoa fica grávida** /=COB= **e há qualquer distúrbio** dentro de [/2]=SCA= **dentro do útero** /=COB= quer dizer /=PAR= a criança pode nascer com qualquer anormalidade /=COB= ou /=DCT= a criança /=SCA= pode prejudicar a mãe durante o parto /=COB= aí o aborto também pode ser feito mas legal //=-COM= (padrão COB/COB) (GB)

No exemplo, ocorre a nominalização da oração temporal *quando uma pessoa fica grávida*, e a subordinada não está desempenhando o seu papel canônico de expressar uma circunstância de tempo entre a matriz e a oração dependente; ao contrário, a construção é nominalizada no COB como objeto do verbo existencial *haver*. Como observado por Diessel (2005), existe uma iconicidade entre a ordem e a tipologia das adverbiais, sendo que aquelas que expressam eventos que ocorreram antes do evento que está na principal são realizadas prepostas e aquelas que exprimem eventos que foram realizados em uma circunstância temporal posterior à da matriz são pospostas.

4.2.2.2 AS ORAÇÕES CONDICIONAIS

As orações condicionais introduzidas por *se* somam 73 ocorrências em configurações padronizadas:

Gráfico 9: Quantidade de orações condicionais nas variedades do português.



Os padrões mais frequentes são TOP/COM (59), combinações de COBs (7) e ocorrências de INT/COB (1), COB/PAR (1) e PAR/COM (1):

(11) efectivamente a mulher africana /=TOP= **se quisermos** /=TOP= e /=DCT= **portanto** /=AUX= **olharmos** /=TOP= **aos estudos e aos levantamentos** que foram [/2]=SCA= **que têm sido feitos** /=TOP= constata-se que oitenta por cento de [/1]=SCA= da mulher /=SCA= e do seu tempo /=TOP= é ocupado com preocupações /=COB= portanto /=AUX= &he /=TMT= no que respeita /=SCA= primeiro /=SCA= ao seu papel como reprodutora // =COM= (padrão TOP/COM) (GB)

(12) eu fui para Porto Alegre /=COB= mas /=DCT= &he /=TMT= eu fiquei em &cida [/2]=SCA= hospedada numa outra cidade vizinha /=COB= **se não me engano** /=PAR= aí de Esteio /=PAR= é perto de Canoas /=COB= é uma coisa assim // (padrão COB/PAR)(BR)

Os exemplos mostram distribuições diferentes das prótases: em (11), a condicional é realizada ao longo de vários Tópicos e a oração principal ocorre no Comentário; em (12) a subordinada ocorre em Parentético, e, no exemplo, não possui escopo dentro da estrofe. Contrariamente ao que prevê a L-AcT, o Parentético, nesse exemplo, é ilocucionário, desempenhando uma ilocução de atenuação/defesa, e a falante faz um comentário à parte do conteúdo locutivo da estrofe. Vejamos alguns exemplos:

(13) porque /=AUX= nós já vimos /=COB= &de [/1]=EMP= devido um certo tempo /=TOP= não é /=AUX= as crianças já estão totalmente /=SCA= assim bem /=SCA= &he /=TMT= reintegrada /=COB= **se eu posso dizer** /=PAR= nós vimos que [/2]=i-COB= não é /=AUX= &he /=TMT= achamos /=SCA= conveniente /=SCA= reinseri-los para a família de [/2]=SCA= para as suas famílias //COM= (padrão COB/PAR)(A)

(14) você não consegue arranjar [/1]=SCA= arrancar um pé' de capim por mais fraquinho que seja /=COB= **se não vier com um torrão de terra agarrado** //COM= (padrão COB/COM)(A)

A ocorrência da subordinada em Parentético, como vemos, não representa uma oração condicional canônica com dependência sintática em relação a uma principal, mas realiza uma ilocução relacionada com a interação pouco frequente nos monólogos. Contudo, em (13), a expressão *se eu posso dizer* parece ser integrada na estrofe, visto que a falante está justificando *reintegrada*, funcionando como “as crianças já estão totalmente assim bem reintegrada, mais ou menos”. Por outro lado, em (14) a condicional é realizada em uma unidade ilocucionária, que integra sintático-semanticamente o enunciado, em uma combinação de COBs. Geralmente, as prótases ocorrem em Tópico, definindo o escopo da ilocução que vem a seguir (se x, então Y), mas no enunciado a prótase se encontra no Comentário. Vejamos outros casos:

(15) porque **se o Estado vem** /=INT= ah nós temos que fazer isso /=COB_r= o povo vai /=INT= não /=AUX_r= não vamos passar a fazer nada /=COB_r= porque vocês não nos pagam //COM_r= (padrão INT/COB_r) (GB)

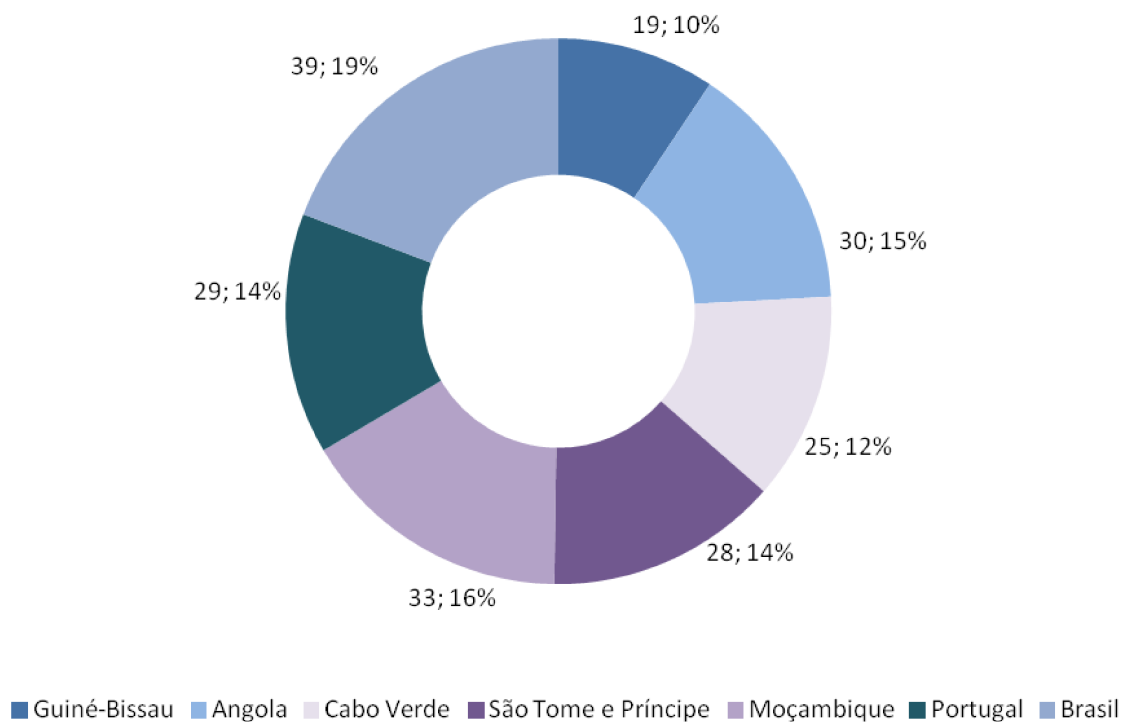
No exemplo, cada Introdutor Locutivo (*se o Estado vem* e *o povo vai*) introduz seu próprio discurso reportado para reproduzir a fala de terceiros. Nesse caso, porém a prótase e a

apódose estão no mesmo plano pragmático-informacional. Devemos considerar também que a etiquetagem desse enunciado foi bastante complicada, devido ao fator prosódico.

4.2.2.3 AS ORAÇÕES CAUSAIS

As orações que expressam causa, motivo ou justificativa foram encontradas em 203 sequências terminadas e o principal subordinador nas orações que ocorrem em padrões informacionais é *porque* (32), seguido por *como* (11):

Gráfico 10: Quantidade de orações causais nas variedades do português.



Os padrões mais frequentes são TOP/COM, para as causais introduzidas por *como*, e combinações de Comentários, para as orações pospostas introduzidas por *porque*. Apesar de as porcentagens serem altas, o número de ocorrências para cada padrão foi relativamente baixo. De todo modo, os padrões não diferem daqueles observados nas outras tipologias. As orações causais podem expressar a causalidade de forma direta ou indireta, além de assumirem diferentes funções pragmáticas (COUPER-KUHLEN, 1996; CHAFE, 1984), considerando que nem sempre o subordinador irá veicular uma função sintática prototípica:

(16) nós hoje /=TOP= podemos dizer que /=SCA= vivemos felizes /=COB= **porque** /=DCT= **graças a Deus** /=COB= eles [/1]=EMP= **vocês deram-nos a independência** /=COM= não é //AUX= (padrão COB/COM) (GB)

(17) eles [os trabalhadores] não acreditam /=SCA= que você conseguiu um direito para eles /=COM= não é //AUX= eu acho que é isso /=COB= porque /=AUX= caso /=TOP= caso não /=COB= **porque** /=DCT= o &q [/2]=EMP= **o advogado trabalhista** /=TOP= **acho que mais resolve é** /=TOP= **coisa assim** /=SCA= **que é corriqueira** /=COB= mas que também não é certo /=COB= não é /=AUX= por exemplo /=PAR= acidente de trabalho /=COB= essas coisas //COM= (padrão COB/COB) (BR)

O exemplo (16) apresenta uma causa direta, ou seja, a justificativa apresentada é sobre o evento descrito na oração. No caso, o falante explica que as pessoas vivem felizes no país devido à independência territorial. Diferentemente, em (17) a causa é indireta, pois o falante justifica o ato de fala anterior, isto é, o porquê ele disse que achava que os trabalhadores não acreditavam que tinham conseguido ajudá-los juridicamente. Em alguns casos, como já mencionamos, a conjunção pode não introduzir orações subordinadas:

(18) &he /=TMT= não /=COB= eu acho que as pessoas que estão a ir para as cidades /=TOP= para mim /=TOP= primeiro /=TOP= **porque** /=DCT= **parece que aqui se trata de uma questão** de &so [/2]=SCA= **de sobrevivência** //COM= (padrão TOP/COM) (PT)

Ao longo das análises e da etiquetagem, observamos que em algumas sequências, como em (18) acima, o *porque* não é usado como conectivo para introduzir uma subordinada, mas sim como conector discursivo (DCT), unidade dialógica que marca a coesão entre ilocuções ou unidades informacionais. Como já apontamos, as unidades dialógicas não compõem o texto do enunciado, portanto não possuem composicionalidade sintática com as demais. Assim, no exemplo acima o falante enumera alguns domínios para a ilocução com a qual ele justifica a migração das pessoas para as cidades. Nesse caso, o DCT parece fazer conexão entre os vários TOPs mas, ao mesmo tempo, também justifica/explica o *para mim* e *eu acho*, funcionando como uma justificativa enunciativa, mas projetada sobre essas unidades de Tópico que possuem uma carga epistêmica. Outros padrões foram encontrados nos dados:

(19) então /=AUX= esse tipo de coisa /=COB= eles querem te mostrar /=SCA= o que tem de bonito na cidade /=COB= &he /=TMT= você vê também até &ques [/1]=SCA= determinadas coisas que /=i-COB= &he /=TMT= para nós /=TOP= &he /=TMT= a nossa organização /=COB= &embo [/1]=EMP= você fica [/2]=EMP= para mim foi uma coisa que na época /=TOP= **principalmente porque eu estava estudando linguística** /=PAR= eu fiquei abismada /=COB= porque /=DCT= eu entendia /=SCA= as coisas teoricamente /=COB= lá eu tive a oportunidade de ver isso na prática //COM= (PAR) (BR)

(20) no entanto ele mandou-me procurar um quarto /=COB= **e como eu já tinha quatrocentos dólares em casa** /=TOP= então /=DCT= estava a andar com um amigo /=COB= e este amigo /=TOP= começou a procurar o quarto para mim durante três meses //COM= (padrão TOP/COM)(A)

O padrão em (19) apresenta uma oração causal realizada em Parentético, unidade informacional que possui função explicativa. Assim, a falante insere uma explicação no meio do enunciado, de modo que a UI está “isolada”, mas mantém as funções metalinguística e explicativa. Já em (20), o subordinador *como*, assim como em todas as ocorrências, é realizado no padrão Tópico-Comentário, no qual a causal é sempre preposta à matriz, pois, ao contrário de *porque*, *como* não é utilizado posposto à principal com valor de explicação (causa indireta).

4.2.2.4 AS ORAÇÕES CONCESSIVAS

As orações concessivas somam 23 enunciados não linearizados, sendo a maioria TOP/COM (6), combinações de unidades ilocucionárias (3 enunciados) e ocorrências isoladas de TOP/PAR e COB/PAR. Os subordinadores mais frequentes são *embora* (10), *mesmo que* (6), *desde que* (5), *apesar que* (3) e :

(21) portanto /=AUX= já isso /=TOP= **mesmo que não haja excesso de populações** /=TOP= mesmo isso já prejudica a própria natureza //COM= (padrão TOP/COM)(A)

(22) porque ela terá /=SCA= direito /=SCA= a que se lhe transmita o direito ao arrendamento /=COB= **desde que o seu &com [/1]=SCA= **companheiro não seja casado**** //COM= (padrão COB/COM) (PT)

(23) até hoje /=TOP= a imagem que eles têm /=i-TOP= **embora** /=SCA= **em conversa depois com eles** /=SCA= **eles digam que** /=SCA= **faz falta alguém assim no liceu** /=PAR= mas /=AUX= sinto que eu fui um bocado dura //COM= (padrão TOP/PAR)(ST)

Em (21), a adverbial é realizada no Tópico e a principal no Comentário; em (22) a distribuição ocorre em COB-COM e em (23), o Parentético assume a função similar que vimos nas causais, sendo “isolado” do restante do enunciado: a falante apresenta que os colegas de profissão têm uma imagem de que ela é muito rígida, mas faz uma concessão ao dizer que, ao mesmo tempo, eles acreditam que ela faz falta. De fato, o Parentético no qual a concessiva ocorre é muito longo e, assim, a falante “esquece” de realizar o predicado de uma possível principal. Portanto, isso mostra que nem as UIs, nem o ato de fala podem preceder a o processamento léxico-semântico-sintático do enunciado. Assim, não pode haver determinação da estrutura informacional sobre a sintaxe, por exemplo.

4.2.2.5 AS ORAÇÕES COMPARATIVAS E CONFORMATIVAS

Foram levantadas 10 conformativas todas introduzidas por *como* em Parentético e TOP/COM (2) e COB/COB (1); e 5 comparativas iniciadas pelo mesmo subordinador em COB/COM (1) e COB/COB (3):

(24) ela &de [/1]=SCA= deve ter alguma foto minha perdida /=COB= **assim como a dela ela deve estar procurando até hoje** /=COM= né //AUX= (padrão COB/COM) (BR)

(25) &he /=TMT= Filosofia /=COB= se não contarmos com a opção seminário /=i-PAR= portanto /=AUX= que tem /=i-PAR= **como sabe** /=PAR(2)= Filosofia e Teologia na sua formação /=PAR= &he /=TMT= nunca houve //COM= (padrão PAR/COM) (A)

(26) e ela /=i-COM= **como eu estava dizendo** /=PAR= me chamou a atenção //COM= (padrão PAR/COM)(BR)

(27) **como sabe** /=TOP= em todos os países isso aconteceu /=COB= também Cabo Verde não podia fugir à regra //COM= que é a questão das mulheres não [/1]=SCA= não estudavam //COM= (padrão TOP/COB)(CV)

A oração em (24) estabelece uma comparação entre o fato de alguém ter perdido tanto a foto dela quanto a de uma terceira pessoa. As conformativas utilizam a mesma conjunção e apresentam funções diferentes em cada padrão. Em (25) a construção *como sabe*, apesar de ocorrer em padronização, é orientada ao interlocutor e não possui dependência sintática com alguma outra oração do enunciado, pois o falante faz um comentário direcionado ao interlocutor, de forma a assegurar que ele está acompanhando o fluxo da conversa. Em (26) a ocorrência em Parentético apresenta função semelhante, porém com a finalidade de retomar algum ponto da conversa e reafirmar o que estava dizendo. Já em (27), a subordinada é realizada em Tópico e orienta o interlocutor para o que será dito a seguir.

4.2.2.6 OUTRAS TIPOLOGIAS

Apesar de os resultados concentrarem-se nas orações causais, condicionais e temporais; outras ocorrências esporádicas foram atestadas, como subordinadas finais (11) introduzidas por *para que* e realizadas em COB/COM (8) e COB/COB (3) (pospostas, v. exemplos (28) e (29) abaixo); e as proporcionais não linearizadas, sendo uma padronizada em TOP/COB em (3):

(28) porque nós temos muitas músicas aqui paradas /=COB= sinfonias de Beethoven /=COB= Mozart /=COB= desses artistas /=SCA= grandes falados /=COB= e nós queremos também meter a música deles /=COB= em acção /=COB= **para que a nossa população comece a distinguir** //COM= (padrão COB/COM) (ST)

(29) &he /=TMT= vê se isso no mundo industrial /=COB= não é /=AUX= em que se criam diferenças entre as pessoas que trabalham /=COB= e /=SCA= se procura que hajam interesses pessoais /=COB= **para que** /=i-COB= &he /=TMT= **cada um tenha qualquer coisa a defender** /=COB= e não haja espírito de que há /=SCA= uma coisa colectiva a defender /=COM= não é //AUX= (padrão COB/COB) (PT)

(30) e também de facto noto que **quanto mais se vai para sul** /=TOP= &he /=TMT= prontos /=AUX= mais há esse tipo de problemas /=COB= mais /1=EMP= &he +=TMT= (padrão TOP/COB) (PT)

Os exemplos (28) e (29) apresentam orações finais realizadas no padrão mais frequente, isto é, combinações de unidades ilocucionárias. É possível notar certa iconicidade

entre a posição da adverbial e a temporalidade dos eventos descritos pelas subordinadas no nível semântico, pois as orações finais denotam eventos que podem ocorrer apenas depois do evento da principal e, do mesmo modo, são realizadas como pospostas à matriz. Já em (30) a oração proporcional é realizada em TOP/COM, no qual o domínio semântico da aplicação da ilocução está no TOP, indicando que há mais problemas na região à medida que vai se aproximando do sul.

4.2.3 A INSUBORDINAÇÃO

Evans (2007) define a insubordinação como orações que, à primeira vista, parecem formalmente subordinadas, mas que funcionam como frases principais. Nesse caso, as orações são usadas como independentes, mesmo que haja índices lexicais, como conjunções, que prototipicamente introduzem construções subordinadas. Do ponto de vista da L-AcT, esse fenômeno é explicado atribuindo a autonomia pragmática de estruturas dependentes à força ilocucionária que elas carregam (BOSSAGLIA; MELLO; RASO, 2020). Nessa abordagem, as insubordinadas tendem a ser realizadas sozinhas dentro do enunciado:

(31) não sei se /=i-COM= &he /=TMT= está interessado um bocado no anonimato /=COM= não é //AUX= em qual aspecto //COM= em todo //COM= até para sua própria defesa /=COM= não é //AUX= **se as pessoas de facto não se comunicam entre si** //COM= &he /=TMT= não podem ser solidárias /=COM= não é //COM= (PT)

(32) é que /=SCA= o vulcão muito raramente mata-se //COM= **se nós não nos acreditarmos nas [1]=SCA= nos seus avisos** /=COB= et cetera //COM= (CV)

No exemplo (31), o falante apresenta uma situação (que as pessoas nas grandes cidades não comunicam entre si) e faz um comentário sobre isso em outra ilocução e pode-se reconhecer uma certa relação entre a condicional e o ultimo enunciado. O enunciado em (32) é um claro exemplo de prótase insubordinada, pois não existe dependência sintática entre a “principal” (*o vulcão raramente mata-se*) e a condicional. Nesse caso, o falante veicula uma ilocução de expressão de obviedade e não há uma relação “se X, então Y”; portanto não existe

a condição direta de “se não acreditarem nos avisos, o vulcão raramente mata”. Outras tipologias também apresentam configuração semelhante:

(33) ele pode comer esse peixe /=COB= comer canoa //COM= **porque há um tubarão no mar** /=COB= que é rei do mar //COM= (ST)

(34) mas problemas tipo /=SCA= complexo de Édipo /=COB= né /=AUX= deve ter alguma coisa /=COB= com certeza tem /=COM= né //AUX= complexo de Édipo /=TOP= assim que eu falo //COM= &he /=TMT= **porque o complexo de Édipo** /=TOP= **é o seguinte** //COM= a [/1]=EMP= &he /=TMT= a [/1]=EMP= a [/1]=EMP= a disputa da [/1]=SCA= a [/1]=EMP= da [/1]=EMP= na menina /=TOP= é o complexo de Electra /=COB= né /=AUX= então /=DCT= é a disputa do pai /=CMM= em relação à mãe /=CMM= entendeu //AUX= (BR)

Em (33), não é possível reconhecer uma relação de dependência sintática entre os dois enunciados. Portanto, o fato de haver um tubarão no mar não é o motivo pelo qual “ele poder comer esse peixe”. No exemplo (34), o falante apresenta uma justificativa enunciativa, na qual explica o porquê proferiu o ato de fala anterior, isto é, porque ele considera que aquele problema pode ser relacionado ao Complexo de Édipo. As causalidades direta e indireta também são verificadas quando as orações ocorrem entre enunciados:

(35) há razões sim //COM= **porque** /=AUX= **primeiro** /=TOP= **é que eu** /=SCA= **tive** /=SCA= **problemas mesmo** /=COM= **com a família** //APC= (PT)

(36) porque ela bate nas folhas /=COB= as partículas de água são [/1]=SCA= são desagregadas /=COB= e portanto /=AUX= quando chega ao solo /=TOP= já vem sem força //COM= **porque houve** <uma &f> [/2]=SCA= **uma primeira parte que amorteceu a queda da água** /=COB= da [/1]=EMP= dessa gota de água //COM= (A)

(37) porque /=AUX= isso também é um dos nossos /=SCA= objetivos /=COB= da [/1]=SCA= de [/1]=EMP= que é da instituição /=COB= promover a vida e os direitos da criança /=COB= com vista à sua autonomia na sociedade /=COB= e à sua reintegração na sociedade //COM= porque /=AUX= nós já vimos /=COB= &de [/1]=EMP= devido um certo tempo /=TOP= não é /=AUX= as crianças já estão totalmente /=SCA= assim bem /=SCA=

&he /=TMT= reintegrada /=COB= **se eu posso dizer** /=PAR= nós vimos que [/2]=i-COB= não é /=AUX= &he /=TMT= achamos /=SCA= conveniente /=SCA= reinseri-los para a família de [/2]=SCA= para as suas famílias //COM= (A)

Nos exemplos (35) e (36), é apresentada a relação indireta entre as “principais” e as adverbiais: em (35), semanticamente, *razões* já introduz uma causa e há uma justificação enunciativa, apesar de o *porque* ser reduzido foneticamente; e em (36), há uma causa direta, pois o falante explica que a queda d’água foi amortecida pelo impacto nas folhas. Diferentemente, em (37) o falante sinaliza que um dos seus objetivos é integrar as crianças na sociedade, mas não existe uma causalidade direta no sentido de que promove a reintegração das crianças porque “as crianças já estão totalmente integradas” - o que, no plano sintático-semântico não seria adequado, afinal não é porque *as crianças estão totalmente assim bem integradas* que o objetivo do projeto é integrá-las. Assim, a relação é no plano pragmático pois a falante justifica o seu ato de fala (*eu disse isso, por causa disso*).

Bossaglia, Mello e Raso (2020) consideram "semi-insubordinadas" as orações que permitem que a principal seja recuperada pelo contexto, mantendo relação textual entre as UIs e ligação textual/pragmática com o enunciado que contém a possível “principal”. As semi-insubordinadas constituem um fenômeno também encontrado no corpus:

(38) sim /=COB= mas eu acho que noutros países /=TOP= por mais que os jogadores percam /=APT= mas são recebidos com um pouco de [/1]=SCA= de carinho /=COM= não é //AUX= quando chegam //COM= mas nós já &rece [/2]=SCA= às vezes recebemos com [/2]=SCA= &des [/1]=EMP= desprezamos //COM= **mesmo quando os Mambas foram perder na África do Sul** //COM= (M)

(39) bom /=AUX= a resposta [/2]=EMP= a resposta é /=SCA= óbvia //COM= **embora careça de algumas explicações** //COM= (PT)

(40) lá a diferença de falares /=COB= lá eles falam por tu /=COB= tu /=COB= tu /=COB= o tempo inteiro /=COB= né /=AUX= e para a gente é superesquisito //COM= **embora a gente saiba** que o português [/3]=SCA= **que o correcto** /=TOP= **seria falar tu** /=COM= não é //AUX=

Em (38), a temporal é focalizada pelo advérbio *mesmo* e veicula uma circunstância temporal do evento descrito no enunciado anterior, isto é, os jogadores foram recebidos com carinho mesmo quando o time perdeu o jogo na África do Sul. Em (39), a concessiva é

realizada depois da principal e veicula um *afterthought*. De acordo com Chafe (1984), nessa construção o falante adiciona uma informação ao enunciado anterior, inserindo-a no enunciado seguinte. Desse modo, ocorre em unidades prosódicas diferentes, pois o falante não conseguiu lembrar o que pretendia dizer a tempo, e o faz no enunciado seguinte (PAWLEY; SYDER, 1983; CHAFE, 1984). No exemplo, a falante “corrige” o anterior, em que disse que a resposta é óbvia: já que algumas precisam de explicação, não existe essa obviedade. Em (40), a concessiva também veicula um *afterthought*. Porém, apesar de as ocorrências que ultrapassam as fronteiras do enunciado revelarem funções comunicativas distintas, nos monólogos a tipologia interacional influencia, visto que os textos se assemelham à escrita. Assim, as relações concessivas entre os enunciados são mantidas, diferentemente do que foi observado em corpus com registros mais acionais (SILVA; BOSSAGLIA, 2019; BOSSAGLIA, 2014).

4.2.4 O TESTE ESTATÍSTICO

O teste utilizado foi o Qui-quadrado e um modelo linear. Assumimos o tipo de subordinada (completiva ou adverbial) como variável resposta (ou dependente), ou seja, a característica observada e analisada no estudo. Como hipótese, buscamos observar se o tipo de subordinada influencia a distribuição das unidades informacionais, ou seja, se o pressuposto de que a distribuição informacional é explicada pelo tipo de oração subordinada, é verdadeiro. Para realizar a análise no R, as ocorrências foram classificadas em linearizada (quando ocorre na mesma unidade prosódica) e não linearizada (em unidade prosódica diferente), assumindo-se a categoria "porcentagem.linearizada" como variável resposta, que é aquela explicada pela variável preditora. Nesse caso, partimos do ponto de vista de que o tipo de subordinada pode influenciar a articulação informacional.

Desse modo, ajustamos um modelo linear com a porcentagem de linearizada como variável resposta e subordinação com variável preditora. A variável resposta foi contrastada usando *treatment coding*, com "porcentagem de linearizada" como nível de referência. Em comparação a um modelo nulo, houve efeito significativo do tipo de subordinação em relação à porcentagem de linearizadas ($F(1,68)=108.89$, $p = 9,307 \times 10^{-16}$) - como parâmetro, assumiu-se, nesse trabalho, $\alpha = 0.05$. Portanto, como o *p-value* é menor que 0,05, há uma diferença estatisticamente significativa e podemos rejeitar a hipótese nula de que não há relação significativa entre a subordinação e porcentagem de linearizadas. Assim, pode-se reconhecer a

significância estatística da porcentagem de completiva em linearização e da porcentagem de adverbial em não-linearização.

5. CONSIDERAÇÕES FINAIS

A análise da subordinação na fala espontânea das variedades do português revelou algumas características interessantes. Por se tratar de textos monológicos e menos acionais, observamos estruturas textualmente/sintaticamente mais complexas que aquelas que se podem encontrar em situações de fala espontânea, com maior acionalidade e variação diafásica. Nesse sentido, notamos que a tipologia interacional influencia na organização sintática, mas com uma complexidade textual e tamanho das sequências terminadas não prototípicos da fala espontâneas.

Embora haja essa característica particular dos monólogos, os resultados obtidos foram semelhantes àqueles observados em corpus orais mais acionais, como aqueles do inglês americano e do italiano da família C-ORAL (SILVA; BOSSAGLIA; BOSSAGLIA, 2015). De modo geral, as completivas funcionam como argumento e tendem a serem realizadas na mesma UI que a principal, enquanto as adverbiais representam eventos autônomos daqueles que ocorrem na matriz, funcionando como adjunto e sendo realizadas em outra UI.

Os padrões informacionais se mostraram semelhantes, com TOP/COM e combinações de Comentários como mais frequentes. Foi observado que as orações temporais e causais introduzidas por *como* tendem a ocorrer no padrão Tópico-Comentário, o que está relacionado ao nível semântico. Além disso, foram encontradas ocorrências menos frequentes, como aquelas em Parentético ou Introdutor Locutivo.

No que tange às tipologias, nas sete variedades o complementizador mais comum foi *que* e os subordinadores adverbiais mais frequentes foram *porque*, *quando* e *se*; o que também foi verificado no inglês americano e italiano (SILVA; BOSSAGLIA, 2019; BOSSAGLIA, 2015). De fato, podemos tecer alguns comentários sobre a insubordinação, que revelou algumas organizações em que a sintaxe é governada pelo nível pragmático, como a justificação enunciativa, porém tais casos não integram a maioria das ocorrências.

Neste trabalho, apesar de não ser possível fazer constatações sólidas a respeito das funções comunicativas, ficou clara a importância da prosódia ao etiquetar e, posteriormente, fazer as análises sintáticas e identificar possíveis entonações e perfis prosódicos que influenciam na interpretação. Além disso, ficou claro que é muito difícil fazer as análises com

materiais de baixa qualidade acústica, e sem ter muita familiaridade com prosódias distintas daquela do português brasileiro ou português europeu, que são mais estudados nesse aspecto. Sendo assim, mostra-se importante explorar os novos recursos que permitam análises de outras variedades. O processo de etiquetagem também precisa ser mais refinado e validado por uma equipe maior e com mais etiquetadores experientes.

Ademais, foi possível observar que o comportamento das tipologias tende a ser o mesmo para as mesmas estruturas em línguas diferentes, mas que, ainda, as variedades do português apresentaram construções interessantes. A omissão do complementizador, por exemplo, foi uma opcionalidade que não havia sido verificada em trabalhos anteriores (BOSSAGLIA, 2015; 2015) sobre o português brasileiro falado. Nesse sentido, do ponto de vista teórico da L-AcT, a sintaxe da fala é fragmentada e de tipo pragmático-informacional, o que não foi observado com frequência no corpus aqui utilizado, justamente devido à menor acionalidade e pragmaticidade dos textos monológicos - o que sugere uma nova exploração da teoria em relação à sintaxe da fala.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- ALEXANDRE, N.; GONÇALVES, R.; HAGEMEIJER, T. A formação de frases relativas de PP no português oral de Cabo Verde e de São Tomé. In: Costa A, Falé I, Brabosa P, organizadores. *XXVI Encontro Nacional da Associação Portuguesa de Linguística – Textos seleccionados*. 21, 22 e 23 de Outubro de 2010; Porto; Portugal. Lisboa: Associação Portuguesa de Linguística. 2011; p. 17-34.
- AGOSTINHO, A. L. *Fonologia e método pedagógico do lung'le*. 2015. Tese (Doutorado em Filologia e Língua Portuguesa), Universidade de São Paulo, São Paulo, 2015.
- ANTHONY, L. *AntConc*. Versão 3.4.3. Tokyo, Japan: Waseda University, 2014. Disponível em <<http://www.laurenceanthony.net/>>.
- ANTÓNIO, Tuaha. Estratégias de ensino da concordância verbal em número à população universitária moçambicana. Tese de Mestrado, Universidade Eduardo Mondlane, Mozambique, 2011.
- ARAÚJO, G. A.; SILVEIRA, A. C. Vogais e Ditongos no Português Vernacular de São Tomé e Príncipe. In Oliveira, Márcia; Araujo, Gabriel Antunes de (orgs). *O Português na África Atlântica*. São Paulo: Humanitas/FAPESP, 2018, p. 261-296.
- AUSTIN, L.J. *How to do things with words*. Oxford: Oxford University Press, 1962.
- BALDUINO, Amanda Macedo; BANDEIRA, Manuele; FREITAS, Shirley. Os processos de elisão e degeminação no português de São Tomé e Príncipe. *Filologia e Linguística Portuguesa*, v. 19, n. 1, p. 163-197, 2017.
- BALDUINO, A. M.; BANDEIRA, M. A ascensão da Língua Portuguesa em São Tomé e Príncipe. *Domínios de Linguagem*, [S. l.], v. 16, n. 3, p. 991–1025, 2022. DOI: 10.14393/DL51-v16n3a2022-4.
- BARRETO, F.; BRANCO, A.; FERREIRA, E.; MENDES, A.; NASCIMENTO, M. F.; NUNES, F.; SILVA, J. Open Resources and Tools for the Shallow Processing of Portuguese. In: International Conference on Language Resources and Evaluation (LREC), 5, 2006, Genoa, Italy. *Proceedings...*, 2006.
- BARTON, D. *Literacy*. Oxford: Blackwell, 1984.
- BARTOLETO, Galaor. *Entenda o que é língua franca*. 02 de janeiro de 2010. Disponível em: <<http://www.galaor.com.br/lingua-franca>>. Acesso em 28-08-2014.
- BENDER, G. J. *Angola sob o domínio português: mito e realidade*. Luanda: Editorial nzila, 2004.
- BIRMINGHAM, David. *A África Central até 1870: Zambézia, Zaire e o Atlântico Sul*. Luanda: ENDIPU/UEE, s/d. file:///C:/Users/adminn/Downloads/137201-Texto%20do%20artigo-264787-1-10-20170818.pdf
- BLANCHE-BENVENISTE, C. *Approches de la langue parlé en français*. Paris: Ophrys, 2000.
- BLANCHE-BENVENISTE, C. Le recouvrement de la syntaxe et de la macro-syntaxe. In: SCARANO, A. (Ed.). *Macro-syntaxe et pragmatique: l'analyse linguistique de l'oral*. Roma: Bulzoni, 2003. p. 53-76
- BLOOMFIELD, L. *Language*. London: Allen and Unwin, 1935.

BOERSMA, P.; WEENINK, D. *Praat: doing phonetics by computer* [Computer program]. Version 6.0.37. Disponível em: <http://www.praat.org/>.

BOSSAGLIA, Giulia. Interface entre sintaxe e articulação informacional na fala espontânea: uma comparação baseada em corpus entre português e italiano. *Caligrama*, v. 19, p. 35-60, 2014.

_____. Orientação pragmática da sintaxe na fala: uma análise corpus-based da subordinação completiva e adverbial no português do Brasil. *Domínios de Linguagem*, v. 9, n. 5, p. 309-335, 2015.

BOSSAGLIA, G.; MELLO, H. Clausal syntactic properties of spoken Brazilian Portuguese. In: GSCP – GRUPPO DI STUDIO SULLA COMUNCAZIONE PARLATA - La comunicazione parlata -Spoken communication., 2017, Napoli. *Anais...*Napoli: Aracne, 2016. p. 115-133.

BOSSAGLIA, G.; MELLO, H.; RASO, T. Illocution as a unit of reference for spontaneous speech: An account of insubordinated adverbial clauses in Brazilian Portuguese. In: In Search of Basic Units of Spoken Language. *John Benjamins*, 2020. p. 221-256.

BRANDÃO, S. Concordância nominal em duas variedades do português: convergências e divergências, *Veredas – Revista de Estudos Linguísticos* 15:1, 164–178, 2011.

CARTER, R.; MCCARTHY, M. *Cambridge Grammar of English: A Comprehensive Guide*. Spoken and Written English Grammar and Usage. Cambridge: Cambridge University Press, 2006. 973p.

CAVALCANTE, F. A. *The topic unit in spontaneous American English: a corpus-based study*. 2015. 184f. Dissertação (Mestrado em Pós-Graduação em Estudos Linguísticos) – Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2015.

CAVALCANTE, F.A. et al. *The information unit of Topic: a crosslinguistic, statistical study based on spontaneous speech corpora*. 2020.

_____; RAMOS, A. The American English spontaneous speech minicorpus: architecture and comparability. *CHIMERA: Romance Corpora and Linguistic Studies*, v. 3, n. 2, 2016, p. 99-124, 2016.

CHAFE, W. How people use adverbial clauses. In: ANNUAL MEETING OF THE BERKELEY LINGUISTICS SOCIETY, 10., 1984, California. *Proceedings...*, California: Berkeley Linguistics Society, 1984. p. 437-449.

_____, W.; DANIELEWICZ, J. 'Properties of written and spoken language', in R. Horowitz and S. J. Samuels (eds.), *Comprehending Oral and Written Language* (New York: Academic Press), 1987, 83-113.

CHAVAGNE, Jean-Pierre. *La langue portugaise d'Angola. Études des écarts par rapport à la norme européenne du portugais*. Dissertação de PHD, Université de Lyon 2, 2005.

CHICUMBA, Mateus Segunda. *A formação de professores de português língua segunda (PL2) em Angola: O caso da Universidade Katyavala Bwila/Benguala*. Master's thesis, Universidade de Lisboa, 2012.

CHIMBUTANE, F. Portuguese and African Languages in Mozambique: A sociolinguistic approach. In: LÓPEZ, L. A.; GONÇALVES, P.; AVELAR, J. O. de (eds.). *The portuguese language continuum in Africa and Brazil*. Amsterdam: John Benjamins, 2018. p. 89-110.

- CHIMUKU, L. *Valores semânticos de "só", "ainda", "ainda só", "já" e "já sim" no português de Angola: proposta de exercícios práticos*. (Dissertação de Mestrado), Évora, 2019.
- CORRÊA, V. R. *Oração relativa: o que se fala e o que se aprende no português do Brasil*. *Sínteses-ISSN 1981-1314* 4, 1999.
- COSTA, T. *Umbundismos no português de Angola: Proposta de um dicionário de umbundismos*. PhD dissertation, Universidade Nova de Lisboa, 2015.
- COUTO, H. H.; EMBALÓ, F. *Literatura, Língua e Cultura na Guiné-Bissau: Um país de CPLP*. *Revista Brasileira de Estudos Crioulos e Similares*, Papia, n. 20, Brasília, 256 p, 2010.
- COUPER-KUHLEN, E. *Intonation and clause combining in discourse: The case of because*. *International Journal of Association*, v. 6, n. 3, p. 386–426, 1996.
- CHIMBUTANE, Feliciano. *Multilingualism in education in post-colonial contexts: A special focus on Sub-Saharan Africa*. In: *The Routledge handbook of multilingualism*. Routledge, 2012. p. 184-200.
- CRESTI, E. *Information and intonational patterning in Italian*. B. Ferguson, H. Gezundhajt, Ph. Martin (eds.), p. 99-140, 1994.
- CRESTI, E. *Corpus di italiano parlato*. Firenze: Accademia della Crusca, 2000.
- CRESTI, E. *La Stanza: un'unità di costruzione testuale del parlato*. In Ferrari, A. (ed). *Sintassi storica e sincronica dell'italiano. Subordinazione, coordinazione e giustapposizione*. *Atti del X Congresso SILFI*, 713-732. Firenze: Cesati, 2010.
- CRESTI, E. *Syntactic properties of spontaneous speech in the Language into Act Theory*. In: RASO, T.; MELLO, H. (Ed.). *Spoken Corpora and Linguistic Studies*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 2014. p. 365-410
- CRESTI, E; MONEGLIA, M. (Ed.). *C-ORAL-ROM*. *Integrated reference corpora for spoken romance languages*. Amsterdam/ Philadelphia: John Benjamins, 2005.
- CRESTI, E.; MONEGLIA, M. *Informational patterning theory and the corpus-based description of spoken language. The compositionality issue in the topic-comment pattern*. In: *Bootstrapping information from corpora in a cross-linguistic perspective*. Firenze University Press, 2010. p. 13-46.
- CRUZ, Arsénio da Silva. *Estudo comparativo entre o perfil linguístico do falante urbano do Lubango e do Huambo e suas implicações no ensino do português*. PhD dissertation, Universidade Nova de Lisboa, 2014.
- CRYSTAL, D. *The Cambridge Encyclopaedia of Language*. Cambridge: Cambridge University Press, 1987. 472 p.
- DEBAISIEUX, J.-M. *Autour de parce que et de puisque*. In: DEBAISIEUX, J. M. (Ed.), *Analyses linguistiques sur corpus: subordination et insubordination en français*. Hermès science publications: Lavoisier, 2013. p. 185-284.
- DIAS, H. N. *Português moçambicano: estudos e reflexões*. Maputo: Imprensa Universitária, 2009.
- DIESSEL, H. *Competing motivations for the ordering of main and adverbial clauses*. *Linguistics*, Cambridge, v. 43, n. 3, p. 449-470, 2005.

- DIK, Simon C. *The Theory of Functional Grammar*. Part 1: The Structure of the Clause. Edited by Kees Hengeveld. 2nd, rev.ed. Berlin; New York: Mouton de Gruyter, 1997.
- DO NASCIMENTO, Maria Fernanda Bacelar. Corpora comparáveis e variação lexical nas variedades africanas do português. *ALFA: Revista de Linguística*, v. 50, n. 2, 2006.
- DOS SANTOS, Eduardo Ferreira. Aspectos da língua portuguesa em Angola. *PAPIA-Revista Brasileira de Estudos do Contato Linguístico*, v. 28, n. 1, p. 25-49, 2018.
- DOS SANTOS, Vinícius Gonçalves; SVARTMAN, Flaviane Romani Fernandes. O padrão entoacional neutro do português de Guiné-Bissau: uma comparação preliminar com o português brasileiro. *Estudos Linguísticos* (São Paulo, 1978), v. 43, n. 01, p. 48-63, 2014.
- DUARTE, I. M. Vantagens de uma gramática de usos para o Português Europeu. Alguns exemplos de análise de expressões extraídas de usos orais informais. *Revista Da Associação Portuguesa De Linguística*, (4), 1-17, 2018. <https://doi.org/10.26334/2183-9077/rapln4ano2018a32>.
- EBERHARD, D. M., GARY F. S.; CHARLES D. F. (eds.). *Ethnologue: Languages of the World*. Twenty-fifth edition. Dallas, Texas: SIL International, 2022.. Online version: <http://www.ethnologue.com>.
- ENKVIST, NILS-ERIK. 'Impromptu speech, structure and process'. In: N.-E. Enkvist (ed.), *Impromptu Speech: A Symposium* (Abo: Abo Akademi Foundation), 1982, 78: 11-32.
- FARIAS, Silvana Silva Araújo de; SILVA, Manoel Crispiniano Alves da. A sintaxe dos pronomes clíticos no português falado em feira de santana-ba: uma comparação com o português luandense. *Macabéa-Revista Eletrônica do Netlli*, v. 8, n. 2, p. 563-584, 2019.
- FERNANDEZ, M. M. J. 'Oralite et ecriture', Paper presented at the Eurotype Plenary Conference, Le Bischenberg, Strasbourg, 1994, 27-31 March.
- FERREIRA, M. *Que futuro para a língua portuguesa em África?* Lisboa- Portugal: ALAC AFRICA, 1988.
- FIGUEIREDO, Carlos; SANTOS, Eduardo Ferreira dos. Construções [FOC+QUE] no português do Município do Libolo, Angola. *Filologia e Linguística Portuguesa*, n. 16, p. 209-231, 2014.
- FIRENZUOLI, V.; SIGNORINI, S. L'unità informativa di topic: correlti intonativi. In: MAROTTA, G.; NOCCHI, N. (Ed.). In: GIORNATE DI STUDIO DEL GRUPPO DI FONETICA SPERIMENTALE, 13, 2003, Pisa. *Anais...*, Pisa: Edizioni ETS, 2003. p. 177-184.
- FLEISS, J. L. Measuring nominal scale agreement among many raters. *Psychological Bulletin* (76), 378-382, 1971.
- FOLEY, W.; VAN VALIN, D. R. Jr. *Functional syntax and universal grammar*. Cambridge: Cambridge University Press, 1984.
- FONSECA, Dagoberto José. As línguas nacionais e o prestigioso português em Angola. *Anais do SIELP*, v. 2, n. 1, 2012.
- GALVÃO, V. C. C. *Evidencialidade e gramaticalização no português do Brasil: Os usos da expressão diz que*. Tese de mestrado. Faculdade de Ciências e Letras. Universidade Estadual Paulista - UNESP, Araraquara: São Paulo, 2001.
- GARTNER, E. Particularidades morfossintáticas do português de Angola e Moçambique. *Confluência*, p. 27-58, 1996.

- GÉNÉREUX, MICHEL/MENDES, AMÁLIA/BACELAR do NASCIMENTO, FERNANDA/PEREIRA, LUÍSA. Lexical analysis of pre and post revolution discourse in Portugal. In: Third Workshop on Building Comparable Corpora, International Conference on Language Resources and Evaluation (LREC 2010), 7, 2010, Malta. *Proceedings*: Malta, 2010, p. 65-71.
- GÉNÉREUX, M; HENDRICKX, I; MENDES, A. Introducing the Reference Corpus of Contemporary Portuguese On-Line, In: *Proceedings of the Eighth International Conference on Language Resources and Evaluation – LREC 2012*, Istanbul, ELRA, 2237-2244. *Português De S.Tomé*. (Dissertação de Mestrado), Lisboa: Universidade de Lisboa, 2010.
- GIANI, D. Le discours directe rapporté dans l’italien parlé et écrit. In: SCARANO, A. (Org.). *Macrosyntaxe et Pragmatique: l’analyse de l’oral*. Roma: Bulzoni, 2003. p. 203-213.
- GIANI, D. Una strategia di costruzione del testo parlato: l’introduttore locutivo. In: LEONI, F. A. (Org.). *Atti del congresso “Il parlato italiano”*. Napoli: D’Auria, 2004. p. 84-97.
- GIVÓN, T. The binding hierarchy and the typology of complements. *Studies in Language*, n. 4: 333-378, 1980.
- GIVÓN, Talmy. *Syntax: an introduction*. Vol. 1. Amsterdam-Philadelphia: John Benjamins, 2001.
- GONÇALVES, Perpétua. *A construção de uma gramática de português em Moçambique: aspectos da estrutura argumental dos verbos*. PhD Dissertation. Faculty of Arts, University of Lisbon, 1990.
- GONÇALVES, R. *Propriedade de subcategorização verbal no português de S. Tomé*. 2010. Dissertação (Mestrado em Linguística), Universidade de Lisboa, 2010.
- GONÇALVES, J. B.; VELOSO, R. Spoken Portuguese: Geographic and Social Varieties. In GÄRTNER, E et al. (Eds.), *Estudos de gramática portuguesa*. Frankfurt am Main: TFM, p. 257- 266, 2000.
- GONÇALVES, Perpétua; CHIMBUTANE, Feliciano O papel das línguas bantu na génese do português de Moçambique: o comportamento sintático de constituintes locativos e direcionais, *Papia*, 14, 7–30, 2004.
- GONÇALVES, R.; HAGEMEIJER, T. O português num contexto multilingue: o caso de São Tomé e Príncipe. *Revista Científica da Universidade Eduardo Mondlane*, v.1, n.1, p. 84-103, 2015.
- GONÇALVES, R.M.G. *Construções ditransitivas no português de São Tomé*. 2016. Tese de Doutorado. Universidade de Lisboa (Portugal), 2016.
- GONÇALVES, P. *Lusofonia em Moçambique: com ou sem glotofagia?*, Comunicação apresentada no II Congresso Internacional de Linguística Histórica–Homenagem a Ataliba Teixeira de Castilho,São Paulo, 2012.
- GONÇALVES, P. Aspectos Morfossintáticos da gramática do português de Moçambique: a concordância nominal e verbal. *Cuadernos de La Alfal*, n.7, 2015, p. 9-16.
- HAGEMEIJER, Tjerk. 2. O português em contacto em África. In: *Manual de linguística portuguesa*, editado por Ana Maria Martins; Carrilho, Ernestina, 43-67. Berlin/Boston: De Gruyter, 2016.

- HALFORD, B. The complexity of oral syntax', in B. Halford and H. Pilch (eds.), *Syntax gesprochener Sprachen* (Tubingen: Gunter Narr), 1990, p. 33-43.
- HART, J.; R. COLLIER; A. COHEN. *A perceptual study of intonation*. Cambridge, Cambridge University Press, 1990.
- HENGEVELD, K. & MACKENZIE, J. L. *Functional Discourse Grammar: A typologically based theory of language structure*. Oxford: Oxford University Press, 2008.
- HLIBOWICKA-WEGLARZ, Barbara. O sistema de tempo, modo e aspecto no crioulo de cabo verde (ilha de santiago). *Romanica olomucensia*, xv, p. 77, 2005.
- HOLM, J. The genesis of the Brazilian Vernacular: insights from the indigenization of portuguese in Angola. *Papia*, v. 19, p. 93-122, 2009.
- HOPPER, P. J.; THOMPSON, S. A. Projectability and Clause Combining in Interaction. In: LAURY, R. (Ed.). *Crosslinguistic Studies of Clause Combining: The Multifunctionality of Conjunctions*. Amsterdam: John Benjamins, 2008. p. 99-124.
- HUNT, K. W. 'Recent measures in syntactic development', *Elementary English*, 1966, p. 43: 732-9.
- INBAR, A. Is subordination viable? The case of Hebrew še 'that'. *Chimera*, v. 3, n. 2, p. 287-310, 2016. Disponível em: <https://revistas.uam.es/index.php/chimera/article/view/6517/7333>.
- INVERNO, L. *Angola's Transition to Vernacular Portuguese*. (Dissertação de Mestrado). Universidade de Coimbra, 2006.
- INVERNO, Liliana. *A transição de Angola para o português vernáculo: Estudo morfosintático do sintagma nominal*. In Ana M. Carvalho (ed.), *Português em contato*, 87-106. Madrid, Frankfurt: Iberoamericana, Editorial Vervuert, 2009.
- JON-AND, Anna. *Variação, contato e mudança linguística em Moçambique e Cabo Verde: A concordância variável de número em sintagmas nominais do português*, tese de Doutorado, Universidade de Estocolmo, 2011.
- JUSTINO, Víctor. *A distribuição e a expressão gramatical do futuro do conjuntivo no Português de Moçambique*, Dissertação de Mestrado, Faculdade de Letras da Universidade de Lisboa, 2011.
- KATO, M. 1993. Recontando a história das relativas em uma perspectiva paramétrica. In. ROBERTS, I. & KATO, M. (orgs.) *Português Brasileiro: uma viagem diacrônica. Homenagem a Fernando Tarallo*. Campinas: Unicamp. ed. 1996.
- KROLL, B. 'Combining ideas in written and spoken English: A look at subordination and coordination', in E. Ochs Keenan and T. L. Bennett (eds.), *Discourse across Time and Space* (Southern California Occasional Papers in Linguistics, 5; Los Angeles: University of Southern California), 1977, p. 69-108.
- LADD, Robert, D. *Intonational phonology*. Cambridge: Cambridge University Press,
- LEIRIA, Isabel. Português língua segunda e língua estrangeira: Investigação e ensino. Idiomatico. *Revista Digital de Didáctica de PLN*. Centro Virtual Camões, 2014.

- LINELL, P. 'The impact of literacy on the conception of language: The case of linguistics', in R. Saljo (ed.), *The Written World* (Berlin: Springer), 1988, p. 41-58.
- LOPES, Francisco João; CAMPOS, Ednalvo Apóstolo. A expressão pronominal no português brasileiro e no português falado em Cabo Verde—trilhando possíveis (as) simetrias. *PAPIA-Revista Brasileira de Estudos do Contato Linguístico*, v. 25, n. 2, p. 319-345, 2015.
- LOPES, Francisco João; OLIVEIRA, Marcia Santos Duarte. Estudos sobre o português falado em Cabo Verde: o 'estado da arte'. *O português na África atlântica*. São Paulo: HUMANITAS/FAPESP, p. 101-138, 2018.
- LOPES, Norma; BAXTER, Alan. A concordância verbal variável no português dos tongas, *Papia*, 21:1, 39–50, 2011.
- LOPES, Ana Keyla Carmo; SOARES, Maria Elias. *O processo de alfabetização em língua portuguesa em angola e em guiné-bissau*. ANAIS DO VI SERGEL, p. 15., 2017.
- LYONS, J. *Semantics I*. Cambridge: Cambridge University Press, 1977.
- MARQUES, Irene Guerra. *Algumas considerações sobre a problemática linguística em Angola*, in: Instituto de Língua e Cultura Portuguesa (ed.), Actas do Congresso sobre a situação actual da língua portuguesa no mundo, Lisboa, ICALP, 205–223, 1983.
- MARTINS, Aracy Alves; GOMES, Silvestre Filipe; CÁ, Virgínia José Baptista. Letramento (s)/Alfabetização em contextos multilíngues de Angola e Guiné-Bissau. *Educação em Revista*, v. 32, p. 391-412, 2016.
- MILLER, J.; WEINERT, R. *Spontaneous spoken language: syntax and discourse*. Oxford: Oxford University Press, 1998. 457p.
- MITTMANN, M. *O C-ORAL-BRASIL e o estudo da fala informal: um novo olhar sobre o tópico no Português do Brasil*. 2012. 248f. Tese (Doutorado em Estudos Linguísticos) – Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2012.
- MOISÉS, L.; CANDE, E.; JESUS, J. Geografia linguística de Moçambique. *Padronização da Ortografia de Línguas Moçambicanas: Relatório do III Seminário*. Maputo: Centro de Estudos Africanos/UEM, p. 279-293, 2012.
- MOLLICA, Maria Cecilia. Anáforas em relativas no português do Brasil. *ALFA: Revista de Linguística*, 1997.
- MONEGLIA, M.; RASO, T. Notes on Language into Act Theory (L-Act). In: RASO, T; MELLO, H. (Ed.). *Spoken corpora and linguistic studies*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 2014. p. 468-494.
- MONEGLIA, M.; CRESTI, E. C-ORAL-ROM–Prosodic Boundaries for Spontaneous Speech Analysis. *Spoken Language Corpus and Linguistic Informatics*. Amsterdam, Benjamins, Usage-Based Linguistic Informatics, v. 5, p. 89-113, 2006.
- NASCIMENTO, M. F. B. do. O Corpus de Referência do Português Contemporâneo e os projectos de investigação do Centro de Linguística da Universidade de Lisboa sobre variedades do português falado e escrito, In Gärtner, E., Hundt, C. and Schönberger, A. (eds.), *Estudos de Gramática Portuguesa (I)*, Biblioteca Luso-Brasileira, Centro do Livro e do Disco de Língua Portuguesa, Frankfurt am Main, 2000, p. 185-200.

- NASCIMENTO, Maria Fernanda Bacelar do et al. Aspectos de unidade e diversidade do português: as variedades africanas face à variedade europeia. *Veredas: Revista da Associação Internacional de Lusitanistas*, n. 9, p. 35-59, 2008.
- NASCIMENTO, M. F. B. do, PEREIRA, L.; SARAMAGO, J. Portuguese corpora at CLUL. In: International Conference on Language Resources (LREC), 2, 2000, Athens, Greece. *Proceedings...*, v. III, 2000, p. 1603-1607.
- NASCIMENTO, M. F. B.; PEREIRA, L.; ANTÔNIA, E. ;GONÇALVES, J. B.; OLIVERIA, S. Aspectos de unidade e diversidade do Português: as variedades africanas face às europeias. *Veredas*, Porto Alegre, n. 9, p. 35-60, 2008.
- NDOMBELE, E. D., & TIMBANE, A. A. O ensino de língua portuguesa em Angola: reflexões Metodológicas em contexto multílingue. *fólio - Revista De Letras*, 12, 1. <https://doi.org/10.22481/folio.v12i1.6604>
- NENCIONI, Goivani. *Di scritto e di parlato: Discorsi linguistici*. Bologna: Zanichelli, 1983.
- NETO, Aquiles Tescari; PEZATTI, Erotilde Goreti. A ordenação dos adverbiais modalizadores epistêmicos no português europeu falado: uma abordagem funcional. *Revista Letras*, v. 65, 2005.
- NHATUVE, Diocleciano João Raúl; FONSECA, Maria do Céu. Aspectos da sintaxe do português falado no sul de Moçambique. *Revista de Letras*, v. 11, p. 145-156, 2012.
- NGUNGA, Armindo. Interferências de línguas moçambicanas em português falado em Moçambique. *Revista Científica da Universidade Eduardo Mondlane*, v. 1, p. 7-20, 2012.
- NOONAN, M. Complementation. In: SHOPEN, T. (ed.). *Language typology and syntactic description*. Cambridge: Cambridge University Press, 42-139, 1985.
- PEIXOTO, C. M. M.; CARIOCA, C. R. *As Representações linguísticas dos académicos guineenses: uma reflexão sobre o estatuto da língua portuguesa fundamentada no corpus do PROFALA*. Congresso Internacional de Dialectologia e Sociolinguística, Belém. Anais... Belém: UFPA, p. 438-450, 2012.
- PERES, J. A.; MÓIA, T. (1995) *Áreas Críticas da Língua Portuguesa*. Lisboa: Editorial Caminho, 2.^a edição, revista, 2003.
- PERL, Matthias. Le portugais et le créole portugais en Afrique. In Massa & Perl (eds.), p. 9-27, 1989.
- PEZATTI, Erotilde Goreti. A ordem de palavras e o caráter nominativo/ergativo do português falado. *ALFA: Revista de Linguística*, 1993.
- RABÊLO, S. F. *Sintagmas locativos no português de Moçambique e do Brasil: o papel do contato de línguas* 2016. 96 f. Dissertação (Mestrado em Linguística) - Universidade de Brasília, Brasília, 2016.
- QUIRK, R., LEECH, G., SVARTVIK, J. *A Comprehensive Grammar of the English Language*. London: Longman, 1985.
- RASO, T. Fala e escrita: meio, canal, consequências pragmáticas e linguísticas. *Domínios de Lingu@Gem*, v. 7, p. 12-46, 2013.

- RASO, T; MELLO, H. (Org.). *C-ORAL BRASIL I: corpus de referência do português brasileiro falado informal*. Belo Horizonte: UFMG, 2012.
- RASO, T.; ULISSES, A. J. Tópico e apêndice no português do Brasil: algumas considerações. *Revista de Estudos da Linguagem*, v. 16, n. 1, p. 247-262, 2008.
- ROCHA, B. *A unidade informacional de introdutor locutivo no português do Brasil: uma análise baseada em corpus*. (Dissertação de Mestrado). Belo Horizonte: Universidade Federal de Minas Gerais, 2011.
- _____; RASO, T. A unidade informacional de Introdutor Locutivo no Português do Brasil: uma primeira descrição baseada em corpus. *Domínios da Linguagem*, v. 5, n. 1, p. 327-343, 2011.
- RUBIO, Cássio Florêncio. *Padrões de concordância verbal e de alternância pronominal no português brasileiro e no português europeu: estudo sociolinguístico comparativo*. Cultura Acadêmica, 2012.
- SANTOS, Vinícius Gonçalves dos. *Aspectos prosódicos do português de Guiné-Bissau: a entoação do contorno neutro*. 2015. Tese de Doutorado. Universidade de São Paulo, 2015.
- SAID; ALI. *Gramática histórica da língua portuguesa*. 7ª ed., São Paulo: Edições Melhoramentos, 1971.
- SANTOS, S. M.; BOSSAGLIA, G. Morphosyntactic, prosodic, functional and distributional description of the information unit of parenthesis in spoken Brazilian Portuguese. *Caletroscópio*, Belo Horizonte, v. 6, p. 37-63, 2018.
- SANTOS, I. da S.; TIMBANE, A. A. Estudo comparativo sobre as escolhas lexicais no português moçambicano e brasileiro: o caso dos verbos e substantivos comuns. *Língu@Nostr@*, Canoas, v. 4, n. 2, p. 23-43, jul.-dez. 2016.
- SCARANO, A. *Frasi relative e pseudo-relative in italiano: sintassi, semantica e articolazione dell'informazione*. Roma: Bulzoni, 2002. 174 p.
- SCHERRE, M. M. P. & NARO, A. J. Sobre a concordância de número no português falado do Brasil. In Ruffino, Giovanni (org.) *Dialettologia, geolinguística, sociolinguística*. (Atti del XXI Congresso Internazionale di Linguistica e Filologia Romanza) Centro di Studi Filologici e Linguistici Siciliani, Università di Palermo. Tübingen: Max Niemeyer Verlag, 5:509- 523, 1998.
- SCHMIDT, T.; K. WÖRNER. EXMARaLDA: Creating, analysing and sharing spoken language corpora for pragmatic research. *Pragmatics* 19: p. 565–582, 2009.
- SCHULZ, G. *Die Bottroper Protokolle: Parataxe und Hypotaxe*. Munich: Max Hueber, 1973.
- SILVA, L. F. L. Focalização e ênfase no português brasileiro. In: 1 Congresso Internacional de Letras, Artes e Cultura: linguagem, memória e arte - interfaces, 2013, São João del-Rei. 1 *Congresso Internacional de Letras, Artes e Cultura: linguagem, memória e arte - interfaces*, 2013. p. 1210-1217.
- SILVA, L. F. L. *Negação verbal no português brasileiro: aspectos teórico-metodológicos em estudo baseado em corpus*. (Dissertação de Mestrado). Belo Horizonte: Universidade Federal de Minas Gerais, 2016.
- SILVA, L. F. L. *Sintaxe da fala, probabilidade e cognição: uma proposta integrada de investigação em estudo sobre NPs baseado em corpus*. (Tese de Doutorado). Belo Horizonte: Universidade Federal de Minas Gerais, 2020.

- SILVA, A. E. M. S. *Estudar a sintaxe da fala com base em corpora: a realização das orações completivas e adverbiais no inglês americano, italiano e português brasileiro*. 2019. 106 f. Trabalho de Conclusão de curso. Belo Horizonte: Universidade Federal de Minas Gerais, Belo Horizonte, 2019.
- SILVA, A. E. M.; BOSSAGLIA, G. A subordinação na fala espontânea do inglês americano: um estudo das orações completivas. *Entrepalavras*, Fortaleza, v. 9, n. 3, p. 13-37, set.dez/2019.
- SOMA ANDRIANO, P. Omissão da marca de plural /s/: uma realidade no Português falado em Angola. *Transversos: Revista de História*. Rio de Janeiro, n. 15, Abril. 2019.
- SORNICOLA, R. *Sul Parlato*. Bologna: Il Mulino, 1981.
- STROUD, Christopher; Perpétua, GONÇALVES (eds.). Panorama do português oral de Maputo (3 vols.). Maputo: Instituto Nacional do Desenvolvimento da Educação, *Cadernos de Pesquisa*, 22, p. 24, 27, 1997.
- SUELELA, David Jorge Lopes. *Complementação Finita No português De Angola: Para Uma gramática Da Frase*. 2019. Tese de Doutorado. Universidade de Evora (Portugal), 2019.
- TIMBANE, Alexandre António. Os estrangeirismos e os empréstimos no português falado em Moçambique. *Via Litterae* (ISSN 2176-6800): Revista de Linguística e Teoria Literária, v. 4, n. 1, p. 5-24, 2012.
- TIMBANE, Alexandre António. *A variação e a mudança lexical da língua portuguesa em Moçambique*, 2013.
- TIMBANE, Alexandre António. Que português se fala em Moçambique? Uma análise sociolinguística da variedade em uso. *Revista Vocábulo*, v. 7, 2014.
- THOMPSON, S. A. “Object complements” and conversation towards a realistic account. *Studies in Language*, v. 26, n. 1, p. 125–163, 2002.
- TUCCI, I. L’inciso: caratteristiche morfosintattiche e intonative in un corpus di riferimento. In: LEONI, F. A. et al. (Org.). *Il parlato Italiano*, 2003, Napoli. *Atti del Convegno Nazionale GSCP*. Napoli: M. D’Auria, 2004. p. 1-14. CD-ROM.
- TUCCI, I. *L’espressione lessicale nella modalità del parlato spontaneo: analisi del corpus C-ORAL-ROM italiano*. Tese de Doutorado – Università degli Studi di Firenze, Firenze, 2006.
- ULISSES, A. J. *A unidade de apêndice no português do Brasil*. Dissertação de Mestrado – Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2008.
- VASCONCELLOS, José Leite de. 1987 [1901]. *Esquisse d’une dialectologie portugaise*. Lisboa: Centro de Linguística da Universidade de Lisboa – INIC [3ed ed].
- VERÍSSIMO, V. O sujeito nulo em duas variedades africanas do Português. *Macapá*, v. 7, n. 2, 2º semestre, 2017.
- VILELA, Mário. Algumas tendências da língua portuguesa em África. In: Mário Vilela. *Ensino e língua portuguesa: léxico, dicionário, gramática*, p. 45-72. Coimbra: Almedina, 1995.
- _____. A língua portuguesa em África: tendências e factos. *Africana Studia* 1: 175-95, 1999.
- UNDOLO, Márcio Edu da Silva et al. *Caracterização da norma do português em Angola*, 2015.