UNIVERSIDADE FEDERAL DE MINAS GERAIS
Cássia Jacqueline Fernandes Oliveira

O APÊNDICE DE COMENTÁRIO NO PORTUGUÊS DO BRASIL – UMA ANÁLISE BASEADA EM *CORPUS*

Belo Horizonte
Faculdade de Letras da UFMG
2012

CÁSSIA JACQUELINE FERNANDES OLIVEIRA

O APÊNDICE DE COMENTÁRIO NO PORTUGUÊS DO BRASIL – UMA ANÁLISE BASEADA EM *CORPUS*

Tese apresentada ao Programa de Pós-Graduação em Estudos Linguísticos da Universidade Federal de Minas Gerais como requisito parcial para obtenção do título de Doutor em Linguística.

Área de concentração: Linguística Teórica e Descritiva

Orientador: Prof. Dr. Tommaso Raso

Belo Horizonte
Faculdade de Letras da UFMG
2012

Tese defendida publicamente no Programa de Pós-Graduação em Estudos Linguísticos e aprovada pela seguinte comissão Examinadora

_____

Prof. Dr. Tommaso Raso – Orientador - UFMG

_____

Profª Drª Helivane de Azevedo Evangelista – UNA

_____

Prof. Dr. Tufi Neder Neto - UFLV

_____

Prof. Dr. César Augusto da Conceição Reis - UFMG

_____

Profª Drª Heliana Ribeiro de Mello – UFMG

Belo Horizonte, 30 de Março de 2012.

_____

Profª Drª Célia Maria Magalhães

Coordenadora do Curso de Pós-Graduação em Estudos Linguísticos da Universidade Federal de Minas Gerais

AGRADECIMENTOS

A Deus, por me ter dado forças para superar mais essa etapa.

À minha família, pela sábia compreensão de minha ausência na presença. Em especial aos meus pais, marido e filho pelo amor incondicional.

Ao Prof. Dr. Tommaso Raso que me acolheu como orientanda e, com toda paciência, entendeu os meus limites. Obrigada pela disponibilidade, pontualidade e sapiência.

À Profª Drª Heliana Mello pelo exemplo de sabedoria, dedicação e delicadeza.

Aos meus amigos do C-ORAL-BRASIL, Bruna, Bruno Alberto, Bruno Rocha e Raissa a quem tantas vezes recorri e que sempre estiveram prontos a me atender; sobretudo, à Heloísa e Maryualê, que me deram tanto conforto em horas tão difíceis!

SUMÁRIO

LISTA DE ABREVIATURAS E SIGLAS **09**
LISTA DE SÍMBOLOS **11**
LISTA DE FIGURAS **12**
LISTA DE GRÁFICOS **15**
LISTA DE QUADROS **16**
LISTA DE TABELAS **17**
RESUMO **20**
ABSTRACT **22**
0. INTRODUÇÃO **23**
1. JUSTIFICATIVA **23**
2. OBJETIVOS DE PESQUISA
2.1 Objetivos Gerais 24
2.2 Objetivos Específicos 24
3. FALA E ESCRITA **25**
CAPÍTULO 1 – A TEORIA DA LÍNGUA EM ATO
1.1 A Teoria da Língua em Ato 28
1.2 A Teoria Perceptiva da entonação - 't Hart ; Collier; Cohen (1990) 32
1.2.1 Primeiro nível da entonação de F0 32
1.2.2 Segundo nível da entonação: perfis prosódicos 35
1.2.3 Terceiro nível da entonação: o padrão tonal 38
1.3 A Teoria de Cresti e o Modelo Perceptual IPO 39
1.4 A identificação dos enunciados de acordo com a Teoria da Língua em Ato 41
1.5 A força ilocucionária 43
1.6 A definição de enunciados versus a definição de sentenças 48
1.7 Enunciados simples e complexos: a estrutura informacional 51
1.8 As unidades Informacionais 54
1.8.1 Unidades Informacionais Textuais 54
1.8.1.1 A unidade de Comentário (COM) 55
1.8.1.2 A unidade de Tópico (TOP) 56
1.8.1.3 A unidade de Apêndice de Comentário (APC) 57
1.8.1.4 A unidade de Apêndice de Tópico (APT) 58

1.8.1.5 A unidade de Parentético (PAR) 59
1.8.1.6 A unidade de Introdutor Locutivo (INT) 62
1.8.2 Unidades Dialógicas ou não-textuais 63
1.8.2.1 A unidade de Incipitário (INP) 64
1.8.2.2 A unidade de Contativo (CNT) 65
1.8.2.3 A unidade de Fático (PHA) 66
1.8.2.4 A unidade de Alocutivo (ALL) 67
1.8.2.5 A unidade de Expressivo (EXP) 70
1.8.2.6 A unidade de Conector Dialógico (DCT) 71
1.9 A perda do isomorfismo 72
1.9.1 A unidade de Escansão (SCA) 72
1.9.1 Comentários Múltiplos (CMM) 73
1.9.2 A Estrofe 76
1.10 O esquema da Língua em Ato 80
1.11 Modalidade versus ilocução e atitude 82
1.12 Resumindo 84

CAPÍTULO 2 – A UNIDADE DE APÊNDICE DE COMENTÁRIO (APC)
2.1 Definição da Unidade de APC 85
2.1.1 Critérios Funcional, Entonacional e Distribucional 87
2.1.2 Classificação Informacional do APC 90
2.1.3 Correlatos Semânticos e Morfossintáticos 92
2.1.4 A Modalização e a Unidade de APC 92
2.2 As discussões sobre a unidade de APC 94
2.2.1 O estudo de Tucci, E (2006) 95
2.2.2 O estudo de Ulisses (2008) 99
2.2.3 Tópico e Apêndice de Comentário: duas unidades distintas? 104
2.2.3.1 A posição de Lombardi Vallauri (2008) 104
2.2.3.2 A posição de Cresti (2000) e Raso; Ulisses (2008) 106
2.2.3.3 Raso; Ulisses versus Lombardi Vallauri 107
2.3 Resumindo 111

CAPÍTULO 3 – METODOLOGIA DA CONSTRUÇÃO DO CORPUS E ANÁLISE DOS DADOS
3.1 O estudo a partir da Linguística de Corpus 112
3.1.1 A composição do C-ORAL-BRASIL 113
3.1.1.1 O cabeçalho 122
3.1.1.2 A qualidade acústica e o equipamento 123
3.1.1.2.1 A qualidade acústica 123
3.1.1.2.2 O equipamento 125
3.1.1.3 A formação dos transcritores 126
3.1.1.4 Treinamento para segmentação prosódica 127
3.1.1.5 A segmentação prosódica 132
3.2 O subcorpus de pesquisa 142
3.2.1 Prospecção e tabulação dos dados 146
3.2.2 A forma entonacional da unidade de APC 148
3.2.3 Detalhamento da metodologia utilizada 149
3.3 Uma comparação com o C-ORAL-ROM 155
3.4 Resumindo 158
CAPÍTULO 4 – CARACTERÍSTICAS DA FALA ESPONTÂNEA
4.1 Unidades de referência da fala espontânea 159
4.2 Estrutura dos enunciados 164
4.3 Resumindo 167
CAPÍTULO 5 – PROCEDIMENTOS DE ANÁLISE DOS DADOS RELATIVOS À UNIDADE DE APC
5.1 A distribuição dos APC na amostra 169
5.2 Análises Funcional, Entonacional e Distribucional 180
5.2.1 Critério Funcional 180
5.2.1.1 As repetições 182
5.2.1.2 As informações tardias 185
5.2.1.3 As retomadas textuais 186
5.2.1.4 Os preenchimentos 187
5.2.1.5 Casos duvidosos em relação à analise funcional 188
5.2.2 Critério Entonacional 190
5.2.2.1 Primeiro caso duvidoso em relação à análise

entonacional 195
5.2.2.2 Segundo caso duvidoso em relação à análise entonacional 199
5.2.3 Critério Distribucional 200
5.2.3.1 Presença de unidades entre COM e APC 204
5.2.3.2 Unidades que podem ser confundidas com a unidade de APC 206
5.2.3.2.1 APC versus PAR 206
5.2.3.2.2 APC versus COB 207
5.2.3.3 Dois APC consecutivos ou um APC escansionado? 211
5.2.4 Coda 213
5.3 Análise Morfossintática 215
5.4 O mini-corpus italiano 217
5.4.1 Distribuição do APC com unidades intercaladas 222
5.4.2 Dois Apêndices 223
5.5 Análise comparativa: APC no PB versus APC no italiano 225
5.6 Resumindo 229
CAPÍTULO 6 – CONCLUSÕES
6.1 A definição de APC e seus critérios de ocorrência 230
6.2 Análise contrastiva entre APC de PB e Italiano 233
REFERÊNCIAS BIBLIOGRÁFICAS 236
LISTA DE ANEXOS
Anexo 1 – Análise Informacional 242
Anexo 2 – Análise Acústica 260
Anexo 3 – Análise Morfossintática 263
Anexo 4 – Total de APC no PB 270

LISTA DE ABREVIATURAS E SIGLAS

ALL Alocutivo
AP Apêndice
APC Apêndice de Comentário
APT Apêndice de Tópico
AUX Auxílio Dialógico
CMM Comentários Múltiplos
COB Comentários Ligados
COEP Comitê de Ética em Pesquisa
COM Comentário
COM_r Comentário de Reportação
CNT Conativo
DCT Conector Discursivo
EMP *Empty*
EXP Expressivo
F0 Frequência Fundamental
INT Introdutor Locutivo
INP Incipitário
LABLITA *Laboratorio Linguistico del Dipartimento di Italianistica dell'Università di Firenze*
PAR Parentético
PB Português do Brasil
PHA Fático
PoS *Part of Speech*
SCA Unidade de Escansão
SN Sintagma Nominal
SP Sintagma Preposicional
SV Sintagma Verbal
SADV Sintagma Adverbial
TMT Tomada de Tempo
TOP Tópico
TPL Lista de Tópicos

LISTA DE SÍMBOLOS

(@) linha de metadados

(*) início de turno

(%) início de linha dependente

(ABC) identificação do informante

(//) quebra entonacional terminal; fim de enunciado

(/) quebra entonacional não-terminal; fim de unidade tonal interna ao enunciado

(+) enunciado interrompido

(< >) sobreposição de fala

([/nº]) *retracting* ou falha na execução do enunciado

(&) início de palavra interrompida

(&he) hesitação ou silêncio preenchido

(“ ”) citação

(hhh) comportamento paralinguístico

(xxx) palavra ininteligível

(yyy) palavra censurada

(*yyyy*) trecho de áudio não-transcrito

LISTA DE FIGURAS

Figura 1.1 - Movimentos de F0 do exemplo 1.1 33
Figura 1.2 - Ressíntese que elimina os movimentos não intencionais de F0 no exemplo 1.1 34
Figura 1.3 - Ressíntese que elimina alguns movimentos intencionais de F0 no exemplo 1.1 34
Figura 1.4 - Configuração de movimentos de F0 da locução "a mãe da Fafica". 35
Figura 1.5 - Configuração de movimentos de F0 da locução "no norte de Minas" 36
Figura 1.6 - Enunciado sem verbo 41
Figura 1.7 - Enunciado com verbo 42
Figura 1.8 - Enunciado sem verbo composto por um sintagma preposicional 42
Figura 1.9 - Ilocução de ordem 44
Figura 1.10 - Enunciado com ilocução de pergunta 44
Figura 1.11 - Enunciado com ilocução de resposta 45
Figura 1.12 - Unidade de Tópico e unidade de Comentário 46
Figura 1.13 - Unidade de Comentário e unidade de apêndice 47
Figura 1.14 - Enunciado complexo com pausa 49
Figura 1.15 - Enunciado complexo sem pausa 50
Figura 1.16 - Enunciado simples sem verbo 52
Figura 1.17 - Enunciado simples com verbo 52
Figura 1.18 - Enunciado complexo sem verbo 53
Figura 1.19 - Enunciado complexo com verbo 54
Figura 1.20 - Unidade de Tópico 56
Figura 1.21 - Unidade de Apêndice de Comentário 58
Figura 1.22 - Unidade de Apêndice de Tópico 59
Figura 1.23 - Parentético curto 61
Figura 1.24 - Parentético longo 61
Figura 1.25 - Introdutor Locutivo com meta-ilocução de discurso reportado 63
Figura 1.26 - Unidade de Incipitário 65
Figura 1.27 - Unidade de Conativo 66

Figura 1.28 - Unidade de comentário seguido por uma unidade de Fático 67
Figura 1.29 - Espectrograma do enunciado do exemplo 1.25 69
Figura 1.30 - Espectrograma do enunciado do exemplo 1.26, destacando o alocutivo (*Rena)* 69
Figura 1.31 - Unidade de Expressivo 70
Figura 1.32 - Unidade de Conector Dialógico 71
Figura 1.33 - Unidade de Escansão 73
Figura 1.34 - Locução de elenco 75
Figura 1.35 - Locução de alternativa 75
Figura 2.1 - Enunciado formado de uma unidade de COM 86
Figura 2.2 - Enunciado formado de uma unidade de COM e uma unidade de APC 87
Figura 2.3 - Unidade de Apêndice de Comentário 89
Figura 2.4 - Unidade de APC com mesma modalidade do COM 93
Figura 2.5 - Unidade de APC sem repetição de modalidade do COM 94
Figura 2.6 - Unidade de COB 109
Figura 2.7 - Unidade de APC 110
Figura 3.1 - Interface Web do LABLITA 146
Figura 3.2 - Visualização de documento CSV em planilha eletrônica 147
Figura 3.3 - Exemplo de Creaky Voice 150
Figura 3.4 - Enunciado com ruído 151
Figura 3.5 - Enunciado sem a realização da síntese 153
Figura 3.6 - Enunciado com a realização da síntese 153
Figura 5.1 - Unidade de APC após unidade de COB em enunciado complexo 195
Figura 5.2 - Unidade de APC após unidade de COB em estrofe 197
Figura 5.3 - Unidade de APC após unidade de COB em estrofe composta de mais unidades informacionais 198
Figura 5.3.1 – Unidade de APC depois de um COB em estrofe 198
Figura 5.4 - Caso duvidoso quanto à análise entonacional 199
Figura 5.5 - Unidade de APC após unidade de CMM 203
Figura 5.6 - Unidade de APC após unidade de COB 204
Figura 5.7 - Unidade de PAR em final de enunciado 207
Figura 5.8 - Representação de ‘e tudo mais’ como COB 208

Figura 5.9 - Representação de ‘e tal’ como COM 209
Figura 5.10 - Representação de ‘e tal’ como APC 210
Figura 5.11 - Representação de dois APC consecutivos 211
Figura 5.12 - Representação de um APC escansionado 212
Figura 5.13 - Exemplo número 1 de Coda 213
Figura 5.14 - Exemplo número 2 de Coda 214
Figura 5.15 - Exemplo número 3 de Coda 215

LISTA DE GRÁFICOS

Gráfico 3.1 - Balanceamento dos informantes por sexo 118
Gráfico 3.2 - Balanceamento dos informantes por idade 119
Gráfico 3.3 - Balanceamento dos informantes por nível de escolaridade 120
Gráfico 4.1 - Proporção de turnos e palavras em monólogos, diálogos e conversações 161
Gráfico 4.2 - Proporção de enunciados simples e complexos em textos dialógicos e monológicos 164
Gráfico 5.1 - Distribuição das unidades textuais, por tipologia, no PB 227
Gráfico 5.2 - Distribuição das unidades textuais, por tipologia, no Italiano 227

LISTA DE QUADROS

Quadro 1.1 - Padrão Prosódico vs Valores Informacionais 40

LISTA DE TABELAS

Tabela 1.1 - Tabela de referência das classes e dos tipos de atos de fala baseada no corpus LABLITA, traduzido de Moneglia (2011) 82
Tabela 2.1 - Frequência de variação informacional do total de APC nos dois modos: informal e formal 95
Tabela 2.2 - Variação morfossintática do total de APC nas tipologias informal e formal 97
Tabela 2.3 - Distinção de expressões isoladas no total de APC 97
Tabela 2.4 - Frequência de sintagmas sobre o total de construções sintáticas 98
Tabela 2.5 - Ocorrência de APC nas Orações Subordinadas 98
Tabela 2.6 - Classificação Informacional dos Textos bpudl10 e bpudl11 101
Tabela 2.7 - Comparação entre Tucci, E (2006) e Ulisses(2008) quanto à frequência de variação informacional do total de APC no modo informal 101
Tabela 2.8 - Ocorrência de expressões sintagmáticas e isoladas nos Textos bpudl10 e bpudl11 101
Tabela 2.9 - Frequência de sintagmas encontrados nos Textos bpudl10 e bpudl11 102
Tabela 2.10 - Percentual de ocorrência de expressões isoladas no bpudl10 e bpudl11 103
Tabela 3.1 - Propriedades das diferentes qualidades acústicas 124
Tabela 3.2 - Qualidade das gravações 124
Tabela 3.3 - Alterações no segmental realizadas na amostra etiquetada informacionalmente do C-ORAL-BRASIL 138
Tabela 3.4 - Diferença no total de quebras prosódicas entre as versões revisadas e etiquetadas das transcrições 138
Tabela 3.5 - Tipos de alterações na anotação de quebras realizadas durante a etiquetagem informacional 140
Tabela 3.6 - Características dos textos do subcorpus 143
Tabela 3.7 - Comparação entre todos os dados do Italiano e do PB quanto ao número de enunciados e palavras 156

Tabela 3.8 - Comparação entre os dados do italiano e do PB quanto aos contextos 156
Tabela 3.9 - Comparação entre os dados do italiano e do PB quanto ao contexto público 157
Tabela 4.1 - Características gerais da amostra 160
Tabela 4.2 - Distribuição de enunciados e padrões ilocucionários complexos com apenas unidades dialógicas, proporções em relação à frequência absoluta e ao total de enunciados complexos 167
Tabela 5.1 - Frequência de APC em comparação às demais unidades textuais presentes na amostra 170
Tabela 5.2 - Distribuição e proporção de enunciados com APC e descrição das situações gravadas na amostra 170
Tabela 5.3 - Características Informacionais do subcorpus de PB 177
Tabela 5.4 - Comparação entre o total de enunciados complexos com APC 178
Tabela 5.5 - Comparação entre o total de Padrões Ilocucionários e Padrões Ilocucionários com APC 179
Tabela 5.6 - Comparação entre o total de estrofes e estrofes com APC 179
Tabela 5.7 - Número de ocorrências funcionais dos APC 181
Tabela 5.8 - Número de ocorrências de APC com diferentes funções de repetição 184
Tabela 5.9 - Valores prosódicos das unidades de COM e APC 190
Tabela 5.10 - Número de ocorrências de APC com curva analisável ou não, sobreposição e ruído 193
Tabela 5.11 - Total de enunciados com APC com curva não analisável 194
Tabela 5.12 - Frequência de categoria informacional nas conversações 200
Tabela 5.13 - Frequência de categoria informacional nos diálogos 201
Tabela 5.14 - Frequência de categoria informacional nos monólogos 201
Tabela 5.15 - Ocorrência de unidades entre COM e APC 205
Tabela 5.16 - Frequência de sintagmas sobre o total de APC 215
Tabela 5.17 - Frequência de expressões nas conversações, diálogos e monólogos 216
Tabela 5.18 - Frequência de construções sintáticas nas conversações, diálogos e monólogos 217

Tabela 5.19 - Ocorrência de APC no minicorpus italiano 218
Tabela 5.20 - Características informacionais do minicorpus italiano 220
Tabela 5.21 - Comparação entre enunciados complexos e enunciados complexos com APC no italiano 221
Tabela 5.22 – Comparação entre padrões ilocucionários com unidades textuais e padrões ilocucionários com APC , no italiano 221
Tabela 5.23 – Comparação entre estrofes e estrofes com APC, no italiano 222
Tabela 5.24 - Número de ocorrências dos APC com unidades intercaladas no italiano 223
Tabela 5.25 – Características informacionais dos corpora de PB e Italiano 225
Tabela 5.26 – Total de unidades terminadas com APC no PB e Italiano 228

RESUMO

Esse estudo consiste em uma análise baseada em corpora da unidade informacional de Apêndice de Comentário e teve como objetivo estabelecer uma análise contrastiva dessa unidade no PB e no italiano. Essa pesquisa é sustentada pela Teoria da Língua em Ato (Cresti, 2000), segundo a qual um enunciado é definido como sendo a menor unidade possível de interpretabilidade pragmática. As fronteiras entre enunciados e suas unidades internas são delimitadas pela entonação. Uma das possíveis unidades internas é a unidade informacional de Apêndice de Comentário (APC), cuja função é a de estabelecer uma relação de integração com a unidade ilocucionária de COM (CMM ou COB). Pesquisas no italiano (Cresti, 2000; Tucci, E, 2006) e no PB (Ulisses, 2008) mostraram que a unidade de APC apresenta um perfil nivelado e descendente e que possui médias de F0 e intensidade inferiores às médias da unidade ilocucionária a qual é integrada. Essa investigação foi conduzida em um subcorpus de PB constituído de 20 textos do C-ORAL-BRASIL, de aproximadamente 1500 palavras cada, bem como no minicorpus italiano, composto também pelo mesmo número de textos, os quais compõem o C-ORAL-ROM. Para as análises prosódicas foram usados os softwares Winpitch e Praat, que nos permitiram identificar e manipular parâmetros de F0 e intensidade, bem como realizar a síntese no processo de verificação das unidades de APC e COB. A pesquisa demonstrou que no italiano há uma presença muito maior de unidades terminadas complexas**,** enquanto o PB apresenta muito mais enunciados simples. Esse é um aspecto muito interessante para a comparação entre as duas línguas. Mesmo considerando como baseline somente os enunciados complexos, a porcentagem de APC em italiano é quase 50% a mais do que em PB**.** O maior número de APC se deve principalmente ao padrões ilocucionários, mas também aos enunciados. Não há diferença quanto às estrofes. É interessante notar, ainda, que nas estrofes dos monólogos, quando a fala necessariamente se complexifica, os APC do PB superam os do italiano nos monólogos. Do ponto de vista entonacional, não aparecem diferenças entre as duas línguas. Do ponto de vista informacional, as proporções entre as várias funções nas duas línguas são

perfeitamente comparáveis. Do ponto de vista morfossintático, a comparabilidade entre as duas línguas é estrita.

Palavras-chave: Estrutura Informacional; Atos de Fala; Apêndice de Comentário.

ABSTRACT

This study is a corpus-based analysis of the information unit Appendix of Comment and aimed to establish a contrastive analysis of this unit in BP and Italian. This research is based on the Informational Patterning Theory (CRESTI, 2000), according to which an utterance is defined as the smallest unit possible of pragmatic interpretability. The boundaries between utterances and their internal units are delimited by intonation. One of the possible internal units is the information unit Appendix of Comment (APC), whose function is to establish a relation of integration with the illocutionary unit COM (CMM or COB). Research in Italian (Cresti, 2000, Tucci, 2006) and BP (Ulisses, 2008) showed that the APC has a flat and falling movement and F0 and intensity values lower than the average of the illocutionary unit to which it is integrated. This investigation was performed on a BP subcorpus consisting of 20 texts of the C-ORAL-BRASIL, of approximately 1,500 words each, and on an Italian subcorpus, also composed of the same number of texts, which make up the C-ORAL-ROM. The softwares WinPitch and Praat were used for the prosodic analyses and allowed us to identify and manipulate parameters of F0 and intensity, as well as to perform the synthesis in the process of verifying the APC and COB units.

Keywords: informational patterning; speech acts; appendix of comment.

0. INTRODUÇÃO

Este estudo se insere na Linguística de Corpus e se propõe a uma análise descritiva da unidade de Apêndice de Comentário (APC), no Português do Brasil (PB), a partir da fala espontânea, como documentado no C-ORAL-BRASIL[1]. A pesquisa fundamenta-se na Teoria da Língua em Ato (CRESTI, 2000).

Embora saibamos que cada vez mais os estudos linguísticos têm-se beneficiado do uso de corpora para a descrição de fenômenos linguísticos ou para a verificação de hipóteses acerca dos mesmos (SARDINHA, 2011, GONÇALVES, 2008, GRABE, 2004, BIBER, CONRAD & REPPEN, 1998, CASTILHO, 1990 entre outros), a ideia de se desenvolver mais uma pesquisa dessa natureza é oriunda da possibilidade de se ter uma observação voltada estritamente à fala espontânea com corpora originais por arquitetura e critério de segmentação, bem como de se poder estabelecer padrões de uso entre o PB e o italiano.

1. JUSTIFICATIVA

Ao longo da história, vários pesquisadores se propuseram a estudar a linguagem, principalmente a escrita, deixando em segundo plano a fala. Ainda que semelhante em alguns aspectos, sabe-se que cada uma delas possui as suas especificidades e que analisar a fala pela lente da escrita é um equívoco.

Sem dúvida, uma questão muito discutida pela linguística atual é compreender como o falante organiza a informação na fala, isto é, como se organiza a sua estrutura informacional. A Teoria da Língua em Ato, que serve de arcabouço teórico para esse estudo, foi desenvolvida para lidar com essas questões, inserindo o estudo da estrutura informacional dentro daquele dos atos de fala (AUSTIN, 1962).

[1] O projeto C-ORAL-BRASIL (Raso e Mello, 2009 e 2010) é coordenado por T. Raso e H. Mello e financiado pelo CNPq, pela FAPEMIG, pela UFMG. O projeto será apresentado no capítulo 3, desse trabalho.

Este trabalho representa mais um passo na descrição da estrutura informacional do PB com base no corpus C-ORAL-BRASIL e uma nova oportunidade de comparação inter-linguística, neste caso, com o italiano.

2. OBJETIVOS DA PESQUISA

2.1 Objetivos Gerais

Essa tese visa ao estudo da unidade informacional de Apêndice de Comentário (APC), na fala espontânea do Português Brasileiro, na diatopia mineira, principalmente em Belo Horizonte e regiões vizinhas. A análise terá como base uma amostra do corpus C-ORAL-BRASIL etiquetada informacionalmente.

É objetivo, também, contribuir para a formação do *corpus* C-ORAL-BRASIL, por meio da coleta, transcrição, segmentação e etiquetagem de seus textos da seção informal do corpus, bem como estabelecer uma análise contrastiva do comportamento da unidade de APC no PB e no italiano.

2.2 Objetivos Específicos

a) Verificar a frequência e as características do padrão informacional de enunciados que apresentam a unidade de APC.
b) Identificar e descrever as características prosódicas da unidade de APC.
c) Identificar e descrever a estruturação morfossintática das unidades de APC.
d) Investigar as diferenças entre diálogos, monólogos e conversações em relação à presença e às características da unidade de APC.
e) Identificar e classificar outras unidades que, em final de enunciado, ocupam a mesma posição da unidade de APC.
f) Comparar o comportamento do APC no PB e no italiano.

Para que se tenha uma visão geral desse estudo, apresentamos, a seguir, o conteúdo de cada capítulo.

O capítulo 1 versará sobre a Teoria da Língua em Ato (CRESTI, 2000), aporte teórico que sustenta essa pesquisa.

O capítulo 2 focalizará a análise da unidade de Apêndice de Comentário, objeto de estudo dessa pesquisa.

O capítulo 3 esboçará a metodologia utilizada nas análises do *corpus;*

O capítulo 4 explicitará as características inerentes à fala espontânea;

O Capítulo 5 terá como propósito a análise dos dados relativos à unidade de Apêndice de Comentário. E

o capítulo 6, apresentará as considerações finais, bem como as conclusões obtidas com base na análise dos dados.

3. FALA E ESCRITA

Na literatura, há várias formas de se conceber a relação entre fala e escrita. Uma delas, segundo Marcuschi (2001), supõe que as diferenças entre fala e escrita se dão dentro do *contínuo* da produção textual, e não na relação dicotômica de dois pólos opostos.

Assim, um determinado texto falado, por exemplo, uma conversação espontânea, representaria o gênero mais representativo dessa modalidade, não sendo aconselhável compará-lo ao gênero mais representativo da escrita, por exemplo, o artigo científico. Este contínuo distingue e, ao mesmo tempo, correlaciona os textos de cada modalidade quanto às estratégias de formulação textual que determinam o

contínuo das características que distinguem as variações das estruturas, as seleções lexicais, etc.

Para Marcuschi, é fundamental ter presente que as duas modalidades, oral e escrita, se relacionam num contínuo de semelhanças e diferenças com algumas preferências, mas não com regras exclusivas.

Outra perspectiva linguística muito difundida, segundo o autor, centra-se nas diferenças de código entre a língua falada e a escrita. Essa visão é responsável pelo preconceito que se tem em relação à fala, considerando-a uma versão desorganizada da escrita.

A perspectiva adotada nesse trabalho, entretanto, ao contrário da opinião defendida por Marcuschi, se opõe a uma visão de entre fala e escrita como contínuo.

Assim sendo, concebe que as diamesias falada e escrita possuem cada uma certas restrições a toda e qualquer produção textual desenvolvidas em cada uma delas, e que essas restrições estão diretamente relacionadas ao canal de veiculação do texto, que, embora atenuadas em alguns casos, continuam operantes.

Do ponto de vista discursivo, a distinção som-grafia é essencial para a relação fala-escrita. Enquanto na fala se tem sinal sonoro (canal: *ar*), na escrita o sinal é visual (canal: vários suportes estáveis); na fala, o falante e o interlocutor estão face a face e interagem entre si, na escrita quem escreve e quem lê estão separados e não interagem; na fala, o falante e o interlocutor compartilham o mesmo espaço e o mesmo tempo e são completamente envolvidos, com ruído; na escrita, quem escreve e quem lê pode estar dividido por continentes e séculos e isso acontece em silêncio; na fala, a motivação é basicamente afetiva e se manifesta em ilocuções, em ações verbalizadas; na escrita, as construções linguísticas são fruto de pensamentos verbalizados, não de ações verbalizadas; na fala, tem-se como referência o enunciado, uma entidade pragmática e prosódica, enquanto na escrita têm-se como referência a sentença, uma entidade sintática; a fala obedece a princípios pragmáticos; a escrita obedece a um princípio prevalentemente sintático;

na fala, o enunciado se coloca como a contraparte linguística da ação: cada enunciado carrega uma intenção comunicativa, uma ilocução, independente da presença de uma configuração sintática reconhecível; o enunciado se define como a menor unidade autônoma pragmaticamente, a menor unidade da comunicação e é identificável com base em um perfil entonacional percebível como terminal (‘t Hart-Cohen-Collier 1990). Na fala, cada enunciado realiza uma ação, visto que a natureza da fala é ilocucionária. Na escrita, o texto como um todo, carrega a potencialidade de uma ação em um dado contexto, um efeito perlocutivo.

Quanto ao processo de textualização, podemos dizer que a formulação textual se manifesta de forma diferenciada na fala e na escrita. Durante um turno de fala, o falante pode mudar de ideia sobre o que está dizendo naquele momento, pode confundir-se na pronúncia de uma palavra ou na concordância verbal ou nominal, pode usar uma palavra inadequada, por exemplo. Para solucionar tais problemas, o falante corrige a sua própria fala. Já na escrita, as correções podem não ser não vistas pelos leitores, pois o leitor tem acesso à versão final e não ao processo de produção. Diferentemente do discurso falado, na escrita temos tempo para ocultar nossas correções.

Outro aspecto típico da fala é a repetição. A repetição é uma das estratégias de formulação textual mais presentes na oralidade. Com a função de explicar, a repetição contribui para a organização discursiva, favorece os interlocutores e exerce uma força argumentativa. Calcula-se que na fala 70% de informação é repetida. Isso é necessário devido ao fato de que a interação falada se baseia na memória de curto prazo, mas seria imperdoável na escrita.

Capítulo 1

A TEORIA DA LÍNGUA EM ATO

Este capítulo tem como escopo a Teoria da Língua em Ato, (Cresti, 2000), aporte teórico que sustentará essa pesquisa, e apresentará também, em linhas gerais, a Teoria da fonética perceptual ('t Hart; Collier; Cohen, 1990), a qual se fundamenta Cresti (2000).

1.1 A Teoria da Língua em Ato

A Teoria da Língua em Ato (Cresti, 2000)[2] fundamenta-se em um estudo empírico da fala espontânea realizado pelo LABLITA (*Laboratorio Linguistico del Dipartimento di Italianistica dell'Università di Firenze*).

A fala espontânea, de acordo com essa perspectiva teórica, é considerada como toda a produção linguística sonora dialogada ou monologada em situação natural, realizada livremente, em contextos e situações comunicativas naturais, formais ou informais.

> A imposição de um molde de segmentação do texto escrito sobre o discurso falado leva o pesquisador a tratar os dados de fala de forma problemática, enviesando especialmente a análise das relações sintáticas no discurso falado. Apesar disso, poucos pesquisadores atentam a este fato e se dão conta da relevância de preservar os aspectos entonacionais da fala em suas transcrições. (Cresti apud Mittmann, em preparação).

Na escrita, todavia, segundo Moneglia (2011) é clara a identificação de unidades linguísticas maiores do que a palavra (unidades da estrutura argumental, sentenças, orações, termos nucleares e dependentes), pois a língua escrita pode ser

[2] Para um resumo da teoria e uma primeira aplicação ao português do Brasil, veja-se MAIA ROCHA, B. A Unidade Informacional de Introdutor Locutivo no Português Brasileiro: uma analise baseada em corpus**. (**Dissertação de Mestrado), 2011.

tranquilamente segmentada de acordo com critérios sintático. Na fala, ao contrario, não é possível utilizar estes mesmos critérios para identificar unidades de referência. Evidências de corpora orais têm mostrado que aproximadamente 30% dos enunciados não apresentam um verbo e não podem ser analisadas conforme parâmetros sintáticos empregados facilmente na escrita.

Em princípio, a unidade linguística que se percebe de maneira mais natural é o turno dialógico. Entretanto, segundo Cresti (2000), o turno dialógico não pode ser considerado como unidade fundamental de referência do discurso falado, porque os turnos apresentam uma ampla variação, podendo ser compostos de apenas uma palavra ou interjeição, ou mesmo de uma longa exposição. O conceito de turno é resultado de uma interpretação antes cognitiva do que linguística.

O exemplo, abaixo, ilustra um trecho de fala espontânea extraído da amostra (bfammm04) que serviu de análise para esse estudo. Como nosso objetivo por ora é apenas mostrar o problema de se considerar apenas o turno dialógico como referência da fala, optou-se por retirar todas as marcas de segmentação prosódica. Essas marcas serão apresentadas mais adiante, quando da apresentação dos critérios de transcrição e segmentação.

*REG: e começou assim então assim eu minha Nossa Senhora toma o Jo não antes né no carro eu ficava co Haroldo corre Haroldo ô meu Deus do céu pega meu filho na sua mão pega meu filho na sua mão e segura porque Nossa Senhora a siora que é mãe siora sabe sio' pega meu filho pega meu filho e cuida porque nũ tinha outro recurso nũ tinha mais ninguém então assim era Nossa Senhora e Jesus Cristo ali e mais ninguém né pega e toma aí entrou lá dentro tem uma incubadora ocupada
*BAL: uhh
*REG: porque tinha um neném que já tinha nascido eles lavaram o João puseram roupa e no bercinho sete meses e meio aí a médica lá no quarto não no na sala de es de cirurgia tal mas nũ tinha né a anestesia já não pegava por causa da adrenalina do coisa he as veia da gente some nessa hora né
*BAL: hum hum
*REG: se ocê nũ pega antes minha filha depois cê muito difícil então nũ

conseguiram pegar a veia mas a enfermeira tentou quinhentas e cinquenta vezes o que eu tive parecia que eu tinha levado uma surra de tão roxa que eu fiquei de trem que es ficaram acho que mais aflito do que eu tentando me furar me detonaram meus braço me arrebentaram comigo sabe a médica virou falou assim ah ela nũ tá tendo hemorragia nem nada desiste nũ precisa de colocar soro nela nem nada não e aí nem anestesia nem nada né porque a gente também nũ teve nũ teve que dar ponto nem nada né

*BAL: hum hum

*REG: só no útero mesmo que fechou o o bura eu nem sei o tanto que abre o é que é nem nada aí ea falou assim cê ia ter neném aonde eu falei lá no Life Center ela falou No' pa quem saiu do Life Center e veio parar no Santa Fé São José eu falei é vim aqui ter filho pelo SUS falei no SUS não minha filha no susto n' é SUS não é susto aí ea tá assim No' que chique nũ sei o quê aí o quarto nũ tinha banheiro não tinha telefone não tinha campainha não tinha nada era um quarto que es tavam iam arrumar aí eu assim depois que passou tudo né cê fica sozim em plena enfermaria Haroldo ainda ficou um to lá comigo e foi levar o Gabriel porque o Gabriel ficou assim super preocupado tal enquanto ele nũ me viu ele nũ ficou tranquilo sabe

*BAL: hum hum

Como se pode observar, ainda que predominantemente monológico, o exemplo acima retrata a interação entre dois interlocutores REG e BAL (dois conhecidos conversando) sobre o nascimento, no carro, de um filho prematuro de Regina. Mesmo não havendo mudanças em relação ao assunto, em um mesmo turno de fala, encontramos passagens de discurso reportado ("minha Nossa Senhora toma o Jo"; "pega meu filho pega meu filho e cuida") para o discurso direto, retificações e retomadas.

Parece ser complicado analisar os turnos de BAL usando a mesma perspectiva adotada para REG. Isso, porém, não exclui o fato de que seja possível realizar uma análise global do evento comunicativo, como proposto pela Análise da Conversação, todavia nosso objetivo é demonstrar aqui a necessidade de se estabelecer uma análise linguística comum a qualquer texto falado, dando ênfase ao caráter pragmático da fala.

A Teoria da Língua em Ato parte do princípio de que a unidade linguística da fala deve corresponder à unidade fundamental da atividade comunicativa, já que é esta atividade que 'sustenta' a fala. A unidade linguística fundamental da fala deve corresponder à unidade fundamental da atividade comunicativa: o ato de fala (Austin, 1962). Partindo do princípio de que a fala espontânea consiste na execução de ações, delimitar a unidade de referência da fala deve corresponder a identificar, no fluxo da fala, as sequências linguísticas que se apresentam como suficientes e autônomas do ponto de vista pragmático, isto é, as entidades linguísticas que veiculam ações. Estas unidades são identificadas com o componente linguístico, o ato locutório, do ato de fala, conforme a perspectiva de Austin (1962). Assim, o enunciado deve ser considerado como a unidade linguística básica da fala, pois corresponde ao componente linguístico de um ato de fala (CRESTI 2000, 2009a; MONEGLIA 2000, 2011; MONEGLIA; CRESTI 1993, 2006).

Esta afirmação fundamenta-se na hipóstese de que seja possível estabelecer uma equivalência entre unidades do domínio das ações humanos (atos) e unidades linguísticas (enunciados). Assim, o enunciado é tido como a "contraparte linguística da ação"; isto é, o ato locutório é a contraparte linguística do ato ilocutório, e é interpretável pragmaticamente em autonomia. Isso significa, entre outras coisas, que um enunciado não precisa necessariamente possuir um verbo, e pode, inclusive, ser composto por uma única interjeição, desde que, entoado de maneira a cumprir uma ilocução. Dessa forma, a identificação dos enunciados se realiza através de uma quebra entonacional percebida como conclusiva.

Antes de apresentarmos as peculiaridades da Teoria da Língua em Ato, acredita-se ser relevante compreendermos, em linhas gerais, a teoria fonética perceptual ('t Hart; Collier; Cohen, 1990), já que a mesma fundamentou os estudos de Cresti (2000).

1.2 A Teoria Perceptiva da Entonação - ‘t HART; COLLIER; COHEN (1990)

Na teoria de fonética perceptual, os autores buscam responder a perguntas como: 1) Quais unidades melódicas o ouvinte distingue? 2) Como o ouvinte estrutura as unidades melódicas na percepção da curva melódica? 3) Como o ouvinte integra informação textual e Mellodia, formando uma única mensagem linguística? A teoria elaborada busca estabelecer uma relação entre a entonação e a percepção, já que acreditam ser a percepção uma espécie de filtro que faz uma redução dos dados acústicos e contém muito mais informações do que os relevantes para a comunicação. Surge, então, o modelo IPO (Instituto de Pesquisa da Percepção) que considera ser percebido pelo falante, dentre todos os movimentos de *pitch* na fala, somente aqueles movimentos produzidos pelo locutor de forma intencional, mesmo que de maneira inconsciente. A teoria perceputal delimita três níveis de análise: o primeiro seria os movimentos de F0; o segundo nível os perfis prosódicos e o terceiro, o padrão tonal.

1.2.1 Primeiro nível de análise da entonação: os movimentos de F0

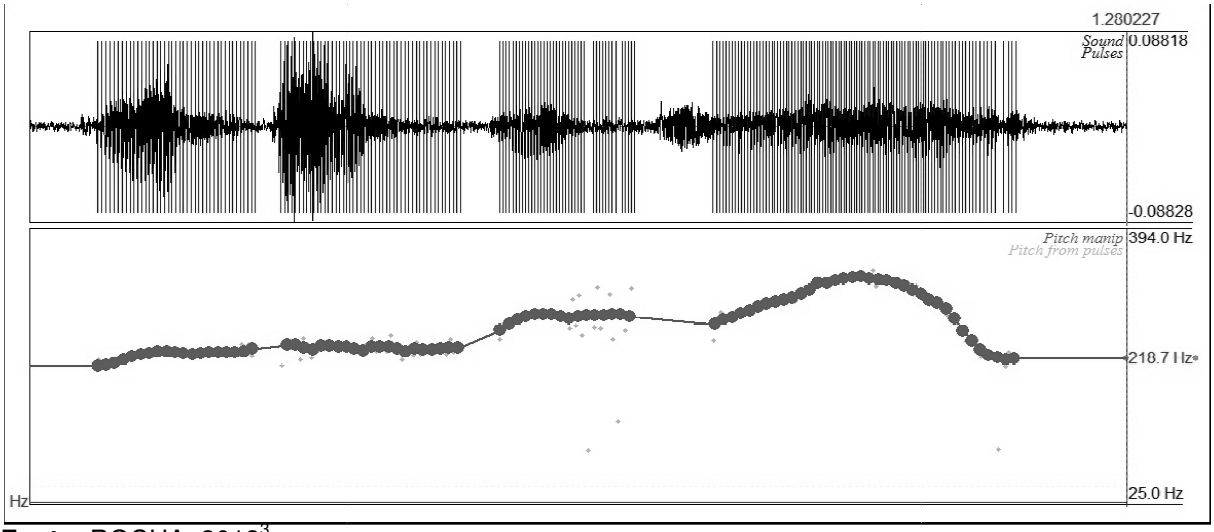
Segundo o modelo IPO, a identificação de movimentos voluntários e involuntários baseia-se na realização de ressínteses em que os movimentos de F0 são simplificados e em testes de percepção para averiguar quais das ressínteses são percebidas como idênticas as originais. As ressínteses que não causam diferenças de percepção são chamadas de *close copies* e os movimentos de F0 que foram eliminados na sua realização são tidos como movimentos involuntários (‘t HART; COLLIER; COHEN,1990).

A título de exemplo, procurou-se ilustrar por meio do software Praat (BOERSMA; WEENINK, 2010), ora movimentos voluntários, ora involuntários, como mostram as figuras abaixo.

Exemplo 1.1 (bfamdl04) (CD áudio 1.1)

*SIL: [29] deu palpite nenhum //=COM=$

Figura 1.1 – Movimentos de F0 do exemplo 1.1



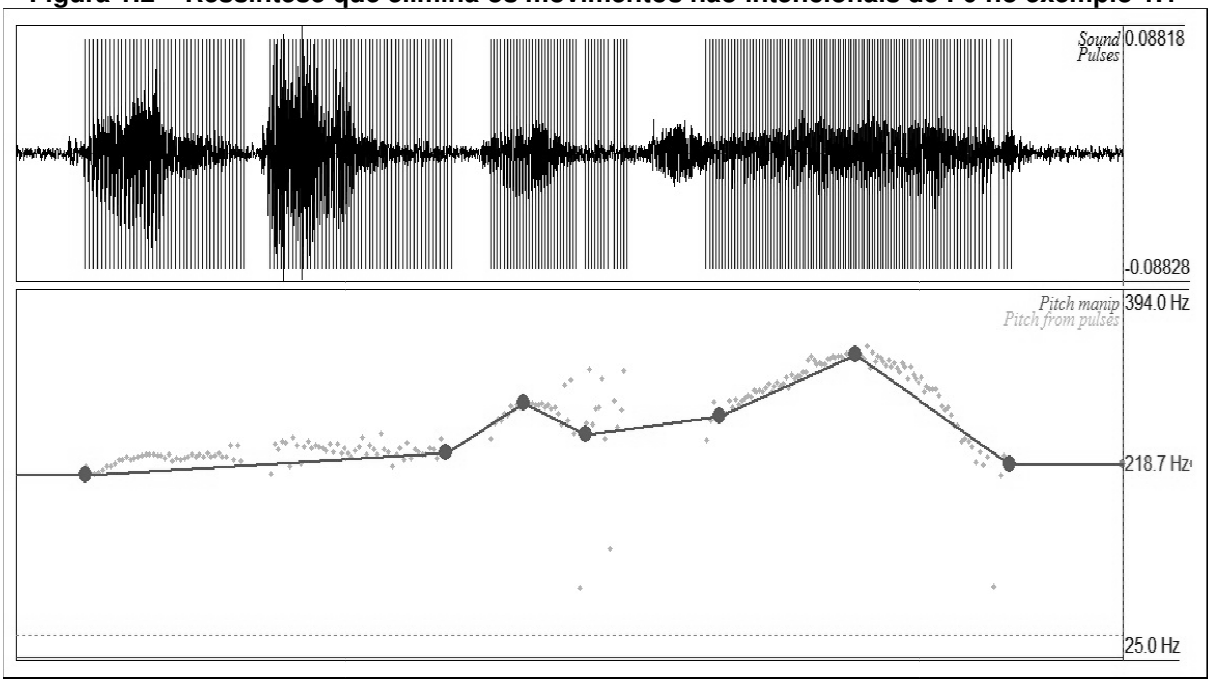
Fonte: ROCHA, 2012[3]

A figura 1.1 ilustra todos os movimentos de F0 do exemplo 1.1. Em cada enunciado, o asterisco (*) marca o início de um turno. A sigla composta de 3 letras maiúsculas identifica o falante (geralmente essa sigla é formada pelas três primeiras letras do nome do falante). As duas barras (//) marcam o final de um enunciado e a barra simples (/) as unidades tonais, conforme se verá mais adiante.

As figuras 1.2 e 1.3 ilustram ressínteses que simplificam a curva de F0 original, eliminando em um caso movimentos intencionais e no outro alguns dos intencionais. Na Figura 1.2 estão presentes somente os movimentos intencionais de F0 do exemplo 1.1, e na Figura 1.3, foram removidos alguns de seus movimentos intencionais. Ao se ouvirem os exemplos, fica evidente que a eliminação dos movimentos na figura 1.2 não causa diferença de percepção na curva melódica do trecho, todavia, a eliminação dos movimentos na figura 1.3 faz com que o falante de PB note uma diferença significativa em relação ao exemplo original.

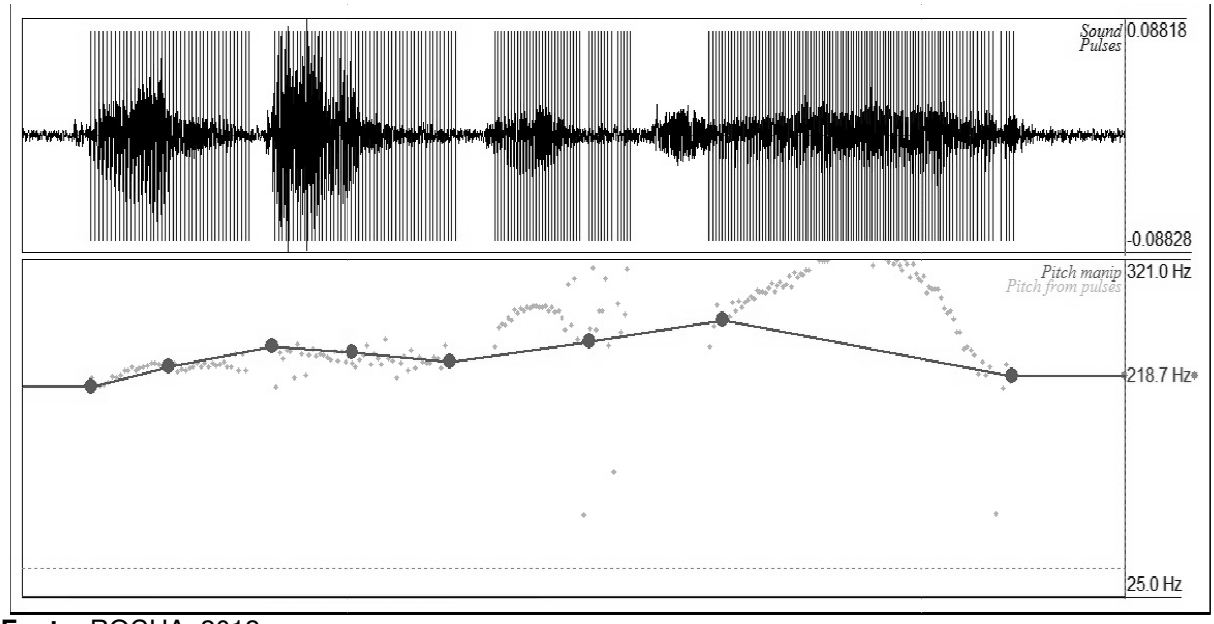
[3] Esse exemplo foi retirado da Dissertação de Rocha (2012).

Figura 1.2 – Ressíntese que elimina os movimentos não intencionais de F0 no exemplo 1.1



Fonte: ROCHA, 2012

Figura 1.3 – Ressíntese que elimina alguns movimentos intencionais de F0 no exemplo 1.1



Fonte: ROCHA, 2012

Considerando os movimentos de *pitch* produzidos de forma voluntária, as pesquisas do IPO mostram ainda que existem classes de movimentos que são avaliados da mesma forma pelos falantes. Ainda aqui não se está falando da função dos

movimentos na articulação da informação, mas simplesmente da curva melódica de partes do enunciado (‘t Hart; Collier; Cohen,1990).

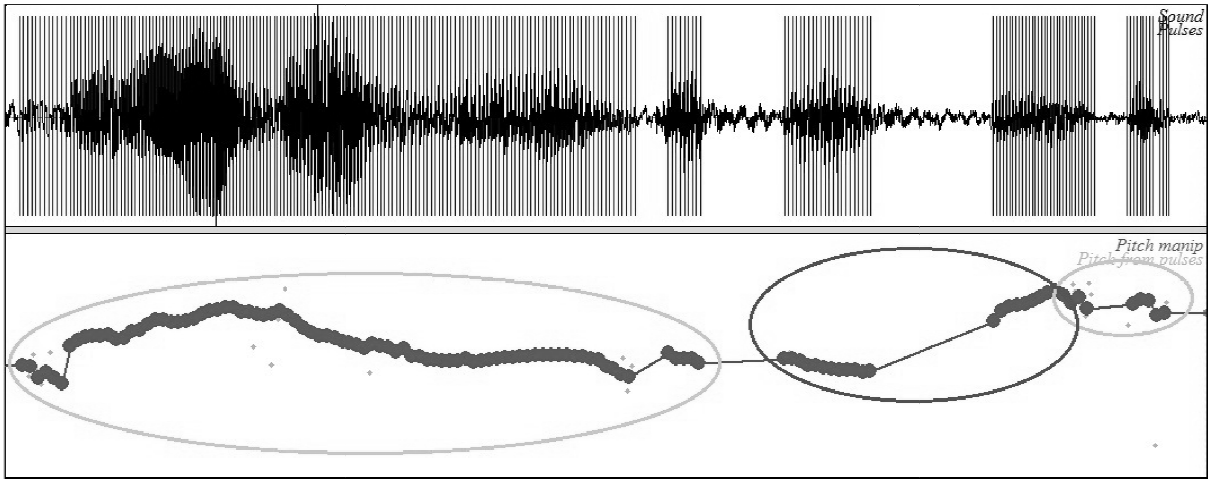
1.2.2 Segundo nível de análise da entonação: perfis prosódicos

As classes teóricas de movimentos, chamadas de *perfís prosódicos*, constituem o segundo nível de análise da entonação. Os perfís podem ser formados por movimentos similares (A, A', A''), configurações de movimentos (A+B, C+D) e movimentos compostos por partes opcionais e obrigatórias (a+A+b, B+a).

Exemplo 1.2 (bfamcv02) (CD áudio 1.2)
*TER: a mãe da Fafica / vai dar / o fogão //

A Figura 1.4 – correspondente à locução "a mãe da Fafica", do enunciado exemplificado em 1.2 – exibe uma configuação de movimentos prosódicos composta por um movimento opcional "a" (primeira elipse), por um movimento obrigatório “A” (segunda elipse) e por um movimento opcional "b" (terceira elipse).

Figura 1.4 – Configuração de movimentos de F0 da locução ´a mãe da Fafica´.



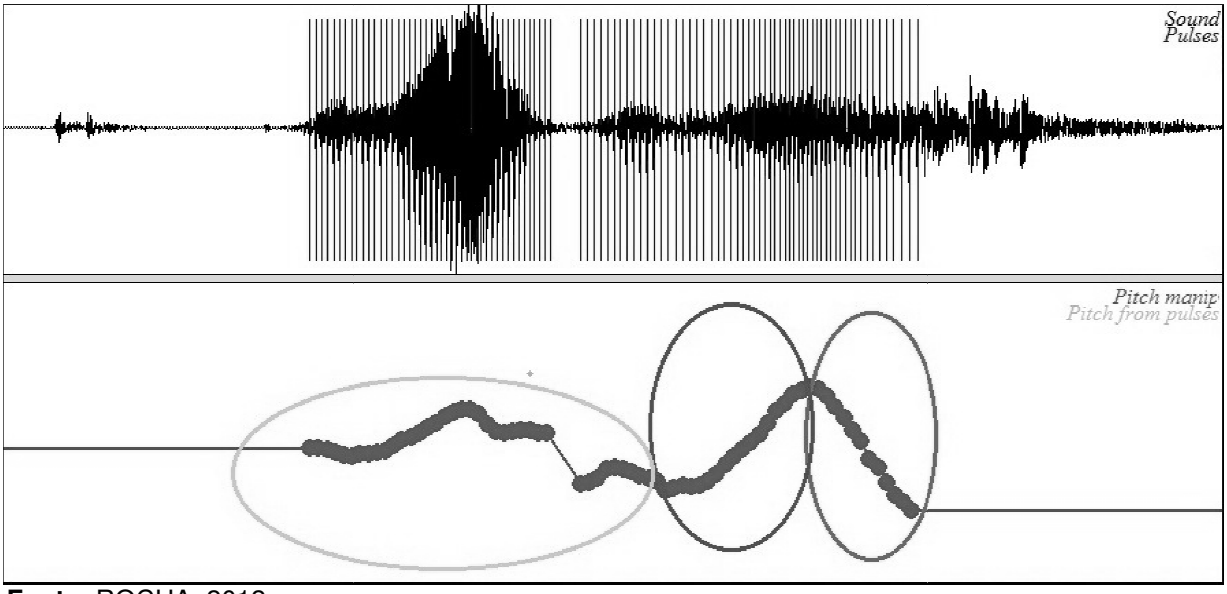
Fonte: ROCHA, 2012

Agora vejamos outro exemplo.

Exemplo 1.3 (bfammn01) (CD áudio 1.3)

*MAI: [3] no norte de Minas / existia um / &su [/1] meio aparentado com a minha esposa //

Figura – 1.5 Configuração de movimentos de F0 da locução "no norte de Minas"



Fonte: ROCHA, 2012

A Figura 1.5, por sua vez, mostra a configuração de movimentos prosódicos referente ao conteúdo locutivo "no norte de Minas" do exemplo 1.3. Como pode ser observado na figura, o perfil prosódico desse conteúdo locutivo é composto por um movimento opcional "b" (primeira elipse) seguida de uma parte obrigatória “A” (segunda elipse) e outra parte obrigatória "B" (terceira elipse). As configurações de movimentos prosódicos dos dois exemplos têm o mesmo perfil prosódico, o que significa dizer que são percebidos pelos falantes nativos do Português Brasileiro como sendo de uma mesma classe.

De acordo com os estudos desenvolvidos pelo LABLITA, as configurações de movimentos que compõem cada perfil podem ser formadas de:

a) um ou dois movimentos relevantes de F0;
b) uma porção opcional chamada *preparação*, anterior ao movimento relevante;
c) uma porção opcional chamada *coda*, posterior ao movimento relevante;

d) uma porção opcional chamada *ligação,* localizada entre o primeiro e o segundo movimentos relevantes.

Até agora, viu-se que movimentos de F0 e configurações de movimentos podem ser avaliados de uma mesma forma pelos falantes, formando *perfis prosódicos* – ou seja, classes teóricas de movimentos e configurações de movimentos avaliados da mesma forma pelos falantes. Todavia, é frequente que um enunciado seja composto por uma combinação de perfís prosódicos, e não por somente um deles. Dentro das possíveis combinações de perfís prosódicos de uma língua, as pesquisas do LABLITA mostram que os perfís se distinguem com base em características composicionais e funcionais[4], formando classes de perfís chamadas *unidades tonais*. Uma *unidade tonal* seria, então, uma classe de perfil prosódico com características funcionais e distributivas, específicas.

Considerando que os enunciados podem ser formados por mais de um perfil prosódico e que os perfis prosódicos se distinguem em função de propriedades distributivas e funcionais dentro do enunciado, existiriam então, para cada língua, modelos de composição de enunciado. Esses modelos seriam as possíveis combinações de perfis prosódicos que uma língua admite levando-se em conta as restrições distributivas e funcionais de cada classe de perfil. Cada modelo composicional de enunciado é chamado de *padrão tonal* e esse seria o terceiro nivel de análise da entonação na Teoria da Língua em Ato.

[4]Imagine-se, por exemplo, um certo perfil prosódico de tipo "X" com as seguintes características: (a) do ponto de vista distributivo, é obrigatório e pode localizar-se em qualquer posição do enunciado; (b) do ponto de vista funcional, presta-se à realização da ilocução. Da mesma forma, imagine-se um perfil de tipo "Y" que: (a) do ponto de vista distributivo, é opcional e deve localizar-se antes do perfil obrigatório; (b) do ponto de vista funcional, estabelece uma referência cognitiva para a realização da ilocução veiculada pelo enunciado.

1.2.3 Terceiro nível de análise da entonação: o padrão tonal

Um padrão tonal é um modelo composicional de enunciado baseado nas propriedades distributivas e funcionais dos perfís prosódicos. Da mesma forma, os perfís prosódicos com características funcionais e distributivas específicas são chamados de *unidades tonais*[5], definida como perfis prosódicos que possuem características funcionais e distributivas específicas.

Assim, com base em Cresti (2000, 2008) e Firenzuoli (2003), existem as seguintes classes funcionais de perfis:

a) unidades nucleares (*root*[6]), chamadas de *comment*, necessárias e suficientes à realização de um padrão;

b) unidades opcionais e de subordinação melódica (*prefix*), chamadas de *topic*, as quais devem vir anteriormente às unidades nucleares;

c) unidades opcionais e de subordinação melódica (*sufix*) à uma unidade *root* ou a uma unidade *prefix*, as quais posicionam-se após a unidade a que são subordinadas;

d) unidades variadas (*parenthesis*, *introducer*, *auxiliary*), subordinadas Mellodicamente e com características funcionais e distributivas variadas, as quais serão explicadas na seção 1.8.

A possibilidade de combinações dessas unidades perceptuais básicas é limitada. Em outras palavras, os padrões entonacionais básicos sempre são os mesmos e baseiam-se no contorno do *pitch*; contudo, é a combinação entre esses padrões básicos que possibilitam a produção de enunciados infinitos. Afirma-se, portanto,

[5] A unidade tonal está associada a características funcionais e distributivas dentro de um perfil prosódico (CRESTI, 2000).
[6] As classes funcionais de *root*, *prefix* e *sufix* foram introduzidas por 't HART; COLLIER; COHEN (1990).

que o número infinito de perfis diferentes é resultado de um número finito de padrões entonacionais básicos e essas regras variam de língua para língua e podem ser expressas em uma gramática da entonação representada pela seguinte fórmula: contorno melódico = (Prefixo) NÚCLEO (Sufixo).

1.3 A Teoria de Cresti (2000) e o Modelo Perceptual IPO

A pesquisa realizada por 't Hart; Collier; Cohen (1990) foi inovadora, uma vez que os autores reconheceram que a correspondência entre prosódia e sintaxe não é obrigatória, tampouco única. A partir desse ponto, Cresti (2000) faz uso da gramática entonacional de 't Hart e propõe a Teoria da Língua em Ato, sinalizando a necessidade de se fazer uma interpretação pragmática dos padrões entonacionais básicos.

Cresti acredita que é preciso entender o ato ilocucionário e para isso somente uma interpretação fonética não é suficiente. Segundo a autora, é necessária uma interação entre a fonética e a pragmática, sendo, portanto, papel da entonação atuar como interface entre os atos locutório e ilocutório. Essa interface se daria por meio da segmentação da fala em unidades mínimas interpretáveis pragmaticamente (os enunciados) e em unidades melódicas internas ao próprio enunciado (unidades informacionais).

Segundo Cresti (2000), o que é comum a cada ilocução é existência de uma forma obrigatória distribucionalmente livre (RAIZ), assim como definida em 't Hart-Collier-Cohen, independente da parte preparatória (PREFIXO), ou conclusiva (SUFIXO), ligada a um diferente tamanho (silábico) da locução (FIRENZUOLI, 2003).

A identificação dos diferentes perfis entonacionais é realizada por meio da análise dos movimentos da curva de F_0, e a cada configuração melódica desses movimentos resulta um perfil característico ao qual é possível atribuir um valor funcional. Logo, as diferenças micromelódicas que caracterizam cada movimento a

partir dos aspectos segmentais não são suficientes para levar a uma alteração do valor perceptivo da função de movimento que é determinado pelo núcleo. Assim sendo, um enunciado pode constituir-se de um padrão informacional simples, se composto apenas por uma unidade tonal, suficiente para a constituição desse padrão, e por isso definida como unidade nuclear (ROOT), ou, informacionalmente, unidade de Comentário (COM). Quanto pode constituir-se de um padrão informacional complexo, quando possui, além da unidade nuclear, outras unidades (PREFIX, SUFFIX ETC) dedicadas ao cumprimento de funções diferentes daquela de veicular a força ilocucionária. Cada unidade tonal tende a veicular uma unidade informacional, originando uma relação biunívoca entre unidade tonal e informacional.[7] Veja-se o esquema proposto por Cresti, baseado na comparação entre os padrões prosódicos de 't Hart e os valores informacionais.

Quadro 1.1 - Padrão Prosódico vs Valores Informacionais

Padrão Prosódico		**Valores Informacionais**
Núcleo	→	Comentário
(Prefixo) (Sufixo)	→	(Tópico) (Apêndice)
(Parentético)	→	(Parentético)
(Incipitário) (Fático)	→	(Incipitário) (Fático)

Fonte: Cresti; Moneglia (2010, p.15 – adaptado)

Para a Teoria da Língua em Ato, proposta por Cresti (2000), as unidades informacionais são identificadas no enunciado através de três critérios distintos: o critério funcional (função exercida pela unidade no enunciado), o critério entonacional (perfil entonacional característico de cada unidade) e o critério distribucional (posição da unidade no enunciado). Dessa forma, a junção desses três critérios possibilita a identificação das unidades informacionais da fala.

[7] Há casos em que isso não acontece, chamados de unidades de escansão. A subseção 1.9.1 tratará desse assunto.

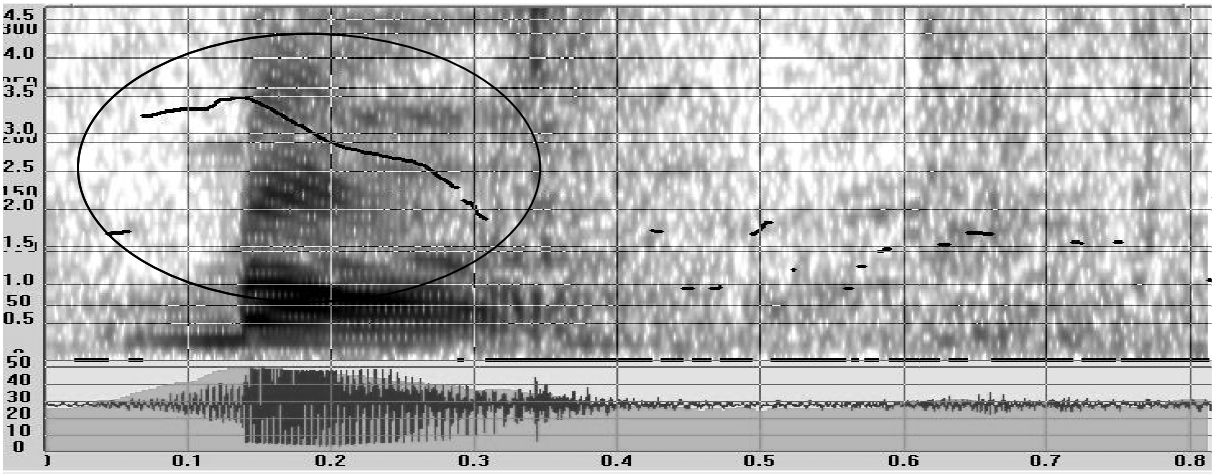
1.4 A identificação dos enunciados de acordo com a Teoria da Língua em Ato

A relação biunívoca entre enunciado e ilocução é baseada em uma interface entonacional. A cada enunciado, unidade mínima de significado pragmático, corresponde-se uma ilocução. Os exemplos a serem apresentados ilustrarão o fato de que a identificação de uma ilocução independe de aspectos sintáticos, podendo um enunciado ora ser composto por um advérbio, ora por uma oração canônica ou mesmo um sintagma preposicional ou até uma interjeição. A característica definidora do enunciado não está na completude semântica ou na expressão de uma predicação, mas sim na sua autonomia pragmática. Afirmar que o enunciado é uma unidade linguística autônoma, pragmaticamente, significa que é uma unidade pragmaticamente interpretável (CRESTI 2000, 2005; CRESTI; MONEGLIA 2010; MONEGLIA 2006, 2011).

Exemplo 1.4 (bfamdl01) (CD áudio 1.4)[8]

*FLA: <que hora que é> / seu estágio //
*REN: **bom** // a gente não marcou //[9]

Figura 1.6 – Enunciado sem verbo[10]

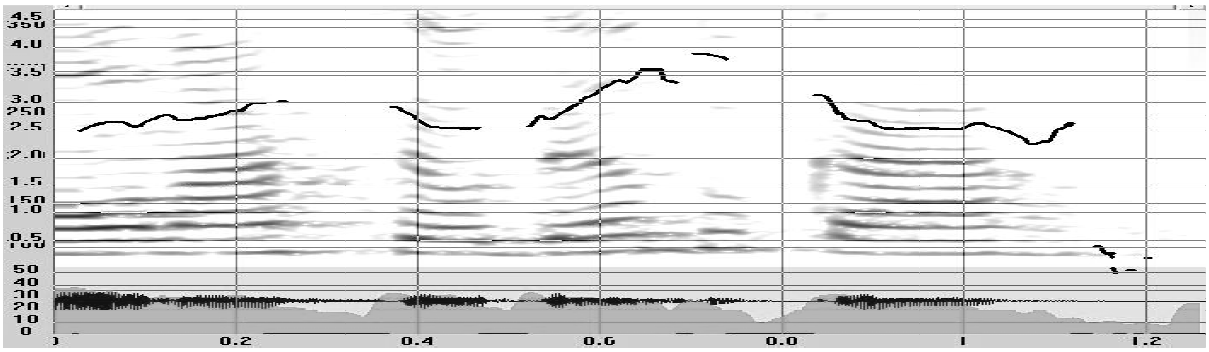


[8] Todos os exemplos em PB, que ilustram esse estudo, foram retirados do C-ORAL-BRASIL.
[9] Esclarecemos que daqui em diante, todas as figuras representadas pelos espectrogramas ilustram os enunciados em negrito, transcritos ortograficamente.
[10] A figura 1 mostra imagem extraída de parte da tela do *software* WinPitch, desenvolvido por Philippe Martin e disponível para *download* em http://www.winpitch.com. Esse *software* foi utilizado nesse trabalho para as análises acústicas e para o alinhamento dos textos. Os arquivos de áudio relativos a cada enunciado apresentado ao longo do trabalho estão disponíveis no CD anexo.

A elipse demarca, na figura, o conteúdo locutivo '*bom*'. Esse conteúdo é realizado em uma única unidade tonal, necessária e suficiente para a atribuição de autonomia pragmática. A Teoria da Língua em Ato a denomina como unidade de Comentário (COM).

Exemplo 1.5 (bfammn04) (CD áudio 1.5)
*REG: **João nasceu dentro do carro** //

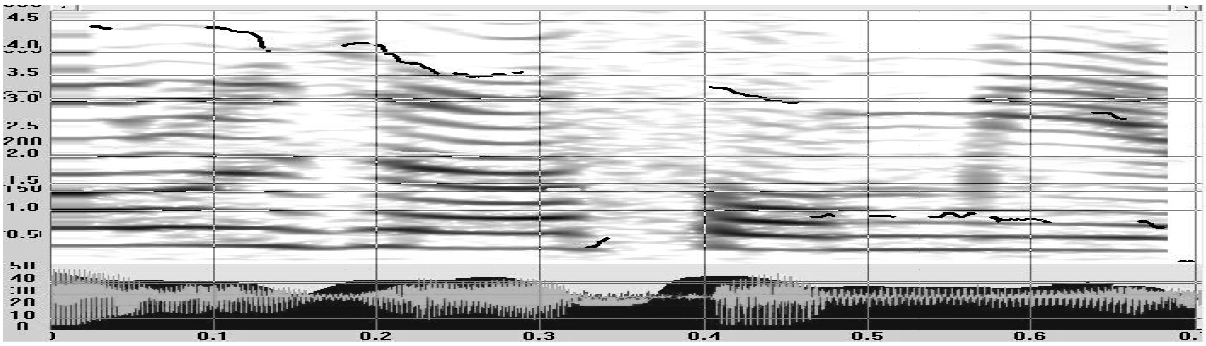
Figura 1.7 - Enunciado com verbo



A figura 1.7 ilustra o conteúdo locutivo '*João nasceu dentro do carro*'. Esse enunciado, assim como o exemplo 1.5, também é realizado através de uma única unidade tonal, cujo perfil lhe confere autonomia pragmática.

Exemplo 1.6 (bfamdl01) (CD áudio 1.6)
*REN: <o que> //
*FLA: **na aula de anatomia** //

Figura 1.8 – Enunciado sem verbo composto por um sintagma preposicional



Na figura 1.8 o mesmo processo se repete com o enunciado *na aula de anatomia.*

É por meio então da segmentação prosódica que se distinguem dois tipos de enunciado: simples e complexos. O enunciado será simples se executado em uma única unidade tonal, e complexo se executado em mais unidades tonais. Cada unidade tonal veicula, em princípio, uma unidade informacional, acarretando em uma relação, também, em princípio, biunívoca entre unidade tonal e informacional. De acordo com Cresti (2000), cada enunciado exprime uma única ilocução, e cada unidade tonal corresponde a uma única função informacional.

Somente um tipo de unidade informacional é obrigatória (Comentário), porque veicula a força ilocucionária e é a única interpretável em autonomia. As demais são opcionais, conforme poderá se verificar a partir da subseção 1.8.1.2. Dessa forma, os enunciados simples são compostos unicamente pela unidade informacional de Comentário (COM), enquanto os enunciados complexos, além da unidade de Comentário, possuem uma ou mais unidades, cuja função não será a de transmitir a força ilocucionária, mas de construir um padrão informacional que ajuda na realização do comentário.

1.5 A força ilocucionária

O tipo de força ilocucionária não está subordinado ao conteúdo locutivo, uma vez que podemos ter o mesmo conteúdo locutivo realizando diferentes ilocuções, como podem existir diferentes conteúdos locutivos realizados de maneira a veicular a mesma ilocução. Desse modo, a curva entonacional (junto a parâmetros pragmáticos, cognitivos e semiológicos)[11] é que permitirá a realização de uma determinada ilocução.

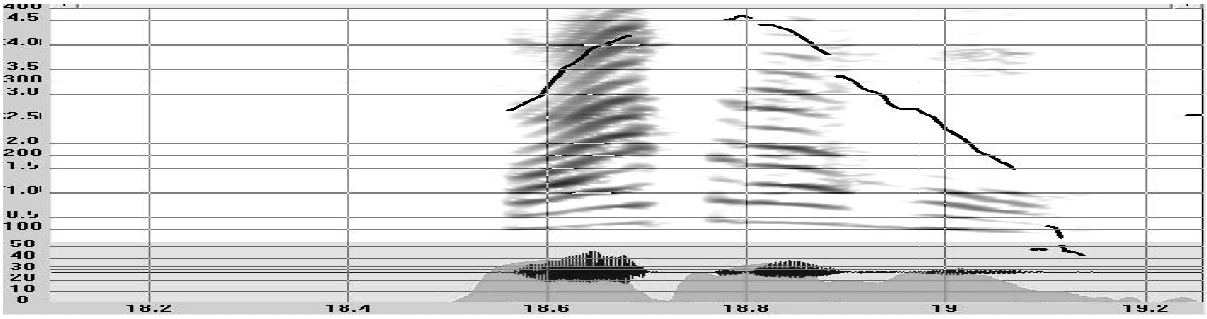
[11] Ver Cresti (2000) e Moneglia (2011).

Vejamos, abaixo, alguns exemplos cuja base do conteúdo locutivo é “vai pra Roma”. Observe-se que a realização de ilocuções de maneiras diferentes, através de diferentes padrões entonacionais, podem veicular estruturas informacionais distintas.

A figura 1.9[12] mostra um perfil entonacional executado em uma única unidade tonal, necessária e suficiente para o cumprimento da ilocução de ordem.

*XYZ: **vai pra Roma** //

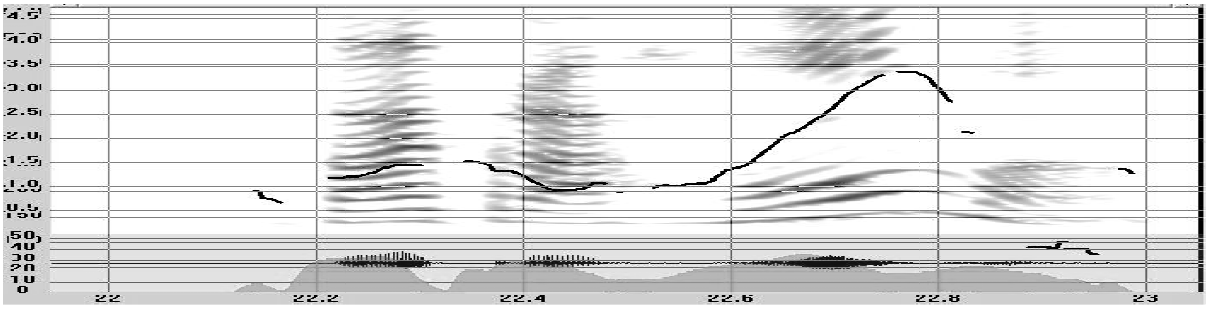
Figura I.9 – Ilocução de ordem



A figura 1.10 representa um enunciado com o mesmo conteúdo locutivo, porém com uma ilocução do tipo pergunta.

*XYZ: **vai pra Roma** //

Figura 1.10 – enunciado com ilocução de pergunta



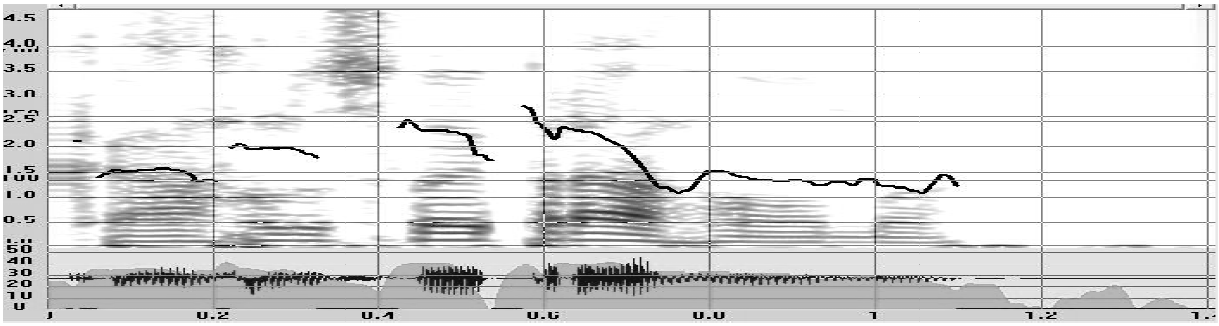
[12] As figuras 1.9 e 1.10 ilustram exemplos produzidos no laboratório de Fonética da UFMG.

Além do enunciado poder conter o mesmo conteúdo locutivo realizado de maneiras diferentes, em uma única unidade tonal, conforme se pode observar nas duas figuras acima, há também a possibilidade desse mesmo conteúdo se manifestar em duas unidades tonais distintas.

A figura 1.11 mostra um perfil entonacional executado em uma única unidade tonal. Essa unidade tonal corresponde à unidade informacional de comentário[13], necessária e suficiente para o cumprimento da ilocução de resposta expressa na letra B. Um possível contexto para realização apropriada desse enunciado poderia ser:

A: Para onde Carlos foi?
B: **Carlos foi pra Roma.**

Figura 1.11 - Enunciado com ilocução de resposta

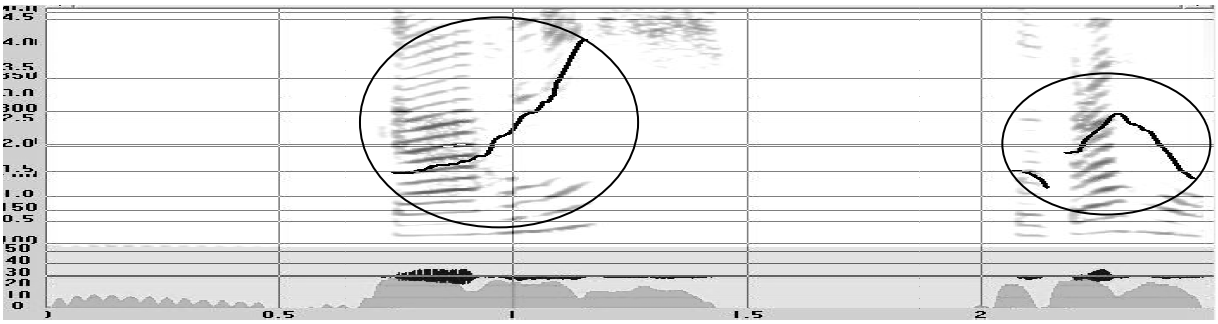


A figura 1.12 apresenta um enunciado executado através de duas unidades tonais, com função respectivamente de tópico[14] e de comentário. A primeira unidade tonal corresponde a um perfil entonacional de tópico, e a estrutura morfossintática correlata a esse perfil é um SN com valor de sujeito, *Carlos*. O segundo contorno, parte central e lado direito da figura, corresponde a um perfil entonacional de comentário de estrutura morfossintática verbal *foi pra Roma*.

[13] A unidade informacional de comentário será melhor explicada na subseção 1.8.1.1 desse capítulo.
[14] Essa unidade será pormenorizada na subseção 1.8.1.2 desse capítulo.

*XYZ: ***Carlos*** */ foi pra Roma //*

Figura 1.12 – Unidade de Tópico e unidade de Comentário



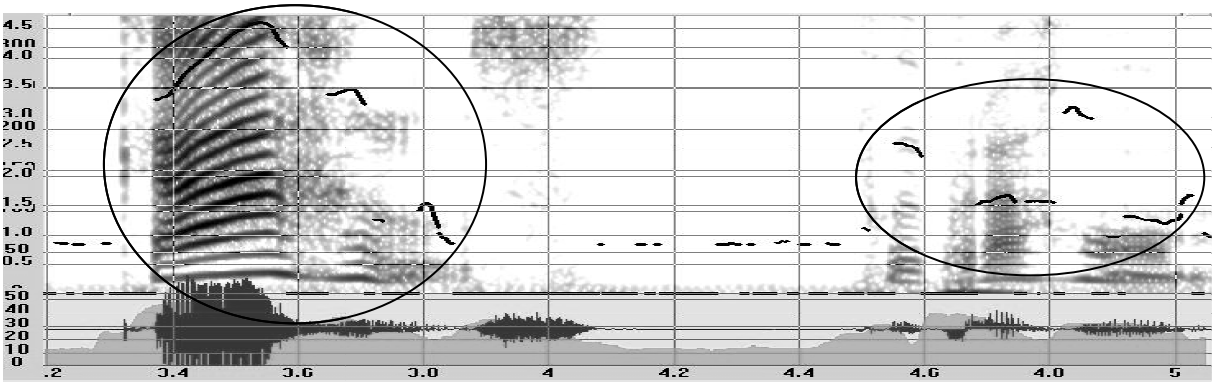
A realização evidenciada na figura 1.12 mostra que o enunciado *Carlos*, delimitado pela 1ª elipse, e *foi pra Roma*, marcado pela 2ª elipse, é constituído por duas unidades tonais (Tópico-Comentário), e, portanto, é classificado como enunciado complexo.

Na figura 1.13, temos outro conteúdo locutivo, também realizado em duas unidades tonais, executando um enunciado complexo. Entretanto, contrariamente ao exemplo anterior, a primeira unidade corresponde agora à unidade de comentário (delimitada pela 1ª elipse), enquanto a segunda unidade corresponde à unidade informacional de apêndice (marcada pela 2ª elipse)[15].

[15] Essa unidade será explicada em detalhes na subseção1.8.1.3 desse capítulo, bem como no capítulo 2, já que é o objeto de estudo em questão.

*XYZ: Carlos /=COM= foi pra Roma //=APC=

Figura 1.13 – Unidade de Comentário e unidade de apêndice



Além da identificação dos enunciados em simples ou complexos, segundo a teoria, é possível identificar a força ilocucionária veiculada pela unidade de comentário, Desse modo, é possível realizar um mesmo conteúdo locutivo com diferentes ilocuções, assim como também é possível, através de conteúdos locutivos distintos, veicular uma mesma ilocução.

É a entonação, portanto, que permite ao interlocutor identificar a ação do locutor, através de um determinado conteúdo locutivo. O interlocutor não atribui valor funcional às diferenças microprosódicas que participam das curvas de diferentes conteúdos locutivos com a mesma força ilocucionária, mas somente identifica as diferentes ilocuções, não implicando o tipo de conteúdo locutivo que a constitui.

Uma forma entonacional é caracterizada por uma preparação, um núcleo e uma coda, das quais apenas o núcleo é a parte obrigatória, pois é a porção que carrega a proeminência perceptual associada com a função informacional. As demais partes são opcionais e definidas com base na distribuição dentro da unidade. A preparação é o segmento que vem à esquerda do núcleo, e a coda é a porção que vem após o núcleo. Quando o material silábico é mais longo do que o necessário para a realização do núcleo, ele é distribuído ao longo da preparação e/ou coda (FIRENZUOLI; SIGNORINI, 2003).

1.6 A definição de enunciados versus a definição de sentenças

Segundo Cresti (2001, 2005), se a entidade de referência da fala é o enunciado, pois corresponde à contraparte linguística de um ato fala, a sentença é entidade de referência da escrita, uma vez que se baseia na predicação e deve possuir autonomia semântica e sintática.

Remetendo-se à Teoria dos Atos de Fala (Austin, 1962), o enunciado é definido pela autora como "toda expressão interpretável pragmaticamente, ou seja, toda expressão através da qual se realize um ato de fala" (Cresti, 2000). Nessa abordagem, cada um dos enunciados proferido por um falante seria uma ação sobre seus interlocutores, ou uma reação a uma ação anterior. Assim, o diálogo constituiria um suceder de ações e reações, em que cada um dos interlocutores age sobre o outro por meio de seus diversos enunciados. A motivação para a realização de cada enunciado seria de natureza psicológica e teria como ponto de partida uma "insurgência pulsional inconsciente" (Cresti 2000, p.43). Segundo a autora, a materialização das ações desempenhadas pelos falantes nos enunciados não ocorre de forma livre, mas é condicionada por esquemas culturais que os impelem a realizar cada ação de uma determinada forma.

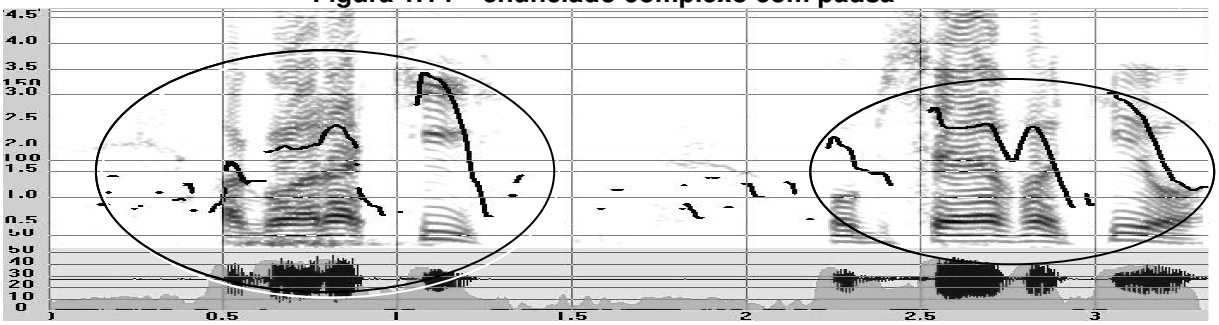
Outra definição tradicionalmente associada ao enunciado é a de sequência existente entre dois silêncios, isto é, uma sequência delimitada por pausas fortes. Contudo, essa definição é por si inconsistente, uma vez que é possível realizar uma sequência de enunciados sem qualquer ocorrência de pausas, como também é possível realizar um enunciado que contenha pausas fortes.

Passamos, então, a apresentar, a seguir, exemplos de execução de enunciados simples e complexos, com ou sem pausa. A pausa, como veremos abaixo, não garante que haja, necessariamente, uma fronteira melódica. Miranda (2001, p.39) diz que as pausas podem ser de dois tipos: pausa silenciosa, que corresponde ao silêncio absoluto em um enunciado e pausa preenchida, que está relacionada a

fenômenos discursivos e possui elementos que preenchem o silêncio, como *eh, hum*. As pausas silenciosas estariam associadas a uma função gramatical, isto é, à organização e estruturação gramatical do enunciado e tenderiam a ser mais breves que as não-silenciosas. Miranda diz ainda que se deve considerar como pausa silenciosa qualquer interrupção do enunciado, independente do valor de sua duração.

Exemplo 1.7 (bpubdl01) (CD áudio 1.7)
*PAU: agora aqui /=COB= observa aqui o' //=COM=

Figura 1.14 – enunciado complexo com pausa

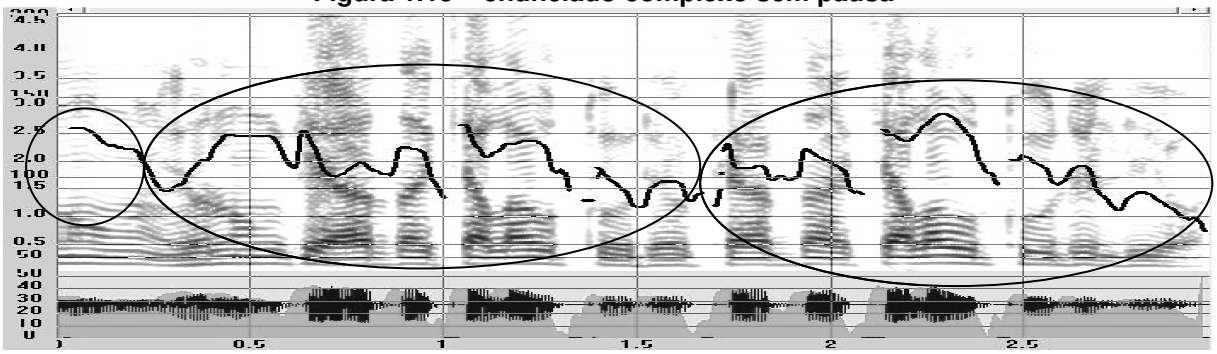


Na imagem acima, os enunciados estão delimitados pelas elipses. Há um enunciado complexo realizado em uma mesma sequência. Entre as unidades de comentário ligado, *agora aqui* (1ª elipse), e unidade de Comentário, *observa aqui ó* (2ª elipse), há um silêncio. Nesse caso, a ocorrência da pausa não coincide com a marcação dos perfis terminais.

Exemplo 1.8 (bpubdl01) (CD áudio 1.8)

*PAU: <não // mas> eu nũ [/2] eu nũ quero daquela altura // eu quero daquela altura ali ó //

Figura 1.15 – enunciado complexo sem pausa



Já na figura 1.15, exemplo 1.8, encontramos a realização de três enunciados, sendo os três simples. Todos eles executados sem qualquer ocorrência de pausa. O primeiro constitui-se de um enunciado curto, porém autônomo; isto é, uma única unidade de comentário *não*, representado pela 1ª elipse, à esquerda da figura. O segundo, representado pela 2ª elipse e o terceiro, representada pela 3ª elipse. As características associadas ao enunciado justificam a sua escolha como unidade analítica fundamental na Teoria da Língua em Ato, pois ao considerar o enunciado como entidade de referência da diamesia falada, tornou-se possível identificar as ilocuções e explicitar os níveis de articulação informacional da fala espontânea, de maneira completa e sistemática.

Resumidamente, podemos dizer que para a Teoria da Língua em Ato a entonação tem como funções:

i. segmentar os enunciados dentro do *continuum* da fala;
ii. segmentar o enunciado em suas unidades internas;
iii. atribuir uma ilocução a cada enunciado;
iv. atribuir uma função informacional a cada unidade tonal do enunciado.

A segmentação da fala em enunciados e em suas unidades internas é feita com base em quebras prosódicas percebidas como *terminais*, conforme a fonética perceptual de 't Hart *et ali* (1990). Ao ouvir uma quebra terminal, o ouvinte tem a percepção de conclusão.

1.7 Enunciados simples e complexos: a estrutura informacional

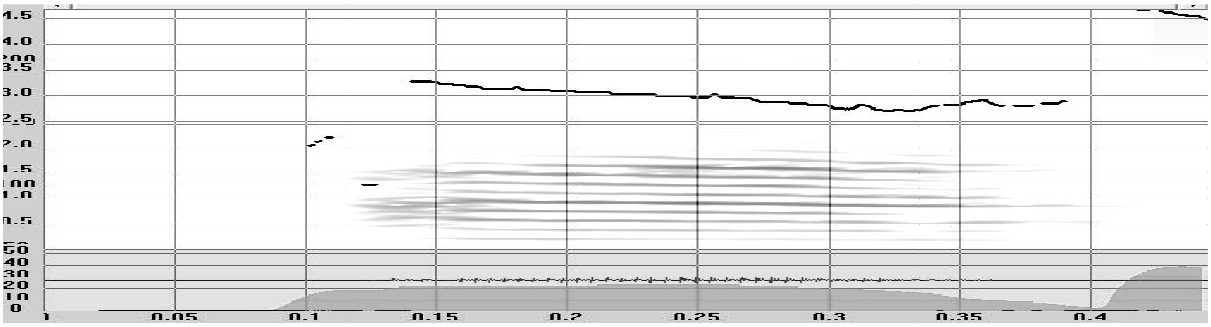
A identificação de um enunciado se realiza através da percepção de uma quebra entonacional percebida como terminal, conforme se verificou nos exemplos anteriores. Isso significa que uma unidade de enunciado (ou a única se o enunciado for simples) deve ser uma unidade de raiz (o comentário) capaz de veicular autonomia pragmática. Esse princípio baseia-se na teoria fonética perceptual (‘t Hart; Collier; Cohen, 1990), acarretando uma relação biunívoca entre enunciado e ilocução. A cada enunciado, ou seja, a cada unidade mínima de significado pragmático, corresponde-se uma única ilocução, uma intencionalidade do falante.

Tanto os enunciados simples, aqueles constituídos por uma única unidade tonal, como os complexos, formados por mais de uma unidade tonal, podem ou não possuir verbo. A figura 1.16, que representa o enunciado em negrito, ilustra um enunciado simples sem verbo e a figura 1.17 um enunciado simples com verbo. A figura 1.18 representa um enunciado complexo sem verbo e a 1.19, um enunciado complexo com verbo.

Exemplo 1.9 (bfamdl01) (CD áudio 1.9)

*BEL: cê tem toc //
*BAL: tenho //
*BEL: **ah** //

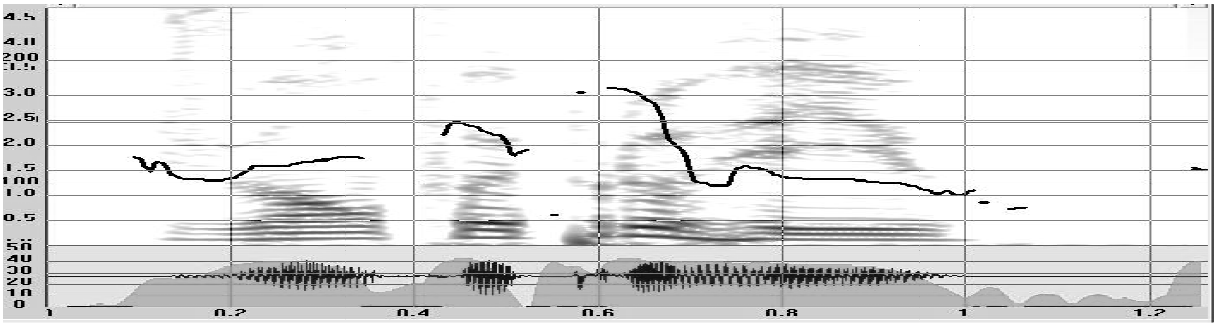
Figura 1.16 – enunciado simples sem verbo



Exemplo 1.10 (CD áudio 1.10)

* LUC: João foi pro Rio[16] //=COM=

Figura 1.17 – enunciado simples com verbo

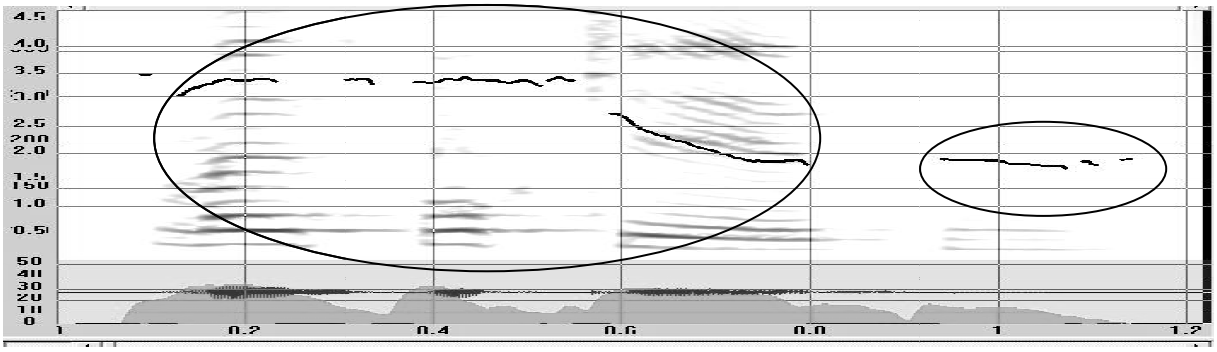


[16] Esse exemplo foi gravado em laboratório.

Exemplo 1.11 (bfamdl02) (CD áudio 1.11)

*BEL: mas porque /=COM= **assim** //**=PHA=**

Figura 1.18 – enunciado complexo sem verbo

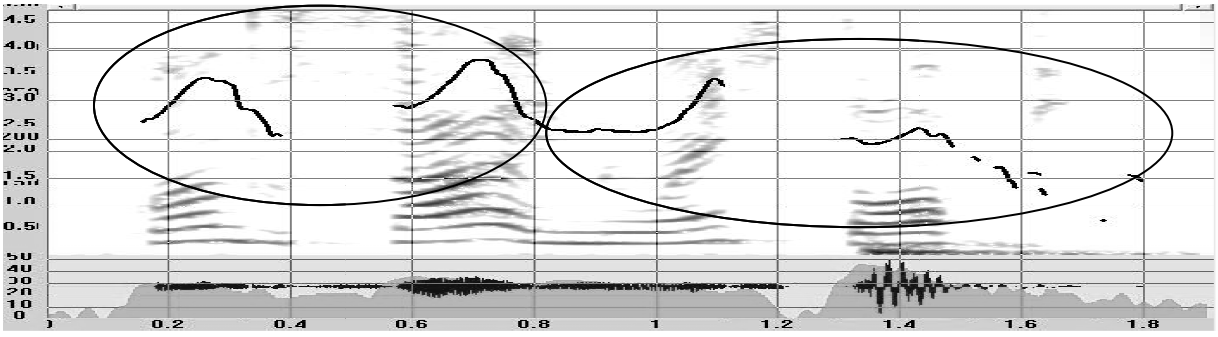


O exemplo 1.11 representa um enunciado complexo de duas unidades tonais. A primeira delas, *mas porque*, registrada na figura 1.18, pela 1ª elipse é uma unidade de comentário. A segunda, *assim,* representada pela 2ª elipse, é uma unidade de fático.

Exemplo 1.12 (bfamdl02) (CD áudio 1.12)

BAL: quando sai /=TOP= não é stop //=COM=

Figura 1.19 – enunciado complexo com verbo



A figura 1.19 ilustra um enunciado complexo com verbo, composto de duas unidades informacionais. A primeira delas *quando sai* (1ª elipse) é uma unidade de tópico. A

segunda unidade *não é stop* (2ª elipse) é uma unidade de comentário, por isso, autônoma.

1.8 As Unidades Informacionais

Há três critérios que definem as unidades informacionais: um critério entonacional, um funcional e um distribucional. O critério funcional baseia-se na função que cada unidade carrega dentro da estrutura informacional; o critério entonacional refere-se ao perfil dedicado à curva que cada unidade informacional apresenta e o critério distribucional diz respeito à posição que cada unidade ocupa no enunciado em relação à unidade de Comentário, unidade central dentro do padrão informacional, por ser aquela que veicula a ilocução.

Segundo a Teoria da Língua em Ato há dois tipos de unidades informacionais: as unidades textuais e as unidades dialógicas. As unidades informacionais textuais são aquelas que compõem o texto do enunciado propriamente dito. Dentre elas encontramos as unidades de Comentário (COM), de Tópico (TOP), de Apêndice de Comentário (APC), de Parentético (PAR), de Introdutor Locutivo (INT), de Apêndice de Tópico (APT) e a Unidade de escansão (SCA)[17]. As unidades informacionais dialógicas ou não textuais, por sua vez, são aquelas que não contribuem para a constituição semântica de um enunciado, mas dedicam-se ao cumprimento pragmático desse enunciado sendo dirigidas ao interlocutor. São elas: Incipitário (INP), Conativo (CNT), Conector Dialógico (DCT), Fático (PHA), Alocutivo (ALL) e Expressivo (EXP).

1.8.1 Unidades Informacionais Textuais

Nesta subseção, apresentamos as unidades informacionais textuais, identificando os critérios funcional, relacionado à função das unidades na interação comunicativa;

[17] Como nessa unidade há perda de isomorfismo, optou-se por descrevê-la na subseção 1.9.1, juntamente com os outros casos dessa natureza.

entonacional, identificado através do perfil da curva de Fo, e distribucional, relacionado à posição dessas unidades no enunciado.

1.8.1.1 A Unidade de Comentário (COM)

A Unidade de Comentário é a mais importante de todas, pois é a única necessária e suficiente para execução de um enunciado. Sua função é a de realização da força ilocucionária, ou seja, a de cumprir um ato de fala.

Entonacionalmente é tida como uma unidade prosódica de raiz que varia conforme o valor ilocucionário; isto é, é interpretável pragmaticamente em autonomia e possui sempre um núcleo, o qual carrega o valor funcional da ilocução. Distribucionalmente pode estar em qualquer posição no enunciado e é com relação a ela que é definida a posição das outras unidades.

No *corpus* italiano foram individualizadas oitenta ilocuções para trinta perfis prosódicos distintos, conforme dados do Lablita (Firenzuoli, 2003), divididas nas cinco classes atitudinais, que serão apresentadas na seção 1.11. Essas classes determinam o tipo de pulsão afetiva que motiva e ancora internamente a ilocução exercida sobre o destinatário, visto que a fala se baseia na ocorrência de ilocuções codificadas convencionalmente em uma língua, a partir de uma pulsão afetiva não convencionalizada.

As expressões que constituem uma unidade de Comentário têm sempre um valor modal[18] intrínseco, que nem sempre é explicitado lexicalmente. Todos esses valores modais podem ocorrer na unidade de Comentário sem que haja qualquer correlação preferencial.

[18] O valor modal diz respeito à atitude do falante sobre sua própria locução.

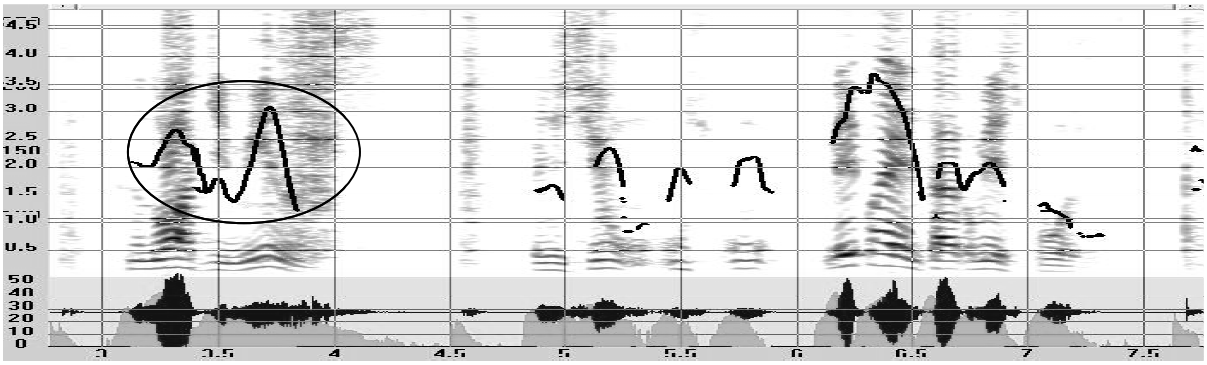
1.8.1.2 A Unidade de Tópico (TOP)

A unidade de Tópico é a unidade textual cuja função é especificar no texto do enunciado o domínio de relevância ao qual a força ilocucionária se refere; isto é, o campo de aplicação da força iloucionária do comentário. Ela tem caráter opcional e é subordinada melodicamente ao comentário, não sendo interpretável autonomamente. Assim como as outras unidades, se define com base nos critérios funcional, entonacional e distribucional. Vejamos um exemplo abaixo:

Exemplo 1.13 (bfammn01) (CD áudio 1.13)

*MAI: **no norte de Minas** **/=TOP=** existia um [/1]=SCA= um [/2]=EMP= meio aparentado com a minha esposa //=COM=

Figura 1.20 – Unidade de Tópico



Na figura 1.20, a elipse sinaliza a unidade de tópico.

Firenzuoli; Signorini (2003) utilizaram os três critérios fundamentais (funcional, entonacional e distribucional) para identificar a unidade de tópico. Funcionalmente o tópico é definido como campo de aplicação da força ilocucionária, isto é, o domínio semântico ao qual o comentário se refere. Entonacionalmente, é uma unidade tonal opcional e prefixal que possui foco funcional à direita (CRESTI, 2000a). Distribucionalmente, o tópico deve estar linearmente ordenado antes do comentário. A unidade de tópico possui foco funcional, cujo perfil, no Italiano, pode ser

identificado em três tipologias: 1) ascendente-descendente; 2) ascendente; 3) descendente-nivelado-descendente.

O tópico pode ser classificado em simples ou complexo. É simples quando o enunciado apresenta apenas uma unidade de tópico, não iterada por outro tópico ou por uma unidade de apêndice (exemplo 1.14). Já o tópico complexo é aquele iterado por uma ou mais unidades de tópico e/ou de apêndice (exemplo 1.15)

Exemplo 1.14 (bpubcv01)
*FLA: **naquelas maquinas** **/=TOP=** elas ja tem balanca propria ai //=COM=

Exemplo 1.15 – (bfammn03)
*ALO: ai /=PHA= **quando e tres anos atras mais ou menos** **/=TOP= chega um filho dela** **/=TOP= perto de mim** **/=APT=** falou assim /=INT= o Aloysio //=COM_r=

Na língua italiana, foram identificadas três diferentes possibilidades de realização prosódica da unidade de tópico. No português brasileiro (diatopia mineira) já foram encontradas 4 formas possíveis[19].

1.8.1.3 A unidade de Apêndice de Comentário (APC)

Funcionalmente, o Apêndice integra textualmente as unidades de Comentário (COM), Comentário Ligado (COB), Comentários Múltiplos (CMM); entonacionalmente é uma unidade tonal sem foco, com uma F0 sempre mais baixa do que a unidade da qual é apêndice, sempre com perfil nivelado ou descendente e intensidade baixa (Cresti 2000; Ulisses 2008; Oliveira 2009, 2009b, 2010). Distribucionalmente deve suceder a unidade informacional de Comentário. É tida como uma unidade de sufixo.

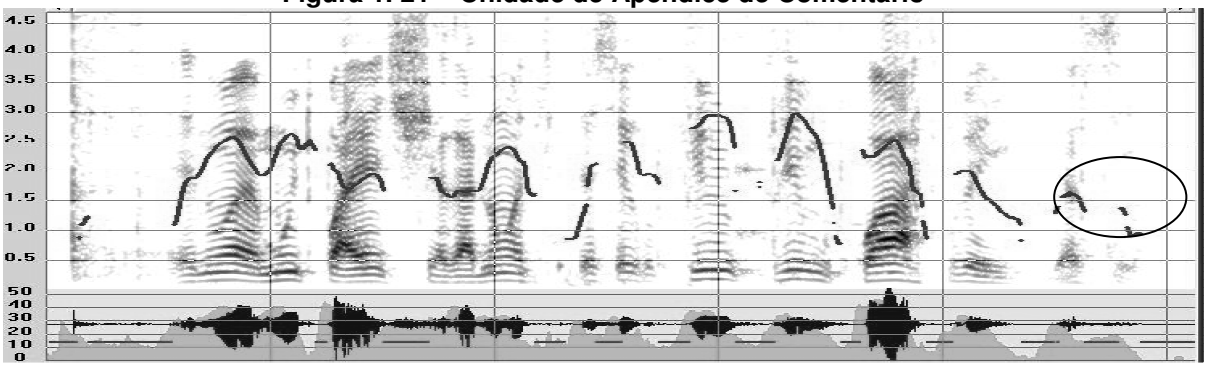
[19] Maiores detalhes ver Mittmann (tese de doutorado, em preparação) e Rocha (2012).

A unidade de Apêndice de Comentário, por ser o foco de análise e discussão do presente estudo, será retomada no capítulo 2. Apresentamos, a seguir, apenas um exemplo de Apêndice de Comentário encontrado no *corpus* coletado.

Exemplo 1.16 (bfammn01) (CD áudio 1.16)

*MAI: [5] ele nũ é muito parente chegado não /=COB= mas &t [/1]=SCA= deve ser /=SCA= primo [/1]=EMP= primo quarto /=COB= por aí /=COM= **deve ser** //=**APC**=$

Figura 1. 21 – Unidade de Apêndice de Comentário



Na figura 1.21, a elipse sinaliza a unidade de Apêndice de Comentário.

1.8.1.4 A Unidade de Apêndice de Tópico (APT)

A unidade de Apêndice de Tópico integra textualmente a unidade de Tópico à qual se refere. Essa unidade pode reproduzir as características de um Tópico, com uma média de frequência fundamental (Fo) mais baixa e sem foco funcional, ou ter um movimento nivelado ou descendente.

Funcionalmente a unidade de Apêndice de Tópico integra textualmente a unidade de Tópico à qual se refere. Distribucionalmente, deve suceder a unidade informacional de Tópico.

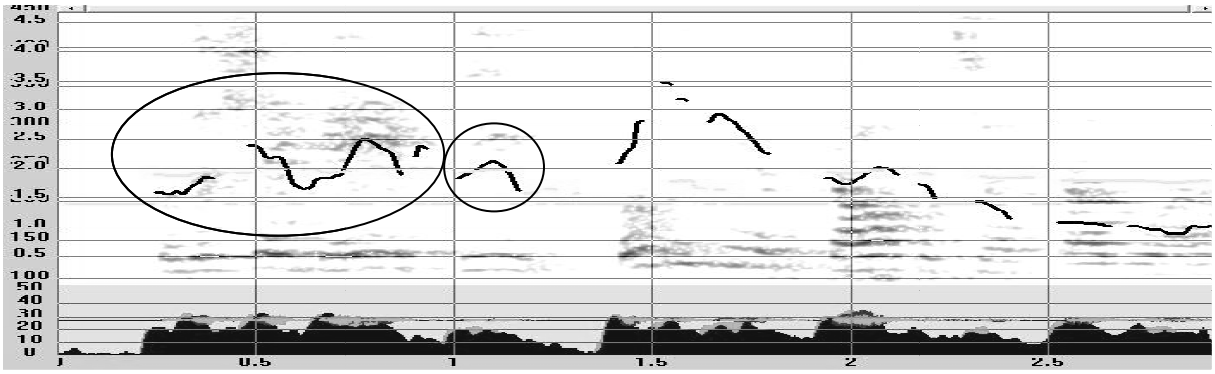
Quanto à modalidade, a unidade de Apêndice de Tópico não possui uma modalidade própria, e depende da modalidade da unidade de Tópico.

A seguir apresentamos um exemplo de Apêndice de Tópico com seu respectivo perfil entonacional.

Exemplo 1.17 (bfamdl02) (CD áudio 1.17)

*BEL: quando eu cheguei /=TOP= **aqui** /**=APT=** todas as minhas calças tinham ficado lá hhh //=COM=

Figura 1.22 – Unidade de Apêndice de Tópico



A figura 1.22 ilustra um enunciado complexo expresso no exemplo 1.17. A primeira unidade é uma unidade de tópico, *quando eu cheguei*, e vem representada na figura por meio da 1ª elipse. A segunda unidade, *aqui*, é o Apêndice de Tópico que integra apenas o conteúdo da unidade de TOP e está demarcado, na figura, pela 2ª elipse.

1.8.1.5 A Unidade de Parentético (PAR)

A unidade de parentético[20] é uma unidade textual que tem como domínio o enunciado como um todo ou parte dele e tem uma função metalinguística. Serve

[20] Maiores detalhes ver VALE (2010).

como instrução para o interlocutor sobre como interpretar o conteúdo proposicional do enunciado e, frequentemente, tem função modalizadora. Pode ser distinguido em pelo menos duas tipologias: o parentético curto e o parentético longo. Esse último caracteriza-se por possuir duração maior que a grande maioria dos outros parentéticos e por expressar um verdadeiro comentário metalinguístico sobre o enunciado.

Entonacionalmente o parentético possui um perfil prosódico uniforme ou descendente, podendo aparecer, às vezes, uma cauda tonal ascendente após uma pausa e ter pequenos movimentos que não constituem focos funcionais. É caracterizada, ainda, por um abaixamento da Fo ou mais raramente por um aumento da F0, com relação ao resto do enunciado, e, frequentemente, por um aumento da velocidade.

Distribucionalmente, a unidade de parentético pode ocorrer em todas as posições dentro do enunciado, inclusive intensificando outra unidade textual, exceto em posição inicial. Há, ainda, a ocorrência de PAR dentro de PAR[21], criando um PAR de nível hierárquico inferior.

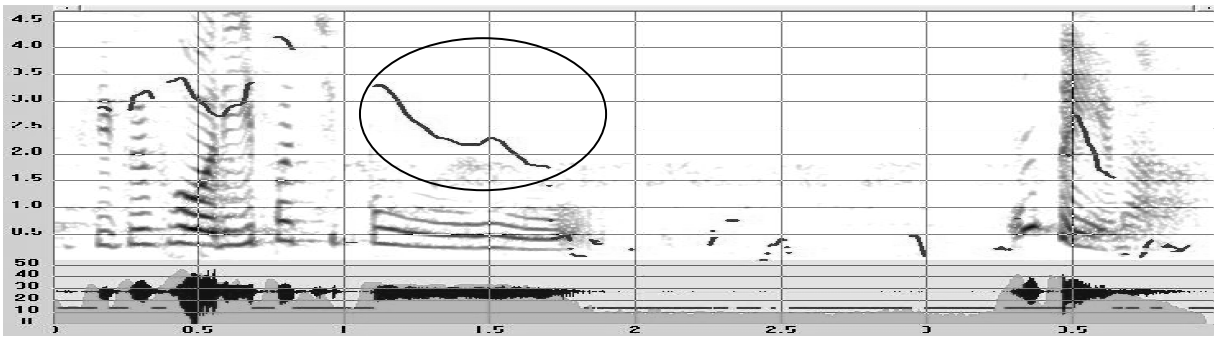
Enquanto o exemplo 1.18 ilustra um parentético curto, o exemplo 1.19 mostra-nos as características de um parentético longo.

[21] No Italiano foram individualizados até três níveis.

Exemplo 1.18 (CD áudio 1.18)

*BRM: e fico em Floripa / **tipo** /=**PAR**= um tempo //[22]

Figura 1.23 – Parentético curto

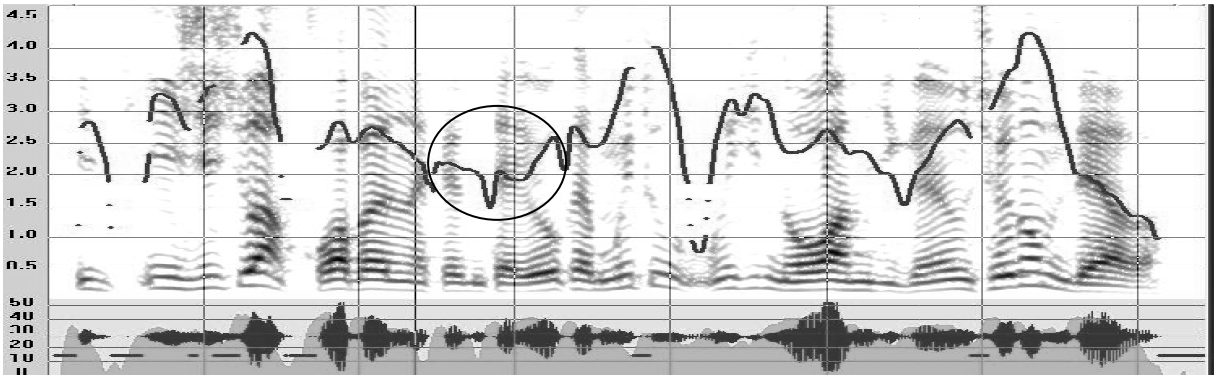


Na figura 1.23, a elipse sinaliza a unidade de Parentético curto.

Exemplo 1.19 (bfammn01) (CD áudio 1.19)

*MAI: &e [/] / existe uma cobra / **lá naquele interior** /=**PAR**= que é muito [/] muito enorme de grande / não sei o nome dela //

Figura 1.24 – Parentético longo



Na figura 1.24, a elipse sinaliza a unidade de Parentético longo.

[22] Esse exemplo não faz parte do subcorpus analisado, mas do C-ORAL-BRASIL.

1.8.1.6 A Unidade de Introdutor Locutivo (INT)

A unidade de Introdutor Locutivo[23] é uma unidade textual que sinaliza para o destinatário que o que segue pertence em bloco a um nível pragmático diferente, sendo quase sempre uma metailocução; ou seja, o que segue sai do aqui e agora da enunciação, anunciando uma suspensão pragmática como no caso de exemplificações, elencos, comparações, instruções e, principalmente, discurso reportado.

A principal função do INT é introduzir uma metailocução (como o discurso reportado, uma lista, uma exemplificação, etc.) marcando a suspensão pragmática. Entonacionalmente o perfil dessa unidade informacional pode variar, mas mantendo sempre algumas características constantes: a F0 do INT é mais baixa do que a da metailocução introduzida, à exceção de metailocuções muito baixas. O que não muda é o forte constraste de F0 entre o INT e o que se segue.

Outra característica é uma velocidade de ilocução muito mais alta, às vezes extremamente mais alta. Enfim, o INT pode ter formas diferentes, mas sempre sem núcleo, e descendentes.

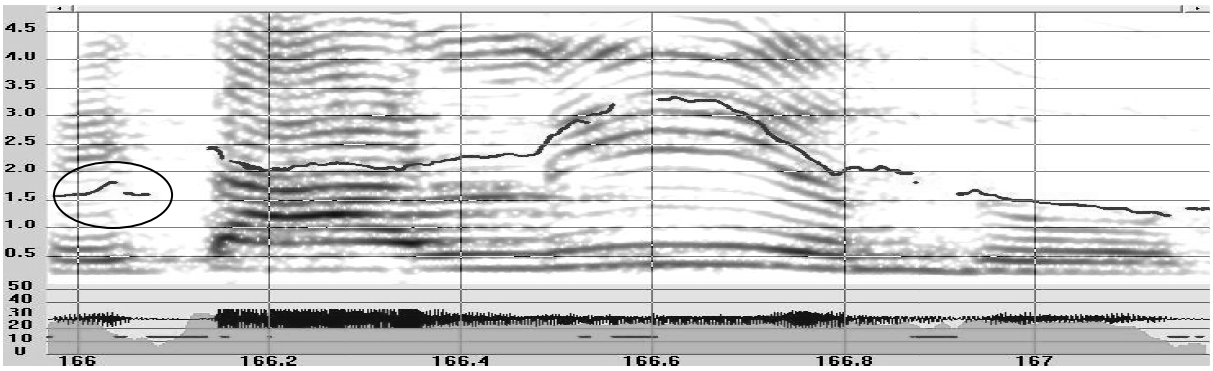
Distribucionalmente, o Introdutor Locutivo ocorre antes da unidade de Comentário ou de um padrão que se quer introduzir (quase sempre em uma estrita continuidade).

[23] Maiores detalhes ver MAIA ROCHA (2011).

Exemplo 1.20 (bfammn04) (CD áudio 1.20)

*REG: **falei** /**=INT**= tá lindo //=COM_r=

Figura 1.25 – Introdutor Locutivo com meta-ilocução de discurso reportado



Na figura 1.25, a elipse sinaliza a unidade de Introdutor Locutivo com meta-ilocução de discurso reportado. Assunto que será tratado a seguir.

A unidade de introdutor locutivo é caracterizada, às vezes, por um valor modal ligado à natureza da metailocução. As chamadas meta-ilocuções são tipos específicos de ilocuções que não operam no mundo. Isso quer dizer que o conteúdo locutivo de tais padrões representa uma ilocução que se refere a coordenadas espaço-temporais diferentes das coordenadas espaço-temporais atuais do enunciado.

Segundo Cresti, as metailocuções são quatro, a saber: Discurso Reportado, Exemplificação Emblemática, Pensamento Falado e Narração. Pesquisas recentes do grupo C-ORAL-BRASIL ampliam esse repertório (MAIA ROCHA, 2011; MAIA ROCHA; RASO, 2011).

1.8.2 Unidades Informacionais Dialógicas ou não-textuais

As unidades informacionais dialógicas ou não-textuais se diferenciam das unidades textuais porque não contribuem para a constituição semântica de um enunciado; todavia dedicam-se ao cumprimento pragmático desse enunciado. Essas unidades

não podem ser escansionadas e também não se referem a um domínio de identificação. Não possuem foco, ponto de vista ou modalidade e sua seleção lexical é mais restrita. Elas correspondem, em outras acepções técnicas, aos são chamados de marcadores discursivos, sem nenhum acordo sobre os que são e como podem ser identificados.

1.8.2.1 Incipitário (INP)

A unidade de **incipitário** (INP) é uma unidade dialógica que sinaliza para o interlocutor o início de um enunciado com valor contrastivo (em termos afetivos, não lógicos) em relação ao enunciado anterior. Isto significa que os incipitários explicitam a vontade do falante em começar um turno ou prossegui-lo. Essa unidade tem a função de introduzir um ato de fala e fazer contraste com o anterior.

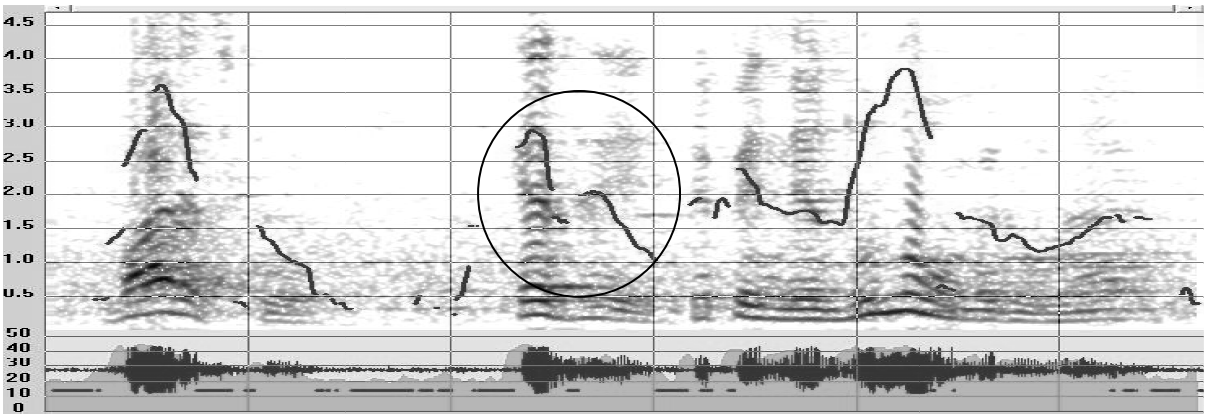
Entonacionalmente é sinalizada através de uma unidade tonal com F0 muito alta, intensa, curta, com perfil ascendente-descendente. Distribucionalmente a unidade de Incipitário acha-se, frequentemente, no início de um enunciado.

A unidade de incipitário não se correlaciona a nenhuma classe lexical específica, portanto pode ser constituída por locuções, advérbios, conjunções coordenadas e subordinadas, pronomes pessoais e interjeições. É uma unidade muito frequente no italiano, mas pouco expressiva no PB. O exemplo a seguir apresenta essas características encontradas na unidade informacional em questão.

Exemplo 1.21 (bpubdl02) (CD áudio 1.21)

*EUG: várias cores // **quer dizer** /**=INP=** dependendo do número / né //

Figura 1.26 – Unidade de Incipitário



Na figura 1. 26, a elipse sinaliza a unidade de Incipitário.

1.8.2.2 Contativo (CNT)

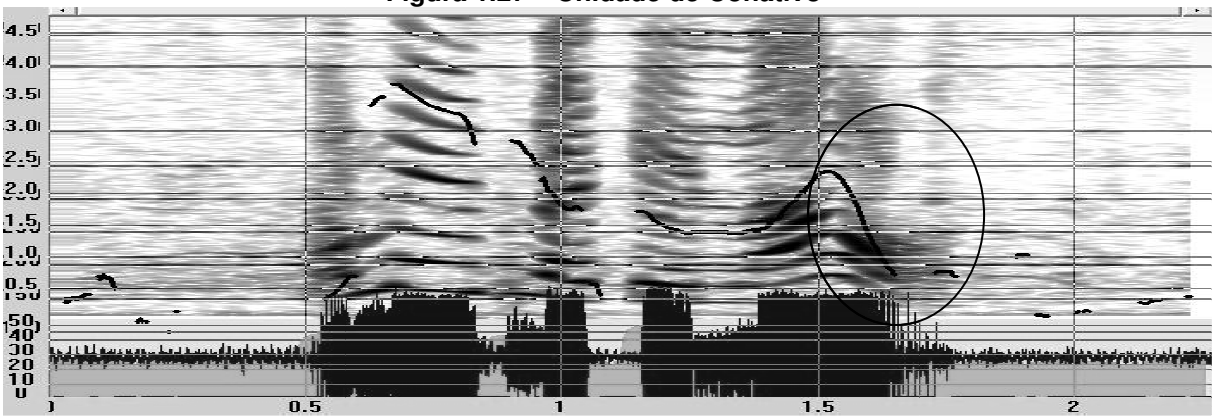
A unidade de **Conativo** (CNT) pressiona o interlocutor, para que ele efetue ou desista de algo. Entonacionalmente é uma unidade dialógica muito intensa e caracterizada por um movimento descendente de F0. Distribucionalmente, a unidade de Contativo não possui uma posição fixa dentro do enunciado, embora apareça de forma mais frequente em posição final de enunciado. É uma unidade pouco recorrente.

Assim como as demais unidades dialógicas, a unidade de Conativo não corresponde a um domínio de identificação e também não possui foco. É lexicalmente variada, podendo ser constituída de nomes, verbos, expressões adverbiais, dentre outras. Observe-se o exemplo abaixo:

Exemplo 1.22 (CD áudio 1.22)

*REN: é o mesmo /=SCA= daquele lá /=COM= o' //=CNT= (bfamdl01)

Figura 1.27 – Unidade de Conativo



Na figura 1.27, a elipse sinaliza a unidade de Conativo.

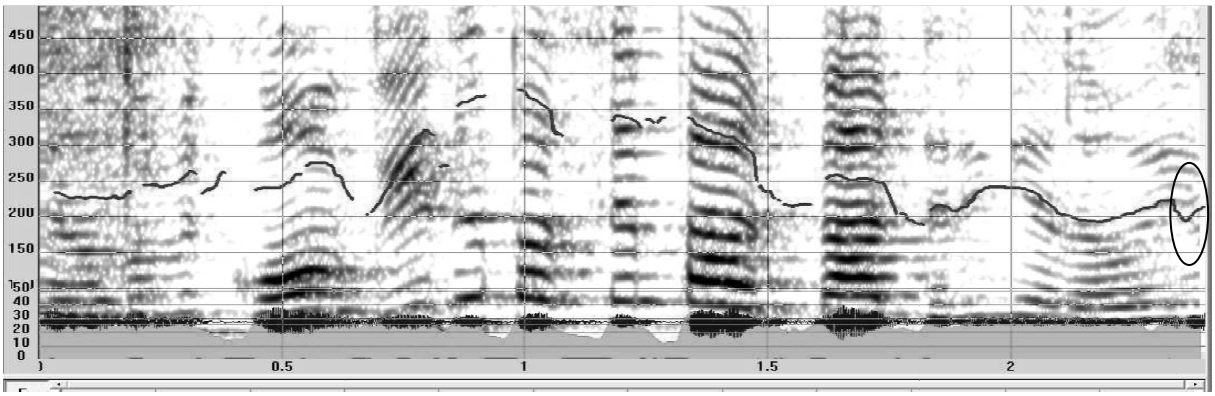
1.8.2.3 Fático (PHA)

O **fático** (PHA) é uma unidade dialógica dedicada ao controle do bom funcionamento da comunicação e assegura a manutenção do canal. Entonacionalmente possui uma intensidade baixa e uma duração muito curta, com movimento nivelado ou descendente de F0 e uma baixa relevância perceptual do conteúdo segmental. Distribucionalmente essa unidade ocorre em qualquer posição no enunciado, sem aparente preferência. Lexicalmente essa unidade é constituída de formas verbais, verbos de percepção, adjetivos, advérbios e interjeição, entre outras.

Exemplo 1.23 (bfamcv02) (CD áudio 1.23)

*RUT: [70] <pelo> amor de Deus /=CMM= nũ me convida pa ser &pa [/1]=SCA= madrinha não /=CMM= hein //=PHA=$

Figura 1.28 – Unidade de comentário seguido por uma unidade de Fático



Na figura 1.28, a elipse sinaliza a unidade de Fático.

1.8.2.4 Alocutivo (ALL)

A unidade de **alocutivo** (ALL) especifica a quem a mensagem é dirigida por meio de seu nome ou de uma atribuição de epíteto. Através dela se mantém a atenção do interlocutor e, simultaneamente, desempenha-se uma função coesiva do ponto de vista social.

Entonacionalmente possui um perfil pouco relevante, levemente modulado ou descendente. Distribucionalmente a unidade de alocutivo pode estar em qualquer posição dentro do enunciado. Segundo Raso; Leite (2010), a posição parece mudar muito nas várias línguas e culturas.

Como todas as unidades dialógicas, a unidade de alocutivo não corresponde à identificação de um domínio e também não possui foco. Realiza o controle da comunicação através de uma interpelação direta, nomeando o interlocutor ao dirigir-

se a ele. É constituída por nomes próprios, nomes de classes que designam um perfil social ou familiar, pronomes pessoais, adjetivos qualitativos.

A unidade de ALL se difere completamente da ilocução de chamamento, com a qual não deve ser confundida, já que sua função informacional primordial está relacionada à ativação afetiva no canal comunicativo entre interlocutores.

O exemplo a seguir ilustra uma unidade de Alocutivo, encontrada no subcorpus.

Exemplo 1.24 (bfamcv02) (áudio 1.24)[24]
*TER: não /=PHA= mas ea tá é brincando /=COM= né /=PHA= Rute //=ALL=

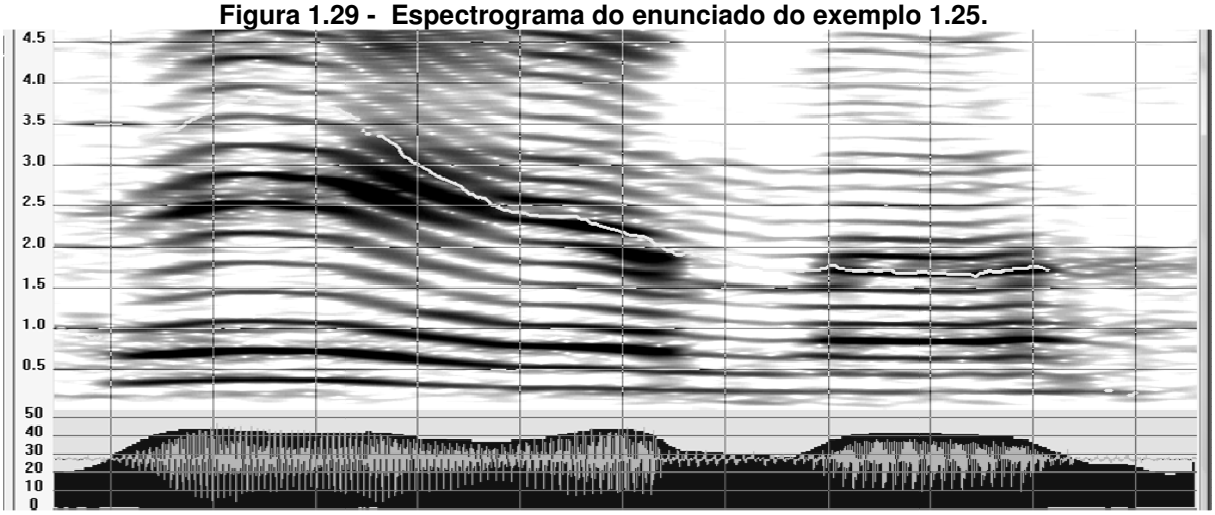
O ALL merece algumas observações ulteriores. Essa unidade seria tradicionalmente considerada dentro da categoria do vocativo. Contudo, a categoria tradicional inclui pelo menos duas unidades completamente diferentes do ponto de vista pragmático: a ilocução de chamamento e a unidade informacional dialógica de alocutivo. No primeiro caso temos uma unidade de comentário, que carrega, portanto, uma força ilocucionária. No segundo, uma unidade dialógica, sem autonomia pragmática e que, portanto, precisa estar em padrão com uma unidade ilocucionária. Essa diferença envolve tanto as características funcionais quanto as características prosódicas das duas unidades. Isso pode ser observado facilmente comparando-se os exemplos 1.25 e 1.26 e os respectivos espectrogramas mostrados nas figuras 1.29 e 1.30. Os dois exemplos mostram o mesmo conteúdo locutivo realizado pelo mesmo falante, em um caso (exemplo 1.25) como ilocução de chamamento e em outro (exemplo 1.26) como alocutivo.

[24] Os exemplos 1.25 a 1.27 foram retirados do capítulo 3 do C-ORAL-BRASIL.

Exemplo 1.25 (bfamdl01) (áudio 1.25)

*FLA: Rena //=COM=

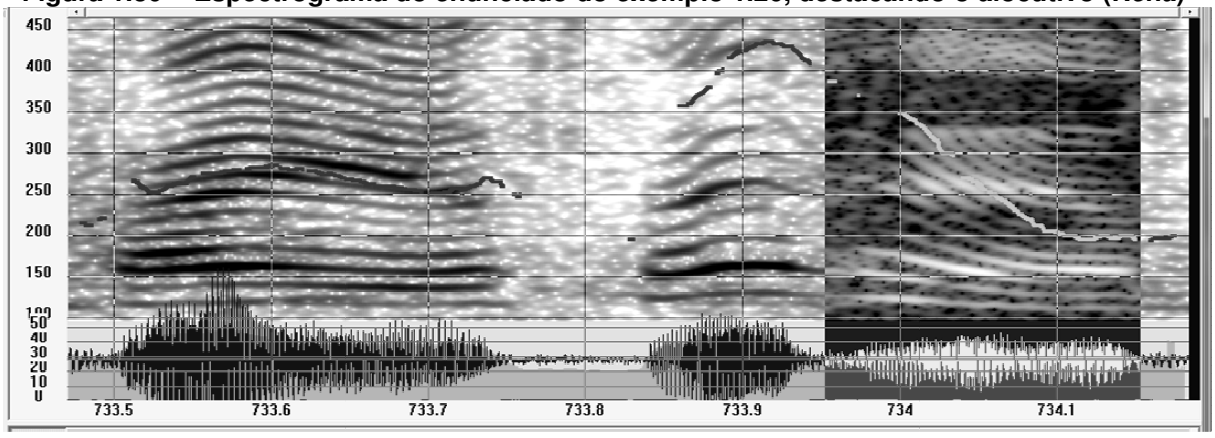
Figura 1.29 - Espectrograma do enunciado do exemplo 1.25.



Exemplo 1.26 (bfamdl01) (áudio 1.26)

*FLA: vai esse /=COM= né /=PHA= Rena //=ALL=

Figura 1.30 - Espectrograma do enunciado do exemplo 1.26, destacando o alocutivo (*Rena*)



A comparação entre o exemplo 1.25 e o exemplo 1.26 mostra que a ilocução possui uma duração quase três vezes maior que aquela do alocutivo, uma intensidade muito mais forte, uma maior variação de F_0 e, principalmente, mostra que na ilocução está presente um foco funcional, enquanto o alocutivo não possui foco. O resultado dessas características prosódicas é que podemos atribuir autonomia prosódica e pragmática à ilocução, mas não ao alocutivo.

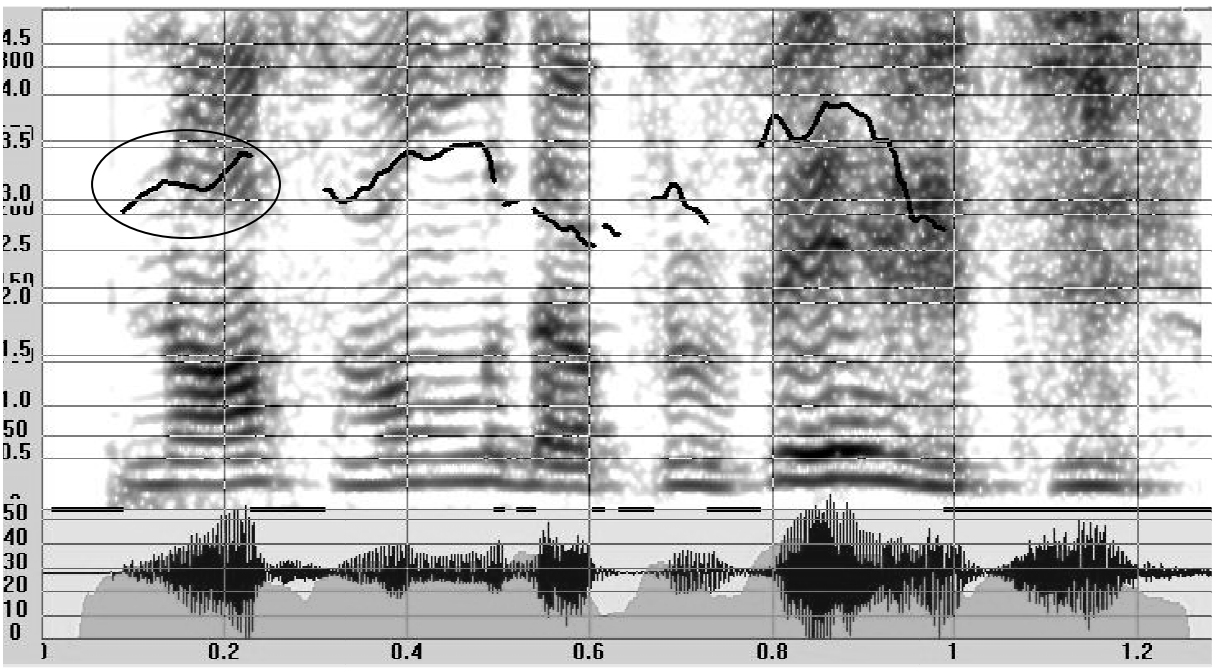
1.8.2.5 Expressivo (EXP)

A unidade de **expressivo** (EXP) é uma unidade dialógica que busca a coesão social. É um suporte emocional para o ato ilocucionário executado pelo Comentário. Entonacionalmente possui um perfil prosódico variado, com intensidade e duração de baixa. Distribucionalmente essa unidade pode ocupar qualquer posição dentro do enunciado.

As unidades de expressivo, em sua maioria, são formadas por nomes, interjeições e advérbios. São culturalmente muito idiossincráticas. São pouco frequentes no italiano e muito frequentes no PB.

Exemplo 1.28 (bfamcv04) (CD áudio 1.28)
*HEL: Nossa /=EXP= vão mudar pro verde hhh //=COM=

Figura 1.31 – Unidade de Expressivo



Na figura 1.31, a elipse sinaliza a unidade de Expressivo.

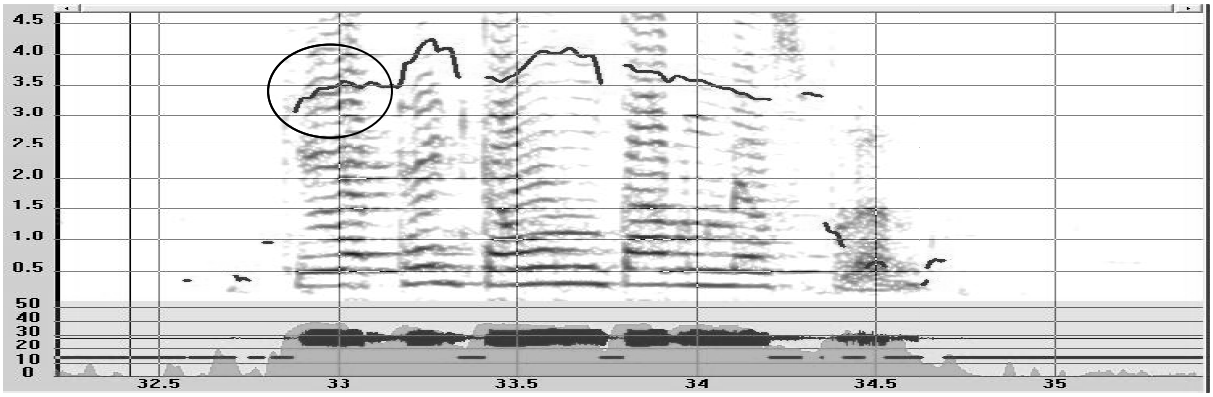
1.8.2.6 Conector Dialógico (DCT)

O **Conector Dialógico** (DCT) é uma unidade dialógica que pode abrir um turno, um enunciado ou um subpadrão em uma estrofe. [25] Sua função, portanto, é a de assinalar que a sequência informacional está em continuidade com a anterior. Entonacionalmente é marcada por um movimento nivelado ou modulado, intensidade alta e longa duração. Distribucionalmente ocorre no início do enunciado ou de um subpadrão de estrofe.

A unidade de Conector Dialógico garante o nexo semântico entre os enunciados através de operadores do tipo coordenativo e/ou subordinativo (conjunções, advérbios).

Exemplo 1.29 (bfammn05) (CD áudio 1.29)
*CAR: [4] porque /=DCT= até então /=TOP= nũ tava no meus plano //=COM=

Figura 1. 32 – Unidade de Conector Dialógico



Na figura 1.32, a elipse sinaliza a unidade de Conector Dialógico.

[25] Assunto a ser tratado na seção 1.9.3.

1.9 A perda do isomorfismo

Até agora mostramos os componentes dos padrões informacionais que podem realizar o enunciado. Nessa exposição consideramos que há uma correspondência biunívoca entre unidade tonal e unidade informacional, assim como entre enunciado e ilocução. Esse isomorfismo se realiza com regularidade alta, mas nem sempre. Ele se quebra em três circunstâncias.

1.9.1 A Unidade de Escansão (SCA)

O primeiro caso de não correspondência biunívoca entre unidade terminada e comentário é o da unidade de escansão. A unidade de escansão se distingue de todas as outras unidades informacionais porque só tem valor entonacional e não informacional. Ela acontece quando o falante não realiza a unidade informacional em uma única unidade tonal por alguma dificuldade particular, por razões enfáticas, ou se o programa melódico da unidade informacional é muito longo em número de sílabas. Quando a locução é extensa ou por alguma razão problemática, o falante segmenta em duas (às vezes mais) unidades com uma entonação neutra e repete esse mecanismo até que a expressão seja completada. O perfil entonacional desejado para realizar a função só é realizado pelo falante na última unidade tonal. Às vezes, a escansão obedece a alguma intenção expressiva.

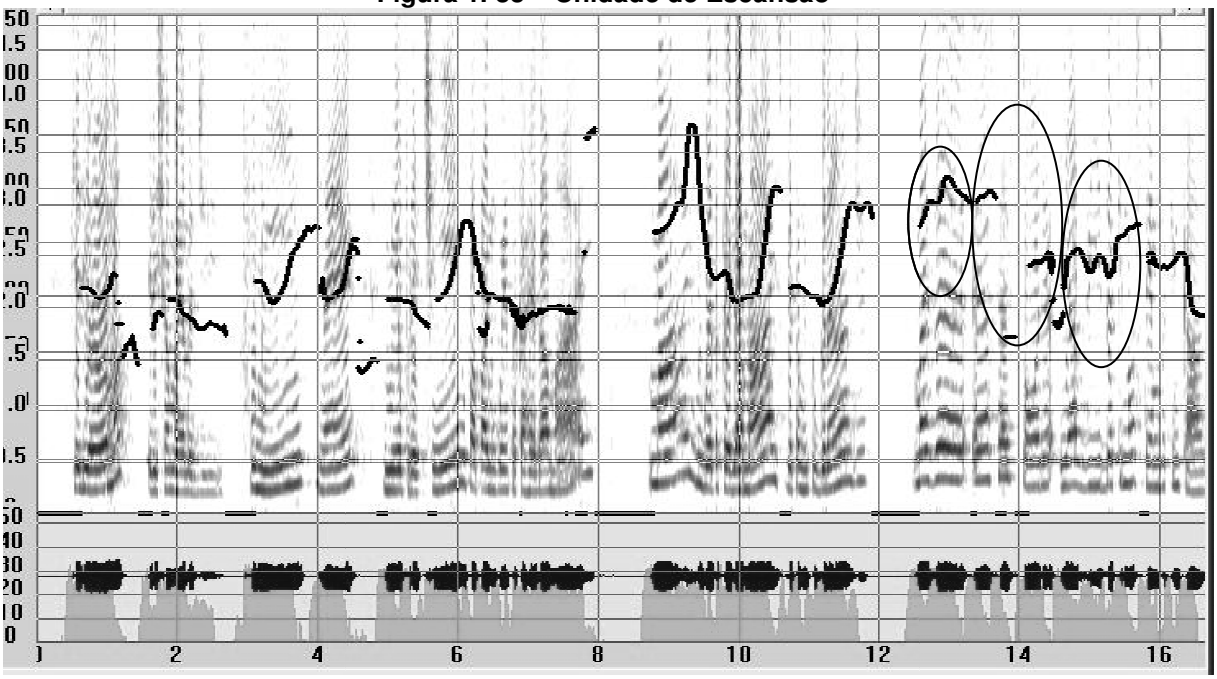
A unidade tonal é sempre marcada por um perfil não terminal. Segundo Cresti (2008, p.36), apenas as unidades textuais de Tópico, Comentário e Parentético foram encontradas escansionadas. Há, também, alguns casos raríssimos de APC.

A unidade de escansão é sempre composicional situacionalmente com as outras unidades tonais da mesma unidade informacional, o que mais uma vez demonstra que o domínio da sintaxe é informacional.

Exemplo 1.30 (bpubdl01) (CD áudio 1.30)

*CAR: porque ele ama /=COB= ele chama ela de /=SCA= fuminho /=COB= né /=COB= meu pretinho /=COB= papai nũ güenta carregar mais /=COB= mas /=DCT= nũ güenta pegar mais /=COB= porque tá muito grande /=COB= **mas é [/1] =SCA= é essa é a história /=SCA= e é a vida /=SCA=** que nós temos aqui em casa //=COM=

Figura 1. 33 – Unidade de Escansão



Na figura 1.33, as elipses indicam as unidades de Escansão nas quais foi dividida a unidade de Comentário.

1.9.2 Comentários Múltiplos (CMM)

O segundo caso de não correspondência biunívoca entre unidade terminada e comentário refere-se aos comentários múltiplos. Os Comentários Múltiplos são padrões ilocucionários que por sua natureza necessitam de mais de um COM. É o próprio programa dessas ilocuções que requer mais de um COM para ser realizado.

Tratam-se de padrões retóricos em que dois ou mais comentários são interpretados holisticamente, com efeito perlocutório diferente da soma das ilocuções consideradas isoladamente. Dentre os tipos de ilocução, há vários padrões que necessariamente se distribuem em mais de uma unidade de comentário. Os mais conhecidos são: 1) a ilocução de elenco, em que cada elemento do elenco localiza-se em uma unidade de comentário diferente e somente o último apresenta uma quebra terminal; 2) a ilocução de comparação, em que cada elemento da comparação também se localiza em uma unidade de comentário diferente; 3) a ilocução de hipótese ou relação de necessidade, em que dois comentários possuem uma relação especialmente estreita, seja de natureza temporal ou de semelhança, ou causal ou consecutiva ou de outra natureza lógica, sem que haja necessariamente um operador lexical que explicite a relação, sendo ela realizada através do padrão prosódico; 4) a ilocução de reforço, em que dois COM se reforçam e 5) a ilocução de confirmação, cujo objetivo é solicitar do ouvinte a confirmação do que fora mencionado no COM anterior.

Os CMM são formados a partir de ilocuções diferentes que são tão padronizadas em uma determinada língua e são interpretadas holisticamente. Pode-se dizer, dessa forma, que os CMM são como uma "gramaticalização" retórica.

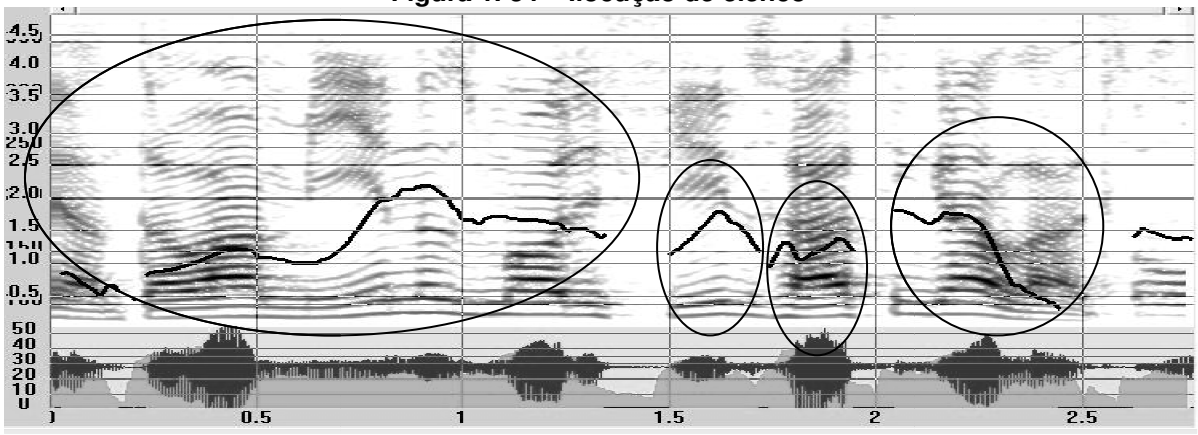
A investigacão da fala espontânea com base no *corpus* C-ORAL-BRASIL mostra-se promissora na descoberta de novos padrões. Inclusive, durante a fase de etiquetagem informacional, diversos padrões emergiram, como o de dupla pergunta, foco duplo e alternativa (MITTMANN, em preparação).

Abaixo ilustram-se dois exemplos. O exemplo 1.29 é um padrão de elenco e o 1.30 um padrão de alternativa. As elipses, nos espectrogramas, delimitam os comentários na ordem em que aparecem os enunciados.

Exemplo 1.31 (bfamcv01) (CD áudio 1.31)

*EVN: **tá o Aminas /=CMM= Mauro e Filhos /=CMM= Racing /=CMM= Galáticos //=CMM=**

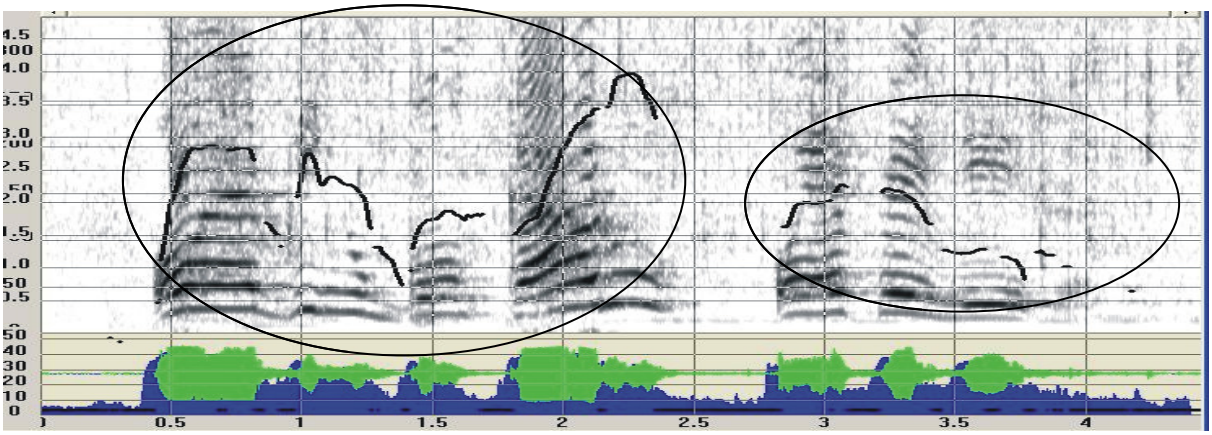
Figura 1. 34 – Ilocução de elenco



Exemplo 1.32 (bfamdl04) (CD áudio 1.32)

*SIL: **ou é vinho bom caro /=CMM= ou é cerveja //=CMM=**

Figura 1. 35 – Ilocução de alternativa



Pesquisas mais recentes mostram que certamente há outros padrões de comentários múltiplos que precisam ainda ser formalizados. O estudo dos comentários múltiplos é um dos aspectos da teoria em maior evolução no momento presente (RASO, 2012).

1.9.3 A estrofe[26]

O terceiro caso de não correspondência biunívoca entre unidade terminada e comentário é a estrofe[27]. As estrofes são unidades terminadas, constituídas por mais de um comentário, e não raramente por vários comentários, cada um ligado ao outro por meio de um sinal prosódico de continuidade. A prosódia sinaliza que, apesar de ter sido realizada uma ilocução, a sequência não deve ser considerada como concluída, mas sim ligada à próxima sequência informacional. Os comentários das estrofes são chamados de *comentários ligados* (COB). Eles não possuem nenhuma relação padronizada, como acontece com os CMM, mas são justapostos um ao outro em um andamento processual. Geralmente esses comentários apresentam um enfraquecimento da força ilocucionária, que é sempre homogênea; isso significa que os COB da mesma estrofe pertencem à mesma classe de ilocuções.

A força ilocucionária é motivada pela interação entre os interlocutores. Nas estrofes a interação se enfraquece, e o mesmo acontece para a força ilocucionária. Por isso, a sequência terminada de uma estrofe não corresponde à realização de um ato de fala ou de atos de fala padronizados, mas a uma atividade de fala que expressa um pensamento em curso através de texto oral. Enfraquecimento da interação e das ilocuções, andamento processual e homogeneidade ilocucionária fazem com que a estrofe como um todo manifeste uma única atitude com relação ao interlocutor. Por outro lado, as ilocuções de enunciados diferentes manifestam cada uma a sua própria atitude.

As estrofes possuem uma estruturação interna na qual à homogeneidade dos valores ilocucionários corresponde uma variação dos valores modais. Dentro de uma estrofe, em torno de cada COB pode-se criar um subpadrão com outras unidades informacionais, principalmente tópicos e parentéticos. Esses subpadrões são frequentemente ligados entre si através de um conector discursivo.

[26] Toda essa parte foi retirada do C-ORAL-BRASIL, capítulo 3.
[27] *Stanza* na terminologia originária da teoria (CRESTI, 2010).

Em princípio, as estrofes podem ser encontradas em qualquer tipo de texto, mas são típicas dos textos monológicos, ou seja, daquelas tipologias textuais em que a interatividade entre os interlocutores é por natureza reduzida. Os exemplos 50-53 mostram quatro estrofes: a primeira e a quarta são argumentativas, a segunda é uma estrofe expositiva, e a terceira, uma estrofe narrativa.

Exemplo 1.33[28] (bpubmn01) (áudio 1.33)
*SHE: agora /=PHA= se eu for olhar tempo mesmo /=TOP= né /=PHA= de [/1]=EMP= de [/1]=EMP= de sentar e escrever /=APT= aquilo /=APT(1)= pra dar aquela aula ideal /=APT= né /=PHA= aquela [/1]=EMP= aquela aula igual se dá no EDUCONLE que cê acha que é &o [/1]=SCA= é tudo de bom /=PAR= né /=PHA= nũ tem /=COB= porque também eu acho que /=SCA= &he /=TMT= &v [/1]=EMP= depende muito dos alunos de tar preparado /=COB= por exemplo /=INT= cê prepara aquela aula /=COB= maravilhosa /=COB= e acha que vai dar tudo certo /=COB= de repente ocê encontra lá o aluno /=COB= né /=PHA= que /=INT= &he /=TMT= &he /=TMT= nũ [/1]=EMP= nũ [/1]=EMP= foi pra escola /=COB= nũ almoçou /=COB= tá [/1]=EMP= tá com fome /=COB= ou brigou com o pai /=COB= ou aquelas questões todas familiares que vão influenciar //=COM=

A estrofe do exemplo 1.33 é muito interessante por vários motivos. Em primeiro lugar mostra que a estrofe pode ter uma dimensão considerável. Nesse caso, sem se considerarem as unidades tonais de tomada de tempo e vazias, tem-se 25 unidades tonais para 23 unidades informacionais, o que significa que duas unidades são escansionadas. Em segundo lugar mostra a complexidade interna da sequência: 5 unidades são de fáticos que servem para sinalizar que o canal está aberto e o turno em curso, importante quando se está processando verbalmente um pensamento complexo; a quantidade de unidades vazias e de tomada de tempo, assim como as unidades de escansão, também reflete o esforço processual do falante. Em terceiro lugar mostra o esforço de contextualização da primeira parte, com um tópico muito complexo, enriquecido por 3 APT, dos quais o segundo é de nível hierárquico

[28] Note-se que EMP é a sigla para *empty*, que indica um segmento que não possui valor informacional, sendo a unidade retratada, ou seja, eliminada pelo falante. O número que segue o sinal de retratação indica a quantidade de palavras canceladas. TMT é a sigla para *time taking* e indica uma unidade tonal usada para tomada de tempo.

diferente (um caso raríssimo), e um parentético, antes de uma segunda parte que realiza as ilocuções. Nesta segunda parte também podem-se notar diferentes níveis: uma ilocução (o primeiro COB) e uma série de metailocuções de primeiro nível (depois do primeiro INT) e de segundo nível (depois do segundo INT). Ao todo, as ilocuções são 10, todas homogêneas, grande parte das quais são metailocuções que exemplificam o conteúdo da primeira.

Exemplo 1.34 (bfammn06) (áudio1.34)
*JOR: e aí /=TOP= eu consegui /=i-COB= a [/1]=EMP= com a experiência que e tinha dentro da multinacional /=PAR= concorrer à vaga e &f isso me facilitou /=COB= e eu passei pra área comercial da empresa pra vender /=COB_s= disjuntores /=COB= transformadores /=COB= motores de /=INT= corrente contínua /=COB= corrente alternada /=COB= isoladores /=COB= e /=DCT= relés de proteção secundária /=COB= e assim foi iniciando a minha vida comercial //=COM=

O exemplo 1.34 é uma estrofe de caráter mais expositivo, com COB homogêneos e com pouca força ilocucionária. Aqui também a hierarquização é muito forte. Dentro das estrofes pode acontecer de se acharem unidades informacionais extremamente raras, geradas pela complexidade estrutural. No exemplo 50 encontramos um APT de segundo nível; no exemplo 1.33 uma lista de parentéticos (PRL); aqui observamos um COB_s, ou seja, um comentário ligado que funciona como um subordinador dos que vêm depois, gerando um nível dependente do COB_s; dentro da mesma sequência, um INT produz uma subordinação ulterior, gerando um terceiro nível de dependência. Note-se também o PAR que interrompe o primeiro COB.

Exemplo 1.35 (bfammn03) (áudio I.35)
*ALO: e /=DCT= e eu /=TOP= cabei com a loja /=COB= mudei pra cá /=COB= tô negociando esses ano todo aí /=COB= eu /=SCA= dei a conta pa [/5]=SCA= mandei /=SCA= pessoal /=SCA= cobrar lá /=COB= mas nũ quis [/2]=SCA= né /=PHA= nũ quis &co [/1]=SCA= nem cobrar dela /=COB= porque eu sabia que era viúva /=COB=

e tal /=COB= e dificuldade /=COB= muitos filho /=COB= dificuldade financeira /=PAR= falei /=INT= ah /=CMM_r= nts //=CMM_r=

O exemplo 1.35 mostra uma estrofe narrativa. Nela podem-se notar 22 unidades tonais, das quais 6 são unidades de escansão, sinal da processualidade da narração. Em uma estrofe narrativa a necessidade de se contextualizar é muito menor que na argumentação. O que vale a pena observar aqui é a maior força ilocucionária dos comentários, sempre muito homogêneos entre eles, e a pouca hierarquização interna. A metailocução de discurso reportado sinalizada pelo INT representa uma maneira de tornar mais vivo o conto, e não uma complexificação como as metailocuções da estrofe argumentativa. Muito interessante também o CMM final, realizado através de uma interjeição e de um ruído com evidente valor comunicativo.

Exemplo 1.36 (bfammn05) (áudio 1.36)
*CAR: quando /=SCA= a pessoa veio buscar ela /=TOP= choramo muito /=COB= chorou eu /=PRL= chorou o pai /=PRL= chorou o irmão /=PRL= né /=PHA= porque a gente nũ queria devolver /=COB= quando chegou lá /=TOP= a gente sentamo lá /=COB= a irmã dela tava com a gente /=COB= que é a tia /=SCA= biológica /=COB= tava com a gente /=COB= a tia deu uma força muito grande /=COB= pa ela que ea devolvesse ela pra nós /=COB= aí ela devolveu /=COB= falou assim /=INT= não quero isso não //=COM_r=

O exemplo 1.36 é uma sequência dramática dentro de uma narração. Não se trata de uma sucessão de eventos, como na estrofe narrativa anterior mais simples, mas de uma estruturação mais complexa que cria uma expectativa que conduz à conclusão final, ulteriormente dramatizada através do discurso reportado.

Esses exemplos de estrofes mostram a grande riqueza estrutural que as sequências terminadas podem alcançar. Trata-se de um aspecto ainda a ser melhor estudado e

constitui um dos principais traços que diferenciam a fala dialógica, mais interativa, e a fala monológica, em que a interação deixa o campo para uma processualização oralizada do pensamento.

1.10 O esquema da Língua em Ato

O termo ilocução foi primeiramente proposto por Austin (1962) que o define como "aquilo que realizamos quando falamos, e o que realizamos com o dizer". Segundo o filósofo, a execução do ato linguístico envolve o cumprimento de três atos simultâneos e distintos entre si: o ato locutório, o ato ilocutório e o ato perlocutório. Esse conceito seria mais tarde retomado por Searle (1969), para quem a ilocução de um ato é sempre identificada pelo reconhecimento de um verbo performativo explícito ou implícito na locução; em outras palavras, para Searle a ilocução tem uma base lexical implícita.

Na teoria da Língua em Ato, a perlocução se fundamenta na pulsão afetiva. Essa por sua vez é caracterizada como pulsão e representação, isto é "a ativação pulsional de um esquema motor que implica em uma intervenção física no mundo" (Cresti 2000, p.43-44). Logo, a manifestação do falante na interação está ancorada à pulsão do locutor, que não é convencionalizada e que se traduz em ações verbais socialmente convencionalizadas. Segundo Cresti apud Cresti & Firenzuoli (2001, p.18) a atitude pulsional pode ser distribuída em cinco classes, que se baseiam nas características afetivas das ações humanas e nas relações dinâmicas entre os falantes. Essas atitudes[29] são inconscientes, isto é, completamente livres e não reguladas por convenções. Tais classes representam diferentes reações linguísticas de um falante a um *input* relacional humano, dependendo da dinâmica relacional entre os participantes da interação. Assim, pode-se dizer que a afetividade ativada torna-se um esquema acional específico social e convencionalmente codificado com uma força pragmática.

[29] A ordem vai da maior ativação afetiva à menor ativação. A recusa ativa mais afetividade pulsional, o rito a menor.

As cinco classes atitudinais caracterizam-se por diferentes graus e níveis de ativação afetiva. São elas:

1. Recusa: atitude ou estado de espírito de liberdade e separação por parte do falante em relação ao seu interlocutor, permitindo um confronto completo entre os participantes da interação.
2. Asserção: uma disposição, atitude ou autoconfiança por parte do falante, como nas próprias realizações de pensamento que permite a manifestação de julgamentos, descobertas, avaliações etc. Ex: explanação, declaração, confirmação de algo.
3. Direção: atitude do falante em tomar em consideração as habilidades, possibilidades, disponibilidade do interlocutor enquanto se espera transformar o mundo por meio das ações, informações ou que o ouvinte transforme a si mesmo com relação a seu horizonte de atenção, seu ponto de vista.
4. Expressão: atitude de manifestação "estética" de estados mentais, na expectativa de que o ouvinte se torne consciente disso e compartilhe desse sentimento. São exemplos a ironia, exclamação e insinuação.
5. Rito: um comportamento externo de execução de tarefas linguísticas que tem valor e efeito legal e social como os cumprimentos, as condolências, um batizado ou uma condenação.

Como se pode perceber, o conjunto dos atos ilocutórios (classes atitudinais) não é passível de definição a partir de critérios lógicos e lexicais[30], mas deve ser identificado na língua com base em pesquisas empíricas realizadas em *corpora*. Inclusive, há ações muito comuns na comunicação natural, como as dêixis e os chamamentos, que não possuem performativos. Assim como há performativos que indicariam ações não encontradas em corpora de fala espontânea ou ainda poderiam classificar da mesma maneira ações diferentes e vice-versa.

Para a língua italiana, por exemplo, o LABLITA[31] identificou, por ora, o repertório apresentado na tabela 1.1 (MONEGLIA, 2011).

[30] Ver Searle (1969)
[31] Ver Firenzuoli (2001) e http://lablita.dit.unifi.it/corpora/

Tabela 1.1 - Tabela de referência das classes e dos tipos de atos de fala baseada no *corpus* LABLITA, traduzido de Moneglia (2011).

REPRESENTATIVOS	Objeção	**DIRETIVOS**	Anúncio	**EXPRESSIVOS**	Reclamação	**RITOS**
Conclusão	Confirmação	Chamamento distal objeto não visível	Conselho	Exclamação	Imprecação	Agradecer
Asserção fraca	Aprovação	Chamamento distal - objeto visível	Advertência	Expressão de contraste	Insinuação	Saudar
Resposta	Desaprovação	Chamamento proximal	Sugestão	Expressão de obviedade	Derisão	Desculpar-se
Comentário	Acordo	Dêixis distal	Proposta	Suavização	Provocação	Dar boas-vindas
Asserção forte	Desacordo	Dêixis proximal	Recomendação	Expressão de surpresa	Repreensão	Congratular
Identificação		Apresentação (objeto/evento)	Convite	Expressão de medo	Dica	Desejar
Verificação		Introdução (pessoa)	Insistir em um pedido	Expressão de ajuda	Encorajar	Elogios
Afirmação		Pedido de informação	Permissão	Expressão de incerteza	Assegurar	Declarações de valor legal
Hipótese/Suposição		Pedido de ação	Autorização	Expressão de dúvida	Ameaçar	*condenação*
Explicação		Ordem	Proibição	Expressão de certeza	Renunciar	*condolências*
Inferência		Pergunta total	Instrução	Expressão de desejo		*batismo*
Definição		Pergunta parcial		Expressão de incredulidade		*promessa*
Narração		Pergunta alternativa		Expressão de pena		*aposta*
Descrição		Pedido de confirmação		Ironia		
Citação		Discurso reportado		Pesar		**RECUSAS**

1.11 Modalidade versus ilocução e atitude

Na literatura, várias são as definições acerca do que vem a ser modalidade. Nessa pesquisa, todavia, nos ateremos apenas à acepção dada pela Teoria da Língua em Ato. (CRESTI 2002b, 2005b; TUCCI 2005, 2008a, 2008b).

A modalidade tradicionalmente se distingue em Alética, Epistêmica e Deôntica. Na modalidade Alética o ponto de vista do falante refere-se à necessidade ou possibilidade de a veracidade das proposições serem verificadas no mundo real ou no mundo possível, por força da lógica factual ou perceptiva. Exemplo: Uma zebra tem de ter listras pretas e brancas.

Na modalidade Epistêmica o ponto de vista do falante baseia-se na possibilidade ou necessidade de uma proposição ser verificada em mundos possíveis, que são especificamente relacionados à crença, opinião ou atitude do falante. Exemplo: Mário pode sair amanhã (ou não).

Na modalidade Deôntica, o ponto de vista do falante (destinado como agente moralmente responsável) diz respeito a dever, obrigação, permissão ou desejos expressos em uma proposição. Exemplo: Você não pode mentir para mim! (poder aqui tem valor diferente do que no exemplo anterior)

Baseando-se nessa tradição, Tucci (2008) diz que o material locutivo (*dictum*), no enunciado, sempre pertence a uma unidade informacional e a modalidade à articulação dessa informação[32]. Portanto, o domínio da modalidade não seria o enunciado, mas a unidade informacional. Os modalizadores são composicionais dentro da mesma unidade informacional, mas não além dela.

Cresti define modalidade como sendo um feito puramente semântico, já que se trata da atitude do falante em relação ao conteúdo proposicional do que ele diz. Diferentemente, a ilocução do falante dirigida ao interlocutor, segundo a autora, é um fato pragmático, uma ação socialmente codificada. Mello; Raso (2011) propõem, também, baseando-se na distinção entre modalidade e ilocução de Cresti, distinção entre essas duas categorias e a categoria de atitude.

[32] "Il materiale locutivo (il *Dictum*), nell'enunciato, appartiene sempre ad una UI. Gli indici di modalità hanno quindi una distribuzione nell'enunciato rispetto alla sua articolazione informativa." TUCCI (2008ª, p.63)

1.12 Resumindo

Procurou-se nesse capítulo apresentar a Teoria da Língua em Ato (Cresti, 2000), que se fundamenta na correspondência entre unidade de ação e unidade linguística (enunciado), por meio de parâmetros entonacionais. Apresentaram-se as unidades informacionais textuais e dialógicas, ressaltando a unidade de COM como a mais importante, por ser a única necessária e suficiente à constituição de um enunciado.

As unidades informacionais textuais são aquelas que compõem o texto do enunciado propriamente dito. Dentre elas encontramos as unidades de Comentário (COM), de Tópico (TOP) e de Apêndice de Comentário (APC), a de Parentético (PAR), de Introdutor Locutivo (INT), Apêndice de Tópico (APT). As unidades informacionais dialógicas ou não textuais, por sua vez, são aquelas que não contribuem para a constituição semântica de um enunciado; todavia dedicam-se ao cumprimento pragmático desse enunciado. São elas: Incipitário (INP), Conativo (CNT), Conector Dialógico (DCT), Fático (PHA), Alocutivo (ALL) e Expressivo (EXP).

Mostramos, também, que há uma correspondência biunívoca entre unidade tonal e unidade informacional, assim como entre enunciado e ilocução. Esse isomorfismo se realiza com alta regularidade, mas não sempre. Ele se quebra em três circunstâncias: Unidade de Escansão (SCA), Comentários Ligados (CMM) e estrofes.

CAPÍTULO 2

A UNIDADE DE APÊNDICE DE COMENTÁRIO (APC)

Este capítulo objetiva apresentar um aprofundamento teórico sobre a unidade de Apêndice de Comentário, segundo a Teoria da Língua em Ato (Cresti, 2000). Para isso serão expostas as considerações de Cresti (2000); Tucci, E (2006); Ulisses (2008); Lombardi Vallauri (2008) e Raso; Ulisses (2008).

2.1 Definição da Unidade de APC

É sabido que na primeira metade do século passado, alguns estudos já se voltavam para os aspectos da diamesia falada (Saussure, 1916; Bally, 1932). As propostas teóricas apresentadas continham certo grau de consenso em relação a alguns construtos e aspectos morfossintáticos típicos do texto oral. Entre os construtos em geral associados à expressão "sintaxe da segmentação" têm-se o deslocamento à direita e o deslocamento à esquerda, as orações clivadas e pseudoclivadas, os anacolutos, as inversões e as frases nominais.

Na literatura tradicional, a funcionalidade do apêndice é caracterizada segundo a estrutura de deslocamento à direita. Entretanto, no quadro teórico proposto por Cresti (2000), este e outros aspectos funcionais devem ser considerados com bases na articulação informacional, sinalizada pela entonação e fundamentados em análises complexas de *corpora*. Essas análises, segundo Cresti, têm revelado que diversos construtos e fenômenos de segmentação podem apresentar-se de forma articulada em diferentes unidades tonais ou em forma linearizada, dentro da mesma unidade tonal.

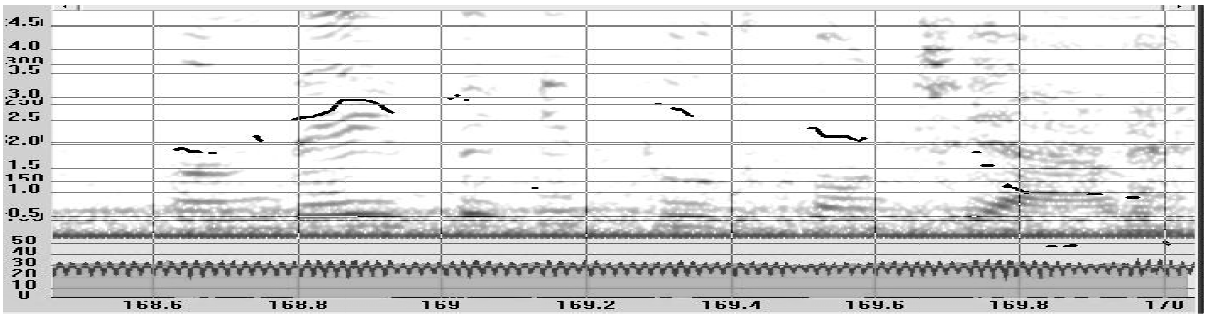
Segundo a Teoria da Língua em Ato, (Cresti, 2000) a unidade de APC é definida como *sendo a integração textual de uma unidade informativa de comentário ou de*

tópico, e como tal deve necessariamente seguir a unidade na qual está interligada; nesse caso, o comentário. (TUCCI, E 2006 p.23)[33].

Um enunciado com mesmo conteúdo locutivo, como veremos logo abaixo, algo extremamente raro em língua espontânea, dependendo de como é enunciado, sofre alteração em sua segmentação prosódica.

Exemplo 2.1 (bfamdl05) **(**CD áudio 2.1)
*ANE: [87] andei isso aqui tudo com o Tiago //=COM =

Figura 2.1 – Enunciado formado de uma unidade de COM



A figura 2.1 ilustra um enunciado simples, constituído por uma unidade de COM. Esse enunciado é completamente autônomo do ponto de vista prosódico e pragmático.

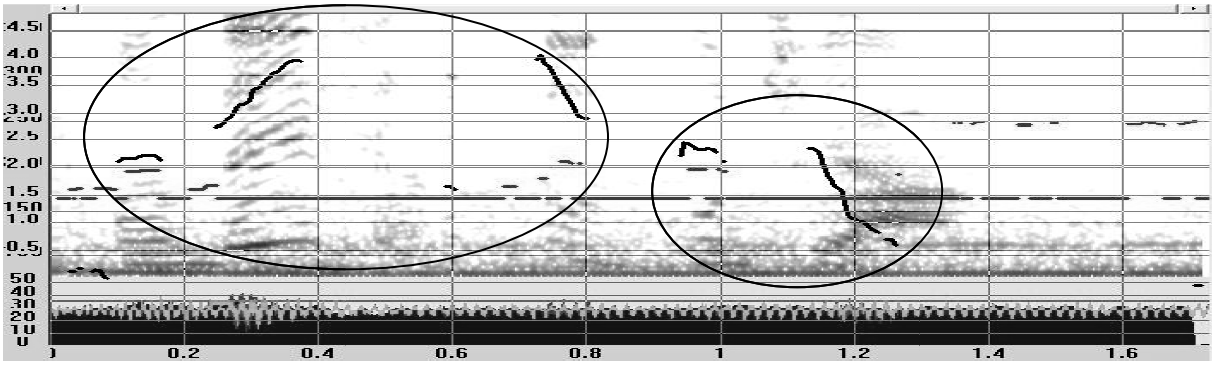
Agora observe o mesmo enunciado proferido com uma entonação diferente. Veja-se que se trata de um enunciado complexo formado por duas unidades tonais. A primeira delas, uma unidade de COM, que possui uma quebra prosódica interpretada como um valor conclusivo; a segunda, uma unidade de APC cuja função é apenas de integração textual à unidade de COM.

[33] [*Le Appendici sono state definite da CRESTI (Cresti, 2000) come "l'integrazione testuale di un'unità informativa di Comment o di Topic", e in quanto tali devono necessariamente seguire l'unità di cui sono integrazione, in questo caso il Comment.]*

Exemplo 2.2 (bfamdl05) **(**CD áudio 2.2)

*ANE: [85] andei isso aqui tudo /=COM= com o Tiago //=APC=

Figura 2.2 – Enunciado formado de uma unidade de COM e uma unidade de APC



A primeira elipse assinala uma unidade de COM e a segunda, a unidade de APC.

Como vimos no capítulo 1, as unidades são identificadas através da aplicação simultânea de três critérios: Funcional, Entonacional e Distribucional.

2.1.1 Critérios Funcional, Entonacional e Distribucional

As características da unidade de Apêndice de Comentário aqui descritas são resultados dos estudos do grupo LABLITA a partir de uma amostra extraída do *corpus* de italiano falado presente no C-ORAL-ROM. As análises feitas em relação à unidade no PB serão explicitadas no capítulo 5 desta pesquisa.

Assim como as outras unidades que fazem parte do padrão informacioanl, a unidade de APC é fundamentada a partir de três critérios: funcional, entonacional e distribucional.

Funcionalmente, a unidade de apêndice de comentário realiza uma mera compilação de um texto, devendo estar posicionada após a unidade da qual faz a integração, ou

seja, uma unidade ilocucionária (COM/CMM/COB). Ulisses explica esse fato dizendo que

> um falante diante da tarefa de realizar uma unidade de COM ou TOP, e percebendo problemas de execução, seja por questões de erros, por mudar de idéia quanto ao que disse, ou por achar que aquilo que disse não é adequado, imediatamente integra novas estruturas linguísticas à unidade que o precede em forma de apêndice, de modo a realizar a sua expansão semântica, correção ou reestruturação. (2008, p.81)

Entonacionalmente, a unidade de APC se configura como uma unidade de sufixo (‘t Hart; Collier; Cohen 1990), subordinada ao COM, e não possui foco funcional, mantendo sempre um perfil nivelado ou descendente identificado por um abaixamento do tom de voz, pela baixa intensidade e pelo fato desse movimento único corresponder à unidade tonal inteira, sem variação de movimento, independente da estruturação silábica (CRESTI 2002a; FIRENZUOLI, 2002).

Distribucionalmente, deve suceder à unidade informacional de comentário, mesmo que não de forma contígua.

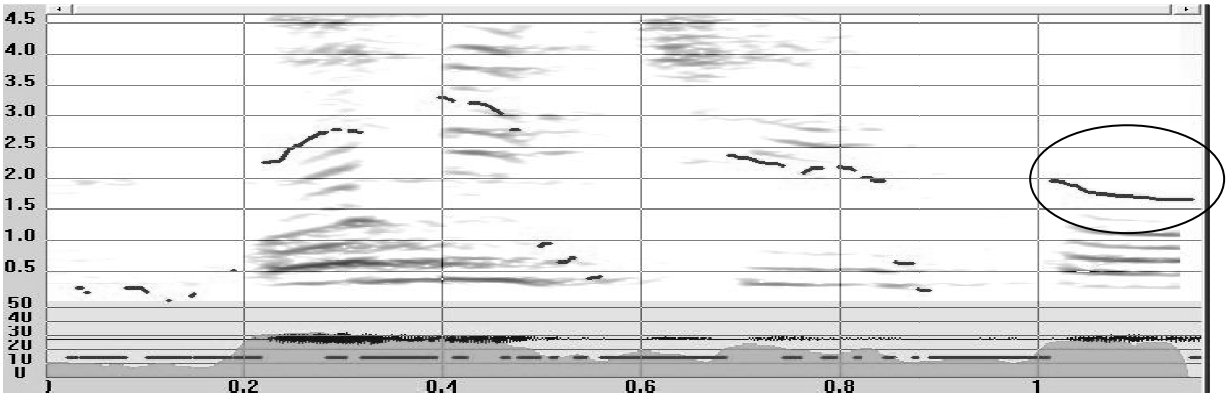
Além disso, não é uma unidade autônoma, isto é, não pode ser interpretável isoladamente.

Outra característica da unidade de apêndice refere-se ao valor de F0, que pode variar conforme o término da unidade de COM. Passemos a mais um exemplo.

Exemplo 2.3 **(**CD áudio 2.3)

*REG: omitir /=COM= **só** //=APC[34]

Figura 2.3 – Unidade de Apêndice de Comentário



A figura 2.3 ilustra um enunciado realizado em duas unidades informacionais. A primeira delas é uma unidade de comentário, *omitir,* e a segunda unidade, que está delimitada pela elipse é uma unidade de apêndice de comentário, *só.*

Se, por exemplo, a unidade de COM é uma pergunta, consequentemente a unidade de APC pode ter uma F0 mais alta do que em outro enunciado em que isso não ocorre, pois a unidade de COM termina alta. Segundo Cresti, essa variação é uma característica peculiar às unidades de tipo sufixo.

Para nos certificarmos de que a segunda unidade é realmente uma unidade de APC e não outra, ocupando a mesma posição em final de enunciado, (esse assunto será detalhado no capítulo 5), procuramos aferir as médias da frequência fundamental e intensidade, em relação à unidade de COM. Os resultados obtidos para o COM foram de F0 média de 271,31 Hz e Intensidade média de 23,77 dB. A segunda unidade, que apenas integra a unidade anterior, com função de preenchimento, tem sua F0 média de 256,39 Hz e possui uma intensidade média de 23,15 dB. O fato de a média de F0 da unidade de APC ser menor que a do comentário permite que a

[34] Esse exemplo não faz parte do subcorpus, mas sim de um texto que irá compor o C-ORAL-BRASIL.

segunda unidade seja um APC. Embora esse critério seja essencial para a identificação da unidade de APC, não é o único. Para a certificação de uma unidade são necessárias também as análises dos critérios distribucional e funcional, apresentadas por Cresti (2000).

2.1.2 Classificação Informacional do APC

A unidade de APC é por definição uma unidade de integração textual. A maior parte das expressões que são usadas funcionalmente como unidade de APC corresponde a um conteúdo "vazio" ou a um conteúdo genérico do ponto de vista semântico. Estudos realizados por Tucci, E (2006), entretanto, com dados do C-ORAL-ROM, nos permitem distinguir entre os diferentes tipos de informação que integram a unidade de comentário. Tucci, E operou uma primeira classificação dessas informações em quatro categorias distintas: repetições de expressões do tema, preenchimento, retomada textual e informação tardia.

As repetições de expressões do tema do discurso são discriminadas por tipologia ou distribuição e podem ser literais (aquelas que não modificam uma dada expressão linguística) ou com variação (a repetição do conteúdo semântico apresenta-se em forma de sinônimos ou perífrases). Distribucionalmente, as repetições podem ocorrer: 1) de forma contígua: quando o conteúdo repetido é expresso no mesmo enunciado; 2) não-contígua: o conteúdo repetido é expresso em outro enunciado de um mesmo turno, ou fora de turno; e 3) *Leit Motiv:* quando ocorrem como um tipo de refrão no interior de uma conversação ou de um monólogo.

Exemplo 2.4 (bfammn01) (CD áudio 2.4)

*MAI: [13] o diâmetro dea deve dar uns [/1]=SCA= uns quarenta cinqüenta centímetro de [/1]=SCA= de &s [/2]=EMP= de grossura /=COM= o diâmetro dela //=APC=$

Embora não seja nosso objetivo propor uma nova classificação informacional para a unidade de APC, acreditamos que as repetições podem ser desdobradas em dois tipos: parcial de forma contígua ou não. No capítulo 5 esse assunto será detalhado.

Os **preenchimentos** realizam a expansão da unidade precedente sem repetir seu conteúdo semântico ou acrescentar informações e geralmente constituem-se de advérbios ou advérbios focalizadores[35];

Exemplo 2.5 (bfammn01) (CD áudio 2.5)
*MAI: [79] é lugar /=SCA= perigoso mesmo o norte de Minas /=COM= lá //=APC=$

A **retomada textual** é a referência ao discurso em si ou a parte do discurso. Pode ser realizada em forma de dêixis discursiva (quando se refere ao discurso em si) ou de recontextualização (quando retoma sinteticamente uma parte do discurso);

Exemplo 2.6 (bfammn01) (CD áudio 2.6)
*MAI: [96] isso aqui /=TOP= em vista de lá /=TOP= é [/1]=EMP= é um cerrado /=COM= aqueas mata lá //=APC=$

Nesse exemplo, a retomada textual se deu por meio de uma reformulação.

A **informação tardia**: refere-se à adição de novas informações, quando a unidade de comentário em si é suficiente para cumprir a ilocução.

Exemplo 2.7 (bfamcv01) (CD áudio 2.7)
GIL: [65] da gente /=TOP= pra alguns setores da organização /=TOP= chamar o pessoal /=COM= dos outro times //=APC=$

Assim como a repetição, acreditamos, também, que a informação tardia possa ser desdobrada no PB. O capítulo 5 tratará desse assunto.

[35] Em Português temos alguns advérbios focalizadores, tais como: exatamente, realmente, claramente etc.

2.1.3 Correlatos Semânticos e Morfossintáticos

De um ponto de vista lexical a unidade de APC apresenta uma alta variabilidade morfossintática, podendo ser composta por palavras isoladas, quase sempre advérbios, (conforme se verificou para o italiano na pesquisa com os dados do C-ORAL-ROM) construções sintáticas ou orações. Entre as construções sintáticas, no Italiano, observou-se uma preferência pelos SN e SP. Quando são orações, há uma predominância, segundo esse estudo, de orações subordinadas relativas e finais. Há, também, algumas ocorrências de orações coordenadas com expressões de obviedade.

No Italiano, encontrou-se a negação nos APC somente na categoria informal, segundo Tucci,E (2006). A negação ocorreu, principalmente, nos APC que funcionavam como repetição de uma informação dada. Muito raramente foram encontrados clíticos na unidade de APC e quando encontrados esses clíticos funcionavam como referenciação anafórica de elementos em comentário.

2.1.4 A modalização e a unidade de APC

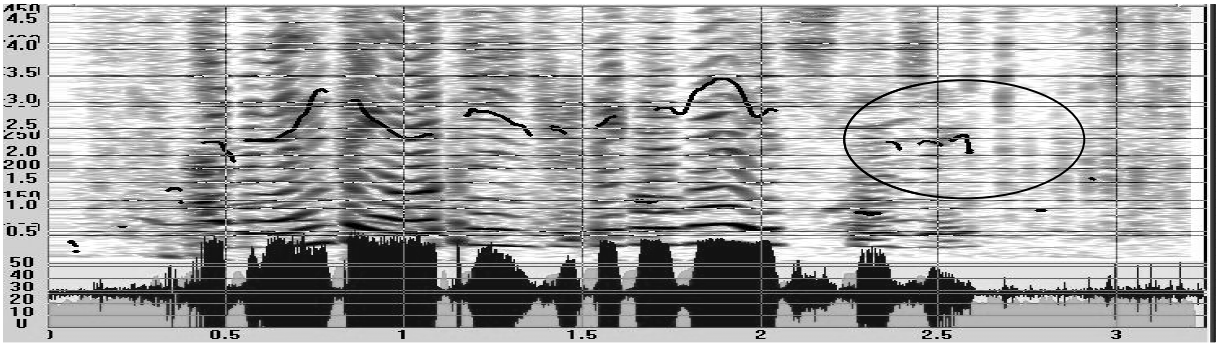
A unidade de APC tem sua modalidade definida pela unidade de comentário, isto é, a modalidade do APC é a mesma da unidade de comentário. Essa ideia é confirmada por Tucci, I (2006), quando seus estudos com dados do italiano mostraram como dentro de um mesmo enunciado apenas algumas unidades informacionais podem ser modalizadas pelos falantes: a unidade de Comentário, a unidade de Parentético, a unidade de Tópico e a unidade de Introdutor Locutivo. Na verdade, Cresti (2000 p.78-80) já havia exposto em sua Teoria da Língua em Ato que a unidade de APC, por ser apenas uma unidade de compilação textual, mantém o ponto de vista da unidade à qual se integra.

No Italiano, não foram encontrados APC com expressão lexical modal (Cresti, 2008). O que se verificou, em alguns casos de APC, fora a repetição literal da modalidade do COM, através de uma retomada lexical.

No PB, os dados até agora analisados reforçam a posição de que o APC segue a mesma modalidade e ponto de vista da unidade de comentário ou simplesmente não tem modalidade. Passemos à análise de dois exemplos:

Exemplo 2.8 (bfamdl01) (CD áudio 2.8)
*REN: [569] ah /=INP= também nũ tem outro /=CMM= a gente vai ter que levar esse /=CMM= **de todo jeito** //**=APC=$**

Figura 2.4 – Unidade de APC com mesma modalidade do COM

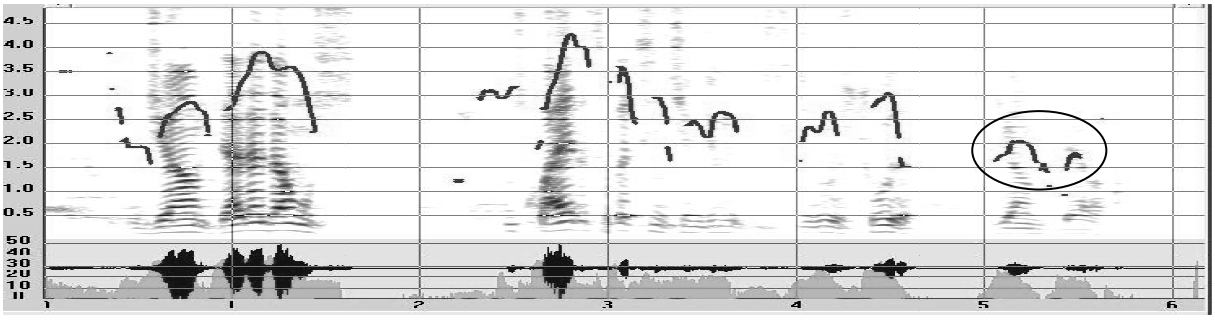


Nesse exemplo, podemos perceber que a unidade de APC, delimitada pela elipse, funciona como integração textual. Ela possui a mesma modalidade da unidade de COM (a gente vai ter que levar esse), em que a expressão “vai ter que” corresponde ao verbo DEVER, dando a ideia de obrigatoriedade, ou seja, uma modalidade epistêmica de necessidade. Ao dizer “de todo jeito”, o falante endossa a necessidade de ter de fazer algo.

Exemplo 2.9 (bfammn01) (CD áudio 2.9)

*MAI: o diâmetro dea deve dar uns [/1]=SCA= uns quarenta cinquenta centímetro de [/1]=SCA= de &s [/2]=EMP= de grossura /=COM= **o diâmetro dela** //**=APC=**

Figura 2.5 – Unidade de APC sem repetição de modalidade do COM



Ao contrário do exemplo 2.9, a unidade de APC, sinalizada pela elipse, não repete a mesma modalidade da unidade de COM, apenas estabelece uma ligação referencial.

2.2 As discussões sobre a unidade de APC

Alicerçados na perspectiva da Teoria da Língua em Ato (Cresti, 2000), quatro trabalhos serão apresentados a seguir. O primeiro, é a monografia de Tucci, E (2006) que analisou a unidade de APC em um *corpus* do italiano; o segundo, uma dissertação de mestrado desenvolvida por Ulisses (2008), sobre o APC no PB; o terceiro, um trabalho de Lombardi Vallauri (2008), que contesta o ponto de vista defendido por Cresti. O quarto e último é um artigo escrito por Raso; Ulisses (2008) em que se referem a dados do PB.

2.2.1 O estudo de Tucci , E (2006)

Tucci, E, em sua monografia, realizou uma pesquisa em um corpus[36] constituído de 5.639 enunciados. Esse estudo analisou textos formais e informais, em italiano, mas não se ateve à especificação das unidades de APC por tipologia textual (monológico vs dialógico). Os resultados mostraram que essa unidade não tem uma frequência muito elevada, tendo sido encontradas em apenas 4,49% de ocorrência dos enunciados. No caso dos textos informais, o percentual foi de 5,12%, caindo para 3,56% dos enunciados que apresentam APC nos textos formais.

Além de classificar as unidades de APC quanto aos critérios funcional, entonacional e distribucional e identificar os correlatos morfossintáticos, Tucci, E (2006) ainda criou uma tipologia para a classificação dessas unidades quanto ao caráter informacional, conforme se verificou na subseção 2.1.2. Observem-se os resultados mostrados através da tabela 2.1 e alguns exemplos utilizados pela pesquisadora, em sua monografia.

Tabela 2.1 - Frequência de variação informacional do total de APC nos dois modos: informal e formal

	Total	***Informal***	***Formal***
Repetições	50%	47%	56%
Preenchimento	5%	7%	3%
Retomada Textual	12%	10%	16%
Informação Tardia	33%	36%	26%

Fonte: TUCCI, E 2006, p.29

Através da tabela 2.1 percebe-se que a ocorrência de APC, em italiano, com função de repetição é maior que as demais categorias, tanto na tipologia informal, quanto formal. As demais categorias também aparecem, mas em menor escala. Contudo, a informação tardia também tem uma relevância significativa, alcançando um terço do total.

[36] Esse corpus faz parte do C-ORAL-ROM

Para uma melhor compreensão, ilustramos com alguns exemplos, em italiano, que aparecem na monografia da pesquisadora.

1) Repetição

*ANT: <se non è bravo /=TOP= allora magari ti verrà da ridere /=COM= **perché non è bravo**> //=APC=
(*ANT: <se não é bom /=TOP= então talvez você vá rir /=COM= **porque não é bom**> //=APC=

Nesse primeiro exemplo, o falante faz uso de uma oração subordinada, "*perché non è bravo*", para repetir algo que já havia sido dito, por isso o exemplo recebe a classificação de "repetição", nesse caso com variação.

2) Preenchimento

*MON: Montgomery era / cauto direi /=COM= **ecco** //=APC=
(*MON: Montgomery foi / cauteloso eu diria /=COM= **ó** //=APC=)

No segundo exemplo, o APC tem característica de preenchimento, já que realiza a expansão da unidade precedente, *Montgomery era / cauto direi,* sem repetir seu conteúdo semântico ou acrescentar informações. Funciona como um apoio para concluir de melhor maneira o enunciado.

3) Retomada Textual

*GAB: cioè /=TOP= che prove ci sono /=COM= **in questo senso** //=APC=
(*GAB: ou seja /=TOP= que provas tem /=COM= **nesse sentido** / /=APC=)

No terceiro exemplo, a unidade de APC é utilizada como recontextualização, retomando sinteticamente uma parte do discurso.

4) Informação Tardia

*ZIA: mi faceva anche rabbia /=COM= **tante volte** //=APC=
(*ZIA: Eu sentia raiva também /=COM= **muitas vezes** //=APC=)

No quarto exemplo, o apêndice funciona como uma informação tardia.

Além de contabilizar as tipologias informacionais das unidades de APC, Tucci, E (2006) também se ocupou dos correlatos morfossintáticos. As tabelas 2.2 e 2.3 ilustram o número de orações (coordenadas e subordinadas). Salientamos que essa diferença é bastante grande se levarmos em consideração a comparação entre os domínios formal e informal. Veja-se que as proporções dos diferentes correlatos morfossintáticos variam muito dependendo dos domínios formal e informal. As orações, no contexto formal, ocorreram 59% a mais do que no contexto informal. As expressões isoladas são presentes quase quatro vezes mais no domínio informal do que no formal. As construções sintáticas diminuíram cerca de 40%, passando do domínio formal para o informal. Esses resultados nos possibilitam afirmar que embora a unidade de APC ocorra tanto nos domínios formal quanto informal, a diferença na distribuição das tipologias informacionais é sempre significativa.

Tabela 2.2 - Variação morfossintática do total de APC nas tipologias informal e formal

	Total	***Informal***	***Formal***
Expressões Isoladas	19%	24,54	6,58
Construções Sintáticas	63%	54,06%	80,26%
Orações	18%	20,86%	13,16%

Fonte: Tucci, E 2006, p.41

As expressões isoladas que ocorrem com muita frequência são os advérbios, seguidos pelos pronomes, num percentual menor, e, às vezes, pelos adjetivos. É significativo o fato de que na categoria de nomes há apenas nomes próprios.

Tabela 2.3 - Distinção de expressões isoladas no total de APC

Expressões Isoladas	***% sobre o total de apêndices***	***% sobre as expressões isoladas***
Advérbio	13%	69%
Adjetivo	1%	5%
Nome Próprio	1%	5%
Pronome	4%	21%

Fonte: Tucci, E 2006, p.41

As construções sintáticas representam 63% do total dos APC presente no *corpus* de italiano e constituem-se de sintagmas nominais, sintagmas verbais, preposicionais, adverbiais e adjetivais. A ocorrência de sintagmas no *corpus* de italiano está melhor detalhada na tabela abaixo.

Tabela 2.4 - Frequência de sintagmas sobre o total de construções sintáticas

Construções Sintáticas	***% sobre o total de apêndices***	***% sobre as construções sintáticas***
SINTAGMA NOMINAL	21%	33%
SINTAGMA VERBAL	4%	6%
SINTAGMA PREPOSICIONAL	34%	54%
SINTAGMA ADVERBIAL	2%	5%
SINTAGMA ADJETIVAL	0,4%	0,01%

Fonte: Tucci, E 2006, p.44

As orações são constituídas de orações subordinadas, coordenadas ou orações principais. No *corpus* de referência, 15% dos APC são orações subordinadas de vários tipos, conforme se poderá verificar na tabela 2.5, abaixo. A frequência mais alta de ocorrência é de orações subordinadas relativas (43%), seguida das orações subordinadas objetivas (26%) e o percentual menor de ocorrência está entre as subordinadas finais, temporais, modais, limitativas e explicativas. As orações coordenadas com as conjunções "e" e "ou" representam 3% dos preenchimentos locutivos presentes nos APC.

Tabela 2.5 - Ocorrência de APC nas Orações Subordinadas

	Total	***Informal***	***Formal***
Relativas	43%	*42%*	44%
Objetivas	26%	31%	11%
Finais	9%	8%	11%
Temporais	6%	8%	0%
Modais	6%	8%	0%
Explicativa	3%	4%	0%
Limitativa	3%	0%	11%
Objetiva r	3%	0%	11%
Objetiva c (e)	3%	0%	11%

Fonte: Tucci, E 2006, p.52-53

Conforme Tucci, E (2006, p.62), nesses estudos não foram encontradas quaisquer expressões lexicais de valor modal e avaliativo, reiterando a ideia de que a unidade de APC não possui autonomia modal.

2.2.2 O estudo de Ulisses (2008)

Ulisses, juntamente a Alves de Deus (2009), fez parte de um projeto-piloto, cujo objetivo era a realização de todas as fases de constituição e análise do corpus (conforme descrito em Cresti, Moneglia (2005)) para posterior constituição de um corpus do PB, o C-ORAL-BRASIL. Nesse piloto foram utilizados três textos, sendo um dialógico (textos I) e dois tendencialmente monológicos (Texto II e III), totalizando 33,32 minutos de gravação, 5337 palavras e 874 enunciados. No texto I havia uma interação entre professor e aluno e nos textos II/III a interação se dá entre professora-coordenadora e professora colega de trabalho.

Com a definição de todos os critérios para a constituição de um *corpus* de língua espontânea e o término do C-ORAL-BRASIL, o texto dialógico passou a ser denominado bpubdl1 e os textos II e III foram unidos e classificados como bpudl10. Um dos participantes do texto bpudl11, MAR,[37], tinha 14 anos de idade na data de realização da coleta de dados e cursava a 7ª série do ensino fundamental, não possuindo outras atividades profissionais. O outro participante/pesquisador era LUC, possuía, na data de coleta de dados, 26 anos de idade e era graduado em Letras.

Já no texto bpubdl10 aparecem a informante FAB e a pesquisadora ADA. FAB possuía 30 anos de idade e graduação em Inglês. ADA, pesquisadora responsável por este estudo, possuía na data de coleta de dados 33 anos e era graduada em Letras.

[37] Os nomes utilizados nesse estudo são fictícios.

Nenhum dos textos analisados por Ulisses é prototípicamente dialógico ou monológico. Pode-se dizer que comparativamente o texto bpudl10 é mais formal e mais monológico do que o texto bpudl11. Ambos devem ser considerados públicos, segundo a classificação do C-ORAL-BRASIL.

O fato de haver uma relação profissional entre a informante dos textos bpudl10 e a pesquisadora, e de o assunto da interação ser sobre trabalho, o tornam tendencialmente mais monológico e formal, com uma fala muito mais controlada. Ao contrário, o texto bpudl11 é de um adolescente que discute seu processo de ensino/aprendizagem escolar. Esse texto é claramente dialógico e informal apesar de se tratar de uma interação entre professor e aluno. Esse aluno é um adolescente em processo de aquisição da norma culta e dos requisitos necessários para uma fala formal. Essas características dos textos, portanto, relativizam a classificação da amostra em formal e informal, além de dialógico e, tendencialmente, monológico (ou seja, com turnos muito longos).

Antes de ilustrarmos as análises realizadas por Ulisses (2008), faz-se necessário enfatizar que o estudo efetivado por ela fazia parte de um projeto piloto, momento esse de menor competência do grupo C-ORAL-BRASIL sobre a Teoria da Língua em Ato, e também havia limitação de dados para pesquisa com a língua espontânea. É nosso objetivo, portanto, não somente expandir as análises de Ulisses, mas também apresentar os avanços realizados pelo grupo do C-ORAL-BRASIL.

As análises mostraram que no texto bpubdl11, o APC ocorre em 14,45% dos enunciados e no texto bpubdl10 em 10,14%. Embora a diferença percentual de ocorrência de APC entre os textos bpudl10 e bpudl11 seja pequena, esses dados indicariam que o APC tende a aparecer mais nos textos dialógicos e informais, confirmando a predisposição mostrada por Tucci, E (2006) (conforme tabela 1), e deixam supor que a unidade de APC seja bem mais frequente em PB do que em italiano.

As tabelas 2.6 e 2.7, a seguir, detalham as funções desempenhadas pelo APC, nesse estudo. A classificação dessas funções não é tão simples, conforme já apontara Ulisses (2008, p.187).

Tabela 2.6 - Classificação Informacional dos Textos bpudl10 e bpudl11

	Total Texto bpudl10	***%***	***Total Texto bpudl11***	***%***
REPETIÇÃO	04	19,04%	11	44,0%
PREENCHIMENTO	05	24,01%	02	8,0%
RETOMADA TEXTUAL	05	23,72%	01	4,0%
INFORMAÇÃO TARDIA	07	33,23%	11	44,0%
TOTAL	21	100%	25	100%

Fonte: ULISSES 2008, p. 187

Tabela 2.7 - Comparação entre Tucci, E (2006) e Ulisses(2008) quanto à frequência de variação informacional do total de APC no modo informal

	TUCCI, E	***ULISSES***
Repetições	47%	32,6%
Preenchimento	7%	15,21%
Retomada Textual	10%	13,04%
Informação Tardia	36%	39,15%

Dada a falta de relevância estatística do estudo de Ulisses (2008), podemos dizer que as proporções não mostram uma grande variação.

Tabela 2.8 - Ocorrência de expressões sintagmáticas e isoladas nos Textos bpudl10 e bpud11

	Total Texto bpudl10	***%***	***Total Texto bpudl11***	***%***
SINTAGMAS	14	66,66%	21	84,0%
EXPRESSÕES ISOLADAS	07	33,34%	04	16,0%
TOTAL	21	100%	25	100%

Fonte: ULISSES 2008, p. 188-189

As proporções das categorias que constituem a unidade informacional de APC no texto bpudl10 e no bpudl11 são variadas, variação essa condicionada, provavelmente pela função realizada pelo APC no enunciado. Nesse estudo, distinguiram-se os sintagmas e as expressões isoladas (advérbios, pronomes,

adjetivos e nomes próprios). A tabela 2.8, acima, ilustra esse percentual de ocorrências.

O sintagma é o principal correlato morfossintático do APC nos textos bpudl10 e bpudl11, com média de 76,08% de ocorrências. Os outros correlatos correspondem às expressões isoladas, em média 23,92%. A categorização morfossintática proposta por Tucci, E (2006) não foi utilizada na íntegra por Ulisses (2008) com o objetivo de preservar as funções sintáticas. A frequência dos vários tipos de sintagmas identificados na amostra foi ilustrada na tabela 2.9.

Tabela 2.9 - Frequência de sintagmas encontrados nos Textos bpudl10 e bpudl11

	% Sobre o total de apêndices dos textos bpudl10	***% Sobre o total de apêndices do Texto bpudl11***
SINTAGMA NOMINAL	14,3%	12,0%
SINTAGMA VERBAL	4,8%	32,0%
SINTAGMA PREPOSICIONAL	52,7%	40,0%
SINTAGMA ADVERBIAL	23,4%	16,0%
SINTAGMA ADJETIVAL	4,8%	0,0%
TOTAL	100%	100%

Fonte: ULISSES 2008, p.190

No estudo de Ulisses (2008), as expressões isoladas que constituem o APC são realizadas em forma de advérbios e pronomes. Na categorização de classificação morfossintática proposta por Tucci, E (2006) são consideradas como expressões isoladas todas as ocorrências de advérbios, adjetivos, pronomes e nomes próprios. O percentual de ocorrência das expressões isoladas está ilustrado na tabela 3.10. Nos estudos de Ulisses, por exemplo, a ocorrência de adjetivos é praticamente nula, dada a pequena amostra analisada.

Tabela 2.10 - Percentual de ocorrência de expressões isoladas no bpudl10 e bpudl11

	% Sobre o total de expressões isoladas no texto bpudl10	*% Sobre o total de expressões isoladas no texto bpudl11*
ADVÉRBIO	83,33%	100%
ADJETIVO	16,67%	0,0%
NOME PRÓPRIO	0,0%	0,0%
PRONOME	0,0%	0,0%
TOTAL	100%	100%

Fonte: ULISSES 2008, p.191-192

Ulisses (2008, p.91), na conclusão de seu trabalho, aponta para o fato de que, na fala espontânea do PB, com a perda dos clíticos, se perde a diferença entre as estruturas deslocadas em unidades tonais diferentes e estruturas deslocadas na mesma unidade tonal, principalmente nos deslocamentos à direita, em que a ordem natural é mantida e nenhum elemento sintático sinaliza o deslocamento. Logo, em PB, contrariamente ao italiano, os deslocamentos à direita correspondem todos a uma unidade tonal suplementar colocada à direita do COM, seja ela uma APC ou um COB.

No italiano, segundo Raso (2009), o deslocamento à direita é necessariamente marcado por uma forma sintática e pode haver também uma marca prosódica. Vejam os exemplos:

1) L'ho mangiato / il pollo. (O como / o frango)
2) L'ho mangiato il pollo. (O como o frango)

No primeiro caso, há tanto a marca sintática, quanto prosódica; no segundo, apenas a marca sintática.

No PB, entretanto, a marca sintática não existe, por isso, para que haja um deslocamento à direita é necessária uma marca prosódica, como no exemplo abaixo.

3) Eu comi / o frango.

2.2.3 Tópico e Apêndice de comentário: duas unidades distintas?

Têm-se discutido se as unidades de TOP e APC são duas unidades informacionais distintas ou se ambas são unidades informacionalmente equivalentes. Mostraremos as posições de Lombardi Vallauri (2008) de um lado e, por outro, a de Cresti (2000) e Raso; Ulisses (2008)

2.2.3.1 A posição de Lombardi Vallauri (2008)

Lombardi Vallauri ao tratar das unidades de APC e TOP tem uma opinião diferenciada de Cresti (2000) e Raso; Ulisses (2008). Ele sustenta a tese de que essas unidades são informacionalmente equivalentes, embora não analise os aspectos prosódicos para endossar sua afirmativa. Utiliza-se dos mesmos exemplos de Cresti (1999), quando a teórica, assim como Raso; Ulisses, defendem uma opinião contrária a dele, para explicar seu ponto de vista.

Vejam-se os exemplos:

1) *Carlo /* va a Roma //
2) Va a Roma / *Carlo* //

Em ambos os exemplos, situações de produção distintas devem ser criadas para que possamos entender que no exemplo 1 “CARLO” é TOP e, no segundo exemplo, que “CARLO” passa a ser APC. Uma pergunta do tipo: “O que Carlo vai fazer?” pode ter tanto o exemplo um, quanto o dois, como resposta. Repare que o conteúdo proposicional é sempre o mesmo, porque semanticamente os dois enunciados dizem que Carlo realizou a ação de ir a Roma. No entanto, segundo Cresti e Raso; Ulisses, no primeiro exemplo “CARLO” é o campo de aplicação da força ilocucionária, portanto o TOP, e delimita semanticamente a ação do comentário. Já em dois, “CARLO” funciona como uma unidade de sufixo depois de uma unidade de núcleo (o COM); ou seja, APC e não pode estabelecer o âmbito de aplicação da força ilocucionária, o qual já deve estar estabelecido.

Lombardi Vallauri, contudo, tem uma opinião distinta dessa quando diz que todos os apêndices de comentário são temas a partir de um ponto de vista informacional. Ele acredita que os apêndices são, na verdade, simplesmente, um "tópico-posposto" com uma diminuição do contorno entonacional, dotados de conteúdo cognitivo.

Lombardi Vallauri continua a sustentar sua argumentação, partindo da premissa apontada por Cresti (2000) e também já vista em Ferrari (2003) e De Cesare (2004), ao manter a distinção entre tópico e apêndice, baseando-se no ponto de vista. Para Cresti, a articulação tópico/comentário é construíd sobre a composição de dois pontos de vista diferentes, enquanto no caso da articulação comentário/apêndice não existe essa diferença, pois o ponto de vista do apêndice decorre daquele do comentário. Cresti diz ainda que geralmente o apêndice acrescenta integrações locutivas, especificações e correções. Lombardi Vallauri discorda dela no que se refere ao apêndice, porque acredita que essa unidade pode manter um ponto de vista diferente daquele do COM; ou seja, tanto o tópico quanto o apêndice, segundo ele, gozam da liberdade de ponto de vista, manifestando uma modalidade diferente do COM.

Lombardi Vallauri enfatiza, ainda, que o nome apêndice sublinha um mero recurso formal; ou seja, o fato de estar à direita do comentário, e sugere que o apêndice seja tido como uma posposição do tópico, já que nem sempre essa unidade irá completar informações do comentário, conforme propõe Cresti.

Lombardi Vallauri finaliza seu artigo dizendo que essa não é uma preocupação puramente terminológica; porém, aponta para um possível equívoco em sua tese: o fato de desrespeitar fatores cruciais como a entonação.

Lombardi Vallauri acredita, portanto, que é simplesmente a posição que gera uma curva diferente.

2.2.3.2 A posição de Cresti (2000) e Raso; Ulisses (2008)

Raso; Ulisses (2008) coadunam o ponto de vista defendido por Cresti (2000), de que as unidades de TOP e APC são duas unidades diferentes. Eles fazem uso dos seguintes argumentos para sustentar a tese de que as unidades de TOP e APC são distintas e não a mesma unidade em posição especular:

- O TOP pode ser novo, mas o APC tem que ser dado. Raso; Ulisses explicam que no estudo realizado por eles não foram tratados a modalidade e o ponto de vista, mas que puderam perceber que muitos TOP servem para instaurar um novo âmbito de aplicação da força ilocucionária, enquanto isso não pode acontecer com o APC, que deve ter já instaurado o referente. Ao contrário, todos os casos em que uma unidade posposta ao COM parece instaurar um novo âmbito de aplicação da força ilocucionária resultam comentários ligados, com um foco funcional claro e, portanto, com força ilocucionária. Tirando esse foco o enunciado perde sua força informacional.
- As duas unidades se distinguem quanto aos perfis entonacionais, sendo que o TOP possui foco funcional, enquanto o APC tem perfil nivelado ou descendente, sem foco.
- O TOP tem modalidade e ponto de vista próprios, enquanto o APC sempre mantém a modalidade e o ponto de vista do COM.
- Os correlatos dos APC se mostram muito diferentes daqueles dos TOP. A maioria de APC é também nominal, mas isso é devido a uma altíssima porcentagem de SP (quase 46%, na amostra de Ulisses (2008)), enquanto nos TOP prevalecem os SN. Entre os SN a proporção entre sujeitos e objetos é invertida: nos APC há uma presença quádrupla dos objetos em comparação aos sujeitos. O mesmo acontece comparando a proporção de SV principais e subordinados: esses últimos chegam a quase 20% de todos os APC, enquanto os principais aparecem com um único caso, na amostra estudada por Raso e Ulisses. Completa o quadro uma porcentagem de quase 20% de advérbios. Logo concluem: a) as unidades de TOP e APC se realizam independente de qualquer configuração

sintática; b) há uma diferença muito grande entre essas duas unidades: o TOP tende a hospedar em prevalência funções sintáticas de SN sujeito e de SV principal; já o APC só aceita excepcionalmente essas funções e prefere claramente as opostas de SN objeto e de SV subordinado.

- O APC não contribui à complexidade informacional do texto com relação à sua tipologia interacional, enquanto o TOP parece ser um dos principais responsáveis da complexidade informacional nos textos monológicos; ou seja, se nos textos monológicos aumenta muito a quantidade de enunciados complexos em relação aos dialógicos, isso não se deve a um aumento das unidades de APC, mas a um aumento das unidades de TOP e, eventualmente, de APT e outras unidades textuais.

2.2.3.3 Raso; Ulisses versus Lombardi Vallauri

Em conclusão do artigo, Raso; Ulisses ponderam sobre os seguintes argumentos apresentados por Lombardi Vallauri, acerca dessa temática:

- Ambas as unidades poderiam ser definidas como âmbito de aplicação da força ilocucionária;
- Os diferentes correlatos entonacionais e morfossintáticos seriam devidos a posição diferente.

O primeiro argumento, segundo Raso; Ulisses parece funcionar com boa parte dos APC, mas não com todos, enquanto a definição do APC como integração textual do COM funciona com todos os APC. Alguns exemplos citados por eles:

*FBA: olha /=INP= a mudança /=TOP= foi [/] foi positiva /=COM= **sim** /=APC né //=FHA=

*GBL: ou então /=AUX= eu tentava &t[/] sozinho /=COM **lá decifrar** //=APC=

*REG: omitir /=COM= **só** //=APC[38]

[38] Esse exemplo foi retirado de OLIVEIRA (2009b).

O segundo argumento atribui à posição um valor superior ao perfil entonacional e aos correlatos morfossintáticos. Isso equivale a dizer que a posição gera uma realização entonacional e morfossintática diferente. Mas para Raso; Ulisses mesmo se isso for verdade, não se explicaria tudo: por que o TOP permitiria a instauração de um novo âmbito da força ilocucionária e o APC não? Seria de novo efeito da posição? Nesse caso, seria necessário distinguir entre um TOP instaurativo, com potencialidades cognitivas de novidade, perfil entonacional com foco, correlatos morfossintáticos e modalidade próprios, e responsável pela complexificação do enunciado; e por outro lado um TOP não instaurativo, com correlatos entonacionais e morfossintáticos diferentes, sem modalidade e não responsável pela complexificação do enunciado com relação ao tipo de interação.

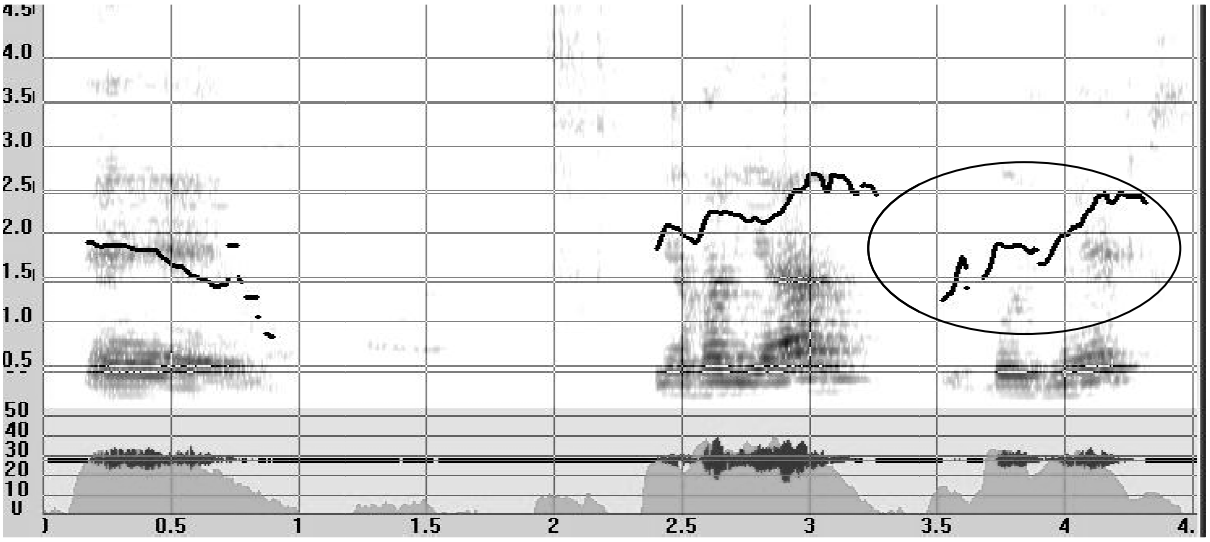
Os autores dizem que com todas essas diferenças, manter o mesmo nome para as duas unidades (ou para as duas realizações de uma suposta unidade única) geraria de fato a necessidade de distinguir entre duas subunidades, o que parece, na verdade, repropor a distinção que se quer negar.

Raso; Ulisses defendem a tese de que a unidade de APC não pode instaurar um âmbito de força ilocuionária novo, já que é uma unidade sempre dada, situacionalmente. Assim sendo, acreditam que a prosódia não tem a ver puramente com o perfil melódico de uma unidade, mas a ela é dada a possibilidade de marcar linguísticamente a sua função na interação verbal. Tanto que para instaurar um âmbito de aplicação da força ilocucionária depois do COM é necessário recorrer a um COB, como no exemplo seguinte:

Exemplo 2.10 (bfammn12) (CD áudio 2.10)

*GER: me avaliaram /=COM= **os dois médicos** //=COB=

Figura 2.6 – Unidade de COB



Nesse exemplo, como se pode observar por meio do espectrograma, o falante, depois de ter realizado a ilocução *me avaliaram*, percebe que o âmbito de aplicação dessa ilocução não havia sido expresso através de um tópico e também não fora dado situacionalmente ou discursivamente. Foi, portanto, a necessidade de expressar o âmbito de aplicação da ilocução que o motivou à nova força ilocucionária, *os dois médicos,* em que a curva de F0 é demarcada pela elipse.

Da forma como foi enunciado esse exemplo, jamais teríamos um enunciado constituído por um COM e um APC, porque o APC não tem capacidade de expressar o âmbito da força ilocucionária, já que apenas serve para reforçar algo já proferido, algo, todavia, que poderia ser descartado sem que houvesse comprometimento semântico. A diferença prosódica com o APC é evidente; aqui temos um foco funcional ausente no APC e este foco é que carrega a força ilocucionária.

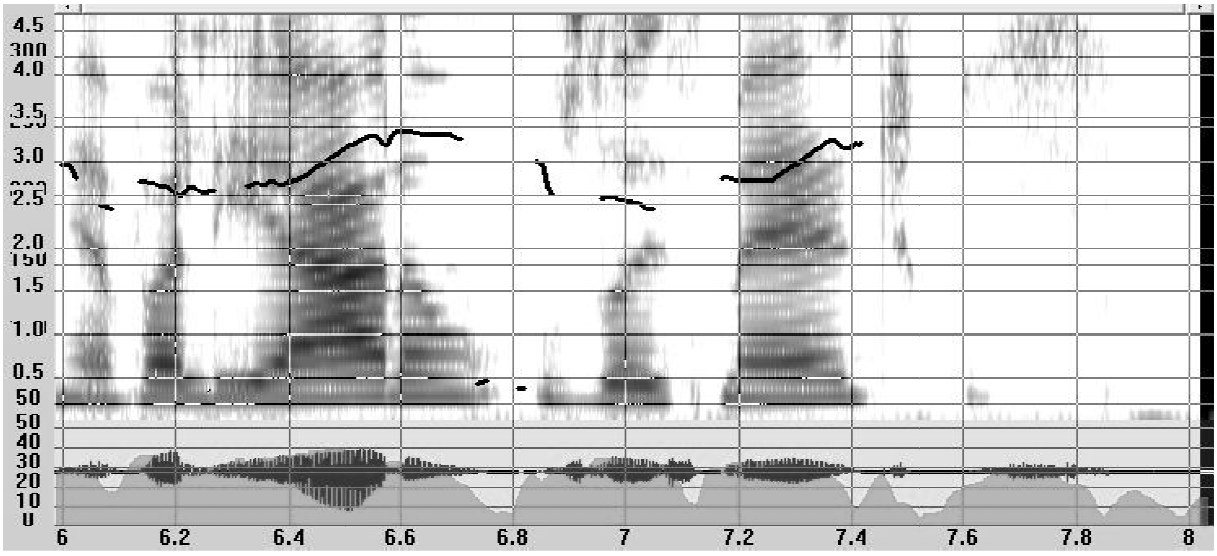
No entanto, se tivéssemos um interlocutor que soubesse de quem se tratavam os dois médicos, isto é, se a informação fosse considerada como dada para ambos os

interlocutores naquela situação, aí sim poderíamos ter o mesmo conteúdo locutivo, com outra ilocução e a primeira unidade (COM) possuiria uma quebra prosódica interpretada como sendo de valor conclusivo, e a segunda, um APC, com função exclusivamente de integração textual.

Veja o exemplo:

Exemplo 2.11 (CD áudio 2.11)
*GER: me avaliaram /=COM= **os dois médicos** //=APC=

Figura 2.7 – Unidade de APC



Embora não possuíssemos esse enunciado em situação real, preferimos reproduzi-lo em laboratório (mesmo sabendo que é artificial) só para que pudéssemos esboçar a possibilidade acima aventada.

2.3 Resumindo

Nesse capítulo apresentamos as pesquisas realizadas em italiano por Tucci, E (2006) sobre a unidade de APC e os estudos de Ulisses (2008), no PB. Além disso, as posições antagônicas de Lombardi Vallauri (2008) e Raso; Ulisses (2008) quanto às unidades de TOP e APC serem duas unidades distintas ou não, foram sustentadas.

Capítulo 3

METODOLOGIA DA CONSTRUÇÃO DO CORPUS E ANÁLISE DOS DADOS

Esse capítulo constitui-se de três seções. A seção 3.1 objetiva justificar a pesquisa dentro do escopo da Linguística de Corpus e apresentar o C-ORAL-BRASIL, expondo todas as etapas de constituição do mesmo. A seção 3.2 visa à apresentação dos critérios de escolha da amostra, bem como os procedimentos adotados para a prospecção dos dados, aplicativos e tecnologias usadas na análise dos dados. E a seção 3.3 procurará estabelecer uma comparação entre os dados do C-ORAL-BRASIL e do C-ORAL-ROM no que se refere à língua italiana.

3.1 O estudo a partir da Linguística de Corpus

Esse estudo objetiva entender como a informação é linguisticamente organizada na comunicação humana, a partir da fala espontânea. Para isso, faz uso de ferramentas para observação e interpretação de dados a partir de uma abordagem da Linguística de Corpus.

Por fazer uso de uma abordagem empirista, a Linguística de Corpus trabalha com a linguagem como um sistema probabilístico. De acordo com essa noção, os traços linguísticos não ocorrem de forma aleatória, sendo possível evidenciar e quantificar regularidades (padrões) linguísticos a partir de contextos situacionais de uso da linguagem: (...) s*istema probabilístico significando que a língua é vista mais como uma questão de probabilidade de possibilidades; ou seja, embora os traços linguísticos sejam possíveis teoricamente, eles não ocorrem com a mesma frequência (SARDINHA, 2005).*

Adotando os pressupostos da Linguística de Corpus, cujo foco é uso linguístico dos falantes, o presente trabalho fez uso de um corpus de fala cujas características são compatíveis com o estudo da estrutura informacional, segundo a visão de Cresti

(2000). Com isso procurou-se garantir que os resultados obtidos na pesquisa possam trazer novas perspectivas para a funcionalidade da unidade de APC.

Segundo Mello (2011), a compilação de um *corpus* oral deve partir da pergunta do para que e para quem ele servirá, depois é necessário que se definam os seus parâmetros de compilação. Os parâmetros de compilação referem-se ao tamanho; número e perfil dos informantes; técnicas para coleta de dados; conteúdo e desenho do *corpus*: gênero (conversação, diálogo, monólogo); variação linguística (geográfica, contextual, individual); tipologia e qualidade dos arquivos de som; unidades de referência da fala, parâmetros de transcrição, acessibilidade; alinhamento texto-som; níveis de anotação; validação; custo; disponibilização.

Além disso, há pontos relacionados aos parâmetros norteadores da compilação de *corpora* orais que devem ser considerados cuidadosamente como: (a) questões legais e éticas relacionadas à identidade dos informantes; (b) Equipamentos e *softwares* a serem utilizados para a gravação e tratamento dos arquivos de som; (c) Treinamento da equipe de trabalho para a execução de todas as etapas do projeto; (d) Técnicas de validação e tratamento estatístico dos dados e (e) Interface de acesso ao *corpus*. Todos esses pontos serão detalhados quando da apresentação das etapas de constituição do C-ORAL-BRASIL.

3.1.1 A composição do C-ORAL-BRASIL

O C-ORAL-BRASIL[39] (Raso; Mello, 2012) é constituído de um corpus de fala espontânea do Português do Brasil coordenado pelos professores Tommaso Raso e Heliana Mello, da UFMG. Esse projeto teve como financiadores a Fundação de Apoio à Pesquisa do Estado de Minas Gerais (FAPEMIG), o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ) e a Universidade Federal de Minas Gerais (UFMG).

[39] (www.c-oral-brasil.org)

O C-ORAL-BRASIL é um *corpus* de fala espontânea informal, representativo das três principais tipologias da fala – diálogos, monólogos e conversações – em contextos públicos e privados, sua amostragem é de cerca de 1.500 palavras por texto. As especificações do *corpus* seguem as orientações internacionais do Text Encoding Initiative (TEI) (http://www.tei-c.org/index.xml) e usufruem das especificações técnicas previamente testadas e recomendadas pelo C-ORAL-ROM para compilação de um *corpus* representativo da fala espontânea (Cresti; Moleglia, 2005). Cada um dos textos gravados é apresentado no *corpus* através de seu arquivo de som, em formato wav, acompanhado de sua transcrição em formato rtf devidamente segmentada em enunciados e unidades tonais (quebras terminais e não terminais) e de seu arquivo de alinhamento texto-som em formato XML, obtido via *software* WinPitch (Martin, 2004). Cada um dos textos apresenta também o seu arquivo de metadados, seguindo os padrões TEI, e suas versões txt e XML etiquetadas morfossintaticamente via anotador sintático PALAVRAS (Bick, 2000), adaptada pelo autor para o C-ORAL-BRASIL (BICK, 2012).

O corpus foi concebido desde o seu início para o estudo da estrutura informacional do Português Brasileiro (PB) e suas ilocuções com base na Teoria da Língua em Ato (Cresti, 2000). Esse corpus compõe o quinto ramo do C-ORAL-ROM (Cresti; Moneglia, 2005), corpus de referência das quatro maiores línguas românicas da Europa.

O C-ORAL-BRASIL[40] propõe somente a metade informal de um *corpus* de fala espontânea. A construção da parte formal está em andamento e não será objeto desta descrição. A parte informal do *corpus* é constituída por 208.130 palavras, distribuídas em 139 textos de, em média, 1.500 palavras cada um. Alguns poucos textos são maiores (até 5.000 palavras) ou menores (desde que apresentem uma autonomia textual); mais adiante analisaremos os motivos dessas escolhas. Os 139 textos foram divididos em dois contextos: familiar/privado (159.364 palavras) e público (48.766 palavras); em cada contexto os textos foram separados em quantidades aproximadamente iguais em três tipologias interacionais: monólogos, diálogos e conversações (ou seja, textos dialógicos com mais de dois participantes).

[40] Todos os dados sobre esse corpus foram retirados de RASO; MELLO (2012).

Um *corpus* de fala realmente espontânea deve poder fornecer um retrato o mais fiel possível da variação situacional. De fato, o que mais condiciona a estruturação da fala, principalmente em uma perspectiva de estudos pragmáticos (mas não somente), como aquela que está na origem dos projetos C-ORAL-ROM e C-ORAL-BRASIL, não é tanto a variação de tópicos, fundamental para *corpora* interessados principalmente no estudo do léxico, nem a variação diastrática, interessante para buscar a variação sociolinguística, nem a variação diatópica, que mostra como falantes de diferentes lugares realizam as mesmas estruturas. Em uma perspectiva pragmática, é fundamental diferenciar o comportamento verbal em situações diversificadas o mais possível. Se por um lado a tradição sociolinguística permite identificar os principais domínios da fala formal, as situações possíveis na fala informal são dificilmente categorizáveis. Por isso, enquanto a fala formal pode ser identificada listando-se uma série de contextos típicos de uso, o âmbito de situações em que é usada a fala informal deve ser deixado aberto, e seu domínio não pode ser representado através de uma lista de contextos de uso, já que nenhum contexto é mais típico do que outros. O objetivo é, portanto, diversificar ao máximo os contextos situacionais para se aproximar mais da representação de um universo aberto. Para que isso seja possível, dado o custo de realização de um *corpus* falado com relação a *corpora* escritos (não só econômico, mas principalmente de tempo para as gravações e as transcrições), é importante fornecer muitos textos diferentes e reduzir o tamanho dos textos. A medida média de 1.500 palavras é suficiente para que uma interação seja textualmente autônoma, e ao mesmo tempo permite que seja representada uma ampla variedade situacional. Em outras palavras, quanto mais ampla é a variedade dos textos, mais apurada é a representação do universo possível; mas os textos devem também possuir uma dimensão suficiente para garantir que neles estejam presentes as características de um determinado contexto que se pretenda representar com eles. Um texto deve, ao mesmo tempo, poder apresentar determinadas propriedades sintáticas (micro-sintaxe) e determinadas estruturas dialógicas (macro-sintaxe e estrutura informacional) (Blanche Benveniste *et al.*, 1990; Scarano, 2003).

Dentro do registro informal, a divisão entre contexto familiar/particular e contexto público documenta o papel com o qual um falante interage com outros: se ele

interage como indivíduo, como acontece nas relações familiares ou com os amigos, ou se ele interage com base em um papel profissional ou institucional, como acontece, por exemplo, exercendo o papel de vendedor ou cliente, de cidadão e funcionário público, de aluno e professor, de garçom e cliente, de colega de trabalho etc. Considerando-se que a maior parte do tempo e a maior variedade de situações comunicativas é exercida em interações de caráter familiar/particular, cerca de 70% do *corpus* foram dedicados a esse contexto, deixando uma porcentagem menor aos textos do contexto público.

Dentro de cada contexto, podemos individualizar três tipologias de interação: i. uma tipologia monológica, em que um falante constrói um texto falado, fruto exclusivamente da própria elaboração, sem, ou quase sem, nenhuma interação; ii. uma tipologia dialógica, em que dois interlocutores interagem entre si; iii. uma tipologia conversacional, em que três ou mais interlocutores interagem entre si. As características dos textos são fortemente condicionadas pela tipologia interacional, principalmente na oposição monológico *versus* interativo. Deve-se, contudo, ressaltar que, contrariamente ao que acontece na fala formal, a fala informal não apresenta, em princípio, uma tipologia monológica perfeita, em que um único falante desenvolve um texto em um único turno de fala, como acontece, por exemplo, em uma conferência. Sempre, ou quase sempre, haverá alguma interação. A medida usada para distribuir um texto dentro da tipologia monológica ou dentro das tipologias interativas é a prevalência, no primeiro caso, da construção de um texto falado que continua se desenvolvendo mesmo depois das eventuais intervenções do interlocutor, que, na maioria dos casos, são desconsideradas pelo monologante. A tipologia monológica será tendencialmente constituída por turnos muito longos e, dentro deles, por enunciado muito articulados estruturalmente, com muitas unidades tonais e com forte andamento processual; a referência ao contexto situacional será escassa, enquanto a contextualização necessária será muito maior. Dependendo da tipologia textual do texto monológico, mudam as ilocuções prevalentes, mas a variação ilocucionária tende a ser escassa dentro da mesma tipologia. Ao contrário, as tipologias interativas serão constituídas por turnos curtos e enunciados padronizados informacionalmente e com poucas unidades tonais; nesse caso a referência ao contexto situacional será muito maior, tornando desnecessária uma

forte contextualização verbal, mas a variação ilocucionária dentro do mesmo texto tenderá a ser ampla[41].

Depois dessa grande variação estrutural entre textos monológicos e textos interativos, o maior fator de variação muda se considerarmos a tipologia monológica ou aquelas interativas. Nos textos monológicos, a estruturação da fala depende principalmente da tipologia textual: conto de vida, explicação profissional, texto argumentativo, piada, receita, fábula etc. Nos diálogos e nas conversações, a variação é basicamente ligada à atividade que os interlocutores estão desempenhando: uma conversa entre amigos em casa será estruturada de maneira muito diferente de uma briga de casal, ou de um diálogo entre um vendedor de sapatos e um cliente, ou de uma interação entre os componentes de um time jogando futebol, ou entre um garçom e os participantes de uma festa aos quais o garçom oferece as bebidas, ou de uma interação entre um professor de ginástica e o cliente de uma academia, que, sob esforço, segue as suas instruções, ou de uma mãe que interage com um bebê que chora, ou de um diálogo entre duas pessoas que estão colaborando para realizar uma tarefa etc. É evidente que em cada atividade as ações que devem ser desempenhadas através da fala mudam radicalmente. Cada situação estimula a emergência de atos de fala diferentes, tamanho diferente dos turnos, tamanho e estruturações diferentes dos enunciados, presença maior ou menor de silêncios etc.

Essas observações deveriam ser suficientes para mostrar a importância capital de uma verdadeira variação diafásica na composição de um *corpus* de fala espontânea. A variação estrutural jamais seria garantida através da variação do tópico das interações, nem através da variação dos falantes. Diferentes falantes cumprem a mesma atividade sem uma apreciável variação estrutural, e a mudança de tópico em bate-papos ou em entrevistas não gera variação ilocucionária ou estrutural.

Apesar de a variação diafásica ter sido o principal objetivo na construção da arquitetura do *corpus*, à variação diastrática foi dedicada também grande atenção. A representação dos sexos é praticamente paritária, e todas as camadas

[41] Para um aprofundamento dessas características, vejam-se os capítulos 3 (Raso) e 5 (Raso; Mittmann), 2012.

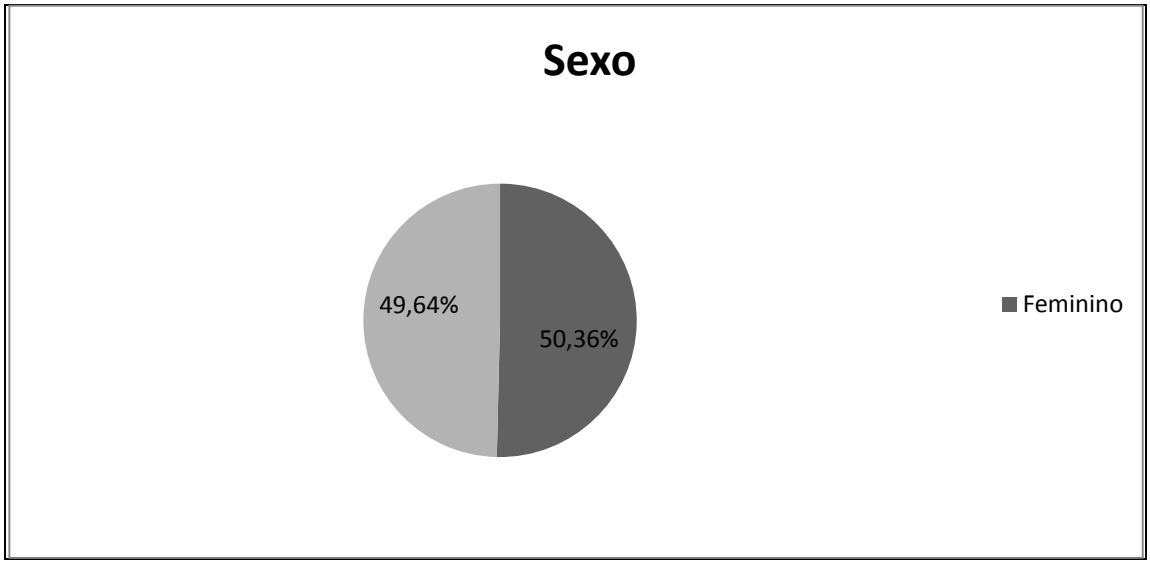
socioculturais e as faixas etárias são amplamente documentadas, com balanceamento muito satisfatório. Essas e outras informações podem ser derivadas nos detalhes dos metadados para cada texto.

O *corpus* C-ORAL-BRASIL informal apresenta 362 falantes. Para 68,23% deles são documentados sexo, idade, origem e nível escolar. O fato de que para pouco mais de 30% não temos a informação relativa à idade, à origem e/ou ao nível escolar se explica porque esses falantes entraram no contexto de gravação sem que a sua presença fosse prevista.

Com relação ao número de falantes, a situação é a seguinte, desconsiderando naturalmente os falantes dos quais não conhecemos a idade, e que representam apenas 1,91% das palavras do *corpus*: 75 pertencem à faixa A; 1 falante aparece uma vez como representante da faixa A e outra como representante da faixa B; 88 falantes pertencem à faixa B; 64 pertencem à faixa C; 15 pertencem à faixa D; 11 pertencem à faixa M.

Os gráficos a seguir documentam até que ponto o *corpus* pode ser considerado balanceado por sexo, idade e nível escolar. As porcentagens dentro dos gráficos se referem ao número de palavras pronunciadas por cada categoria.

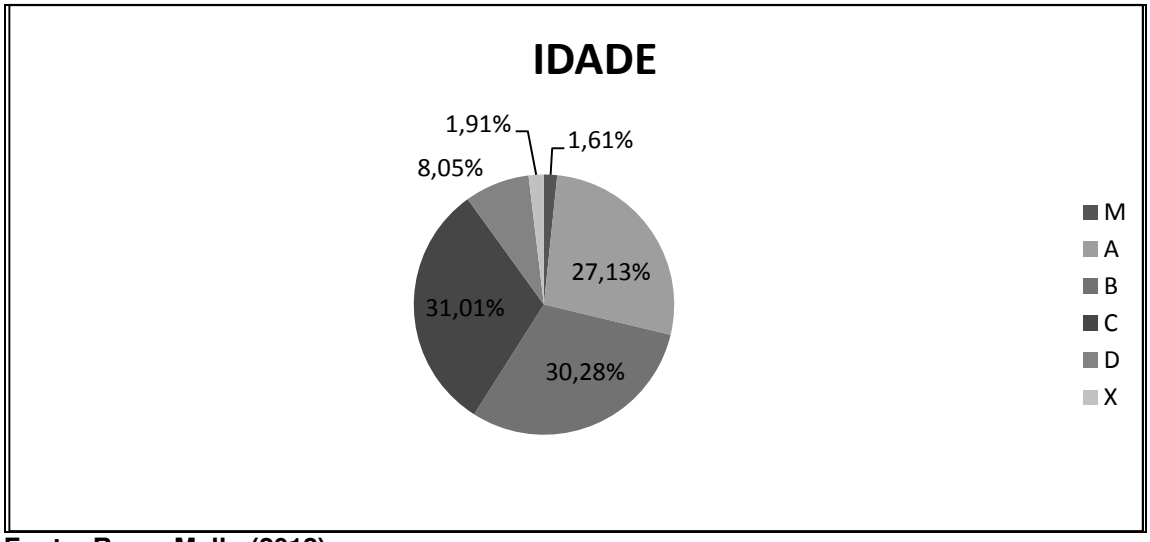
Gráfico 3.1 – Balanceamento dos informantes por sexo



Fonte: Raso; Mello (2012)

O balanceamento por sexo é praticamente perfeito em número de palavras. Quanto ao número de informantes, 203 são do sexo feminino e 159 do sexo masculino, ficando apenas um informante, que pronuncia apenas uma palavra, com sexo não identificado.

Gráfico 3.2 – Balanceamento dos informantes por idade



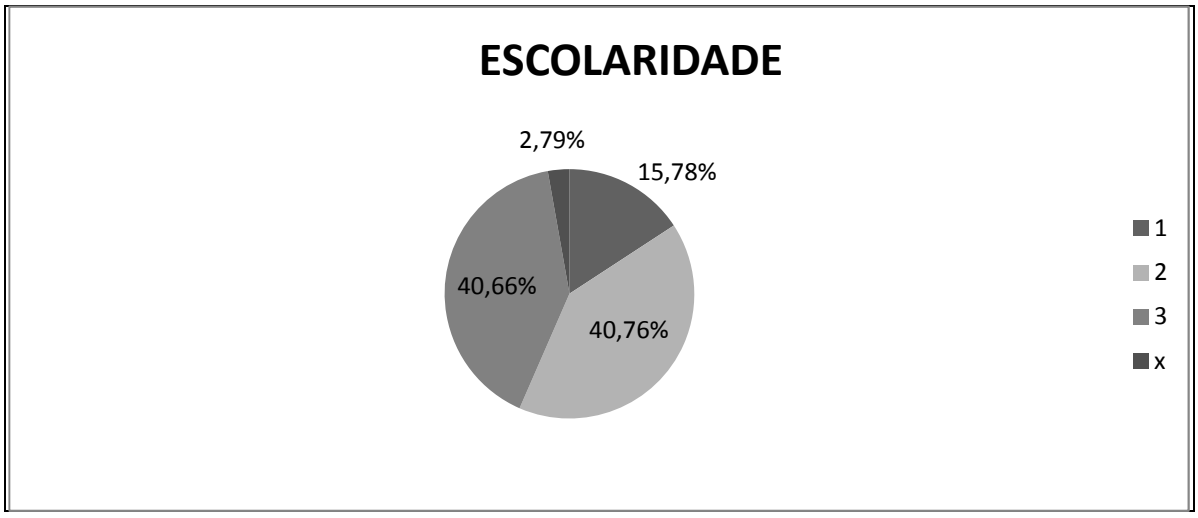
Fonte: Raso; Mello (2012)

Quanto à idade, o grupo A, constituído por falantes entre os 18 e os 25 anos de idade, representa 27,13% do número de palavras do *corpus*; o grupo B, formado pelos falantes entre 26 e 40 anos, é formado por 30,28% das palavras; o grupo C, formado pelos falantes entre 40 e 60 anos, é representado em 31,01% das palavras; o grupo D, formado pelos falantes superiores a 60 anos, é representado em 8,05% das palavras. Ainda temos 1,61% das palavras pronunciadas por menores de idade e 1,91% das palavras pronunciadas por falantes de idade desconhecida. Podemos concluir que o *corpus*, em número de palavras, é muito bem balanceado quanto aos falantes de idade entre os 18 e os 60 anos, e provavelmente, mais em geral, quanto a falantes acima dos 18 anos de idade, considerando que a incidência de falantes acima de 60 anos é menor na sociedade brasileira.

Com relação ao número de falantes, a situação é a seguinte, desconsiderando naturalmente os falantes dos quais não conhecemos a idade, e que representam apenas 1,91% das palavras do *corpus*: 75 pertencem à faixa A; 1 falante aparece uma vez como representante da faixa A e outra como representante da faixa B; 88 falantes pertencem à faixa B; 64 pertencem à faixa C; 15 pertencem à faixa D; 11 pertencem à faixa M.

O gráfico 3.3 mostra a situação relativa ao nível escolar, sempre com base no número de palavras do *corpus*.

Gráfico 3.3 – Balanceamento dos informantes por nível de escolaridade



Fonte: Raso; Mello (2012)

A faixa 1, equivalente à ausência de escolarização ou escolarização até o nível primário incompleto (não mais de 7 anos de escolarização), corresponde a 15,78% das palavras do *corpus*; a faixa 2, correspondente a uma escolarização até o terceiro grau completo, desde que se exerça uma profissão que não requer o terceiro grau, é relativa aos falantes responsáveis por 40,76% do *corpus*; a faixa 3, relativa aos falantes que possuem título de terceiro grau e exercem uma atividade que exige esse título, cobre os falantes que pronunciam 40,66% das palavras. De alguns falantes não conhecemos o nível escolar (faixa X no gráfico). Estes últimos são responsáveis por 2,79% do *corpus*.

Com base nesses dados, podemos dizer que o *corpus* representa melhor o nível escolar médio-alto, mas não deixa de representar também o nível de escolaridade baixa ou ausente. Portanto, o *corpus* deve ser considerado representativo da variedade padrão do português na diatopia mineira, com uma significativa representação também da variedade mais popular. Quanto aos falantes, 46 pertencem à faixa 1; 101 pertencem à faixa 2; 104 pertencem à faixa 3; e 1 falante é registrado como pertencente à faixa 2 na gravação mais antiga e à faixa 3 na mais recente[42].

Olhando então para a variação diastrática do C-ORAL-BRASIL, podemos concluir que ela, apesar de não constituir o objetivo principal na arquitetura do *corpus*, é muito bem balanceada: seu balanceamento é próximo ao ideal quanto à divisão por sexos e por idades; no nível escolar o balanceamento não é ideal, porque o *corpus* é mais representativo dos níveis escolares médios e altos do que do nível escolar baixo. Esse último nível, contudo, é representativo de quase 15% do *corpus*.

Um último aspecto relativo à variação diastrática é constituído pela ocupação profissional. Nesse caso trata-se de uma categoria aberta; portanto, não é possível fazer generalizações e fornecer estatísticas como para as categorias anteriores. Contudo, vale a pena olhar para a variedade de ocupações dos informantes do *corpus* e para a prevalência de algumas delas. Olhando para os metadados, aparece imediatamente clara uma forte incidência no *corpus* das ocupações ligadas ao mundo da educação: em primeiro lugar porque muitos dos pesquisadores que realizaram as gravações são alunos de graduação ou pós-graduação e frequentemente participam das interações; em segundo lugar porque o círculo de conhecidos entre os quais foram procurados os informantes contempla pessoas ligadas ao ambiente dos pesquisadores, com muitos professores, tanto de disciplinas quanto de nível de ensino muito variados, diretores de escola e pessoas que trabalham, com funções muito diferentes, em instituições de ensino; em terceiro lugar porque a ocupação mais frequente nos informantes de faixa etária até os 25 anos (mas também muito frequente em falantes mais velhos) é o estudo. Mas além desse domínio, estão presentes informantes de outros domínios: muitos

[42] Isso foi devido ao momento cronológico da gravação, quando o falante pode ter mudado o seu nível escolar ou atividade profissional.

comerciantes ou vendedores, artistas, funcionários públicos, profissionais liberais de domínios muito diferentes (promotores, médicos, psicólogos, dentistas, farmacêuticos, fisioterapeutas, engenheiros etc.), donas de casa, artesãos, técnicos, gerentes, agricultores, operários, pedreiros, empregadas domésticas, auxiliares de várias ocupações, corretores, e uma grande variedade de ocupações representadas por um único informante.

Como já dito, a variação diatópica do C-ORAL-BRASIL é essencialmente mineira. Um *corpus* dessas dimensões deve se concentrar em representar as outras variações dentro de uma única diatopia. Assim é também para os *corpora* do C-ORAL-ROM, que representam para o espanhol a região de Madri, para o francês a região de Marselha, para o italiano a região de Florença e para o português a região de Lisboa (Cresti *et al.*, 2002). Em todos os *corpora* há informantes externos a essas regiões e até ao país, já que a composição de uma grande área urbana pressupõe uma porcentagem de pessoas de fora, mas é mandatório que mais de 50% do *corpus* informal sejam representativos diretamente da diatopia escolhida. Para o C-ORAL-BRASIL foi escolhida a diatopia relativa à região de Belo Horizonte, com extensão até Minas Gerais inteira.

3.1.1.1 O cabeçalho

Os dados nos cabeçalhos seguem os mesmos critérios adotados pelo C-ORAL-ROM, à exceção da classificação do grau de instrução, já mencionado, e da qualidade acústica, como será explicado na próxima subseção. Veja um exemplo:

@Title: Daughter
@File: ifammn06
@Participants: CAR, Carmosina (woman, C, 1, house-keeper, narrator, Alpercata (MG)) MAR, Maryualê (woman, B, 3, professor, intervenient, Florianópolis)
@Date: 12/04/2008
@Place: Belo Horizonte
@Situation: narration about how CAR adopted her youngest daughter, CAR's kitchen, CAR makes lunch, not hidden, researcher participant (CAR works as housekeeper at the researcher's house)
@Topic: daughter's adoption

@Source: C-ORAL-BRASIL
@Class: informal, familiar/ private, monologue
@Length: 9’51”
@Words: 1508
@Acoustic_quality: AB
@Transcriber: Maryualê M. Mittmann
@Revisor: Heloisa P. Vale
@Comments: text collected and recorded by Maryualê M. Mittmann. CAR pronounces "dócia" and "vivendos" when it should be "dócil" and "vivendo". Sometimes CAR calls the researcher Mara and not Mary.

Os dados do cabeçalho serão importantes para levantamentos estatísticos referentes a sexo, idade, escolaridade, ocupação, papel desempenhado pelo falante, bem como o de fornecer informações sobre as atividades desenvolvidas ao longo de toda a gravação, o assunto e o local em que a mesma ocorreu.

3.1.1.2 A qualidade acústica e o equipamento

3.1.1.2.1 A qualidade acústica

Conseguir gravações de boa qualidade acústica em contextos naturais muito variados é uma tarefa difícil, mas importantíssima em *corpora* de fala espontânea. Não é suficiente que as gravações permitam a transcrição dos aspectos segmentais; um *corpus* que possa ser estudado não somente quanto ao léxico e à morfossintaxe, mas também pragmaticamente e foneticamente, deve possuir uma qualidade acústica muito mais elevada, o que é fácil em laboratório, mas muito mais difícil em situações naturais. É essencial em primeiro lugar poder contar com um equipamento sem fio de alta qualidade, mas é fundamental também ter uma boa estratégia de gravação, planejar as situações a serem gravadas, gravar muito mais tempo e muito mais textos do que se pretende utilizar, de maneira a se poder escolher o material de melhor qualidade. A tabela 3.1 mostra as propriedades das diferentes qualidades acústicas consideradas no *corpus*; a tabela 3.2 mostra as qualidades das diferentes gravações agrupadas com base nas diferentes ramificações do *corpus*. Todas as gravações são em formato wav.

Tabela 3.1 - Propriedades das diferentes qualidades acústicas

Propriedades	Etiqueta
Qualidade muito alta. Ótima resposta dos microfones. Quase toda a gravação possui qualidade apropriada para quase todo tipo de análise fonética. Quase nenhuma sobreposição. Quase nenhum ruído de fundo. Computação da F_0 possível em (quase) todo o arquivo.	A
Qualidade alta. Ótima resposta dos microfones. Grande parte da gravação possui qualidade apropriada para quase todo tipo de análise fonética. Poucas sobreposições. Pouco ruído de fundo. Computação da F_0 possível em (quase) todo o arquivo.	AB
Qualidade média. Boa ou média resposta dos microfones. Muitas partes da gravação possuem qualidade apropriada para análise fonética. Computação da F_0 possível em grande parte do arquivo. Não há excesso de sobreposições e de ruído de fundo.	B
Qualidade baixa. Média resposta dos microfones. Computação da F_0 possível em pelo menos 60% do arquivo. Mesmo quando não é possível a computação de F_0, a gravação continua clara para a escuta.	BC
Qualidade baixa. Média ou baixa resposta dos microfones. Computação da F_0 possível em pelo menos 60% do arquivo. Algumas partes da gravação podem não ser claras para a escuta.	C

Fonte: Raso; Mello (2012)

Tabela 3.2 – Qualidades das gravações

	A	AB	B	BC	C	total
bfamcv	8	11	4	6	5	34
bpubcv	1	2	---	1	5	9
bfamdl	7	14	6	5	3	35
bpubdl	5	2	2	1	1	11
bfammn	13	10	10	1	2	36
bpubmn	3	4	4	2	1	14
Total	40	43	25	14	18	139

Fonte: Raso; Mello (2012)

A tabela 3.2 mostra que 60% das gravações possuem qualidade acústica alta ou muito alta, enquanto somente 23% possuem qualidade acústica baixa. Inevitavelmente a qualidade baixa é mais frequente nas conversações, que por sua natureza possuem mais sobreposições e uma gestão mais difícil dos microfones. Em princípio, a qualidade baixa foi aceita somente quando algum aspecto da gravação era de interesse alto, ou do ponto de vista diafásico ou diastrático. Observe-se que algumas situações interessantes diafasicamente vêm necessariamente com ruído de fundo, como o caso da compra em supermercado e várias outras. Pretender uma qualidade acústica alta nesses casos equivaleria a renunciar à documentação daquela diafasia.

3.1.1.2.2 O equipamento

O tipo de microfone utilizado nas gravações do C-ORAL-ROM de Português Europeu – microfones omnidirecionais – tem como característica central captar sinais sonoros oriundos de toda e qualquer posição ao redor do microfone. Como ponto positivo, microfones omnidirecionais conseguem captar, simultaneamente, a voz de uma grande quantidade de pessoas que esteja ao seu redor, sendo necessário um só microfone para vários falantes. Como consequência, o volume de equipamentos necessários para que se realize uma gravação é reduzido em relação aos casos em que cada participante deve dispor de um microfone.

Todavia, os microfones omnidirecionais captam uma serie de ruídos ambientes, os quais diminuem a qualidade acústica da gravação[43]

O C-ORAL-BRASIL, diferentemente, apresenta uma quantidade reduzida de gravações feitas com a mesma configuração de equipamentos do C-ORAL-ROM de Português Europeu. Em sua maior parte, as gravações do *corpus* de Português Brasileiro foram feitas utilizando um microfone de lapela *wireless* para cada participante, os quais conferem uma alta qualidade de áudio com pouca interferência de ruídos ambientes.

Além disso, o uso de microfones *wireless* tem outra consequência significativa: garantem ao participante de uma gravação a possibilidade de se movimentar livremente sem que perca qualidade de audio, aumentando significativamente o tipo de situações comunicativas passiveis de serem gravadas. Essa caracteristica contribui amplamente ao intuito de se ter uma melhor representação diafásica do PB – um dos objetivos centrais do C-ORAL-BRASIL, em consonância com o projeto C-ORAL-ROM.

[43] Para capturar a voz de uma pessoa que está longe de um microfone omnidirecional, deve-se aumentar o volume do gravador. Em contrapartida, aumenta-se também a quantidade e o volume dos sons ambientes gravados. Grosso modo, pode-se dizer que a quantidade acústica de uma gravação feita com um microfone omnidirecional é diretamente proporcional ao quanto os participantes estão próximos ao microfone e inversamente proporcional ao volume dos sons ambientes.

3.1.1.3 A formação dos transcritores

A primeira fase de trabalho relacionada à transcrição consistiu-se da formação do time de transcritores. Este grupo foi formado com alunos de graduação e pós-graduação vinculados ao projeto de pesquisa. Os transcritores que trabalharam no projeto C-ORAL-BRASIL passaram por um processo de formação acadêmica que lhes permitiu conhecer a base teórica e os procedimentos adotados na elaboração do *corpus*. Esta formação contou com cursos e *workshops*, alguns deles oferecidos pelos coordenadores do projeto C-ORAL-ROM, que atuaram como parceiros do projeto brasileiro.

Em uma primeira fase do projeto, um grupo de transcritores em potencial frequentou cursos sobre a base teórica dos critérios de segmentação, a Teoria da Língua em Ato. Depois desta primeira fase, dez alunos (entre doutorandos, mestrandos e bolsistas de iniciação científica) foram convidados a realizar um treinamento mais específico relativo à segmentação prosódica. Este treinamento teve como objetivo não apenas capacitar os transcritores para a tarefa de segmentar prosodicamente os textos por eles transcritos, mas especialmente tinha como propósito validar a consistência da anotação da segmentação prosódica do *corpus*.

Ao longo do desenvolvimento do projeto, houve mudanças no time de transcritores (saída de alguns membros e inclusão de novas pessoas), resultando assim em um grupo heterogêneo no que se referia ao grau de formação teórica e experiência com a tarefa da transcrição e segmentação. Para lidar com este fato, decidiu-se que os transcritores menos experientes trabalhariam somente nas fases iniciais do processo transcritório, e que seu trabalho seria sempre revisado por um dos transcritores com mais experiência. Antes de iniciarem as transcrições, todo membro da equipe realizou exercícios de transcrição que foram avaliados pelos coordenadores do projeto.

3.1.1.4 Treinamento para a segmentação prosódica

Uma das características importantes do C-ORAL-BRASIL é o fato de que o texto transcrito recebeu uma anotação que indica as fronteiras prosódicas perceptíveis no fluxo da fala. As fronteiras prosódicas podem ser de diferentes tipos e cada tipo é assinalado na transcrição com um sinal próprio. Para a implementação deste tipo de anotação no *corpus*, a equipe de transcritores passou por um treinamento que teve dois objetivos:

I. capacitar os transcritores na identificação das pistas entonacionais que marcam as diferentes fronteiras prosódicas no fluxo da fala e diferenciá-las de outras informações acústicas que não implicam a segmentação da fala em unidades prosódicas; o ponto fundamental era garantir que todos os transcritores marcassem da mesma maneira o mesmo fenômeno percebido e não se confundissem na marcação fenômenos prosódicos diferentes;
II. validar a segmentação prosódica das transcrições do *corpus*, garantindo um alto grau de coerência e consistência na anotação das fronteiras prosódicas nas transcrições.

A constituição do time de transcritores do projeto C-ORAL-BRASIL foi efetivada de modo a construir um time coeso e com alta capacitação para a realização das transcrições do *corpus*. O fato de a transcrição apresentar um nível de anotação prosódica além do segmental consistiu em um desafio na formação e treinamento. É importante assinalar que as transcrições e revisões foram realizadas por indivíduos diferentes, o que fatalmente resulta em interpretações, mesmo ligeiramente, diferentes de um mesmo estímulo. Além disso, é sabido que um mesmo indivíduo também pode realizar julgamentos distintos em momentos diferentes. Com isso, houve a necessidade de elaborar um processo que minimizasse as inconsistências na anotação da segmentação do fluxo da fala em unidades prosódicas. Este

treinamento estabeleceu o patamar de confiabilidade na anotação de fronteiras prosódicas do *corpus* e ao mesmo tempo validou o processo (RASO 2011, p. 5)

Para se chegar a esta metodologia de trabalho, baseou-se na experiência do projeto C-ORAL-ROM e também nos prazos, recursos e pessoal específicos do projeto C-ORAL-BRASIL. Com a experiência e a orientação dos coordenadores do grupo europeu, decidiu-se não replicar o mesmo processo aplicado no C-ORAL-ROM, mas sim implementar procedimentos diferentes, e, com isso, foram desenvolvidos avanços metodológicos. Com a implementação desta nova metodologia, verificaram-se resultados muito bons quanto à formação dos transcritores, e isso refletiu-se em uma qualidade muito alta nas transcrições do C-ORAL-BRASIL.

Assim como no C-ORAL-ROM, os textos são transcritos no formato CHILDES-CLAN (MacWhinney, 2000) implementado para a anotação prosódica (Moneglia; Cresti, 1997). A anotação prosódica prevê a segmentação da fala em enunciados (barra dupla) e em unidades tonais (barra simples); são marcados também os enunciados interrompidos (+) e os *retractings* ([/]). A transcrição é de base ortográfica, mas com uma ampla série de adaptações importantes para capturar fenômenos da fala em curso de lexicalização ou de gramaticalização e podê-los estudar através de ferramentas informáticas e estatísticas.

A maioria dos critérios adotados para a transcrição[44] foram os mesmos estabelecidos para o C-ORAL-ROM, porém muitos casos não estavam previstos naqueles *corpora* e foram sendo acrescentados à medida que os transcritores se deparavam com situações novas, tais como:

1. ruídos paralinguísticos: tosse e riso, bem como outros sons paralinguísticos, são transcritos como *hhh* e, como regra geral, não recebem marca de segmentação prosódica, por entender-se que tais sons não apresentam valor informacional de natureza linguística (não constituem uma unidade informacional).

[44] Todos esses critérios foram retirados do C-ORAL-BRASIL. Para maiores detalhes ver: RASO; MELLO (orgs). *C-ORAL-BRASIL I: Reference Corpus for Informal Spoken Brazilian Portuguese/C-ORAL-BRASIL I: Corpus de Referência do Português Brasileiro Falado Informal* (2012).

Exemplo 3.1 (*bfamcv04*)
*HEL: *hhh barril //*

Entretanto, se o ruído possuir algum valor ilocucionário, a segmentação se torna necessária. Isso acontece, por exemplo, se um falante pede dinheiro emprestado e recebe como resposta um riso que vale como recusa frente ao absurdo do pedido. Nesse caso, é atribuído valor comunicativo ao riso, que é, portanto, transcrito e segmentado:

*ABC: hhh //

São transcritos de maneira especial o clique que convenciona manifestação de incômodo, grafado como *nts*, e o ruído *psiu*, usado frequentemente como chamamento.

Algumas vezes o falante realiza uma leitura e substitui parte do trecho lido por um som que sinaliza desinteresse pelo referido trecho. Convencionou-se transcrever o som como *nanananã*, ainda que este não seja exatamente o que foi produzido pelo falante.

2. Hesitações e palavras interrompidas: hesitações por parte do falante, bem como tomadas de tempo preenchidas por um som vocálico, são transcritas como *&he*, independentemente da qualidade da vogal produzida pelo falante.

3. Onomatopeias: são transcritas conforme a pronúncia. Por exemplo, a onomatopeia para o bater na porta pode ser transcrita *toque toque toque* ou *toc toc toc*, dependendo de como for pronunciada. A imitação do choro pode ser transcrita *ué*, caso seja esta a pronúncia.

4. Interjeições e exclamações: As interjeições *ah*, *eh*, *ih*, *oh* e *uh* são transcritas ortograficamente. Os casos *oh* e *eh* apresentam alguma dificuldade na transcrição. A interjeição *oh* deve ser distinguida do vocativo *ô* e de *o'* (vocativo e redução da forma *olha*). Quando se torna difícil distinguir entre *oh* e *o'*, recorre-se a uma prova

de comutação: se é possível a substituição por *ah*, significa que deve ser transcrita *oh*. Algumas vezes é difícil a distinção entre *eh* (interjeição), *&he* (tomada de tempo) e *é* (verbo *ser*). A distinção entre *eh* e *&he* também pode ser feita por uma prova de comutação, mudando-se a vogal da exclamação.

Certas exclamações são usadas com extrema frequência para afirmar ou negar, e também para outras ilocuções. Em caso de afirmação (ou ironia) temos *hum hum*, *ham ham*, *hum* e *ham*.

Em casos de negação temos as formas *uhn uhn* e *ahn ahn*, e para a interrogação temos *uhn* e *ahn*.

Exclamações de cunho religioso são transcritas com inicial maiúscula, e representam-se as formas abreviadas e reduzidas, especialmente comuns na fala mineira. Para a exclamação *Nossa Senhora* são possíveis as variações *Nossa*, *No'*, *Nu'* e *Nusga.* Para *Virgem Maria* são previstas *Vixe'* e *Vix'.* Para a redução de *Ave Maria* convencionou-se a forma *Aff'*.

5. Siglas e acrônimos: são transcritos ortograficamente e com letras maiúsculas caso pronunciados como palavras.
6. Numerais: quando formados por apenas uma palavra, são transcritos conforme a ortografia. Os elementos que compõe numerais compostos são separados por hífen:

7. Palavras estrangeiras e erros de pronúncia: sempre que existe uma tradição gráfica em português, as palavras estrangeiras são transcritas segundo essa tradição, ou seja, conforme a ortografia da língua portuguesa (ex.: *estresse*, *estéreo*), a não ser que fique claro que a pessoa esteja falando com a pronúncia estrangeira (ex.: *stereo*). Em caso de inexistência de uma forma aportuguesada, respeita-se a grafia em língua estrangeira.

Nomes de produtos estrangeiros (*YouTube*, *Google Video*) são transcritos com a grafia original, ainda que haja a forma aportuguesada para certas palavras, como no caso de *Google Video*, em que existe a grafia portuguesa *vídeo*.

Se a pronúncia da palavra estrangeira é incorreta (ex.: *Big Brogher* ao invés de *Big Brother*), transcreve-se a forma conforme pronunciada e registra-se o erro nos comentários do cabeçalho. Quanto aos erros de pronúncia em palavras do português (ou de qualquer língua), se o falante erra e se corrige, registra-se simplesmente a presença do erro e da correção. Se o falante erra sem se corrigir, transcreve-se a forma pronunciada e registra-se o erro nos comentários do cabeçalho.

8. Formas transcritas conforme a ortografia padrão: a) i*nfinitivos***:** a marca de infinitivo (*r* final), mesmo se não perceptível, é mantida em todos os verbos; b) l*etras do alfabeto*: são transcritas ortograficamente (*á*, *bê*, *cê*, *xis*, *jota* etc.); c) a*s formas que*, *quê*, *por que*, *porque*, *porquê*: são transcritas conforme a ortografia. Se *quê* está em final de enunciado, recebe acentuação; d)t*ítulos e nomes próprios*: são transcritos sempre com inicial maiúscula.

9. Palavras não transcritas ou censuradas: quando o transcritor é capaz de identificar que existe uma palavra, porém não consegue compreendê-la, transcreve-se *xxx*. No caso de o transcritor não ser capaz de transcrever um trecho maior do que uma palavra, a convenção é *yyyy*.

Há, ainda, o uso de alguns critérios não-ortográficos que tratam principalmente de alguns fenômenos em via de gramaticalização ou lexicalização como a) *aférese* (*brigado*<obrigado); b) *fenômenos relativos à conjugação verbal* (*vamos*, *vamo* e *vão; a*s formas apocopadas equivalentes a *pode*, do tipo *po'* (como em *po'* fazer); a forma *deixa* do verbo *deixar* é transcrita ortograficamente quando é pronunciada como *deixa* ou como *eixa*, mas é transcrita *xá;* a forma *tó*, do verbo *tomar* no imperativo (equivalente a *toma/tome*), com sentido de transferência de posse; o verbo *olhar* no indicativo é passível de redução na expressão *olha lá>a' lá; a* primeira pessoa do plural transcreve-se respeitando a pronúncia, por exemplo, *empurramos* ou *empurramo*, conforme for pronunciada. Analogamente, é respeitada na transcrição a forma não-padrão *empurremo*, assim como as outras formas não-padrão, como *seje* por *seja*, *envem* por *vem*. A redução das desinências nos paradigmas verbais é sempre respeitada na transcrição. Portanto são mantidos os

tipos *nós vai*, *eles foi*, *vocês faz*, *tu é* etc.); o plural *(ques menino bonito* (= que meninos bonitos); fenômenos relativos a pronomes *(ocê* e *ocês* para a segunda pessoa); c) p*reposições articuladas e reduzidas(pra*, *pro*, *pras*, *pros*). A preposição *em* é transcrita como *ni* quando assim é pronunciada (*ni mim*, *ni entrevista* equivalem a *em mim*, *em entrevist*a); *Articulação de preposições e pronomes* (são previstas as combinações das preposições *para* (*pra*, *pa*, *pr'*, *p'*), *em* (*ni*, *n'*), *de* (*d'*), com (*c'*) e os pronomes *e'*, *ea*, *es*, *eas*, *ocê*, *ocês*, *cê*, *cês*); d) *negação* (frequentemente a negação *não* se apresenta sob forma clítica ou enfraquecida. Decidiu-se criar um elemento não-ortográfico (*nũ*) para registrar o fenômeno); e) *construções interrogativas, o pronome relativo e pseudo-relativo* (em construções interrogativas do tipo *que que*, *por que que*, *quando que*, *quanto que* etc., não se transcreve o verbo *é* se não for pronunciado. Se a pronúncia do primeiro *que* for com vogal *e* aberta, considera-se que o verbo é pronunciado e transcreve-se *que é que*); f) *as formas Senhor e Senhora (senhor*, *sior*, *seu*, *sô*); g) d*iminutivos (Sozinho*, por exemplo, é transcrito *sozim* quando é esta a pronúncia; h) *o intensificador mó* (a forma, tanto com o sentido de *maior* quanto com o de *muito*, é transcrita tal como se pronuncia, ou seja, *mó*.) e i) *Rotacismo (Planta* é pronunciada como *pranta).*

3.1.1.5 A segmentação prosódica

De modo a evitar o uso artificial para a fala dos sinais de pontuação normalmente empregadas na escrita (vírgula, ponto final, interrogação etc.), a transcrição do C-ORAL-BRASIL representa os fenômenos de segmentação da fala através de um sistema de anotação de fronteiras prosódicas (Moneglia; Cresti, 1997) desenvolvido com base na perspectiva da Teoria da Língua em Ato e que já havia sido implementado no projeto C-ORAL-ROM e, antes ainda, no *corpus* LABLITA (Cresti, 2000a). Este sistema de anotação pretende representar, de maneira bastante simples para o transcritor e para o leitor, a segmentação natural da fala em unidades delimitadas prosodicamente e que apresentam autonomia pragmática, os enunciados; bem como a estruturação interna destes em unidades menores,

também delimitadas prosodicamente mas que não têm autonomia pragmática (RASO 2012, p. 19).

A premissa fundamental que orientou os critérios de segmentação no C-ORAL-ROM e no C-ORAL-BRASIL é a da existência de variações prosódicas perceptivelmente relevantes no contínuo da fala natural, que o dividem em unidades prosódicas discretas (Moneglia, 2006; Moneglia *et al.*, 2005). Estas variações, que causam o efeito de segmentação (*parsing*) da fala, são chamadas de quebras prosódicas. Uma quebra prosódica pode ser interpretada como transmitindo um valor conclusivo ou não conclusivo. As quebras prosódicas que delimitam enunciados são percebidas pelo interlocutor como conclusivas e assinalam, portanto, que o enunciado está terminado. Este tipo de quebra prosódica é chamada de quebra prosódica terminal. As quebras que não marcam a conclusão do enunciado são chamadas de quebras não terminais. Quebras prosódicas terminais (e certos tipos de quebras não terminais) têm alta saliência perceptual, conforme demonstrado em testes de acordo quanto à identificação de quebras prosódicas terminais e não terminais no contínuo da fala, tanto em anotadores experientes quanto não experientes (BUHMAN *et al.*, 2002; MONEGLIA; CRESTI, 2006; NESTERENKO, 2006; MO; COLE; LEE, 2008; RASO; MITTMANN, 2009; MONEGLIA *et al.*, 2010).

É importante não confundir quebra prosódica com pausa, entendida como um silêncio do falante. A fronteira entre enunciados não é determinada pela existência de silêncio entre eles. Verificam-se muitos casos de quebra percebida como terminal sem que haja, necessariamente, silêncio entre enunciados adjacentes. Naturalmente, silêncios no fluxo da fala constituem certo tipo de quebras prosódicas, mas não necessariamente terminal. Nas abordagens atuais a recursos multimídia de fala, o critério mais comum para segmentar o fluxo de fala é a identificação de pausas, provavelmente também pela possibilidade oferecida pela tecnologia de uma detecção automática. Um estudo realizado sobre o C-ORAL-ROM (Moneglia, 2005) observou que 63% das quebras terminais são seguidas por pausa, enquanto 37% não apresentam alguma pausa depois. Ao contrário, cerca de 42% das quebras não terminais é seguida por pausa. Isso mostra como a pausa não pode ser considerada um critério apropriado para a segmentação da fala espontânea.

A segmentação prosódica do texto é realizada durante a transcrição deste, ou seja, os transcritores anotam as quebras prosódicas percebidas por eles enquanto transcrevem o texto gravado, já que ambas as tarefas são de natureza perceptual. Os transcritores dispunham de um arquivo de áudio em formato digital e realizaram a transcrição e a segmentação prosódica com base totalmente perceptual. As transcrições foram feitas diretamente em formato digital, utilizando para tal um aplicativo de edição de texto e um aplicativo de edição de áudio que permitisse a fácil navegação no arquivo de som da gravação a ser transcrita.

Após ser transcrito/segmentado, cada texto do *corpus* passou pela revisão de pelo menos um outro transcritor, antes de passar para a próxima etapa, a de alinhamento. A revisão incluiu a conferência da grafia correta das palavras, a aplicação adequada dos critérios não-ortográficos e a marcação apropriada de quebras prosódicas terminais e não terminais.

O alinhamento do texto com o som também se constituiu como uma fase de revisão, especialmente da segmentação prosódica. A unidade básica para o alinhamento é uma sequência delimitada por // ou +; então, durante o procedimento de alinhamento, o responsável pelo alinhamento (sempre um transcritor) realizou correções sempre que julgou que a marcação das quebras prosódicas não estava adequada.

Todos os textos, após serem alinhados, passaram por uma nova revisão, realizada por um grupo de três transcritores experientes, que apresentou grau de acordo na marcação de quebras prosódicas (avaliado por teste Kappa – FLEISS, 1971).

Durante a revisão do alinhamento, foi dada especial atenção à colocação das quebras prosódicas terminais, isso por duas razões principais. Primeiramente, o grau de saliência perceptual deste tipo de quebra é mais alto, e ao mesmo tempo as quebras terminais têm o valor de delimitar os enunciados, que são considerados a unidade básica da comunicação. Assim, a comparabilidade entre C-ORAL-BRASIL e C-ORAL-ROM seria prejudicada caso a unidade de referência não fosse adequadamente representada no primeiro. Em segundo lugar, durante a avaliação

do grau de acordo na anotação de quebras prosódicas realizadas antes do início do trabalho de transcrição, percebeu-se que o maior fator de desacordo estava justamente na distinção entre quebras terminais e não terminais em certos textos de caráter monológico.

Quanto à validação de um *corpus*, o processo consiste, de modo geral, na verificação do *corpus* em relação a um conjunto de especificações técnicas preestabelecidas para sua construção e garante um padrão mínimo de qualidade do produto final, especialmente em corpora que passam por transcrição, como também daqueles que recebem algum tipo de anotação ou de etiquetagem. No caso do C-ORAL-BRASIL foram realizadas duas validações internas que avaliaram aspectos distintos do conteúdo do *corpus*: a transcrição e a segmentação. Por se tratar de validações de conteúdo, foram realizadas manualmente. As validações foram realizadas com o objetivo central de controlar a qualidade do *corpus*. A validação da transcrição é necessária porque é a fidelidade da transcrição que irá sustentar os achados quantitativos de fenômenos morfossintáticos a serem investigados (RASO 2012, p. 25).

No que se refere à validação da segmentação, também se procurou assegurar a comparabilidade com o C-ORAL-ROM, uma vez que estudos comparativos em relação à estrutura informacional só podem ser realizados caso as unidades de referência dos *corpora* comparados sejam rigorosamente as mesmas.

O acordo entre transcritores foi mensurado através do teste estatístico Kappa[45] (Fleiss, 1971) de 0,84 de acordo no total de quebras prosódicas e um índice de 0,90 de acordo nas quebras prosódicas terminais, resultado que é considerado excelente, estatisticamente. Após as primeiras revisões e antes das segundas, um novo teste alcançou um Kappa geral de 0,86 e de 0,91 para as quebras terminais.

[45] O Kappa é uma medida de concordância entre observadores, com valores entre o e 1; 0 significa falta de acordo ou acordo esperado pelo acaso; 1 significa acordo total. Um valor Kappa maior que 0 indica alguma concordância. Segundo a tradição encontrada na literatura (Landis; Koch, 1977): 0-0,19 = acordo fraco; 0,2-0,39 = acordo vago; 0,4-0,59 = acordo moderado; 0,6-0,7 = acordo substancial; >0,8 = acordo quase total. (Para maiores detalhes, ver Raso;Mittman, 2011)

Os critérios de transcrição também foram validados. Os resultados parciais apontam uma margem de erro de 1,4%, o que estatisticamente é excelente[46]. A segmentação dos enunciados em unidades tonais é feita somente com base na identificação de quebras prosódicas com perfil entonacional percebido como terminal (final de enunciado) ou não-terminal (final de unidade interna ao enunciado).

O processo de validação garante um padrão mínimo de qualidade do produto final, especialmente em corpora que passam por transcrição, como também daqueles que recebem algum tipo de anotação ou de etiquetagem.

Comparativamente ao C-ORAL-ROM, nota-se que o acordo obtido para os *corpora* que o compõem é tendencialmente maior do que o que se chegou no C-ORAL-BRASIL. Contudo, os resultados podem ser comparados apenas parcialmente, em virtude da aplicação de metodologias de validação diferentes em cada projeto. No projeto C-ORAL-ROM, o acordo quanto à segmentação prosódica foi avaliado por uma instituição externa, através de um grupo de avaliadores não *experts*. Esta avaliação foi realizada depois que o *corpus* já havia sido transcrito e segmentado prosodicamente.

Uma amostra aleatória de 1/3 dos enunciados de cada *corpus* do C-ORAL-ROM (italiano, francês, espanhol e português europeu) foi utilizada para fins de validação. Os avaliadores externos passaram por duas sessões de treinamento antes de iniciarem a tarefa, que consistia em conferir a anotação prosódica realizada em cada um dos enunciados da amostra. Neste caso, os avaliadores tiveram acesso ao texto transcrito e segmentado, e deveriam apenas indicar seu acordo quanto à segmentação a eles apresentada. O acordo obtido em todas as amostras foi sempre maior do que 0,6 (Kappa), e o acordo total percentual obtido foi sempre maior do que 95% (MONEGLIA *et al.*, 2005).

No caso do C-ORAL-BRASIL, uma validação externa com transcritores não *experts* ainda deve ser realizada. Naturalmente, a metodologia empregada na avaliação externa realizada no C-ORAL-ROM induz a um grau de acordo superior ao que é

[46] Mello, Raso, Vale e Goulart (em preparação).

obtido quando se comparam anotações realizadas independentemente, em que cada avaliador efetivamente anota as quebras percebidas em um texto não segmentado, como a que foi realizada no C-ORAL-BRASIL. As particularidades da metodologia empregada no C-ORAL-BRASIL trouxe vantagens do ponto de vista da execução do *corpus* brasileiro, uma vez que elevou muito o nível de habilidade dos transcritores que trabalharam no *corpus*.

Outro aspecto interessante relaciona-se à análise dos casos de desacordo na anotação das quebras prosódicas. Estes casos fornecem material para a realização de estudos mais detalhados sobre o tema, que ainda poderiam ser mais explorados (Moneglia *et al.*, 2010). Em especial, o estudo dos contextos de acordo e desacordo pode revelar que parâmetros prosódicos são determinantes na produção dos diferentes tipos de quebra prosódica, o que poderia contribuir para o avanço da compreensão sobre a estrutura prosódica da fala espontânea, mas também ser de grande valia para áreas relacionadas à tecnologia da fala, tais como o Processamento de Linguagem Natural.

Depois de resolvidas essas questões, os textos passaram por um processamento automático para a prospecção de dados e análise estatística por meio do programa R. Em seguida, os textos da primeira versão foram alinhados, palavra a palavra, com a segunda versão, utilizando para isso planilhas eletrônicas. Como era esperado, as versões de cada texto apresentaram um numero de palavras diferentes, devido a inclusões e exclusões referentes a correções no segmental. Como a inclusão ou exclusão de palavra pode alterar a marcação de uma quebra prosódica, as alterações no segmental também foram controladas.

Em relação às quebras prosódicas, foram controladas todas as alterações possíveis, que são:

I. Inclusão de quebra: terminal, interrupção, não terminal, retratação.
II. Exclusão de quebra: terminal, interrupção, não terminal, retratação.
III. Alteração de quebra: terminal para interrupção, terminal para não terminal, terminal para retratação; não terminal para interrupção, não terminal para

terminal, não terminal para retratação; interrupção para terminal, interrupção para não terminal, interrupção para retratação; retratação para interrupção, retratação para não terminal e retratação para terminal.

Para análise, foram utilizados os 20 textos da amostra etiquetada, em duas versões (total de 40 textos), com um total de 31.748 tokens. Os tokens correspondem a todas as posições em fronteira de palavra, considerando-se as palavras incluídas durante a etiquetagem, com isso, a soma do numero de palavras dos textos e menor do que o numero de tokens analisados. Deste total, foram registradas alterações de tipo segmental em 1300 tokens, o que corresponde a 4% de alterações no segmental.

Tabela 3.3 - Alterações no segmental realizadas na amostra etiquetada informacionalmente do C-ORAL-BRASIL

TIPO DE ALTERAÇÃO	FREQUÊNCIA	%
Inclusão de palavra	528	41
Exclusão de palavra	211	16
Correção de palavra	561	43
Total	1300	100%

Fonte: Raso; Mello (2012)

As alterações realizadas no segmental foram pequenas, indicando alto grau de fidelidade das transcrições revisadas. Além desses resultados, foi proposta uma comparação da anotação de quebras prosódicas entre as versões revisadas e etiquetadas, que estão representados na tabela 3.4 a seguir.

Tabela 3.4 – Diferença no total de quebras prosódicas nas versões revisada e etiquetada das transcrições

TIPO DE QUEBRA	TRANSCRIÇÃO REVISADA	TRANSCRIÇÃO ETIQUETADA	DIFERENÇA ABSOLUTA	DIFERENÇA %
Quebra terminal	5239	5040	-199	-4%
Interrupção	470	465	-5	-1%
Quebra não terminal	4033	4346	313	7%
Retratação	827	877	50	6%
TOTAL	10569	10728	159	1%

Fonte: Raso; Mello (2012)

Como se pode notar, a diferença entre as quebras na transcrição e após a etiquetagem é proporcionalmente pequena, havendo apenas um acréscimo de 1% de quebras no total.

No nível segmental o aumento é de 4% de tokens na versão etiquetada e 1% de quebras na versão etiquetada. Pôde-se concluir, portanto, que as mudanças em nível segmental influenciaram pouco a anotação prosódica "*Em outras palavras, inclusões e exclusões de palavras provavelmente afetaram muito pouco a inclusão ou exclusão de quebras prosódicas*". (MITTMANN, 2011 - em preparação)

Quanto às quebras terminais, observou-se que foram superespecificadas no momento das transcrições e revisões. Isto é, diversas quebras terminais foram substituídas por quebras não terminais. Nos textos monológicos houve uma diminuição de -2,8% de quebras; nos textos dialógicos, -1,5%; as conversações, todavia, apresentaram um aumento de 0,3% de quebras terminais.

> Estes resultados são compatíveis com aqueles obtidos na validação da segmentação, pois o acordo em relação as quebras terminais em monólogos foi um dos valores de maior flutuação entre os testes.Proporcionalmente, a maior diferença está na atribuição de quebras não terminais. Isso se deve a dois fatores: a transformação de quebras terminais em não terminais (de modo coerente com o resultado descrito acima) e a inclusão de quebras não terminais onde não havia antes nenhuma quebra. É importante assinalar que, dos 31.748 tokens que constituem a população analisada, apenas 1.815 representam alterações, o que corresponde a somente 6% do total de tokens. (MITTMANN, 2011 - em preparação)

A tabela 3.5 detalha os tipos de alterações realizadas durante todo o processo. Os resultados são muito importantes porque ratificam o fato de haver maior frequência de alterações em relação às quebras não terminais. Esse tipo de quebra está relacionada à realização de padrões informacionais dentro dos enunciados, estrofes e padrões ilocucionários; em outras palavras, dentro de qualquer sequência linguística pragmaticamente autônoma, assinalada por um perfil prosódico conclusivo.

Tabela 3.5 – Tipos de alterações na anotação de quebras prosódicas realizadas durante a etiquetagem informacional

TIPO DE ALTERAÇÃO	FREQUÊNCIA	%
Inclusões de Quebra	**622**	**34,3%**
terminal	155	8,5%
interrupção	26	1,4%
não terminal	410	22,6%
retratação	31	1,7%
Alterações de Quebra	**699**	**38,5%**
terminal →interrupção	39	2,1%
terminal →não terminal	369	20,3%
terminal →retratação	5	0,3%
não terminal →interrupção	20	1,1%
não terminal →terminal	94	5,2%
não terminal →retratação	32	1,8%
interrupção → terminal	33	1,8%
interrupção → retratação	17	0,9%
interrupção → interrupção	52	2,9%
retratação → interrupção	6	0,3%
retratação → não terminal	26	1,4%
retratação → terminal	6	0,3%
Exclusões de Quebra	**494**	**27,2%**
terminal	79	4,4%
interrupção	19	1,0%
não terminal	364	20,1%
retratação	32	1,8%
TOTAL	**1815**	100%

Os resultados apresentados na tabela estão divididos entre inclusões de quebra, alterações de quebra e exclusões de quebra. Em relação às inclusões de quebra houve um total de 34,3% de alterações. A maior parte refere-se à inclusão de quebras não terminais com 26,6% de ocorrências.

Quanto às alterações de quebras (38,5%), o que se destaca é a mudança de quebra terminal para quebra não terminal com 20,3% de ocorrências. As exclusões de quebra representam 27,2% das mudanças e, novamente, são as quebras não terminais que mostraram a maior taxa de alterações com 20,1% de ocorrências.

A título de exemplo, segue um trecho de uma transcrição etiquetada informacionalmente.

Exemplo 3.2 – bfamdl02 [1]-[16]:
*BEL: [1] que que fica aqui //=COM=$
*BAL: [2] eu tô pensando ainda //=COM=$

*BEL: [3] mas tipo <assim /=INT= cê nũ> [/2]=EMP= cê nũ sabe /=CMM= ou cê comprou e nũ <tem> //=CMM=$
*BAL: [4] <&he /=TMT= olha só> //=COM=$[5] <não> /=INP= isso aí veio da mochila //=COM=$
*BEL: [6] ah //=COM=$[7] <uai> //=COM=$
*BAL: [8] <veio como parte> +=EMP=$[9] olha só //=COM=$
*BEL: [10] <talvez as pilhas /=COM= não> //=PHA=$
*BAL: [11] <que que eu tava pensando &a [/6]=EMP= que que> eu tava pensando aqui //=COM=$[12] &a [/1]=EMP= não //=COM=$[13] as pilhas /=TOP= eu coloquei aqui //=COM=$[14] as recarregáveis /=TOP= tão aqui //=COM=$
*BEL: [15] ah /=CMM= tá //=CMM=$
*BAL: [16] tem essas aqui /=COB= que foram usadas /=COB= só que a gente não sabe ainda se elas acabaram /=COB= ou não /=COB= <se elas tem muito> tempo de vida /=COB= etecétera //=COM=$

Esse trecho é parte de um discurso dialógico entre dois interlocutores BEL e BAL que conversam sobre a instalação de um equipamento. As etiquetas são cunhadas em letras maiúsculas e representam as abreviaturas em inglês, de cada uma dessas unidades.

Como se pode notar, as vantagens do C-ORAL-BRASIL para os estudos linguísticos são várias, entre elas a constituição e divulgação de um *corpus* que represente a variedade de fala de Belo Horizonte e região; a constituição de um *corpus* representativo do ponto de vista diafásico, dado que o *corpus* está sendo construído com a preocupação de conter as mais diversas situações comunicativas como já mencionado; a disponibilização de arquivos sonoros alinhados com o texto e o espectrograma; a análise da estrutura informacional do PB e a possibilidade de comparação dessa estrutura com a estrutura informacional das quatro línguas européias presentes no C-ORAL-ROM, dado o uso dos mesmos critérios de segmentação dos textos. Além disso, trabalhos de natureza léxicomorfossintática

serão possíveis através de um *tagger* preparado para esse *corpus* por E. Bick, autor do *tagger* Palavras[47].

3.2 O Subcorpus de pesquisa

Para os estudos da estrutura informacional e das ilocuções, foi selecionado um *subcorpus*, comparável ao *subcorpus* italiano (Raso, 2012) que é formado por 20 textos (6 conversações, 7 diálogos e 7 monólogos), 31.465 palavras e 5.484 enunciados. A escolha dos textos foi feita de maneira a representar uma ampla variação diafásica e diastrática, garantindo a qualidade acústica e a diversificação dos falantes. O *subcorpus* foi etiquetado informacionalmente por meio de um complexo sistema manual. A etiquetagem foi validada através de um acordo entre três etiquetadores e depois comparada, e novamente validada, através do confronto com os etiquetadores do *subcorpus* italiano.

Para a seleção do subcorpus alguns parâmetros foram estabelecidos (RASO; MELLO, 2009, 2010):

a) Representatividade de cada ramificação (diálogos, conversações e monólogos);
b) Maior variação possível de situações comunicativas;
c) Alta qualidade acústica, determinada pela clareza do espectrograma, ausência (total ou parcial) de ruído, ausência de retorno do sinal, clareza da voz, bom ganho de áudio e baixa porcentagem de falas sobrepostas;
d) Diversidade de informantes, equilíbrio em relação ao número de vozes masculinas e femininas e de faixas etárias;
e) Não marcação quanto à diastratia dos informantes;
f) Interesse do conteúdo de modo a aumentar a atenção de transcritores e segmentadores e o nível de informatividade.

[47] Bick (2000, 2012) e Bick; Mello; Raso (2010).

A tabela 3.6 ilustra todos os textos que compõem a amostra, evidenciando a situação de produção, o número de informantes quanto ao gênero, o tempo de duração de cada gravação e o número de palavras em cada texto.

Tabela 3.6 - Características dos textos do subcorpus

Texto	Situação	Masc.	Fem.	Duração h:min:s	Nº palavras
Total.....		**28**	**27**	**03:58:36**	**31318**
Conversações.....		**15**	**9**	**01:07:28**	**9774**
bfamcv01	amigos avaliam um campeonato de futebol organizado por eles e planejam o próximo	4	0	00:07:00	1467
bfamcv02	senhoras conversam sobre os preparativos do casamento de uma parente	0	3	00:07:51	1725
bfamcv03	amigos jogam sinuca	5	0	00:06:50	1390
bfamcv04	amigos jogam "Imagem e Ação", após explicar as regras do jogo para uma das participantes	2	2	00:07:30	1766
bpubcv01	funcionários de banco de sangue explicam como o sangue coletado é armazenado*	1	3	00:08:30	1798
bpubcv02	reunião ordinária em uma sede regional de partido político	3	1	00:29:47	1628
Diálogos.....		**6**	**8**	**01:45:28**	**11331**
bfamdl01	colegas de apartamento fazem as compras do mês*	0	2	00:14:39	2131
bfamdl02	colegas de faculdade batem papo enquanto organizam o material de gravação	1	1	00:07:26	1572
bfamdl03	casal faz uma viagem de carro*	1	1	00:10:30	1637
bfamdl04	domésticas, mãe e filha, fazem a limpeza da cozinha após o almoço*	0	2	00:19:32	1249
bfamdl05	corretor de imóveis leva a irmã para visitar apartamento*	1	1	00:11:28	1736
bpubdl01	engenheiro e pedreiro trabalham em uma obra	2	0	00:26:08	1568
bpubdl02	cliente e vendedor interagem durante a compra de calçados*	1	1	00:15:45	1438
Monólogos.....		**7**	**10**	**01:05:40**	**10213**
bfammn01	senhor narra história fantástica sobre uma cobra	2	0	00:05:02	1086
bfammn02	sobrinha de Carlos Drummond de Andrade conta histórias da família ao neto	1	1	00:07:23	1677
bfammn03	narrativa de "causos" divertidos para a família*	3	3	00:07:08	1206
bfammn04	senhora conta sua experiência no hospital após ter dado à luz no carro*	0	1	00:06:57	1450
bfammn05	senhora fala sobre a adoção da filha após a morte de sua filha biológica*	0	2	00:09:52	1580
bfammn06	pai conta seu percurso profissional à sua filha	1	1	00:10:02	1600

	bpubmn01	entrevista de avaliação sobre aulas de inglês na rede pública de ensino	0	2	00:19:16	1614

* presença de pequenas intervenções de um ou mais informantes que estavam fora da situação.

Os nomes dos textos indicados na tabela acima são siglas que indicam a qual seção do corpus pertence cada um, conforme o padrão estabelecido para o C-ORAL-ROM: a primeira letra indica a língua, as próximas três letras apontam o contexto, e as últimas duas letras designam a tipologia textual. Assim, temos “b” para português do brasileiro; “fam” para contexto familiar/privado e “pub” para contexto público; “cv” para conversação, “dl” para diálogo e “mn” para monólogo. Por fim, cada texto recebe um número sequencial de dois dígitos que o identifica dentro da seção a que pertence.

O subcorpus de pesquisa tem um total de 31.318 palavras em 3h 58min e 36s de gravação, da qual participam um total de 55 informantes, 28 do sexo masculino e 26 do sexo feminino. As conversações compreendem 1h 7min e 28s de gravação com 24 informantes (15 homens e 9 mulheres), os diálogos compreendem 1h 45min e 28s de gravação com 14 informantes (6 homens e 8 mulheres), e os monólogos correspondem ao todo a 1h 5min e 40s de duração e 17 informantes (7 homens e 10 mulheres). Para o cálculo do número de palavras não foram consideradas as ocorrências de “hhh” (convenção utilizada para som paralinguístico) e as palavras hifenizadas foram consideradas como apenas uma, o que inclui todos os numerais compostos por mais de uma palavra, como por exemplo “trinta-e-nove”.

A tabela 3.6 trouxe a descrição das situações em que os informantes foram gravados. A variação diafásica é um parâmetro importante para este trabalho, pois é ela que garante, em primeiro lugar, que os textos analisados sejam realmente espontâneos, produzidos em contextos naturais; e em segundo lugar, variando as situações varia também a acionalidade presente nos diferentes textos e segmentos da amostra.

A título de exemplo, ilustramos primeiramente um trecho de um texto monológico.

Exemplo 3.3 (bfammn02)

*DFL: [11] e então /=DCT= tinha muito texto do tio Carlos /=TOP= então ele falava /=INT= ah /=EXP_r= ele é tio da minha tia //=COM_r=$ [12] ele passou a ter o apelido de tio da minha tia hhh //=COM=$ [13] os menino falava assim hhh /=INT= ô louro /=CMM_r= tio da minha tia hhh //=CMM_r=$ [14] e então /=DCT= uma vez veio o tio Carlos /=COB= eu /=SCA= mandei uns livros pra ele /=SCA= pr' ele dar uns autógrafos /=COB= <e ele até> /=SCA= me censurou /=COB= ah /=EXP_r= você devia ter me falado /=COB_r= que eu /=SCA= tenho livro aqui /=COB_r= mandaria /=COB_r= nũ precisava d' cê ter mandado //=COM=$

*LUC: [15] <hum hum> //=COM=$

*DFL: [16] então /=DCT= e pro Laurinho ele pôs /=SCA= &he /=TMT= a [/1]=EMP= a dedicatória /=COB= e pôs “do tio da sua tia” hhh //=COM=$ [17] eu contei o caso <pra ele /=COM= né> //=PHA=$

*LUC: [18] <hum hum> //=COM=$

*DFL: [19] e o outro caso interessante foi quando /=INT= há muitos anos atrás /=TOP= eu fui [/2]=EMP= pela primeira vez que eu fui no angiologista /=TOP= ele &pedi [/1]=SCA= perguntou meu nome //=COM=$

*LUC: [20] hum hum //=COM=$

Todo texto monológico é formado majoritariamente por narrativas, mesmo sendo ele constituído por dois interlocutores, como é o caso do exemplo acima. O que o caracteriza como monólogo é o fato de um dos interlocutores (LUC) apenas ratificar o que é dito por DFL. Em momento algum LUC interfere na história contada por DFL, ele apenas a ouve e demonstra esse fato por meio do ‘hum hum’. Ao contrário, os grupos das conversações e dos diálogos compreendem textos nos quais a fala é extremamente situada e ancorada no contexto imediato, e estes consequentemente são muito acionais. Passemos a um exemplo.

Exemplo 3.4 (bpubdl03)

*GUI: [19] vão lá //$

*GUI: [20] pode ir //$

*TOM: [21] pode //$

*GUI: [22] pera aí //$
*GUI: [23] aí //$
*GUI: [24] pode abrir um pouquinho as mãos //$
*GUI: [25] isso //$
*GUI: [26] porque senão cê vai sentir muito tríceps //$
*GUI: [27] a cabeça nũ tá tremendo não / né //$
*TOM: [28] não //$

Nesse caso, o interlocutor GUI preocupa-se com o que diz o interlocutor TOM. A interação é construída à medida que o diálogo se estabelece. Assim sendo, pode-se afirmar que, substancialmente, todo texto dialógico é interacional, enquanto todo texto monológico é mais textual.

3.2.1 Prospecção e tabulação dos dados

Toda a parte de prospecção dos dados, conforme se verá abaixo, foi realizada na Itália, no laboratório *LABLITA* (Panunzi; Gregori, 2011). Para a prospecção dos dados foi utilizada uma técnica que mesclou o uso da interface de buscas do LSLD, abaixo representada pela figura 3.7, com o uso de planilhas eletrônicas, que permitiram a manipulação de dados.

Figura 3.1 – Interface web do *LABLITA*



O LABLITA Spoken Language Database[48] é um banco de dados no formato XML desenvolvido por Panunzi e Gregori (2011) no qual se permite que se estudem as unidades informacionais em corpora de fala anotados, seguindo os princípios da teoria dos padrões informacionais - Information Patterning Theory (Cresti, 2000; Moneglia; Cresti, 2006; Scarano, 2009). O esquema XML permite a anotação em três níveis: fronteiras prosódicas, estrutura informacional, anotação morfossintática e lematização.

Definidos os critérios de busca, o sistema retorna às unidades terminadas (enunciados, padrões ilocucionários, estrofes) que se associam aos critérios e apresenta as ocorrências na tela. Os resultados podem ainda ser exportados em um documento no formato CSV, permitindo a manipulação dos dados por meio de uma planilha eletrônica. O documento CSV apresenta em cada linha: a identificação do texto, o número da unidade no arquivo de alinhamento Winpitch (Martin, 2004), sigla do informante que produziu a ocorrência, enunciado correspondente à ocorrência, e por fim as siglas das unidades informacionais que formam o padrão na ocorrência. A figura 3.8 ilustra o resultado da busca no formato CSV, conforme visualizado em uma planilha eletrônica.

Figura 3.2 – Visualização de documento CSV em planilha eletrônica

bfamdl04	170 HRM	é /=CMM= né /=CMM= Silvia //=ALL=	CMM	CMM	ALL			
bfamdl04	177 SIL	mas não /=PHA= é da Ernesta /=CMM= nũ vou tomar não //=CMM=	PHA	CMM	CMM			
bfamcv01	48 EVN	ô /=EXP= mas tem outros lugares /=CMM= o negócio é que a gente não procurou //=CMM=	EXP	CMM	CMM			
bfamcv01	54 EVN	< o' /=CNT= o Arnaldinum > é caro /=CMM= e tem aqueles problemas //=CMM=	CNT	CMM	CMM			
bfamcv01	220 GIL	Aminas /=CMM= e /=DCT= Mauro e Filhos //=CMM=	CMM	DCT	CMM			
bfamcv02	44 TER	ah /=EXP= a Fafica não /=CMM= uai /=PHA= a Fafica tem quarenta-e [/1]=SCA= < quarenta-e-pouco > //=CMM=	EXP	CMM	PHA	SCA	CMM	
bfamcv02	70 RUT	< pelo > amor de Deus /=CMM= nũ me convida pa ser &pa [/1]=SCA= madrinha não /=CMM= hein //=PHA=	CMM	SCA	CMM	PHA		
bfamcv02	222 TER	é mesmo /=CMM= né /=CMM= < Rute > //=ALL=	CMM	CMM	ALL			
bfamcv02	231 TER	não /=CMM= mas nũ é não /=CMM= &Ru [/1]=EMP= Jael //=ALL=	CMM	CMM	EMP	ALL		
bfamcv02	282 RUT	não /=PHA= eu < gosto assim /=CMM= sabe > //=CMM=	PHA	CMM	CMM			
bfamcv02	312 TER	&he /=TMT= &he /=TMT= e a Dani também vai ficar &b linda /=CMM= né /=CMM= < Rute > //=ALL=	TMT	TMT	CMM	CMM	ALL	
bfamcv03	45 CEL	não /=CMM= é /=CMM= uai //=PHA=	CMM	CMM	PHA			

[48]Por enquanto esta ferramenta não foi disponibilizada para o público. O acesso só é permitido aos pesquisadores do C-ORAL-BRASIL.

Esta ferramenta permitiu 1) gerar de todos os enunciados do subcorpus, necessários ao levantamento de todas as estruturas informacionais da amostra; 2) listar todos os enunciados que continham pelo menos uma unidade de APC; 3) enumerar todos os enunciados que continham um ou mais auxílios dialógicos, mas que não continham outras unidades informacionais que não o COM ou CMM e assim verificar a frequência de ocorrência do APC proporcionalmente ao total de enunciados que possuíam uma ou mais unidades textuais além do COM.

As planilhas eletrônicas possibilitaram que os dados fossem organizados, classificados e tabulados. Durante este processo a interface web do LSLD foi usada com o objetivo de conferir os resultados obtidos e minimizar os erros.

3.2.2 A forma entonacional da unidade de APC

Com base nas pesquisas do IPO, a Teoria da Língua em Ato admite que alguns movimentos de F0 (e algumas configurações de movimentos) são percebidos da mesma forma pelos falantes, constituindo classes teóricas de movimentos ou perfis prosódicos.

Os perfis prosódicos são formados por um ou dois movimentos obrigatórios de F0 e podem ser integrados por porções opcionais de *preparação*, *ligação* ou *coda*. Também foi visto que, dentro dos *padrões entonacionais* (composições formadas por um ou mais perfis prosódicos), observa-se que alguns perfis compartilham características distributivas e funcionais em comum: alguns, chamados de *root*, são necessários e podem se posicionar livremente no padrão entonacional. Outros, chamados *prefix*, são opcionais e devem aparecer antes de um perfil *root* ou de um outro perfil *prefix*. Os *sufix* são igualmente opcionais, mas devem vir sempre depois de um *root* ou de um *prefix*. Sendo assim, existem classes de perfis entonacionais, chamadas *unidades tonais*, as quais compartilham características distributivas e funcionais.

Segundo a Teoria da Língua em Ato, existe uma correspondência sistemática entre o nível de analise da entonação e o nível de analise da informação no enunciado. Dessa forma, cada unidade tonal e, no *padrão informacional*, uma *unidade informacional*. As unidades com perfil de *root* tem, no padrão informacional, a função de realizar a ilocução do enunciado e denominam-se Comentário. As unidades de *prefix* constituem, no padrão informacional, o âmbito de aplicação da forca ilocucionária do enunciado e são chamadas de Tópico. As unidades de *sufix* são integrações textuais das unidades precedentes e são chamadas de Apêndice.

As pesquisas do LABLITA (Cresti, 2000) mostram que a unidade de APC é definida como sendo a integração textual de uma unidade informativa de comentário e como tal, deve necessariamente seguir a unidade na qual está interligada, o comentário.

Entonacionalmente, o APC se configura como uma unidade de sufixo, subordinada ao COM, e não possui foco funcional, mantendo sempre um perfil nivelado ou descendente identificado por um abaixamento do tom de voz, pela baixa intensidade e pelo fato desse movimento único corresponder à unidade tonal inteira, sem variação de movimento, independente da estrutura silábica (CRESTI, 2002a; FIRENZUOLI, 2002).

3.2.3 Detalhamento da metodologia utilizada

A identificação e descrição da forma entonacional de APC pressupõe uma análise acústica apurada, por isso foi fundamental que todos os textos do subcorpus apresentassem qualidade acústica igual ou superior a de textos classificados como de qualidade ‘B’ no C-ORAL-BRASIL (Raso, 2012). Isso significa que os textos devam apresentar pouco ruído de fundo, pouca sobreposição da fala dos falantes, parte considerável do áudio com qualidade para análise fonética, sendo possível aferir as curvas de F0.

Na análise de todos os enunciados que continham pelo menos uma unidade de APC, verificamos que havia dois enunciados que manifestavam a creaky voice e 15 em que os ruídos externos se sobrepunham às falas dos falantes ou a fala de mais informantes estava sobreposta.

Como é sabido, na creaky voice "as cartilagens aritenóides estão firmemente unidas, de tal forma que as cordas vocais vibram apenas no lado oposto, gerando uma Mellodia baixa" (Ladefoged, 1975), impossibilitando o cálculo de sua média. A título de exemplo, ilustramos uma dessas ocorrências.

Exemplo 3.5 (bfammn01) (CD áudio 3.5)
*MAI: [22] uma /=SCA= tora /=COM= da madeira //=APC=$

Figura 3.3 – Exemplo de Creaky Voice

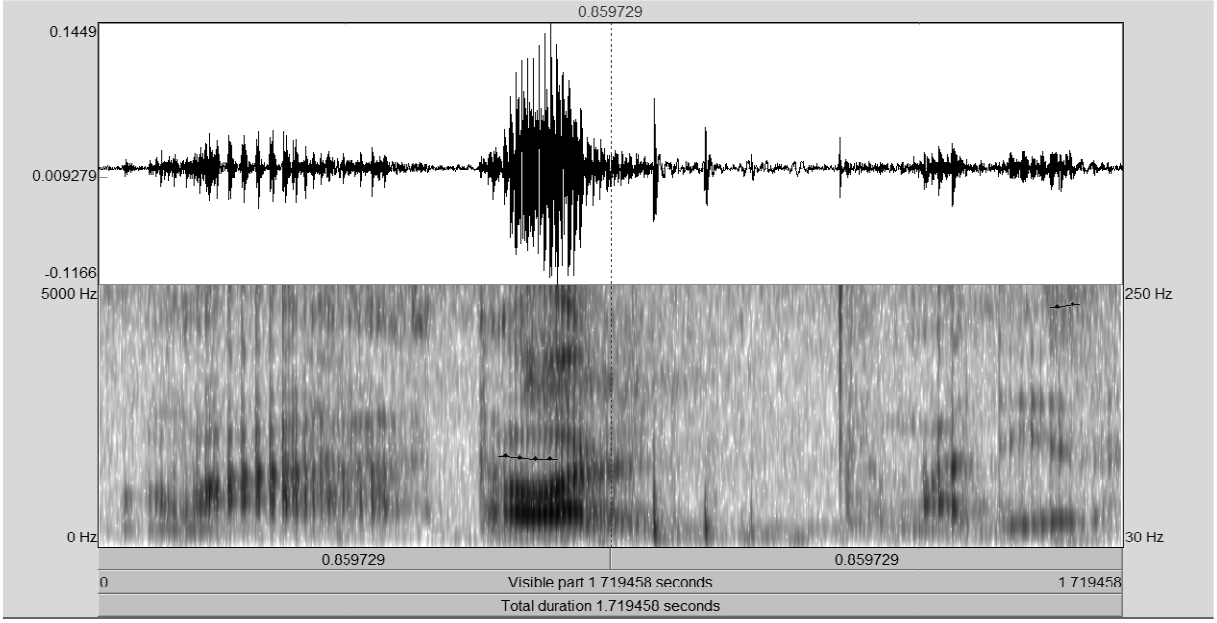
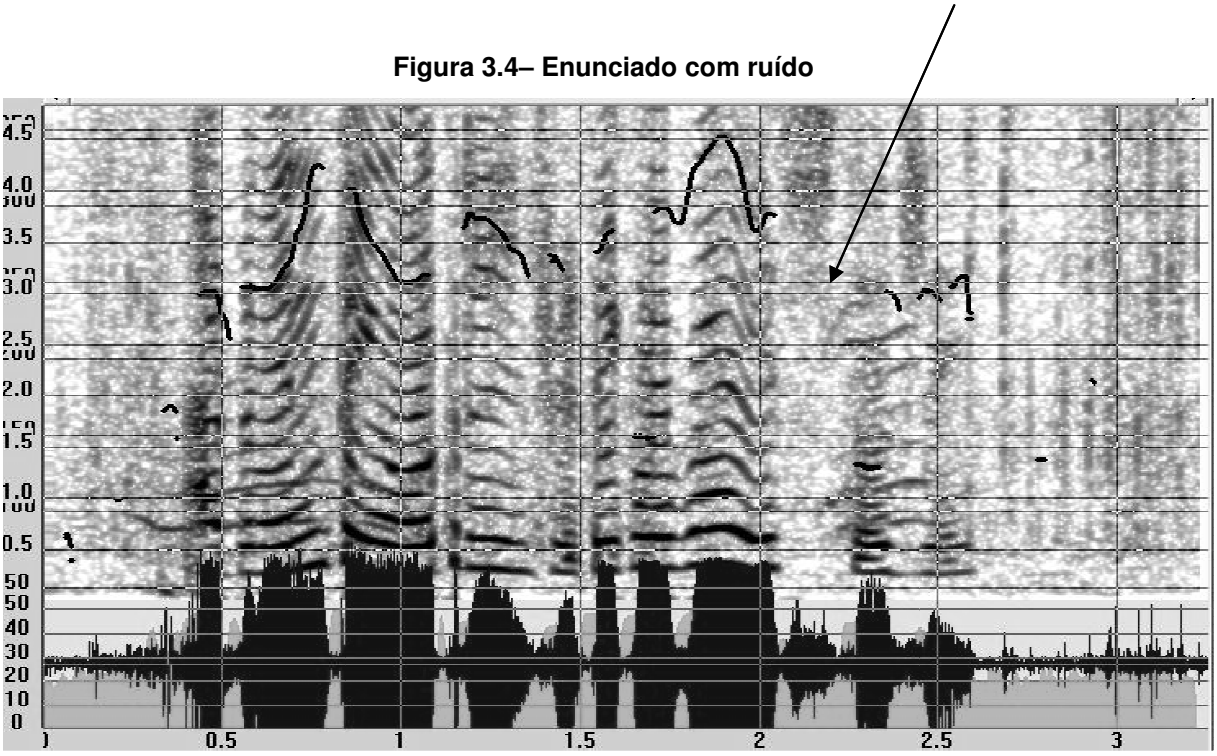


Figura 3.4– Enunciado com ruído



Conforme prevê a Teoria da Língua em Ato, a unidade de APC é fundamentada a partir de três critérios: funcional, entonacional e distribucional. Assim sendo, em primeiro lugar procuraram-se todas as unidades que realizavam uma mera integração da unidade de COM (critério funcional), depois que se localizasse após a unidade de COM (critério distribucional).

Quanto ao critério funcional, procuramos classificar a unidade de APC, conforme proposto por Tucci, E (2006), em que são categorizadas em repetições de expressões do tema, preenchimento, retomada textual e informação tardia.

Com base nessa classificação, expandimos duas categorias: a repetição e a informação tardia. As repetições foram subdivididas em literal, parcial e com variação. A informação tardia foi fracionada como somente interpretável com composicionalidade sintática, com possível ou não interpretabilidade sintática e com reformulação. Todos os critérios para essa reclassificação, bem como os exemplos serão apresentados no capítulo 5.

Para a análise entonacional fez-se uso de dois softwares Winpitch (Martin, 2004) e Praat[49] (Boesma; Weenink, 2011). O *Winpitch* é um analisador do sinal da fala, capaz de proporcionar ao pesquisador uma análise em tempo real de frequência dos formantes e dos parâmetros prosódicos - Fo, intensidade e duração. Fornece espectrogramas de alta resolução coloridos e monocromáticos e marcadores de tempo que facilitam a segmentação do sinal. Esse programa, ainda, permite que se selecione, de uma frase, um único segmento, possibilitando ouvi-lo reiteradas vezes, fornecendo, detalhadamente, suas qualidades físicas. Todo o subcorpus teve o som alinhado ao texto por meio desta ferramenta, bem como:

a) corte de trechos do sinal de áudio, a fim de verificar como se combinam Mellodicamente unidades prosódicas não adjacentes;
b) alteração do movimento da F0, com o objetivo de apurar se dado movimento em questão é relevante para a atribuição funcional da unidade prosódica;

Através do WinPitch foram extraídos os enunciados complexos que continham APC.

As médias relativas a F0 e intensidade dos APC e dos COM que antecedem a unidade de APC foram calculadas por meio do Praat. Nos casos em que havia dúvida quanto à classificação da unidade de APC, foi utilizada a ferramenta de Síntese do software WinPitch. A síntese visou à verificação do movimento, buscando saber se se tratava de um foco ou se era apenas um movimento não perceptualmente relevante. A título de exemplo, ilustramos um enunciado em que ora não foi feita a síntese, ora em que esse procedimento foi utilizado. A síntese, como se poderá notar por meio dos áudios, nos permitiu a alteração do padrão prosódico da unidade de APC, com vistas a garantir a certificação dessa unidade.

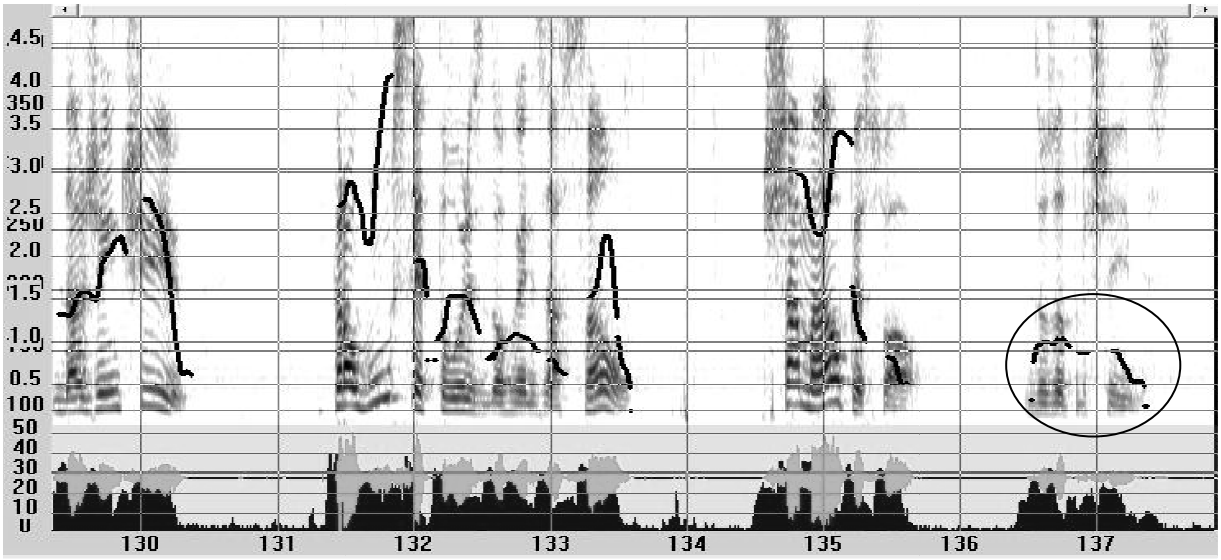
A elipse demarca a unidade de APC.

[49] O Praat (Boersma; Weenink) é um programa destinado a pesquisas em Fonética e pode ser baixado gratuitamente em http://www.praat.org/.

Exemplo 3.6 (bfamcv01) (áudio 3.6)

GIL: da gente /=TOP= pra alguns setores da organização /=TOP= chamar o pessoal /=COM= **dos outro times** //=APC=$

Figura 3.5– Enunciado sem a realização da síntese.

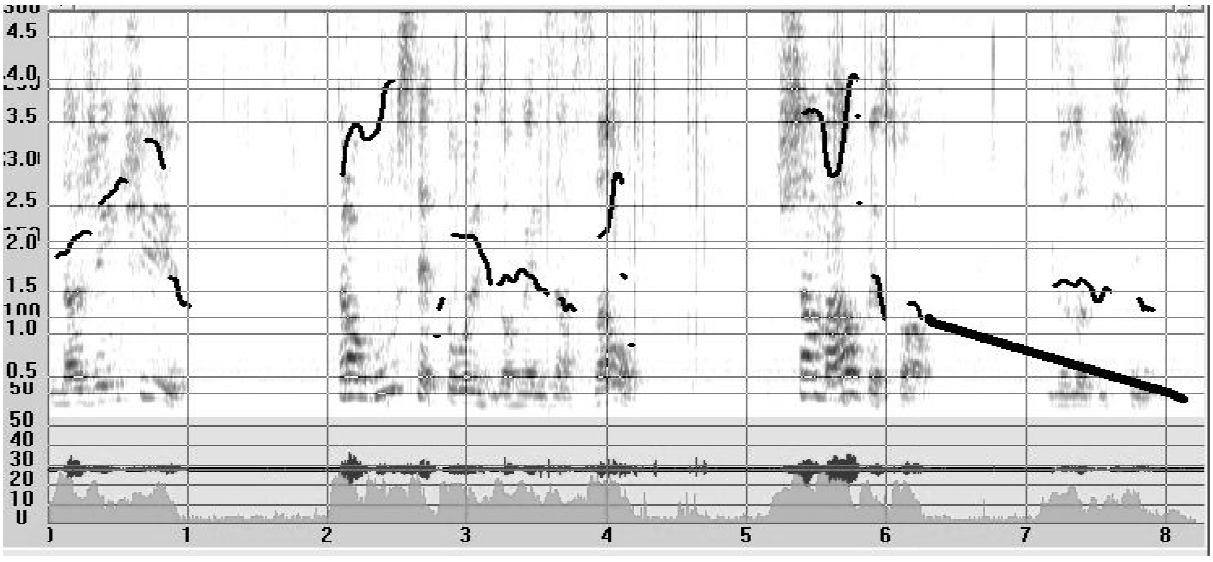


Exemplo 3.7 (bfamcv01) (áudio 3.7)

GIL: da gente /=TOP= pra alguns setores da organização /=TOP= chamar o pessoal /=COM= **dos outro times** //=APC=$

A figura 3.6, abaixo, representa o exemplo 2.25 depois de sintetizado.

Figura 3.6 – Enunciado com a realização da síntese.



As manipulações foram avaliadas com base nos resultados que produzem na percepção auditiva e na interpretação cognitiva da unidade tonal manipulada. A percepção auditiva diz respeito a capacidade de reconhecer diferenças acústicas entre a unidade tonal original e a unidade manipulada com base na percepção de suas características prosódicas. As categorias de analise referentes à percepção auditiva são: *idêntico*, *semelhante* e *diferente*, entendidas como um *continuum*.

Desse modo, *semelhante* e *diferente* não se sobrepõem, mas descrevem níveis de percepção da diferença. Segundo essa classificação, somente uma *close copy* (ou seja, a eliminação de movimentos não relevantes de F0) pode ser percebida como idêntica a unidade tonal original. A manipulação da unidade informacional de APC será sempre percebida como semelhante. A manipulação semelhante é aquela cujas diferenças acústicas são poucas e podem ser percebidas, sobretudo, se o falante sabe que o áudio foi manipulado. No segundo caso, as diferenças acústicas são mais evidentes e o falante percebe claramente a alteração.

A síntese também foi utilizada no processo de verificação das unidades de APC e COB. O Objetivo era saber se determinado movimento constituía um traço fonologicamente funcional ou não, a fim de discernirmos se se tratava de uma ou outra unidade.

Na análise ditribucional foram enumerados todos os APC que ocorriam depois de uma unidade ilocucionária (COM, CMM e COB). Além disso, verificou-se que há casos em que as unidades de ALL, PAR e PHA podem vir entre as unidades de COM e APC.

Além de todos os procedimentos acima citados, reconheceu-se que havia duas outras unidades (PAR e COB) que, dependendo do contexto, ocupavam a mesma posição da unidade de APC. A primeira tem como características a)poder ser retirada do enunciado sem que o mesmo, do ponto de vista prosódico, altere a realização completa do enunciado; b) possuir valor modal e c) funcionar como uma intervenção metalingüística. A segunda possui a) foco funcional; b) valor cognitivo de 'novo' e c) média de F0 e intensidade superiores à locução que o antecede.

Foram detectados, ainda, um único caso de APC consecutivo e outro de APC escansionado.

Além disso, outra situação foi mencionada, a coda. A coda ocorre quando há uma unidade tonal que parece ter todas as características prosódicas de um APC, mas com características informacionais distintas.

No capítulo 5, serão exemplificados todos os casos acima citados.

Por fim, classificaram-se as unidades de APC morfossintaticamente.

3.3 Uma comparação com o C-ORAL-ROM

Um dos objetivos dessa pesquisa é verificar como a unidade de APC se comporta no PB e no italiano. Assim sendo, faz-se necessário entendermos como se configurou a arquitetura do C-ORAL-ROM para depois estabelecermos um paralelo entre as duas línguas.

Conforme já fora mencionado, por ora o C-ORAL-BRASIL apresenta somente a metade informal do *corpus*, podendo assim ser comparável unicamente com a metade correspondente dos *corpora* C-ORAL-ROM. Contudo, a metade informal do C-ORAL-BRASIL é maior e melhor balanceada do que a parte correspondente dos *corpora* do C-ORAL-ROM. A tabela 3.7 mostra a comparação entre o C-ORAL-BRASIL e a parte informal dos *corpora* C-ORAL-ROM, relativo ao italiano, quanto às dimensões enunciados e palavras.

Tabela 3.7 – Comparação entre os dados do italiano e do PB quanto ao número de enunciados e palavras

	Enunciados	Palavras
PB	34.167	208.130
Italiano	23805	154.967

Fonte: Raso (2012) - adaptada

A tabela mostra que o tamanho do C-ORAL-BRASIL ultrapassa amplamente aquele do C-ORAL-ROM, com variações de porcentagens entre uma língua e outra e do parâmetro de comparação. O aumento com relação ao número de textos é devido a um aumento das dimensões do *corpus*, mas também ao menor uso de textos de tamanho superior a 1.500 palavras, de maneira a melhor preservar a diversificação diafásica.

O grande aumento no número de enunciados (mais do que 40% a mais do que o italiano) também tem várias causas: o aumento das dimensões do *corpus* pode explicar a diferença com relação ao italiano; De fato, os textos mais interativos e nos quais os informantes estão realizando atividades diferentes da simples ação de falar tendem a apresentar mais enunciados simples do que textos em que duas ou mais pessoas discutem sobre um tópico. Nesses últimos casos, é mais fácil que a interatividade se dilua em turnos mais longos e enunciados mais complexos. Esse aspecto pode facilmente ser verificado investigando-se as situações comunicativas dos vários *corpora*.

A diversificação diafásica, que é um aspecto crucial para o *corpus* italiano e ainda mais para o C-ORAL-BRASIL, vem representada na tabela abaixo.

Tabela 3.8 – Comparação entre os dados do italiano e do PB quanto aos contextos

	famdl	%	famcv	%	fammn	%	Total
PB	55.361	34,73%	51.887	32,55%	52.116	32,70%	159.364
Italiano	37.345	29,02%	46.119	35,84%	45.212	35,13%	128.676

Fonte: Raso (2012) - adaptada

Faz-se necessário lembrar que a regra para a constituição do C-ORAL-BRASIL é de que haja uma proporção tendencialmente de 2/3 para as tipologias interativas (conversações e diálogos) e de 1/3 para a tipologia monológica. Considerando-se isso, pode-se notar que o italiano respeita essa proporção, quando comparado ao PB. Uma proporção igual entre conversações e diálogos é, contudo, considerada desejável. Nesse aspecto, nenhuma língua se aproxima tanto quanto o PB de uma representatividade paritária, já que o italiano privilegia as conversações. Entre as línguas do C-ORAL-ROM, aquela que mais se aproxima de uma proporção ideal é o italiano.

Além da proporção em palavras, é útil considerar a proporção em número de textos, pois esse número diz algo a respeito da variação de situações, falantes e tópicos representados. No PB temos 34 conversações, 35 diálogos e 36 monólogos, com uma proporção quase perfeita. O italiano, o mais equilibrado dentro do C-ORAL-ROM, apresenta uma proporção de 28 conversações, 20 diálogos e 25 monólogos.

Tabela 3.9 – Comparação entre os dados do italiano e do PB quanto ao contexto público

	pubdl	%	pubcv	%	pubmn	%	Total
PB	17.997	36,90%	14.547	29,83%	16.222	33,26%	48.766
Italiano	9.995	38,01%	10.246	38,97%	6.050	23,01%	26.291

Fonte: Raso (2012) - adaptada

Olhando para os dados relativos ao contexto público, observamos que a proporção mais importante (conversações + diálogos *versus* monólogos) é respeitada completamente só em PB; o italiano possue uma quantidade baixa de palavras nos monólogos. Quanto ao número de textos, o PB apresenta 9 conversações, 11 diálogos e 14 monólogos. No C-ORAL-ROM, o italiano, 5 conversações, 5 diálogos e 4 monólogos.

As estatísticas relativas aos metadados presentes no C-ORAL-ROM não permitem documentar estatísticamente a distribuição dos falantes por nível escolar; uma

sondagem, sem valor estatístico, feita em 30 textos de italiano, mostrou que a presença de falantes com alto nível escolar é especialmente representada em italiano. Contudo, parece que a distribuição nos vários níveis de escolarização seja mais equilibrada em PB. O fato de que os níveis escolares médio-alto ou alto sejam mais representados é, a nosso ver e dentro de certos limites, uma escolha desejável para documentar o padrão linguístico sincronicamente (RASO 2012, p. 30).

Quanto à diatopia dos informantes, a comparação é mais difícil porque os dados das partes informal e formal não são desagregados nas tabelas do C-ORAL-ROM. De fato, a quantidade de informantes que não possuem a documentação da origem aumenta muito na parte formal dentro da ramificação de *media*, que inevitavelmente é menos controlável. Contudo, mesmo considerando-se que a comparação sofra limite, o *corpus* C-ORAL-BRASIL apresenta uma porcentagem de dispersão diatópica extremamente baixa, menor do que no italiano.

Podemos dizer, então, que a comparabilidade entre o *corpus* C-ORAL-BRASIL é garantida dentro dos limites do que foi dito aqui. Sem dúvida pode-se dizer que o *corpus* de italiano é o mais comparável ao C-ORAL-BRASIL, principalmente pela riqueza na variação diafásica, mas também por outras características apontadas. O capítulo 5 apresentará uma comparação entre PB e italiano de todas as medidas referentes à unidade de APC.

3.4 Resumindo

Esse capítulo procurou justificar o estudo da fala espontânea por meio da perspectiva da Linguística de Corpus. Além disso, apresentou os procedimentos de construção e validação do projeto C-ORAL-BRASIL e ilustrou-se a prospecção dos dados e análise estatística por meio do Programa R, realizado para toda a amostra. Estabeleceu, ainda, uma comparação entre os dados do C-ORAL-BRASIL e do C-ORAL-ROM no que se refere ao italiano.

Capítulo 4

CARACTERÍSTICAS DA FALA ESPONTÂNEA

Esse capítulo esboçará as medidas que caracterizam a amostra. Estas medidas se subdividem em termos da distribuição de turnos, enunciados e unidades tonais em relação a cada tipo interacional da amostra. Além disso, fundamentam a análise dos dados especificamente relacionados ao Apêndice de Comentário, a serem apresentados no capitulo 5.

4.1 Unidades de referência na fala espontânea

A fala espontânea pode ser dividia em turnos dialógicos, enunciados e unidades tonais, conforme o escopo teórico que sustenta essa pesquisa. As primeiras medidas de nossa amostra a serem observadas serão a distribuição dos turnos dialógicos, do número de enunciados e as unidades tonais. As médias de enunciados por turno e de unidade tonal por enunciado permitem avaliar o grau de interatividade presente no texto. São valores que expressam bem as diferenças entre as tipologias da amostra. A média de enunciados por turno é uma medida que indica a freqüência de alternância de turnos na interação: quanto maior a média, menor a alternância de turnos e menor a interatividade. A média de unidades tonais por enunciado é um indicador do tamanho e complexidade dos enunciados do texto. Quanto maior a média de unidades tonais por enunciado, maior a complexidade dos enunciados do texto, indicando menor interatividade, já que a acionalidade expressa pelo COM se contextualiza em maior número de unidade não ilocucionárias. Conversações e diálogos tendem a apresentar médias menores do que monólogos (RASO; MITTMANN, 2012).

A tabela 4.1 ilustrará as características gerais da amostra. Serão apresentados os valores para cada texto, bem como os valores acumulados por tipologia interacional. Os valores médios de enunciados por turno, calculados com base apenas nos

enunciados concluídos, e os valores médios de unidades tonais por enunciado, calculado com base apenas nas unidades tonais concluídas, também serão esboçados.

Tabela 4.1 – Características gerais da amostra

Texto	Turnos dialógicos	Enunciados concluídos	Enunciados por turno (média)	Interrupções	Retratações	Unidades Tonais	Unidades tonais por enunciado (média)
Conversações	**1333**	**1848**	**1,39**	**191**	**253**	**3164**	**1,71**
bfamcv01	159	307	1,3	41	46	441	2,13
bfamcv02	239	256	1,49	29	36	579	1,63
bfamcv03	185	296	1,6	10	38	467	1,58
bfamcv04	323	422	1,31	43	28	645	1,53
bpubcv01	265	323	1,22	32	35	611	1,89
bpubcv01	162	244	1,51	36	70	421	1,73
Diálogos	**1371**	**2275**	**1,66**	**176**	**228**	**3513**	**1,54**
bfamdl01	338	542	1,6	24	19	781	1,44
bfamdl02	176	247	1,4	35	56	453	1,83
bfamdl03	172	300	1,74	38	41	505	1,68
bfamdl04	123	244	1,98	9	11	367	1,5
bfamdl05	239	391	1,64	40	46	566	1,45
bpubdl01	158	262	1,66	14	32	407	1,55
bpubdl02	165	289	1,75	16	23	434	1,5
Monólogos	**250**	**920**	**3,68**	**74**	**388**	**2707**	**2,94**
bfammn01	19	98	5,16	8	70	245	2,5
bfammn02	95	171	1,8	13	57	477	2,79
bfammn03	48	135	2,81	9	59	353	2,61
bfammn04	26	181	6,96	8	21	446	2,46
bfammn05	31	135	4,35	18	56	401	2,97
bfammn06	6	72	12	4	47	328	4,56
bpubmn01	25	128	5,12	14	78	457	3,57
TOTAL	**2954**	**5043**	**1,71**	**441**	**869**	**9384**	**1,86**

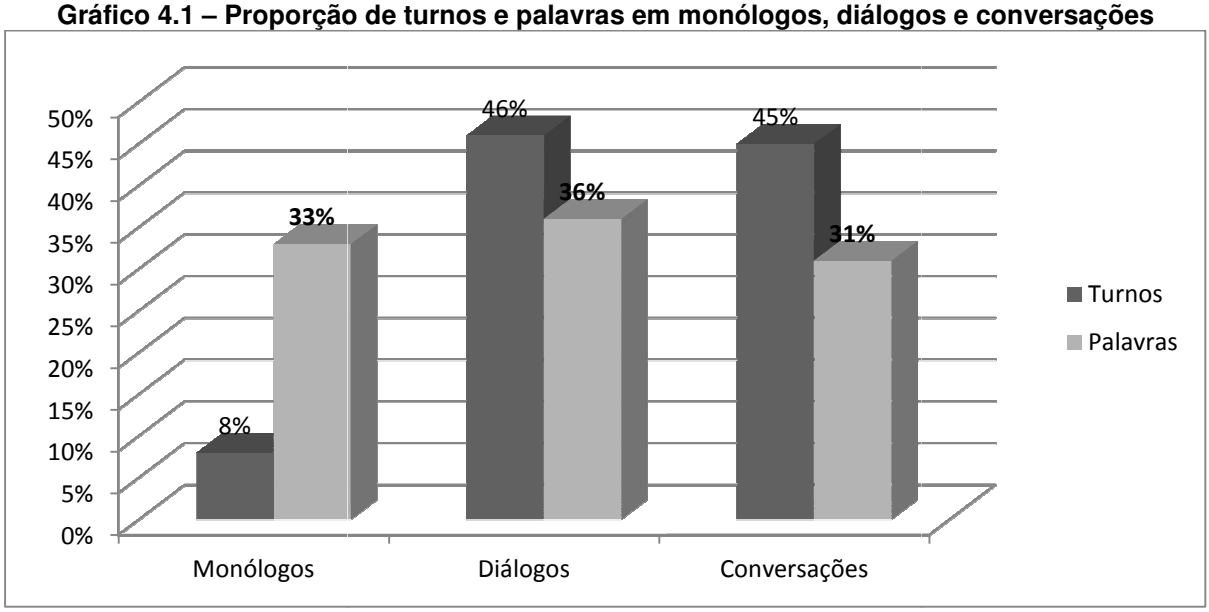
Fonte: RASO (2012)

Comparando as tipologias textuais, com base na tabela acima, podemos afirmar que as conversações e os diálogos estão mais próximos entre si do que dos monólogos. Enquanto a média de enunciados por turno nas conversações é de 1,39 e nos diálogos 1,66, nos monólogos é de 3,68.

Quanto à complexidade dos enunciados, também se nota que diálogos e conversações se opõem aos monólogos. Nas conversações os enunciados têm em média 1,71 unidades tonais, nos diálogos, 1,54; já os monólogos possuem em média 2,94 unidades tonais por enunciado.

Além dessas medidas, o número de palavras em relação à proporção de turnos de cada tipologia também foi comparado, conforme ilustra o gráfico abaixo.

Gráfico 4.1 – Proporção de turnos e palavras em monólogos, diálogos e conversações



Com base nesse gráfico pode-se observar que os monólogos representam 33% das palavras do subcorpus e estão distribuídas em apenas 8% dos turnos dialógicos totais. As conversações e diálogos perfazem 67% do corpus em número de palavras, que estão distribuídas em 92% dos turnos. As conversações apresentam 45% dos turnos e 31% das palavras, enquanto os diálogos têm 46% dos turnos e 36% das palavras.

Como se pode notar, os diálogos tendem a apresentar um número um pouco maior de turnos e de palavras do que as conversações pois, com menos falantes, ocorrem menos sobreposições e interrupções. A competição pelo turno é menor em iterações um a um do que em casos em que mais do que dois falantes interagem. Estes números apontam para semelhanças estruturais entre diálogos e conversações de um lado, em contraste com os monólogos de outro. *São um primeiro indicador que nos permite tratar os dados em termos de uma oposição entre interações dialógicas (conversações e diálogos) versus monológicas* (RASO; MITTMANN, 2012).

Agregam-se a estes valores a média de palavras por enunciado, considerando o total de palavras contáveis de cada texto, dividido pelo total de enunciados concluídos mais as interrupções e obtiveram-se os seguintes resultados: os diálogos como as conversações apresentam uma média de menos de 6 palavras por enunciado e os monólogos apresentam 8 ou mais palavras. A média geral das tipologias, é de 4,79 palavras por enunciado para as conversações, 4,62 palavras por enunciado para os diálogos e 10,27 palavras por enunciado para os monólogos. Como se pode notar, os monólogos apresentam o dobro da média de conversações e diálogos.

Na tipologia conversação há um texto 'bfamcv01' cujas características diferenciam-se dos outros textos que compõem esse grupo. É um texto pouco ancorado no contexto situacional imediato, por isso há a necessidade de os falantes se situarem linguisticamente.

Outras características desse texto são a presença de enunciados mais complexos e longos, possuir uma média de unidades tonais por enunciado de 2,13 e ainda uma maior média de palavras por enunciado.

Abaixo um trecho que ilustra esta situação.

Exemplo 5.1 – bfamcv01
*EVN: acho <que a gente> tem / que olhar direito // eu falei isso naquea reunião lá / &he / &he / la do [/1] na casa do Artur / que tava o Juninho / nũ sei que / ai eu fiz um [/1] uma comparação <com aulas de [/1] de língua hhh> //
*LUI: <eu acho que / o nome / do> [/1] do [/1] do [/1] do [/1] do [/1] do [/1] da modalidade esportiva que a gente criou / que foi Futebol Arte / ficou / queimado / por causa dessa taca / pra mim e uma <coisa que> <e [/1] foi triste> / e triste> falar / mas / cê fala que e Futebol Arte / a galera começa a zoar / já //
*GIL: <e> // todo mundo <encheu o saco> / <por causa disso> //
*LEO: "zoar" em que sentido //
*LUI: no sentido de que / e tosco //
*EVN: ah / que arte o que / umas coisa assim //

Com o intuito de estabelecer uma analogia entre um texto ancorado em contexto extralinguístico imediato e outro não. Apresentam-se um exemplo, que o reproduzimos abaixo, com o mesmo número de turnos e o comparamos ao exemplo 5.1.

Exemplo 5.2 – bfamcv04

*CEL: a gente <devia ter jogado e Uno> //
*BRU: <zerinho ou um> //
*CEL: <Uno> e mais legal jogar //
*LUC: Uno e sensacional / bicho //
*CEL: <No'> // que burrice / hein //
*HEL: <o que> //

Como se pode notar, o exemplo 5.1 reproduz uma situação fortemente situada e ancorada no contexto extralinguístico imediato, pois trata-se de uma conversa entre rapazes sobre um campeonato de futebol de salão. O exemplo 5.2, todavia, não necessita de um contexto imediato. Todo texto ancorado em contexto extralinguístico imediato tende a apresentar uma grande quantidade e variedade de atos de fala. Ao contrário, um texto pouco ancorado no contexto tende a apresentar elaboração textual mais complexa, pois ha necessidade de preencher linguisticamente aquilo que não tem como ser interpretado com base no contexto imediato.

Dessa forma podemos dizer que a narrativa tende a ter pouca acionalidade, pois a maioria dos enunciados repetem os mesmos atos de fala; isto é, enquanto que em um texto de grande acionalidade são cumpridas várias e diferentes ações (cada enunciado carrega uma ilocução diferente), em um texto de baixa acionalidade pode-se dizer que, de modo geral, um conjunto pequeno e restrito de ações é realizada.

Pode-se dizer, ainda, que as tipologias de textos falados se dividem, em realidade, em duas categorias principais: dialógicos e monológicos, pois estas categorias não

dependem exclusivamente do número de participantes de uma interação, mas sim da situação comunicativa.

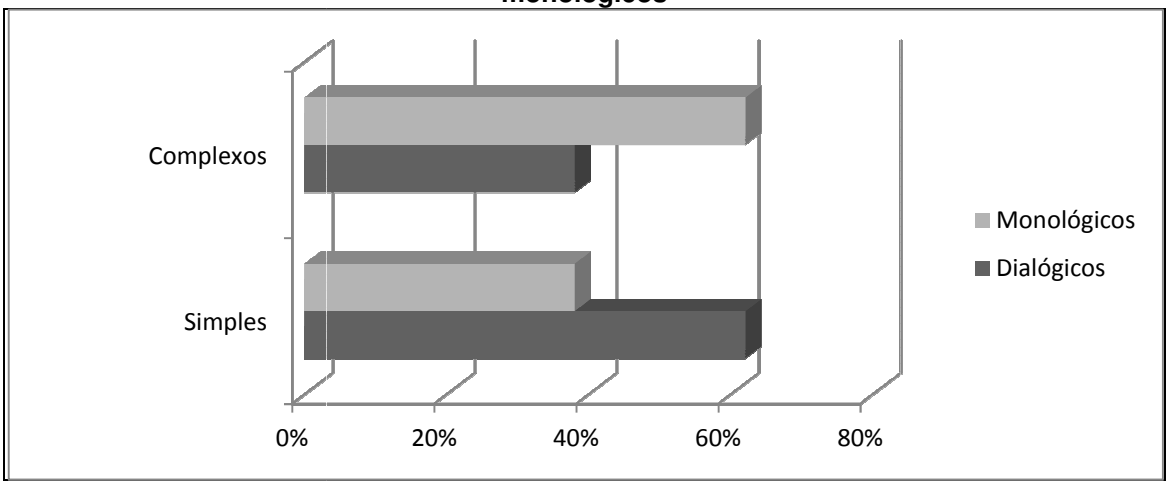
4.2 Estrutura dos enunciados

Os enunciados, segundo a Teoria da Língua em Ato, podem ser classificados como simples e complexos. Se simples, são compostos por apenas uma unidade prosódica, se complexos, constituem-se de mais de uma unidade prosódica. Outra medida é a distribuição das unidades tidas como concluídas. Elas são classificadas em unidades terminadas, padrões ilocucionários e estrofes.

A diferença quanto à estrutura, nos textos monológicos e dialógicos diz respeito à extensão das unidades concluídas e ao número de sequências complexas. Os textos monológicos são compostos por unidades concluídas mais longas, bem como por um número maior de sequências complexas; nos textos dialógicos ocorre o inverso.

O gráfico 4.2 ilustra a proporção de enunciados simples e complexos de cada grupo.

Gráfico 4.2 – Proporção de enunciados simples e complexos em textos dialógicos e monológicos



O gráfico evidencia o contraste entre os enunciados das tipologias diálogo e monólogo. Enquanto nos monólogos encontram-se 62% de enunciados complexos e 38% de enunciados simples, nos diálogos há exatamente o oposto; isto é, 38% de enunciados complexos e 62% de enunciados simples.

O fato de haver nos textos monológicos vários enunciados simples do tipo “hum hum //” (ou equivalentes) por parte do interlocutor, simplesmente representam um sinal de acompanhamento de quem escuta o texto. Se considerarmos somente os enunciados simples preenchidos por “hum hum”, verifica-se que os mesmos representam 11% de todos os enunciados simples nos textos monológicos, contra apenas 2% nos dialógicos.

Podemos, então, enumerar as seguintes características, em relação às tipologias analisadas.

1) os textos monológicos apresentam uma incidência bem menor de enunciados simples do que os dialógicos. Ao contrario, a porcentagem de enunciados complexos é superior nos monólogos. Isso se justifica porque é menor o número de enunciados em textos monológicos do que os dialógicos (considerando-se, naturalmente, que os textos de ambas as tipologias tem tamanhos (em numero de palavras) comparáveis).
2) observa-se que cada enunciado do tipo dialógico possui uma média de 1,62 unidades tonais, enquanto nos monológicos a média é de 2,94 unidades tonais por enunciado, considerando-se apenas as unidades concluídas. Ao verificar a média de unidades tonais considerando apenas o universo dos enunciados complexos, a diferenca aumenta: enunciados complexos de textos dialógicos têm em média 2,49 unidades tonais, enquanto os enunciados complexos dos textos monológicos têm uma média de 3,9.
3) proporcionalmente, a média de unidades tonais por enunciados em textos dialógicos é equivalente a 56% da média dos monológicos, ou seja, a tipologia monológica tem pouco mais que o dobro de unidades tonais por enunciado que a dialógica (em média). Se apenas os enunciados complexos de cada tipologia forem considerados, vê-se que a média de unidades tonais por enunciado complexo nas tipologias dialógicas é 64% da média dos textos

monológicos. Mesmo considerando apenas enunciados com mais de uma unidade tonal, a média dos monológicos ainda supera a dos dialógicos. Além disso, a diminuição da diferença entre as tipologias explica-se devido a diferença quantitativa de enunciados simples que existe entre elas.

4) em textos dialógicos, as unidades informacionais de auxilios dialógicos tem uma participação mais importante na constituição de enunciados complexos do que nos textos monológicos. Este resultado indica que a complexidade textual (devida a presenca de unidades informacionais textuais) é mais recorrente em textos monológicos, e é um produto da baixa acionalidade presente em tal tipologia.

Quanto aos enunciados, diferenciam-se não somente pelo fato de serem simples ou complexos, mas podem se distinguir quanto às unidades terminadas complexas da seguinte maneira: i) unidades terminadas complexas formadas por uma unidade de COM ou CMM e uma ou mais unidades dialógicas; ii) Unidades terminadas complexas porque contém pelo menos uma unidades informacional textual , podendo ou não apresentar também unidades dialógicas e iii) unidades terminadas complexas porque contém, pelo menos, dois ou mais comentários em relação não padronizada (COB), podendo ou não apresentar outras unidades textuais e/ou dialógicas).

Ainda tratando de complexidade textual, o primeiro tipo de enunciado complexo enumerado acima deve ser considerado parecido com os enunciados simples, porque as unidades de auxilio dialógico (fáticos, conativos, alocutivos e expressivos) são dirigidas ao interlocutor e não contribuem para a composição do texto propriamente dito, então estas unidades não tornam a estruturação textual do enunciado mais complexa.

A distribuição dos enunciados complexos do tipo (COM/CMM+AUX), em termos de proporção relativa em diálogos e monólogos é semelhante a distribuição dos enunciados simples. Se observamos a incidência desse tipo de enunciado complexo sobre o total de enunciados complexos, observamos que nos dialógicos ela é mais

que o dobro (28%) do que nos monológicos (12%), conforme se pode observar na tabela 4.2.

Tabela 4.2 – Distribuição de enunciados e padrões ilocucionários complexos com apenas unidades dialógicas, proporções em relação à frequência absoluta e ao total de enunciados complexos

Tipologias	Complexos com somente AUX	%	Complexos totais	Complexos com somente AUX em relação a todos os complexos
DIALÓGICOS	481	87	1717	28
Conversações	219	40	821	27
Diálogos	262	47	896	29
MONOLÓGICOS	71	13	617	12
TOTAL	**552**	**100**	**2334**	**24**

Fonte: MITTMANN, 2011 (em preparação)

Fica evidente nessa tabela que 87% dos enunciados complexos com apenas unidades dialógicas estão nos diálogos. Assim sendo, pode-se afirmar que uma das características dos textos dialógicos é apresentar uma alta porcentagem de unidades dialógicas que cumpre uma função pouco significativa em textos cujo contexto seja pouco interativo, como ocorre nos monológicos.

Faz-se necessário, entretanto, afirmar que um enunciado complexo que além da unidade ilocucionária contém somente um auxilio dialógico não pode ser considerado como um enunciado de estrutura informacional complexa.

4.3 Resumindo

Nesse capítulo viu-se que a fala espontânea pode ser dividia em turnos dialógicos, enunciados e unidades tonais, conforme a Teoria da Língua em Ato (CRESTI, 2000).

Observou-se que os monólogos representam 33% das palavras do subcorpus e estão distribuídas em apenas 8% dos turnos dialógicos totais. As conversações e diálogos perfazem 67% do corpus em número de palavras, que estão distribuídas

em 92% dos turnos. As conversações apresentam 45% dos turnos e 31% das palavras, enquanto os diálogos têm 46% dos turnos e 36% das palavras.

Notou-se, também, que os diálogos tendem a apresentar um número um pouco maior de turnos e de palavras do que as conversações pois, com menos falantes, ocorrem menos sobreposições e interrupções. A competição pelo turno é menor em iterações um a um do que em casos em que mais do que dois falantes interagem. Estes números apontam para semelhanças estruturais entre diálogos e conversações de um lado, em contraste com os monólogos de outro.

Capítulo 5

Análise dos dados relativos à unidade de APC no PB e no Italiano

Este capítulo tem como escopo os resultados referentes à análise da unidade informacional de APC na fala espontânea brasileira, mais especificamente na variedade mineira. É nosso objetivo expor os resultados da distribuição da amostra e apresentar as análises informacional, entonacional e morfossintática dessa unidade. Além disso, pretende-se estabelecer uma comparação entre os resultados do subcorpus brasileiro e o italiano, publicado pelo Lablita em 2011.

5.1 A distribuição dos APC na amostra

A unidade de APC é uma unidade informacional de tipo textual. Isto significa que a unidade de APC compõe o texto falado, diferentemente das unidades com função dialógica, cujo objetivo é regular a interação. Ainda que seja uma unidade textual, sua função é de apenas integrar a unidade de COM, já que não serve de âmbito para aplicação da força ilocucionária como a unidade de TOP, não possui função metalinguística como o PAR, nem introduz uma metailocução como o INT, tampouco possui autonomia pragmática como a unidade de COM.

A tabela 5.1 tem por objetivo ilustrar a proporção total de APC presentes na amostra em comparação às demais unidades textuais: tópico, apêndice de tópico, parentético e introdutor locutivo.

Tabela 5.1 – Frequência da unidade APC em comparação às demais unidades textuais presentes na amostra (adaptada Mittmann , em preparação)

Unidades Textuais	Quantidade	Percentual
Apêndice de Comentário (APC)	112	10%
Tópico (TOP)	604	53%
Apêndice de Tópico (APT)	23	2%
Parentético (PAR)	158	14%
Introdutor Locutivo (INT)	236	21%
TOTAL	1136	100%

Comparando as frequências relativas, a unidade de APC ocupa o quarto lugar de ocorrência no subcorpus brasileiro, com um percentual de 10% sobre o total de unidades textuais, o que nos permite afirmar não se tratar de uma unidade muito recorrente. A tabela 5.2 exibe uma lista com todos os textos da amostra, seguida da descrição da situação, do número de enunciados com APC, do número total de unidades terminadas e da proporção de enunciados com APC em relação ao total de enunciados de cada texto.

Tabela 5.2 – Distribuição e proporção de enunciados com APC e descrição das situações gravadas na amostra

Texto	Situação	Enunciados com APC	Total de unidades terminadas	Proporção de enunciados com APC por enunciado
		110	5484	**2%**
Conversações		32	2039	**2%**
bfamcv01	amigos avaliam um campeonato de futebol organizado por eles e planejam o próximo	09	248	**4%**
bfamcv02	senhoras conversam sobre os preparativos do casamento de uma parente	03	385	**1%**
bfamcv03	amigos jogam sinuca	03	306	**1%**
bfamcv04	amigos jogam "Imagem e Ação", após explicar as regras do jogo para uma das participantes	03	465	**1%**
bpubcv01	funcionários de banco de sangue	05	355	**1%**

	explicam como o sangue coletado é armazenado			
bpubcv02	reunião ordinária em uma sede regional de partido político	09	280	**3%**
Diálogos		37	2451	**1,5%**
bfamdl01	colegas de apartamento fazem as compras do mês*	05	566	**1%**
bfamdl02	colegas de faculdade batem papo enquanto organizam o material de gravação	06	282	**2%**
bfamdl03	casal faz uma viagem de carro	07	338	**2%**
bfamdl04	domésticas, mãe e filha, fazem a limpeza da cozinha após o almoço	06	253	**2%**
bfamdl05	corretor de imóveis leva a irmã para visitar apartamento	06	431	**1%**
bpubdl01	engenheiro e pedreiro trabalham em uma obra	01	276	**0,4%**
bpubdl02	cliente e vendedor interagem durante a compra de calçados	06	305	**2%**
Monólogos		41	994	**4%**
bfammn01	senhor narra história fantástica sobre uma cobra	16	106	**15%**
bfammn02	sobrinha de Carlos Drummond de Andrade conta histórias da família ao neto	06	184	**3%**
bfammn03	narrativa de "causos" divertidos para a família	04	144	**3%**
bfammn04	senhora conta sua experiência no hospital após ter dado à luz no carro*	08	189	**4%**
bfammn05	senhora fala sobre a adoção da filha após a	02	153	**1%**

	morte de sua filha biológica			
bfammn06	pai conta seu percurso profissional à sua filha	00	76	__
bpubmn01	entrevista de avaliação sobre aulas de inglês na rede pública de ensino	05	142	**3,5%**

Quanto à proporção geral de enunciados com APC em relação ao total de enunciados de todos os textos da amostra, o resultado é de 2%, o que significa dizer que a cada 100 unidades terminadas da amostra, aproximadamente 2 deles terão APC.

A proporção relativa às conversações é de 2%, nos diálogos de 1,5%, e nos monólogos 4%. Os monólogos têm mais que o dobro de ocorrências de APC se comparado às demais tipologias.

Observando em detalhe a proporção de cada texto, repara-se que na tipologia conversação os textos bfamcv01 e bpubcv02 apresentam uma proporção de enunciados com APC maior do que o geral dessa tipologia, bem como o texto bfammn01, nos monólogos. Nos diálogos, o texto bpubdl01 apresenta proporção inferior aos demais textos dessa tipologia. Os dois tipos de desvio podem ser explicados com base na situação comunicativa.

Como já demonstrado no capítulo 4 pelas medidas de números de enunciados por turno, número de enunciados complexos e média de unidade tonal por enunciado, os textos bfamcv01, bpubcv02 e bpubdl01 apresentam características de um texto de maior interatividade relativamente aos outros textos de mesma tipologia. Esta característica é dada pela situação comunicativa. Já o texto bfammn01 é um texto de baixa acionalidade (que é dada pela situação comunicativa) e também possui traços de um falante com baixa diastratia. Assim sendo, pode-se afirmar que quanto maior a interatividade, menor será o número de enunciados complexos com unidades textuais, e, consequentemente, um menor número de ocorrências de APC.

O primeiro texto, bfamcv01, é uma conversa entre amigos avaliando um campeonato de futebol organizado por eles, ao mesmo tempo em que planejam o que desejam fazer para o próximo campeonato. Esta situação leva os participantes a produzirem enunciados mais longos, cujo assunto não é situado em contexto imediato. Um enunciado representativo desta situação pode ser observado no exemplo 5.1.

Exemplo 5.1 (bfamcv01) (CD áudio 5.1)
*EVN: [54] <o´ /=CNT= o Arnaldinum> é caro /=CMM= e tem aqueles problemas //=CMM=$ [55] existem outros lugares /=INT= tão bons quanto o Arnaldinum /=CMM= e mais baratos //=CMM=$ [56] tem o SESC /=COB= que é bom pa caramba //=COM=$ [57] quadra é igual à do <Arnaldinum> //=COM=$
*LUI: [58] <e o> Sagrado //=COM=$
*GIL: [59] velho //=COM=$
*EVN: [60] Sagrado //=COM=$ [61] agora /=PHA= o Sagrado é mais caro /=i-COM= eu acho /=PAR= que o Arnaldinum //=COM=$
*GIL: [62] sabe que que eu penso /=COM= velho //=ALL=$ [63] <eu nũ sei> que cês acham //=COM=$
*EVN: [64] <não> //=COM=$
*GIL: [65] da gente /=TOP= pra alguns setores da organização /=TOP= chamar o pessoal /=COM= **dos outro times** //=APC=$

Nesse exemplo, a unidade de APC teve por objetivo adicionar uma informação que até o momento não havia sido proferida.

O segundo texto bpubcv02 mostra uma reunião entre os membros de um partido político, na sede desse mesmo partido.

Exemplo 5.2 (bpubcv02) (CD áudio 5.2)
*CAR: [24] então /=PHA= e tá ali ainda naquele [/1]=SCA= naquele fio ainda /=SCA= <ligado> /=SCA= contrário //=COM=$
*JEA: [25] <tá> //=COM=$
*OSV: [26] <tá> /=CMM= tá //=CMM=$

*JEA: [27] só <vieram fazer a leitura> //=COM=$
*OSV: [28] <mas ali nũ tem problema não> /=CMM= que ali é só extensão /=CMM= né <hhh> //=PHA=$
*CAR: [29] <n' é> não //=COM=$ [30] nũ é gato /=COM= **nem nada não** //=APC=$

Nesse caso, a unidade de APC realiza a expansão da unidade precedente sem repetir seu conteúdo semântico ou acrescentar informações.

O terceiro texto bpubdl01 retrata a conversa entre um engenheiro e um pedreiro durante a realização de uma obra.

Exemplo 5.3 (bpubdl01) (CD áudio 5.3)
*PAU: [45] porque +=EMP=$ [46] ah /=EXP= eu tenho uma aqui //=COM=$ [47] pequenininha //=COM=$ [48] então aqui /=TOP= quarenta centímetros +=COB=$
*ROG: [49] que ali e' ficou com quarenta-e-cinco //=COM=$
*PAU: [50] aonde /=CMM= que <ficou /=CMM= com quarenta-e-cinco> //=APC=$

A unidade de APC nesse exemplo apenas repete literalmente algo que já fora proferido no discurso.

O quarto texto bfammn01 é o mais representativo em relação ao número de ocorrências de APC. Esse é um texto em que o falante possui baixa diastratia e ao narrar uma história fantástica, marcas dessa variedade ficam evidentes. O exemplo 5.4 ilustrará essa afirmativa.

Exemplo 5.4 (bfammn01) (CD áudio 5.4)
*MAI: [85] na Amazonas tem muito dessa cobra aí //=COM=$[86] só ni lugar /=SCA= que tem /1 =SCA= tem mata mata muito /=SCA= forte /=COB= mata /=SCA= perigosa /=COB= que tem <esses tipo> de cobra /=COM= né //=PHA=$
*DUD: [87] <uhn> //=COM=$

*MAI: [88] no norte de Mina /=TOP= tinha esse /2 =SCA= antigamente /=PAR= tinha esse tipo de cobra tudo /=COM= né //=PHA=$[89] talvez agora já acabou /=COB= porque já desmataram muito /=COM= né //=PHA=$[90] nũ existe muita mata mais igual /1 =SCA= igual era /=COB= mas tem muita mata forte aí /=COB= nũ é igual essas /1 =SCA= essas capoeirim /=COB= cê fala assim /=INT= ah /=EXP_r= isso aqui é uma mata //=COM_r=$[91] isso aqui /=TOP= em vista de lá /=TOP= é /1 =EMP= é um cerrado /=COM= **aqueas mata lá** //=APC=$

O APC aqui faz referência a parte do discurso por meio de uma retomada textual, embora haja uma pequena reformulação.

É interessante destacar, também, que nesse texto encontramos o único enunciado da amostra composto por um APC duplo, assunto a ser detalhado na seção 5.2.3.3.

Outra medida importante para a análise da amostra é a ocorrência e distribuição de unidades terminadas simples, constituídas de apenas uma unidade prosódica (enunciados simples) e unidades terminadas complexas (enunciados complexos, enunciados com unidades textuais, enunciados com unidades dialógicas, padrão ilocucionário e estrofes) compostas de mais de uma unidade prosódica e sua relação com a tipologia do texto.

Segundo Cresti (2005b), a opção por uma estrutura simples ou complexa não está ligada à formalidade dos textos, mas sim à estrutura do evento comunicativo (tipologia interacional), que opõe diálógicos (informais e formais) a monólogos (informais e formais).

É possível aprofundar o raciocínio sobre a complexidade dos enunciados, pois é importante levar em consideração diferentes tipo de enunciados complexos.

Enunciados complexos são aqueles que apresentam mais de uma unidade tonal, e, portanto, mais de uma unidade informacional – a unidade de Comentário e uma ou

mais unidades textuais e/ou dialógicas. Por sua vez, os enunciados simples são aqueles que têm apenas uma unidade tonal/informacional, a unidade de Comentário. Assim, um enunciado complexo pode ser composto por:

a) comentário e uma ou mais unidades dialógicas;
b) comentário e uma ou mais unidades textuais;
c) comentário e uma ou mais unidades dialógicas em conjunto com uma ou mais unidades textuais.

Segundo Raso (2012), torna-se interessante desagregar os dados dos enunciados complexos e não tratá-los como uma única categoria, pois a natureza informacional dessa complexidade pode ter motivações diferentes nos enunciados de textos dialógicos e monológicos. Devemos tratar de maneira diferente os enunciados que são complexos porque há unidades que acrescentam informação ao conteúdo locutivo do enunciado (tipos b e c) daqueles que são complexos unicamente do ponto de vista dialógico, mas não informacional (tipo a).

De acordo com o C-ORAL-BRASIL, no grupo dialógico (diálogos e conversações) há 62% de enunciados simples, contra 38% de enunciados simples no monológico. A comparação entre enunciados complexos evidencia o oposto, isto é, 38% de enunciados simples e 62% de enunciados complexos nos monólogos.

A tabela 5.3 esboçará o número de unidades terminadas na amostra, excluíndo-se os enunciados simples e os enunciados complexos composto por uma ou mais unidades dialógicas.

Tabela 5.3 – Características Informacionais do subcorpus de PB

Tipologia Informacional	Conversação	Diálogo	Monólogo
Total de unidades terminadas	**1855**	**2304**	**950**
Total de enunciados	**1534**	**1972**	**633**
Total de enunciados simples	1095	1452	351
Total de unidades escansionadas simples (COM + SCA, TMT, EMP)	91	121	63
Total de enunciados complexos com unidades dialógicas	196	232	63
Total de enunciados complexos com unidades textuais	**108**	**125**	**100**
Junção de enunciados com COM + unidades textuais e dialógicas	**44**	**42**	**56**
Total de Padrão ilocucionário	**202**	**225**	**77**
Padrão ilocucionário simples	147	148	34
Padrão ilocucionário simples com unidades escansionadas	13	19	10
Padrão ilocucionário com unidades dialógicas	24	30	8
Total de padrão ilocucionário com unidades textuais	**14**	**20**	**21**
Junção de padrão ilocuionário + unidades textuais e dialógicas	**4**	**8**	**4**
Estrofes	**119**	**107**	**240**

Esses resultados nos permitem afirmar que há mais incidência de enunciados complexos nos textos monológicos do que nos diálogos e conversações, porque os monólogos possuem unidades terminadas mais longas, além de possuírem um maior número de sequências complexas. Com base nessas características, é de se esperar a maior incidência de APC nessa tipologia textual.

Faz-se necessário esclarecer, entretanto, que o número de enunciados complexos é superior nos monólogos, considerando-se que os textos de ambas as tipologias – monólogos versus (diálogos e conversações) têm tamanhos (em número de palavras, conforme visto no capítulo 4) comparáveis.

Tendo esboçado as características informacionais do subcorpus do PB, passamos a ilustrar as unidades terminadas com unidades textuais, padrões ilocucionários e estrofes que são aquelas onde podem haver incidência da unidade de APC.

Na tabela 5.4 será demonstrada a proporcionalidade entre os enunciados complexos da amostra e os enunciados complexos com APC.

Tabela 5.4 – comparação entre o total de enunciados complexos e total de enunciados complexos com APC

Tipologias	Total de enunciados complexos c/ unidades textuais	Total de enunciados complexos com APC
Conversação	152/1855 – 8,2%	25/152 – 16,4%
Diálogo	167/2304 - 7,2%	32/167 – 19,2%
Monólogo	156/950 – 16,4%	31/156 – 20%

Como se pode notar, o número de enunciados complexos com APC na amostra é maior nas conversações e diálogos juntas (35,6%) - pois apresentam características textuais semelhantes, isto é, o falante não tem um programa mental pré-construído, interagindo à medida que o discurso vai se construindo - , porque em ambas as tipologias, há quase o dobro de enunciados que os monólogos, conforme dito no capítulo 4.

Todavia, se desconsiderarmos essa medida veremos que são os monólogos os responsáveis pela maior incidência de enunciados complexos com APC, assim como ocorre nos enunciados complexos com unidades textuais da amostra. Esses dados corroboram o fato de que os monólogos são textual e informacionalmente mais complexos do que os diálogos.

Passemos à tabela 5.5 para a comparação entre o total de padrões ilocucionários da amostra e os padrões ilocucionários com APC.

Tabela 5.5 – comparação entre o total de padrões ilocucionários e padrões ilocucionários com APC

Tipologias	Total de padrões ilocucionários	Total de padrões ilocucionários c/ APC
Conversação	18/1855 – 1%	4/18 – 22,2%
Diálogo	28/2304 - 1,2 %	4/28 – 14,3%
Monólogo	25/950 – 2,6%	2/25 – 8%

Tendo por base a tabela 5.5, o que se verifica é que não há muitas ocorrências de padrões ilocucionários na amostra. Na verdade, sua maior incidência acontece nos monólogos, com 2,6% sobre o total analisado. Nos diálogos (conversações e diálogos) o percentual é de 2,2% sobre o total de padrões ilocucionários analisados.

Todavia, se formos analisar a incidência das unidades de APC nos padrões ilocucionários veremos que se somarmos os resultados das tipologias conversação e diálogo teremos 36,7% sobre o total de ocorrências de APC nos padrões ilocucionários, em detrimento aos 8% ocorridos nos monólogos.

Agora vejamos a incidência de APC, nas estrofes.

Tabela 5.6 – comparação entre o total de estrofes e estrofes com APC

Tipologias	Total de estrofes	Total de estrofes c/ APC
Conversação	119/1855 – 6,4%	2/119 – 1,7%
Diálogo	107/2304 - 4,6%	1/107 – 0,9%
Monólogo	240/950 – 25,3%	9/240 – 3,8%

Como os monólogos são textos pautados principalmente no desenvolvimento semântico de um pensamento verbalizado construído por um único falante, é de se esperar que haja uma maior incidência de estrofes nessa tipologia textual (25,3%). Logo após, mas com pouca frequência estão as conversações (6,4%) e os diálogos (4,6%).

Quanto à incidência das unidades de APC nas três tipologias, o resultado é similar; isto é, há 3,8% de ocorrências de APC nos monólogos, dada a complexidade informacional dessa tipologia. Isso porque, embora o falante de um monólogo possa

ter certo grau de interação com um interlocutor, sua principal atividade é a execução processual do pensamento durante a interação.
As conversações e diálogos representam juntas 2,6 % sobre o total de estrofes com APC.

5.2 Análises Funcional, Entonacional e Distribucional

Em conformidade com a Teoria da Língua em Ato, há três critérios que devem ser respeitados ao se analisar uma unidade informacional: o critério funcional, o entonacional e o distribucional. Nas subseções que se seguem, cada um desses critérios será analisado a partir da unidade de APC.

5.2.1 Critério Funcional

Segundo a Teoria da Língua em Ato, funcionalmente a unidade de APC realiza uma mera compilação de um texto, devendo estar posicionada após a unidade da qual faz a integração, a unidade de COM.

Tucci, E (2006), conforme apresentado no capítulo 2, propõe que as unidades de APC, como unidades integradoras que são, sejam categorizadas em repetições de expressões do tema, preenchimento, retomada textual e informação tardia. As repetições e as informações tardias são semanticamente plenas, enquanto as retomadas textuais e os preenchimentos são semanticamente vazios.

Embora a autora não tenha realizado seus estudos fazendo a distinção entre as tipologias monólogo e diálogo no corpus de italiano falado, o que se viu foi que a repetição é o tipo de função mais frequente para os APC com 50% de ocorrências, e em segundo lugar estaria a informação tardia, com 33% das ocorrências sobre o total de APC analisados.

A tabela a seguir detalha, por tipologia, as funções desempenhadas pelo APC em nossa amostra.

Tabela 5.7 - Número de ocorrências funcionais dos APC

TEXTOS	REP	%	P	%	RT	%	IT	%	Total por texto	%
bfamcv01	-		4		3		3		10	31
bfamcv02	-		1		-		2		3	9
bfamcv03	2		-		-		1		3	9
bfamcv04	1		-		-		1		2	6
bpubcv01	-		2		2		1		5	16
bpubcv02	1		4		-		4		9	28
	4	**12,5%**	**11**	**34,4%**	**5**	**15,6%**	**12**	**37,5%**	**32**	**100**
bfamdl01	-		2		-		3		5	14
bfamdl02	-		5		1		-		6	17
bfamdl03	2		4		-		1		7	19
bfamdl04	1		1		-		4		6	17
bfamdl05	-		2		-		3		5	14
bpubdl01	1		-		-		-		1	3
bpubdl02	-		3		1		2		6	17
	4	**11,1%**	**17**	**47,2%**	**2**	**5,5%**	**13**	**36,1%**	**36**	**100**
bfammn01	5		1		1		11		17	41
bfammn02	-		2		-		3		5	12
bfammn03	1		2		-		1		4	10
bfammn04	-		8		-		-		8	20
bfammn05	1		1				-		2	5
bfammn06	-		-		-		-		-	-
bpubmn01	-		5		-				5	12
	7	**17%**	**19**	**45%**	**1**	**2,4%**	**15**	**35,7%**	**42**	**100**
TOTAL GERAL	**15**	**14%**	**47**	**43%**	**8**	**7%**	**40**	**36%**	**110**	**100**

Legenda

REP – Repetição

P – Preenchimento

RT – Retomada Textual

IT – Informação Tardia

É importante salientar que na tabela acima só foram evidenciados 110 unidades de APC, embora tenham sido encontradas 112 unidades na amostra. Isso ocorreu porque encontramos dificuldade em classificar duas unidades terminadas. Esse assunto será abordado no final dessa seção.

Os resultados apresentados na tabela 5.7 reforçam a premissa de que a unidade informacional de APC é geralmente utilizada para realizar a correção ou o acréscimo de material lexical da unidade de COM.

Na tipologia conversação, a informação tardia é a classificação informacional que mais se destaca com 37,5% de ocorrências. Em seguida estão os preenchimentos (34,4%), as retomadas textuais (15,6%), e por último as repetições (12,5%).

Nos diálogos, há a predominância dos preenchimentos (47,2%), em seguida estão as informações tardias (36,1%), repetições (11,1%) e retomadas (5,5%).

Os monólogos procedem de maneira semelhante aos diálogos. Os preenchimentos destacam-se com 45,2% de ocorrências, as informações tardias com 35,7%, as repetições com 16,6% e as retomadas textuais com 2,4%.

Em âmbito geral, as unidades de APC no PB desempenham mais a função de preenchimento, com 43% de ocorrências sobre o total de APC, seguido pela informação tardia com 36%, depois as repetições com 14% e, por último, as retomadas textuais incidindo em apenas 7% sobre o total de APC na amostra.

Por se tratar de uma classificação bastante coerente, optamos apenas por tentar aproximar as categorias propostas por Tucci (2006) do PB, propondo uma subdivisão de duas categorias: as repetições e as informações tardias.

5.2.1.1 As repetições

Segundo Tucci, E (2006), as repetições de expressões do tema do discurso são discriminadas por tipologia ou distribuição, podendo ser literais ou parciais. Distribucionalmente, as repetições podem ocorrer de forma contígua, não-contígua e *Leit Motiv*[50].

Nossa proposta concentra-se na inclusão de mais uma discriminação por tipologia, a repetição parcial de forma contígua ou não. Vejamos os três únicos exemplos encontrados na amostra.

[50] Em nossa amostra não foi encontrado nenhum APC com função informacional de *Leit Motiv.*

Exemplo 5.5 (bpubcv02) (CD áudio 5.5)
*CAR: [212] eu nunca fui na casa lá dela não /=CMM= mas diz que é <muito ruim> /=CMM= **lá o local** //=APC=$

Exemplo 5.6 (bfamdl03) (CD áudio 5.6)
*LUZ: [1] porque /=DCT= eu só soube que eu nũ +=EMP=$ [2] eu tive certeza absoluta que eu nũ era daqui quando eu saí //=COM=$ [3] que eu senti que então /=SCA= eu tava no meu lugar /=COM= né //=PHA=$ [4] porque eu [/1]=SCA= eu me senti /=SCA= respirando /=i-COB= né /=PHA= adequada /=COM= né /=PHA= **no lugar** //=APC=$ [5]

Exemplo 5.7 (bfammn01) (CD áudio 5.7)
*MAI: [20] é um trilho dentro da mata /=COB= a mata virgem //=COM=$ [21] n' é matinha igual essas capoeirinha aqui não /=COB= é mata mesmo /=COB= de /=SCA= madeira /=SCA= da grossura que /=SCA= quato homem nũ abarca um pau //=COM=$ [22] uma /=SCA= tora /=COM= **da madeira** //=APC=$

Nos três exemplos listados acima o que se viu na unidade de APC foi a reformulação de uma informação dada pelo contexto e não apenas uma repetição não-literal dessa informação, embora deva-se admitir que, às vezes, essa informação é quase insignificante, quando tomada apenas semanticamente.

Proposta a reclassificação da função informacional repetição, passamos a esboçar na tabela 5.8 o número de ocorrências de APC, com diferentes funções de repetição, na amostra.

Tabela 5.8 - Número de ocorrências de APC com diferentes funções de repetição

Textos	Total de ocorrências de APC com função de repetição	Repetição Literal		Repetição Parcial		Repetição com Variação
		Contígua	**Não-contígua**	**Contígua**	**Não-contígua**	
bfamcv01	-	-	-	-	-	-
bfamcv02	-	-	-	-	-	-
bfamcv03	2 – 50%	-	1	1	-	-
bfamcv04	-	-	1	-	-	-
bpubcv01	1- 25%	-	-	-	-	-
bpubcv02	1- 25%	-	-	-	-	1
Total Parcial	**4- 100%**	**-**	**2 – 13%**	**1 –7%**	**-**	**1 – 7%**
bfamdl01	-	-	-	-	-	-
bfamdl02	-	-	-	-	-	-
bfamdl03	2 – 50%	-	1	-	-	1
bfamdl04	1 – 25%	-	-	-	1	-
bfamdl05	-	-	-	-	-	-
bpubdl01	1 – 25%	-	1	-	-	-
bpubdl02	-	-	-	-	-	-
Total Parcial	**4 – 100%**	**-**	**2 – 13%**	**-**	**1 – 7%**	**1 – 7%**
bfammn01	5- 72%	2	-	1	1	1
bfammn02	-	-	-	-	-	-
bfammn03	1- 14%	-	-	-	1	-
bfammn04	-	-	-	-	-	-
bfammn05	1-14%	-	1 -	-	-	-
bfammn06	-	-	-	-	-	-
bpubmn01	-	-	-	-	-	-
Total Parcial	**7 – 100%**	**2 – 13%**	**1 – 7%**	**1 – 7%**	**2 – 13%**	**1 – 7%**
TOTAL GERAL	**15- 100%**	**2 – 13%**	**5 – 33%**	**2 – 13%**	**3 – 20%**	**3 – 21%**

Embora os percentuais de ocorrência desse tipo de classificação informacional seja muito pequeno, podemos dizer que as repetições literais são as mais comuns entre as possíveis funções do APC, com um percentual de 46% sobre o total de APC. Em seguida, estão as repetições parciais com 33% de ocorrências e, em último, as repetições com variação (21%).

Quanto ao fato de as repetições ocorrerem dentro ou fora do enunciado, achamos ser desnecessário pontuar o percentual, já que se tratam de valores pouco representativos. Entretanto, classificamos no anexo 1 todos esses enunciados, conforme propõe a tabela acima.

5.2.1.2 As Informações Tardias

As informações tardias, segundo Tucci, E (2006), referem-se à adição de novas informações à unidade de COM. Após análise de nossos dados, vimos que a informação tardia é caracterizada também pela composicionalidade sintática com o conteúdo expresso na unidade de COM. Faz-se necessário esclarecer, entretanto, que nem sempre é possível interpretar se o APC compõe sintaticamente com a unidade de COM, mas não se dá nenhum caso de informação tardia sem que seja possível a composicionalidade sintática. Analisemos alguns exemplos para melhor ilustrar a afirmação acima.

Exemplo 5.8 (bfamcv01) (CD áudio 5.8)
*GIL: [65] da gente /=TOP= pra alguns setores da organização /=TOP= chamar o pessoal /=COM= dos outro times //=APC=$

A unidade de APC nesse caso integra a informação da unidade de COM, procurando elucidar de que 'pessoal' se trata (o pessoal 'dos outro times'). Veja-se que a unidade de APC passa a compor sintaticamente com a unidade de COM ('chamar o pessoa'. Que pessoal? 'O dos outros times') . Agora, passemos a outro exemplo.

Exemplo 5.9 (bfamcv01) (CD áudio 5.9)
*GIL: [137] <é> //=COM=$ [138] todo mundo <encheu o saco> /=COM= <**por causa disso**> //=APC=$

Nesse exemplo, o APC pode ou não ser interpretado com composicionalidade sintática.

Dessa forma, podemos dizer que se o APC só puder ser interpretado por meio da composicionalidade sintática, então será obrigatoriamente classificado como sendo uma informação tardia.

Na amostra, há, também, um único exemplo de APC com função de informação tardia e valor de glosa. Vejamos.

Exemplo 5.10 (bfammn01) (CD áudio 5.10)
MAI: [9] e quando chegou lá /=TOP= &he /=TMT= montou uma casinha pa ele /=COB= pa família dele /=COB= e tal /=COB= e e' vinha na cidade pa compar alguma coisa &dif [/1]=SCA= diferente /=COM= **que nũ era da roça** /**=APC**= né //=PHA=$

Repare que essa glosa visa a reforçar o que fora mencionado na unidade de COM, funcionando como uma espécie de nota explicativa, já que se diz 'compar alguma coisa diferente', diferente de quê? De 'que nũ era da roça'.

5.2.1.3 As Retomadas Textuais

A retomada textual é a referência ao discurso em si ou a parte do discurso. Pode ser realizada em forma de deixis discursiva, quando se refere ao discurso em si, ou de recontextualização, ao retomar sinteticamente uma parte do discurso. A título de exemplo, ilustramos alguns casos.

Exemplo 5.11 (bfamcv01) (CD áudio 5.11)
*EVN: [151] é /=INP= a <gente tem que> <restringir também /=COM= **isso**> //=APC=$

Nesse caso, a unidade de APC, por meio do pronome 'isso', é um encapsulador, cuja função é retomar algo mencionado.

Agora vejamos um exemplo de retomada com reformulação.

Exemplo 5.12 (bfammn01) (CD áudio 5.12)

*MAI: [85] na Amazonas tem muito dessa cobra aí //=COM=$[86] só ni lugar /=SCA= que tem /1 =SCA= tem mata mata muito /=SCA= forte /=COB= mata /=SCA= perigosa /=COB= que tem <esses tipo> de cobra /=COM= né //=PHA=$
*DUD: [87] <uhn> //=COM=$
*MAI: [88] no norte de Mina /=TOP= tinha esse /2 =SCA= antigamente /=PAR= tinha esse tipo de cobra tudo /=COM= né //=PHA=$[89] talvez agora já acabou /=COB= porque já desmataram muito /=COM= né //=PHA=$[90] nũ existe muita mata mais igual /1 =SCA= igual era /=COB= mas tem muita mata forte aí /=COB= nũ é igual essas /1 =SCA= essas capoeirim /=COB= cê fala assim /=INT= ah /=EXP_r= isso aqui é uma mata //=COM_r=$[91] isso aqui /=TOP= em vista de lá /=TOP= é /1 =EMP= é um cerrado /=COM= **aqueas mata lá** //**=APC**=$

Nesse exemplo, a unidade de APC funciona como paráfrase daquilo que fora mencionado no enunciado 85.

5.2.1.4 Os Preenchimentos

Os APC com função de preenchimento foram os mais encontrados na amostra. Eles representam 43% sobre o total de APC analisados e tem como objetivo realizar a expansão da unidade precedente sem repetir seu conteúdo semântico ou acrescentar informações.

Exemplo 5.13 (bfammn01) (CD áudio 5.13)
MAI: [79] é lugar /=SCA= perigoso mesmo o norte de Minas /=COM= lá //=APC=$

Exemplo 5.14 (bfammn03) (CD áudio 5.14)
*ALO: [37] sio' /=SCA= precisa de ir lá p' siora despedir //=COM_r=$ [38] do pai lá /=COM_r= **e tal** //**=APC**_r=$

5.2.1.5 Casos Duvidosos em relação à análise funcional

Embora tenhamos ampliado a proposta de Tucci, ainda assim há dois casos que nos deixaram muito inseguros quanto à sua classificação informacional. Vejamos:

Exemplo 5.15 (bfamdl05) (CD áudio 5.15)

*ANE: [207] <só tem esse> apartamento pra vender //=COM=$

*CES: [208] obrigado> //=COM=$

*ENC: [209] eu creio que sim //=COM=$ [210] que +=EMP=$

*ANE: [211] tá //=COM=$

*ENC: [212] tá //=COM=$

*ANE: [213] esquecemo de perguntar <pa ele a> previsão /=COM= né //=PHA=$

*CES: [214] <bom> //=COM=$ [215] é //=COM=$ [216] deixa eu <perguntar pra ele /=COB= a> previsão //=COM=$

*ANE: [217] <pergunta lá> //=COM=$

*CES: [218] eu acho que a previsão dele tá pra entregar em dezembro //=COM=$ [219] a previsão de entrega do prédio //=COM=$

*ENC: [220] no começo yyyy //=COM=$

*CES: [221] ahn //=COM=$

*ENC: [222] no começo yyyy //=COM=$

*CES: [223] então /=INP= até final de novembro //=COM=$

*ENC: [224] isso //=COM=$

*ANE: [225] &he /=TMT= essa localização aqui /=TOP= eu nũ acho ruim //=COM=$

*RAQ: [226] eu gostei também foi por causa do construído já ao redor /=COM= <né> //=PHA=$

*ANE: [227] <é> //=COM=$

*RAQ: [228] <é sim> //=COM=$

*ANE: [229] <tá tudo> <construído> //=COM=$

*CES: [230] <Anete> //=COM=$ [231] <ele> [/1]=EMP= e' disse que é &d +=EMP=$ [232] final de novembro //=COM=$ [233] <é /=INP= porque agora é rápido> /=COM= **isso <aqui**> //=APC=$

Em princípio, havíamos classificado o APC acima como sendo uma retomada textual, porque acreditávamos que o pronome ‘isso’ referia-se à entrega do prédio. Todavia, ao fazermos a revisão de toda a classificação vimos que esse APC também poderia ser classificado como preenchimento, já que é semanticamente vazio.

Outro caso que nos deixou indecisos foi um enunciado retirado do texto bfammn02.

Exemplo 5.16 (bfammn02) (CD áudio 5.16)

DFL: [165] e às vezes ele tava lendo jornal /=TOP= de vez em quando ele [/1]=SCA= ele interrompia /=COB= ele tava sempre atento às nossas brincadeira /=PAR= e' falava /=INT= o que é o que é //=COM_r=$ [166] a gente <assustava /=COB= e tinha que +=EMP=$ [167] às vezes +=EMP=$ [168] se sabia /=CMM= falava /=CMM= <mas se nũ sabia /=CMM= né> //=PHA=$
*LUC: [169] <uma charada /=COB= **ali**> /=APC= <na hora> //=COM=$

Parece que esse APC, a priori, seria uma retomada textual, desde que se considerasse que o advérbio ‘ali’ refere-se ao discurso em si, ao fato de ‘ali’ coincidir com o tempo em que ocorriam as brincadeiras; mas se não interpretarmos dessa maneira, poderíamos dizer que se trata de uma informação tardia, algo inserido entre as unidades de COM apenas para reafirmar o lugar puramente discursivo em que ocorriam as brincadeiras.

Embora tenhamos dúvidas quanto à classificação informacional da unidade de APC em apenas dois casos, pensamos poder afirmar que nem sempre a unidade de APC apresenta uma única função; pode, por vezes, ser interpretada de maneiras distintas dado o contexto enunciativo.

5.2.2 Critério Entonacional

Uma análise acústica bem sucedida deve ocorrer de duas maneiras: por meio da percepção e através de uma documentação instrumental. O ideal é que para cada enunciado realizado possam-se efetuar esses dois tipos de análise. Todavia, conforme poderemos verificar abaixo, nem sempre isso ocorre; em alguns casos, somente a percepção é que nos permitirá dizer se a unidade analisada trata-se de um APC, porque o sinal acústico não está limpo e, contrariamente ao ouvido humano, os instrumentos não conseguem diferenciar sons (ruído,sobreposição de vozes) que se sobrepõem, tornando a curva de Fo não confiável.

Caracterizar acusticamente a unidade de APC implica em perceber que a média de F0 e a média da intensidade do COM do qual são apêndices sejam maiores do que sua própria média de F0 e a média de intensidade. Além disso, é preciso que o perfil entonacional dessa unidade seja nivelado e descendente.

Para a realização da análise acústica utilizamos dois softwares: o Winpitch e o Praat. Através do WinPitch foram extraídos todos os enunciados que continham as unidades de APC, bem como a realização da síntese de alguns enunciados duvidosos. Por meio do Praat, realizaram-se as análises relativas à média de F0 e média de intensidade do APC, também a média de F0 e média da intensidade das unidades de COM, CMM e COB que podem anteceder a unidade de APC.

Abaixo, a tabela 5.9 esboça todos os valores aferidos por meio da análise instrumental. Aqueles exemplos cuja análise foi apenas perceptual serão comentados mais a diante.

Tabela 5.9 - Valores prosódicos das unidades de COM e APC

Texto	Numeração do exemplo	Média F0 – COM (HZ)	Média F0 – APC (HZ)	Média Int. – COM (dB)	Média Int. – APC (dB)
bfamcv01	2	189,9	100	55,21	52,56
	3	270,4	237,7	65,28	56,54
	4	114,2	107,1	56,68	51,41
	5 e 6	COM1= 142Hz COM2=	APC1= 120,1 Hz	COM1= 53,18 dB COM2=	APC1= 49,47 dB

		132,7 Hz	APC2= 114,8 HZ	57,73dB	APC2= 48,59 dB
	7	140,1	113,4	62,79 dB	55,65 dB
	9	152,2	50	52,22	51,06
	10	142,6	128,5	62,41	61,64
bfamcv02	11	352	184,6	54,98	53,97
	12	222,3	172,5	63,24	54,14
	13	248,5	173,2	60,71	48,91
bfamcv03	14	216,8	144,3	65,68	60,46
	16	146,6	90,9	46,34	34,22
bfamcv04	17	236	209,3	61,55	50,46
	18	289,7	203,8	62,74	59,37
	19	293,5	267,3	73,35	69,75
bpubcv01	21	160,6	130,8	66,67	63,46
	24	239,95	50	63,91	59,39
bpubcv02	25	148	93,26	63,21	59,94
	26	130,4	118	69,98	58,92
	27	193,3	138,8	77,61	67,17
	28	126	72,45	64,17	54,89
	29	109,7	99,97	67,86	58,83
	32	146,3	128,5	69,62	68,16
	33	162,5	124,2	73,77	71,01
bfamdl01	37	259,8	197,1	70,86	65,82
bfamdl02	39	205,8	185,5	56,56	48,8
	40	224,5	150,8	52,67	44,46
	41	216,4	153	66,94	59,64
	42	300,4	83,89	53,13	37,33
	43	161,15	131,2	64,87	61,99
	44	246	215,2	71,5	70,63
bfamdl03	47	282,4	191,2	63,8	62,16
	48	221,9	198,8	59,94	55,77
	49	228,6	95,81	63,03	61,43
	50	83,77	71,89	56,13	51,98
	51	75,81	66,79	63,1	60,37
bfamdl04	52	255,8	165,3	61,17	50,39
	53	256,9	196,8	64,44	61,52
	54	216,1	132,3	58,89	53,35
	55	253,1	200,8	74,74	67,31
	56	235,7	173,3	59,37	52,48
	57	234,6	128,6	58,8	52,88
bfamdl05	58	75,22	52,1	62,95	57,88
	59	175,2	116,6	48,44	44,02
	60	240,3	138,3	61,15	50,71
	61	258,3	141,9	59,95	58,69
	63	60,46	55,19	68,35	34,25
bpubdl01	64	102,6	97,71	60,84	61,78[51]
bpubdl02	65	273,2	157,3	76,76	73,33
	66	187	109,6	69,87	63,21
	67	135,3	115,8	63,42	63,09
	70	165,5	139,2	75,2	74,11
bfammn01	71	137,1	57,88	92,42	49,18
	72	163,5	133,3	59,55	58,03
	74	250	57,63	122,9	56,91
	77	134,7	108,4	67,06	58,43
	82	164,1	112,2	52,25	47,37

[51] Os exemplos 64, 104 e 109 merecem uma análise mais acurada, porque as diferenças entre as médias são praticamente nulas. No caso do exemplo 64 ainda há outro problema, a média de intensidade do APC é superior ao da unidade de COM.

	83	179,4	109,9	61,7	56,06
	84	209	180,3	61,08	49,65
	87	100,2	92,51	56,64	48,58
bfammn02	88	123	102,5	53,61	49,75
	89	212,1	93,63	48,38	44,95
	90	247,4	120,5	56,34	55,9
	91	250,8	175,9	50,84	56,5
	92	157,15	143,4	67,25	66,59
	93	227,5	174,5	55,64	52,79
bfammn03	94	108,5	72,82	62,76	49,24
	95	139,8	88,73	59,69	57,36
	96	121,2	85,77	60,79	51,42
	97	137,8	129,19	57,29	52,36
bfammn04	98	275,6	244,7	57,2	56,49
	99	212,6	205,5	58,28	57,49
	100	449,89	346,6	58,47	55,98
	101	141,8	132	61,73	60,83
	103	374,2	223,2	58,27	52,37
	104	269,1	268,8	61,62	54,93
	105	241,8	137,3	55,73	48,97
bfammn05	106	221,1	194,7	60,49	53,2
	107	232,6	221,6	42,82	39,14
bfammn06	-				
bpubmn01	108	254,8	132,8	66,8	63,18
	109	226,4	224,5	67,70	57,24
	110	293,4	155,9	70,24	56,84
	111	207,8	199,40	67,47	53,66
	112	309,2	288,8	74,47	71,9

Como se pode observar, a média de F0 da unidade de COM é sempre mais alta que a da unidade de APC. Isso porque, conforme já fora mencionado no capítulo 1, a unidade de COM sempre realiza uma força ilocucionária e possui um foco entonacional. Na unidade de APC, nenhum movimento funcional é encontrado, uma vez que essa unidade visa à integração de novas estruturas linguísticas de modo a realizar uma expressão semântica, uma correção ou mesmo uma reestruturação de um enunciado. O abaixamento do tom de voz ou a média baixa de intensidade são, também, características dessa unidade.

A tabela 5.10, a seguir, procura delinear as unidades de APC quanto à curva de F0.

Tabela 5.10 - Número de ocorrências de APC com curva analisável ou não, com sobreposição e com ruído

Texto	Total de APC por texto	APC c/ curva analisável	APC c/ curva não analisável
bfamcv01	10	8	2
bfamcv02	03	3	-
bfamcv03	03	2	1
bfamcv04	03	3	-
bpubcv01	05	3	2
bpubcv02	09	8	1
Total Parcial	**33 – 29%**	**27– 24%**	**6 – 5%**
bfamdl01	05	3	2
bfamdl02	06	6	-
bfamdl03	07	5	2
bfamdl04	06	6	-
bfamdl05	06	5	1
bpubdl01	01	1	-
bpubdl02	06	6	-
Total Parcial	**37 – 33%**	**32- 29%**	**5 – 4%**
bfammn01	17	14	3
bfammn02	06	6	-
bfammn03	04	4	-
bfammn04	08	7	1
bfammn05	02	2	-
bfammn06	0	-	-
bpubmn01	05	5	-
Total Parcial	**42 – 37,5%**	**38 – 34%**	**4 – 4%**
TOTAL	**112 – 100%**	**97 –87 %**	**15 - 13%**

Do total de 112 APC encontrados na amostra, 97 (87%) possuem curva analisável e apenas 15 (13%) não analisável. Para garantir a validação desses enunciados como APC, além da análise entonacional, acima apresentada e da análise funcional anteriormente exibida, realizou-se a análise distribucional, como se poderá verificar na subseção 5.2.3.

Entre os casos de curva não analisável, o que se observou foi a presença ora de ruídos, ora de sobreposição de vozes ou até mesmo de creaky voice em alguns casos, que acabaram por invalidar a análise acústica de 15 enunciados. A tabela 5.11 ilustrará esses enunciados.

Tabela 5.11 – Total de enunciados com APC com curva não analisável

Texto	APC c/ curva não analisável	APC com ruído e sobreposição	APC com sobreposição	APC com creaky voice
bfamcv01	2	2		-
bfamcv02	-	-		-
bfamcv03	1	1		-
bfamcv04	-	-		-
bpubcv01	2	2		-
bpubcv02	1	-	1	-
bfamdl01	2	2	-	
bfamdl02	-			
bfamdl03	2	2	-	
bfamdl04	-			
bfamdl05	1	-	1	
bpubdl01	-	-	-	
bpubdl02	-	-	-	
bfammn01	3	1	-	2
bfammn02	-	-	-	-
bfammn03	-	-	-	-
bfammn04	1	1	-	-
bfammn05	-	-	-	-
bfammn06	-	-	-	-
bpubmn01	-	-	-	-
TOTAL	**15 - 13%**	**11 – 74%**	**2 – 13%**	**2 – 13%**

O maior problema com os APC com curva não analisável, como se pode notar, foram aqueles casos em que a gravação apresentava tanto ruído, quanto sobreposição de vozes. Como os instrumentos de análise acústica não conseguem diferenciar os tipos de som, como o ouvido humano, tivemos que descartar esses onze casos, quanto à análise acústica.

A creaky voice também nos causou problemas. Nos dois casos em que foi detectada, igualmente não se puderam aferir os cálculos da média de F0 e de intensidade.

No caso das sobreposições, quando foi possível separar os sons (de estéreo para mono) por meio do software Audacity[52], as médias foram calculadas. Infelizmente, há dois casos, conforme esboçado na tabela 5.11, em que isso não foi possível.

Além desses problemas, há também dois casos duvidosos para nós, quando da análise acústica. Vejamos:

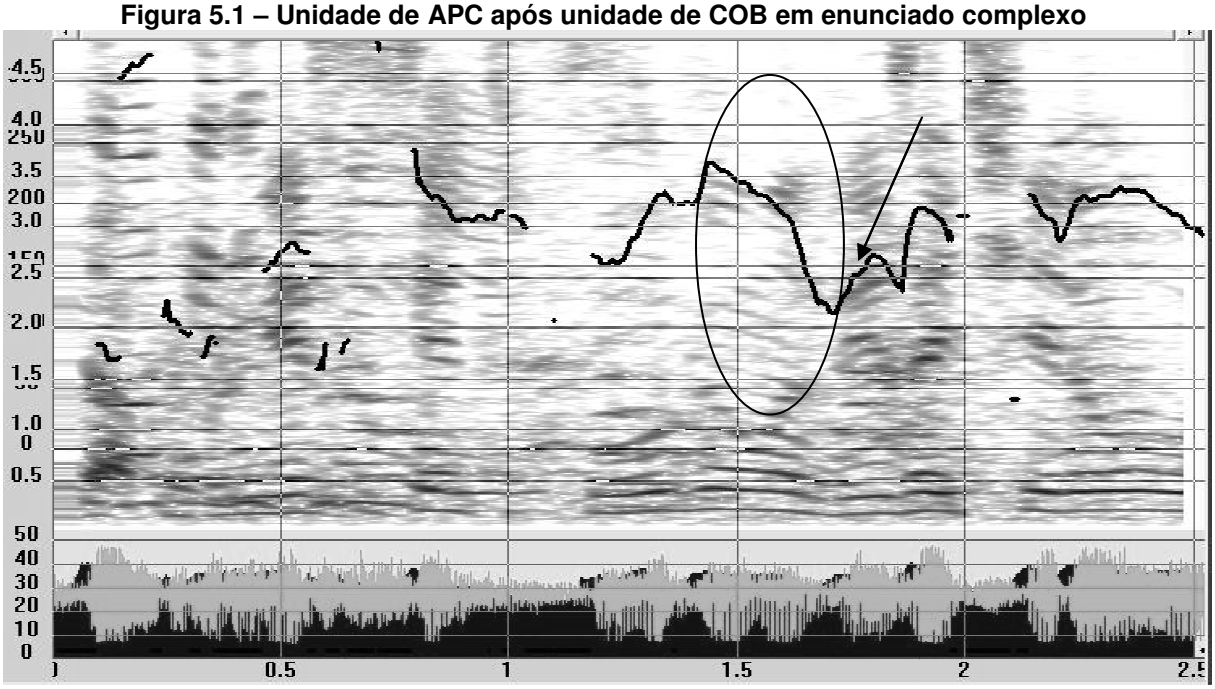
[52] http://audacity.sourceforge.net

5.2.2.1 Primeiro caso duvidoso em relação à análise entonacional

O primeiro caso duvidoso é quando após uma unidade de APC vem uma unidade de COB. Passemos ao exemplo.

Exemplo 5.17 (bpubcv01) (CD áudio 5.17)
*MAR: [167] <é o> mesmo processo //=COM=$
*EMM: [168] <centrifugação> //=COM=$
*MAR: [169] <você o centrifuga> /=COB= a plaqueta é mais pesada /=COB= vai pro fundo /=COB= e o plasma <xxx> //=COM=$
*FLA: [170] <que são aqueles gruminhos> //=COM=$
*BRU: [171] <ah /=CMM= tá //=CMM=$ [172] porque primeiro cê centrifuga> /=COB= **uma vez** /**=**APC= aí tira o +=EMP=$

Figura 5.1 – Unidade de APC após unidade de COB em enunciado complexo



Como se pode reparar, por meio do espectrograma, logo após o APC, delineado pela elipse, há uma subida (onde a seta aponta), algo até então só verificado nesse exemplo. A média de F0 da unidade de COM é 162,2 Hz e a média de intensidade

69,93 dB. A unidade de APC apresenta uma média de F0 de 170,5 Hz e uma média de intensidade de 70,76 dB. O que é comum de acontecer, como já fora dito, são as médias de F0 e de intensidade do COM serem maiores do que as do APC. O fato de isso não ocorrer especificamente nesse caso, pode ser justificado, acredita-se, pelo fato de haver uma unidade interrompida logo após a unidade de APC, que em princípio realizaria uma força ilocucionária.

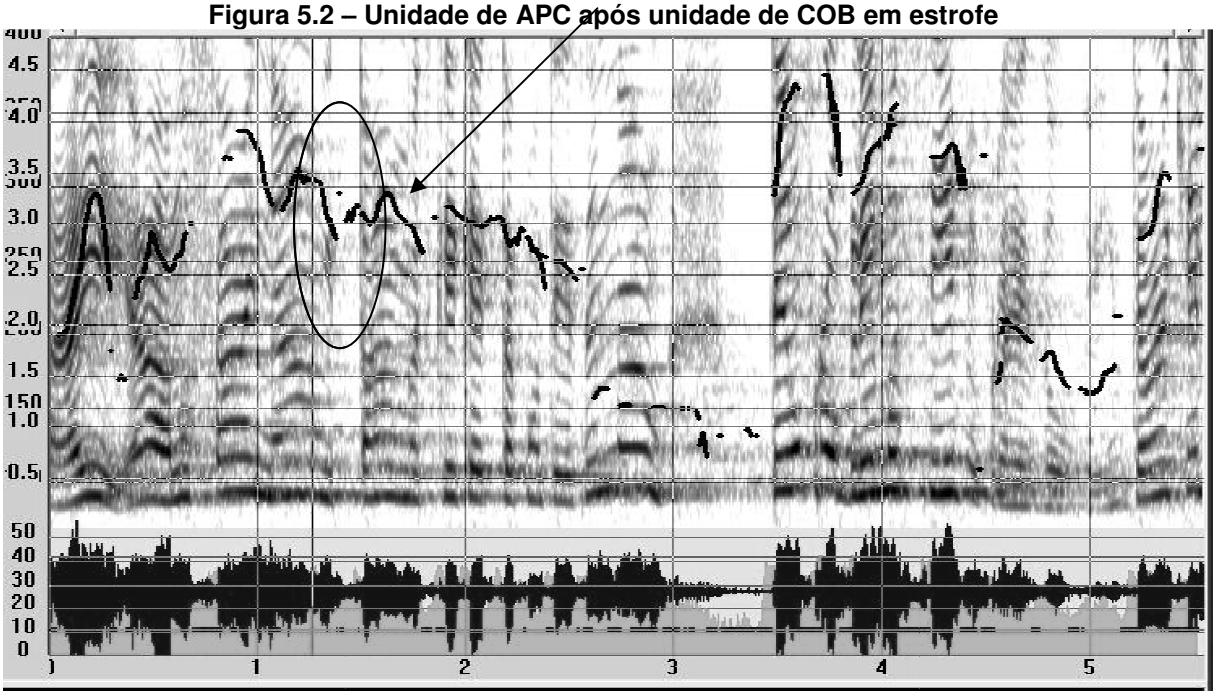
O fato de a unidade de APC integrar a unidade de COB só foi verificado em três enunciados da amostra. O primeiro foi citado acima, os outros dois, como se poderá conferir abaixo, refletem um comportamento prototípico da unidade de APC; isto é, apresentar média de F0 e intensidade inferior à unidade ilocucionária a qual se integra. Como se trata apenas de uma hipótese, pensa-se que esse deva ser um comportamento investigado em todos os textos do C-ORAL-BRASIL.

Em sequência, ilustramos os dois exemplos cujo comportamento da unidade de APC está em conformidade com o proposto pela Teoria da Língua em Ato. As elipses delimitam as unidades de APC.

Exemplo 5.18 (bpubmmn01) (CD áudio 5.18)

*SHE: [112] ah /=EXP_r= nũ sai não /=COB_r= **e tal** /=APC_r= mas que que é que tá acontecendo /=COB_r= não /=COB= tem um problema sempre pra resolver //=COM=$

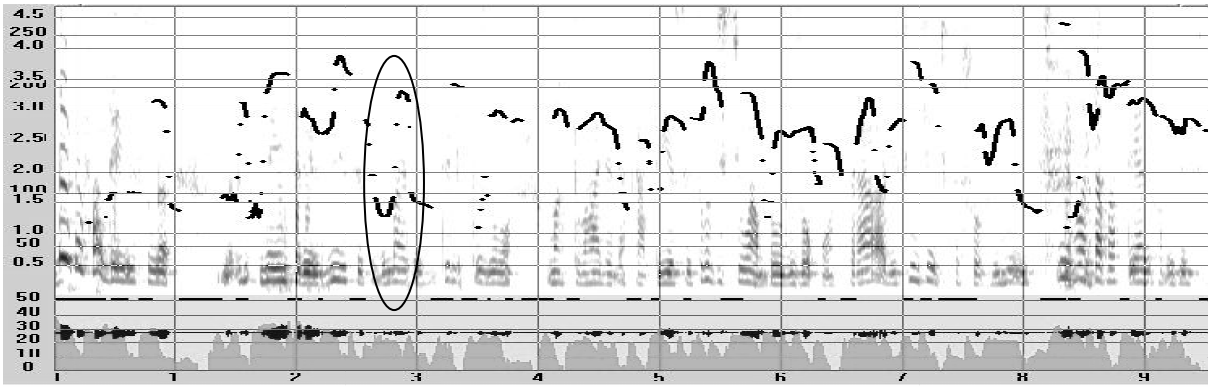
Figura 5.2 – Unidade de APC após unidade de COB em estrofe



Exemplo 5.19 (bfammn02) (CD áudio 5.19)

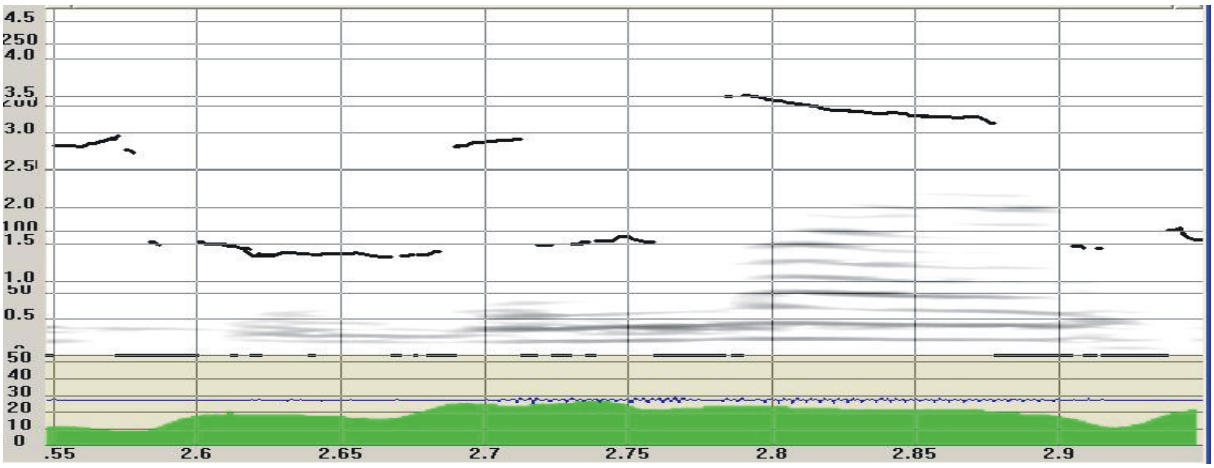
*DFL: [173] <eu até já> falei isso na [/1]=SCA= &n [/1]=EMP= ni entrevista /=COB= **e tudo mais** /=APC= às vezes ele punha uma=SCA= uma toalha na cabeça /=COB= ou então ele vestia uma blusa ao contrário /=COB= e fingia de cachorrinho /=COB= e levantava a perna pra fazer xixi //=COM=$

Figura 5.3 – Unidade de APC após unidade de COB em estrofe composta de mais unidades informacionais



Por ser esse exemplo uma estrofe constituída de várias unidades informacionais, optamos por ilustrar abaixo, somente a unidade de APC depois de um COB, para melhor visualização das curvas entonacionais.

Figura 5.3.1 – Unidade de APC depois de um COB em estrofe

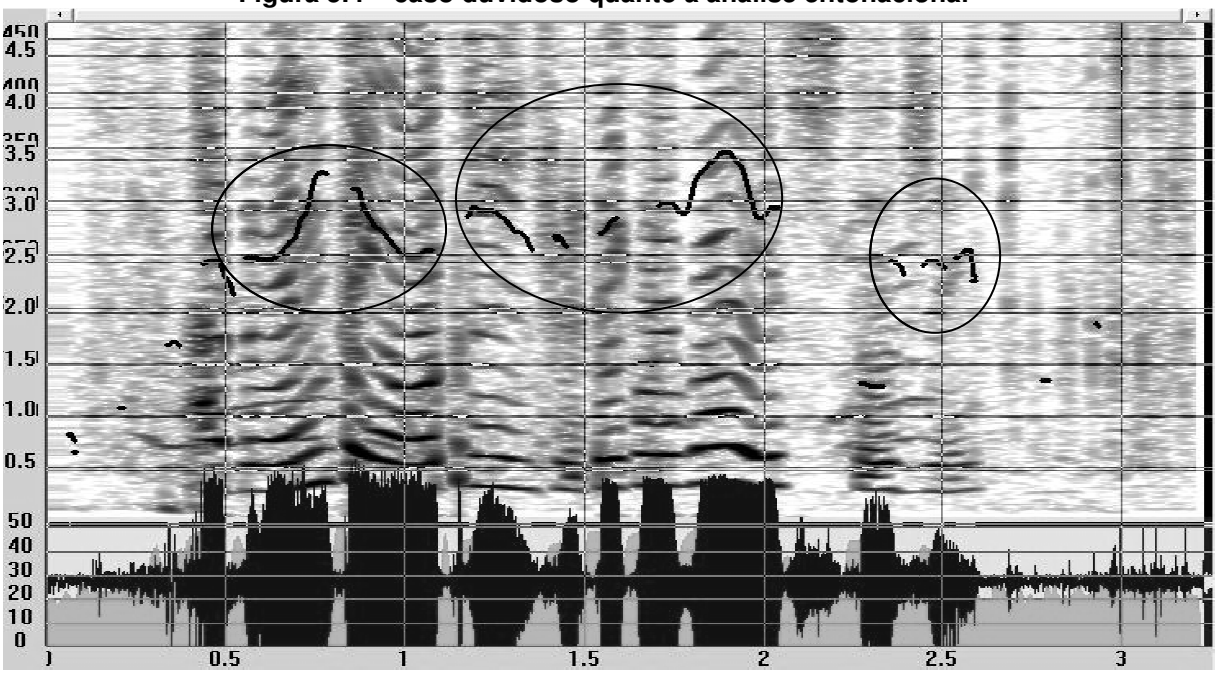


5.2.2.2 Segundo caso duvidoso em relação à análise entonacional

Esse segundo caso é ainda mais duvidoso, porque apresenta tanto características peculiares da unidade de APC, quanto da unidade de PAR. Vejamos:

Exemplo 5.20 (bfamdl01) (CD áudio 5.20)
*REN: [568] a culpa é sempre do arroz //=COM=$
*REN: [569] ah /=INP= também nũ tem outro /=CMM= a gente vai ter que levar esse /=CMM= **de todo jeito** //=(APC)

Figura 5.4 – caso duvidoso quanto à análise entonacional



A unidade de COM, subdividida em dois CMM, está representada pelas 1ª e 2ª elipses, com uma média de F0 de 281,9 Hz e uma média de intensidade de 79,67 dB. A segunda unidade, delineada pela 3ª elipse, embora tenha as médias de F0 (126,5 Hz) e intensidade (71,65 dB) normais para uma unidade de APC e estar distribucionalmente à direita da unidade de COM, bem como possuir o mesmo valor modal (deôntica - de todo jeito) da unidade de COM (ter que), nos causou dúvida porque o perfil de sua curva entonacional não é totalmente nivelado ou descendente, como se espera de uma unidade de APC; há um pequeno movimento. Além disso,

do ponto de vista prosódico, se eliminada do enunciado, nada faltaria à realização completa para o entendimento desse enunciado.

Assim como no primeiro caso duvidoso, acreditamos que esse deva ser um problema investigado em todos os textos do C-ORAL-BRASIL, a fim de se estabelecer um padrão de comportamento para casos dessa natureza.

5.2.3 Critério Distribucional

Para a Teoria da Língua em Ato, a unidade de APC só pode ocorrer depois de uma unidade ilocucionária do tipo COM (CMM e COB). Como a incidência dos APC nessas unidades ilocucionárias já foram apresentadas na seção 5.1, optamos por ressaltar aqui somente as características informacionais mais salientes em cada tipologia, para verificar se o tipo de unidade terminada condiciona de alguma maneira o valor informacional da unidade de APC.

Tabela 5.12 – Frequência de categoria informacional nas conversações

Categoria Informacional	COM	CMM	COB
Repetição	3/32 – 9,4%	1/32– 3%	-
Preenchimento	8/32 – 25%	1/32 – 3%	2/32 – 6%
Retomada Textual	3/32 – 9,4%	2/32 – 6%	-
Informação Tardia	12/32 – 37,5	-	-

Como se pode notar, após a unidade de COM, o tipo de função informacional mais comum dos APC nas conversações é a informação tardia, com 37,5% de ocorrências. Em seguida, estão os preenchimentos (25%) e as repetições e retomadas textuais juntas totalizam 18,8%.

Em relação aos padrões ilocucionários (CMM), a função mais recorrente é a de retomada textual com 6% de ocorrências. Repetições e preenchimentos juntas representam também 6%.

Nas estrofes (COB), a única função informacional encontrada foi a de preenchimento, com 6% de ocorrência.

Tabela 5.13 – Frequência de categoria informacional nos diálogos

Categoria Informacional	COM	CMM	COB
Repetição	3/36 – 8,3%	1/36 - 2,7%	-
Preenchimento	15/36 – 41,7%	2 – 5,5%	-
Retomada Textual	1/36 – 2,7%	-	1/36 – 2,7%
Informação Tardia	12/36 – 33,3%	1/36 – 2,7%	-

Nos diálogos, após a unidade de COM, a função informacional mais comum é a de preenchimento com 41,7% de ocorrências. Em seguida encontramos as informações tardias com 33,3%, as repetições com 8,3% e as retomadas textuais com 2,7%.

Nos padrões ilocucionários a função informacional de APC mais comum é a de preenchimento (5,5%). Logo após estão as repetições e informações tardias que juntas totalizam 5,4%.

Nas estrofes (COB) a única categoria informacional encontrada foi a de retomada textual (2,7%).

Tabela 5.14 – Frequência de categoria informacional nos monólogos

Categoria Informacional	COM	CMM	COB
Repetição	7/42 – 16,7%	-	-
Preenchimento	11/42 – 26,2%	3/42 – 7,1%	5/42 – 11,9%
Retomada Textual	1/42 – 2,4%	-	-
Informação Tardia	11/42 – 26,2%	1/42 – 2,4%	3/42 – 7,1%

Nos monólogos, após a unidade de COM, tanto as informações tardias, quanto os preenchimentos são mais freqüentes, totalizando 52,4% de ocorrências. As

repetições incidem com 16,7% e as retomadas textuais são menos comuns com apenas 2,4% de ocorrências.

Após os padrões ilocucionários (CMM), os APC exercem maior função de preenchimento (7,1%). Em seguida encontramos as informações tardias (2,4%).

Nas estrofes (COB), a categoria informacional mais saliente é a de preenchimento (11,9%), seguida apenas pela informação tardia (7,1%).

Tendo como referência os resultados esboçados nas três tabelas acima, pode-se dizer que não há como estabelecer uma padrão de recorrência entre o tipo de unidade ilocucionária (COM, CMM e COB) e a categoria informacional.

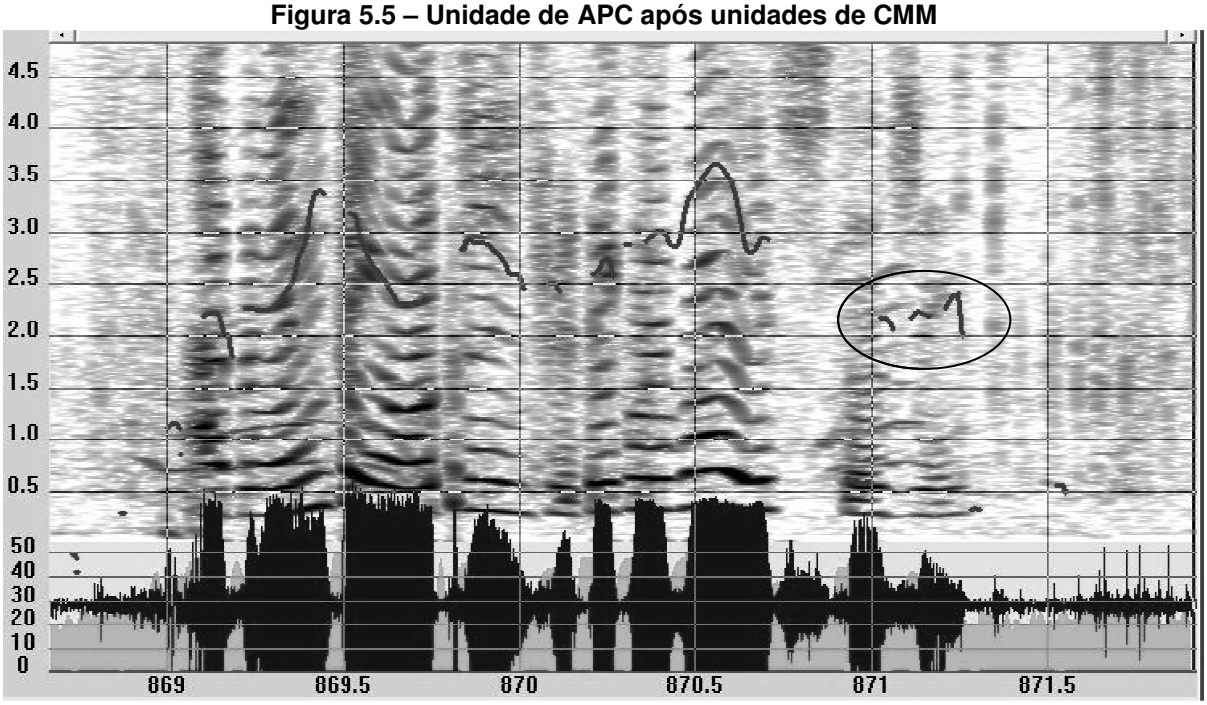
Abaixo seguem alguns exemplos para ilustrar a ocorrência de APC após as unidades ilocucionárias de CMM e COB.

Na figura, a elipse delimita a unidade de APC.

Exemplo 5.21 (bfamdl01) (CD áudio 5.21)

*REN: [600] ah /=INP= também nũ tem outro /=CMM= a gente vai ter que levar esse /=CMM= **de todo jeito** //**=APC=$**

Figura 5.5 – Unidade de APC após unidades de CMM



No exemplo 5.21 há um enfraquecimento do princípio ilocutivo, formando um padrão prosódico por meio de um modelo acional que permite ligar duas ilocuções de reforço, resultando em um efeito retórico convencionado. Nesse caso, a unidade de APC serve apenas para preencher o que fora mencionado no padrão ilocucionário.

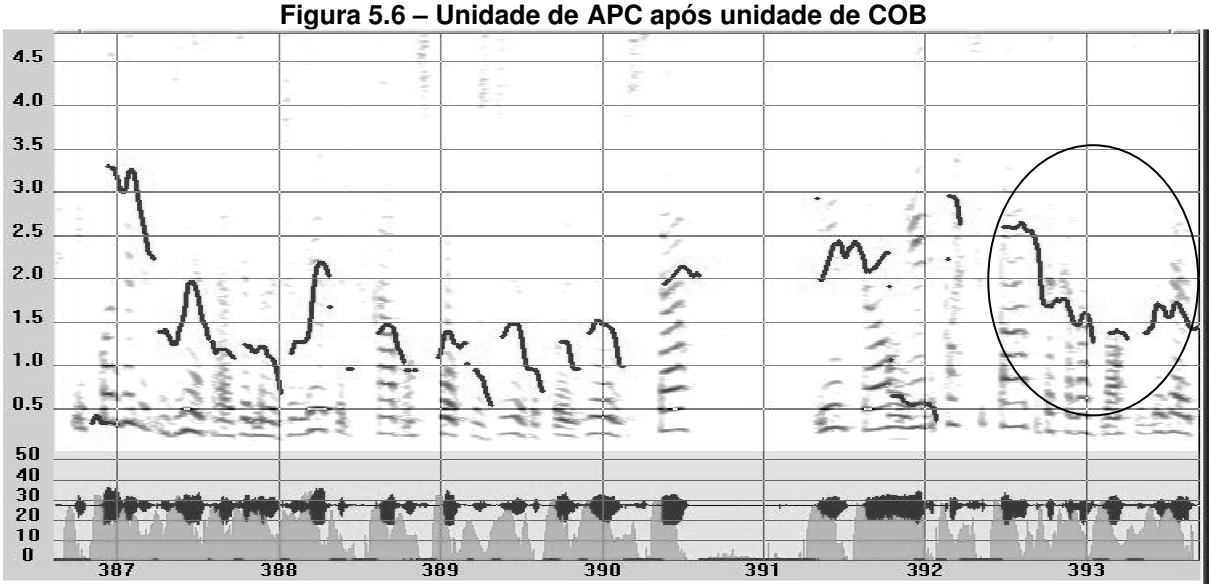
O exemplo a seguir ilustra uma unidade de APC ocorrendo depois de um COB, dentro de uma estrofe.

Na figura, a elipse delimita a unidade de APC.

Exemplo 5.22 (bfammn02) (CD áudio 5.22)

*DFL: [164] e quando eles /=SCA= vinha trazer a Maria Julieta /=TOP= ele ficava uns poucos dias depois vinham buscar /=PAR= e então /=DCT= ele participava /=COB= **daquelas nossas brincadeira /=APC=**

Figura 5.6 – Unidade de APC após unidade de COB



Nesse exemplo, a unidade de COB é percebida como tendo um sinal de continuação do enunciado, cuja quebra prosódica terminal só é percebida no final da estrofe. A unidade de APC funciona como informação tardia de possível interpretação da composicionalidade sintática.

5.2.3.1 Presença de unidades entre COM e APC

Entre a unidade de Comentário e a unidade de APC podem aparecer intercaladas as seguintes unidades: alocutivos (ALL), fáticos (PHA), conativos (CNT) , expressivos (EXP) e parentéticos (PAR). Porém, em nossa amostra só foram verificadas as presenças das unidades de alocutivo, parentético e fático. A tabela 5.15 ilustrará o

número de ocorrência dessas unidades entre os 110 enunciados com APC analisados.

Tabela 5.15 - Ocorrência de unidades entre COM e APC

TIPOLOGIA	COM/ALL/APC	COM/PAR/APC	COM/PHA/APC
Conversação	2	1	1
Diálogo	0	1	2
Monólogo	0	2	0

Como se pode notar, tendo por base a tabela acima, não é muito comum que essa inserção ocorra no PB, ainda que seja possível. Seguem-se os exemplos identificados na amostra.

Exemplo 5.23 (bfamcv02)
*RUT: [92] eu nũ fui não /=COM= **minha filha** /=**ALL=** lá não //=APC=$

Exemplo 5.24 (bfamcv03)
*TON: [247] é ocê /=COM= **Onofre** /=**ALL**= que tá cagando assim //=APC=$

Exemplo 5.25 (bfamcv04)
*HEL: [317] <não> /=COB= mas aí hhh +=EMP=$ [318] <e se precisar de um esclarecimento> /=COM= <**assim** /=**PAR**= no> meio do negócio hhh //=APC=$

Exemplo 5.26 (bpubcv02)
*OSV: [271] se for o caso só dela mesmo /=TOP= esses caso que /=SCA= não tiver jeito /=TOP= da [/1]=EMP= da gente fazer /=APT= tira ela então /=COM= **né** /=**PHA**= no caso //=APC=$

Exemplo 5.27 (bpubdl02)
*JAN: [43] <e uma> dessa /=COM= **aqui** /=**PAR**= também <o'> //=APC=$

Exemplo 5.28 (bfamdl03)
*LUZ: mas tem o Ronan /=SCA= também /=COM= **né** /**=PHA**= < aqui > //=APC=

Exemplo 5.29 (bfamdl03)
*LUZ: [4] porque eu [/1]=SCA= eu me senti /=SCA= respirando /=i-COB= né /=PHA= adequada /=COM= **né** /=**PHA**= no lugar //=APC=$ [5] <falei> /=INT= Nossa //=COM_r=$

Exemplo 5.30 (bfammn01)
*MAI: ele nũ é muito parente chegado não /=COB= mas &t [/1]=SCA= deve ser /=SCA= primo [/1]=EMP= primo quarto /=COM= **por aí** /**=PAR**= deve ser //=APC=

Exemplo 5.31 (bfammn02)
*LUC: [48] desde o começo /=COM= **assim** /=**PAR**= e tudo o mais //=APC=$

5.2.3.2 Unidades que podem ser confundidas com a unidade de APC

Nas análises, observou-se que há determinados contextos em que duas outras unidades podem ocupar a mesma posição do APC, com um perfil entonacional às vezes parecido, e serem confundidas com o APC. Tratam-se das unidades de PAR e COB. As subseções seguintes tratarão desse aspecto.

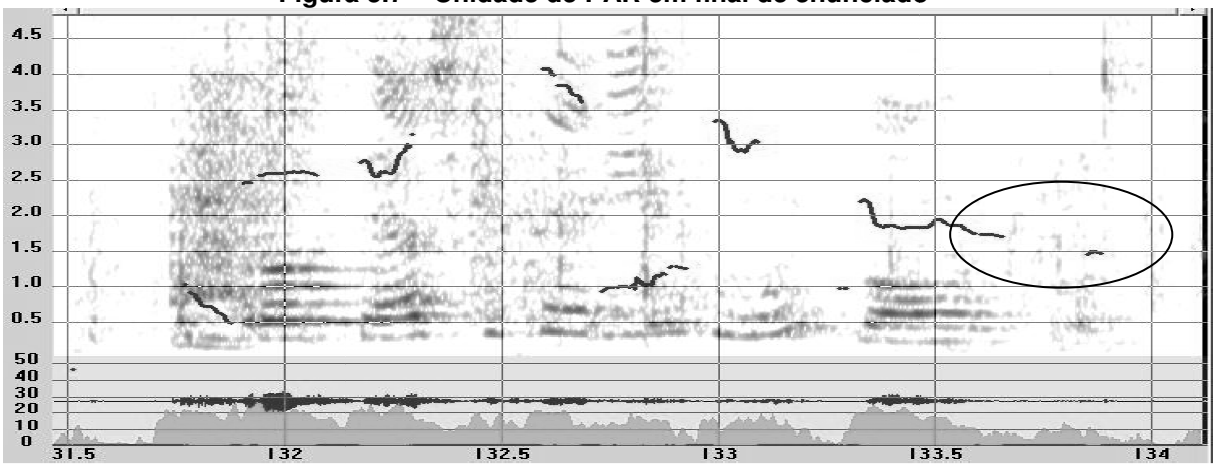
5.2.3.2.1 APC versus PAR

A unidade de PAR, quando da ocorrência em final de enunciado, isto é, em uma mesma distribuição possível para um APC, pode, às vezes, provocar dúvidas quanto à sua interpretação, uma vez que o PAR pode também ter um perfil nivelado. Para distingui-las, primeiramente, deve-se observar que se se retirar a unidade de APC percebe-se, frequentemente (do ponto de vista prosódico), a falta de algo para a realização completa do enunciado; o mesmo não ocorre quando da eliminação da

unidade de PAR. Segundo, a unidade de PAR possui sempre valor modal ou, pelo menos, constitui uma intervenção metalinguística, cujo ponto de vista é externo àquele do resto do enunciado. No caso do APC, pode, às vezes, haver uma co-articulação da primeira sílaba do APC com a unidade de COM, algo que não ocorre nas unidades de PAR. Caso haja um elemento modalizador no APC, esse elemento deve repetir o mesmo valor modal de COM. Passemos ao exemplo.

Exemplo 5.32 (bfamdl02) (CD áudio 5.32)
*BEL: [91] então /=DCT= pr' ocê acho que seria /=SCA= muito bom /=COM= **assim** //**=PAR=**

Figura 5.7 – Unidade de PAR em final de enunciado



5.2.3.2.2 APC versus COB

Quando temos, em posição de um possível APC, a presença de um COB (marcado como COM, com valor de F0 baixo) é possível confundi-lo com o APC.

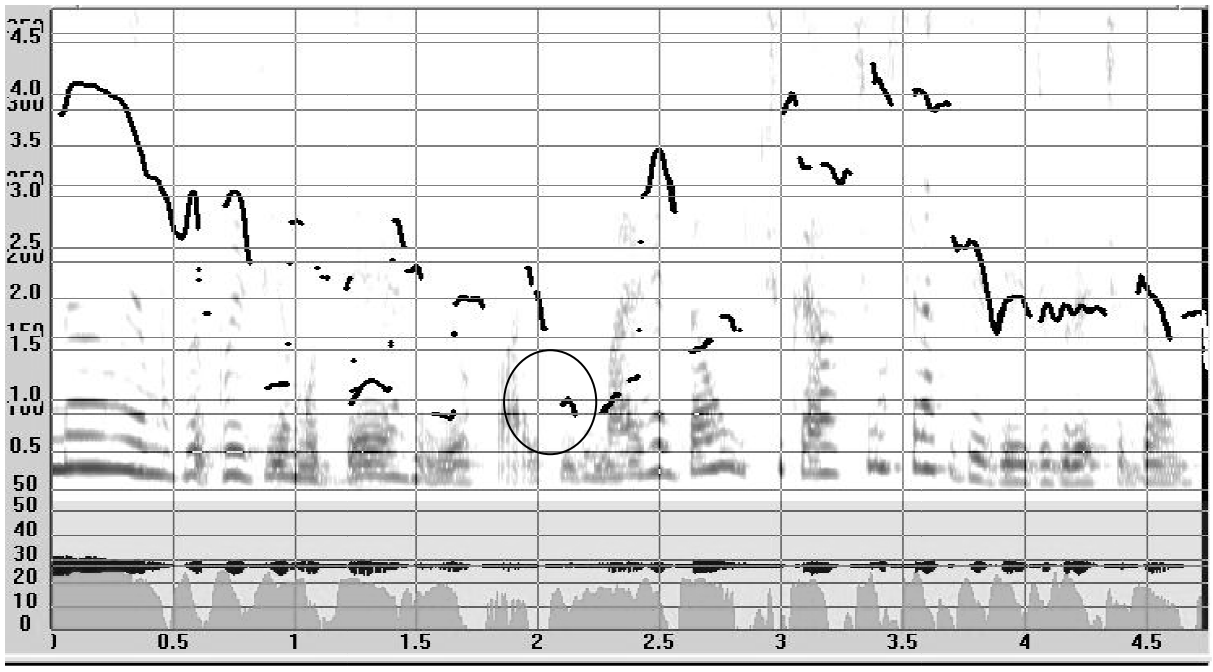
Os COB são unidades ilocucionárias e possuem, portanto, um foco funcional. Às vezes, com unidades pequenas em final de estrofes, esse foco pode não ser claro. Do ponto de vista puramente semântico também pode haver dúvidas, em casos de sintagmas pequenos, se eles devam ser interpretados ilocucionalmente ou não. As

principais pistas para decidir se tornam, portanto, os valores cognitivos de 'dado' e 'novo' e as medidas de F0 e intensidade, eventualmente suportadas perceptualmente. De fato, são características do APC não poder ser cognitivamente novo, bem como de ter médias de F0 e intensidade inferiores àquelas do COM (CMM/COB) do qual são APC. Contudo, não é necessário que um COB seja novo e que tenha valores de F0 e intensidade superiores à locução que o antecede. E, alguns poucos casos pode não ser possível tomar uma decisão certa. Os exemplos a seguir mostram casos em que a decisão é tomada facilmente e casos duvidosos.

Exemplos 5.33 (bfammn02) (CD áudio 5.33)

*DFL: [125] papai /=INT= não /=CMM_r= é &mov [/1]=SCA= poesia moderna /=CMM_r= e meu pai [/3]=EMP= **e tudo mais** /=COB_r= então /=INP= o tio Carlos tinha um certo dever de gratidão /=SCA= com o papai /=COB= porque /=DCT= o papai é que /=SCA= foi <estudar direito no Rio> +=COB=$

Figura 5.8– Representação de 'e tudo mais' como COB

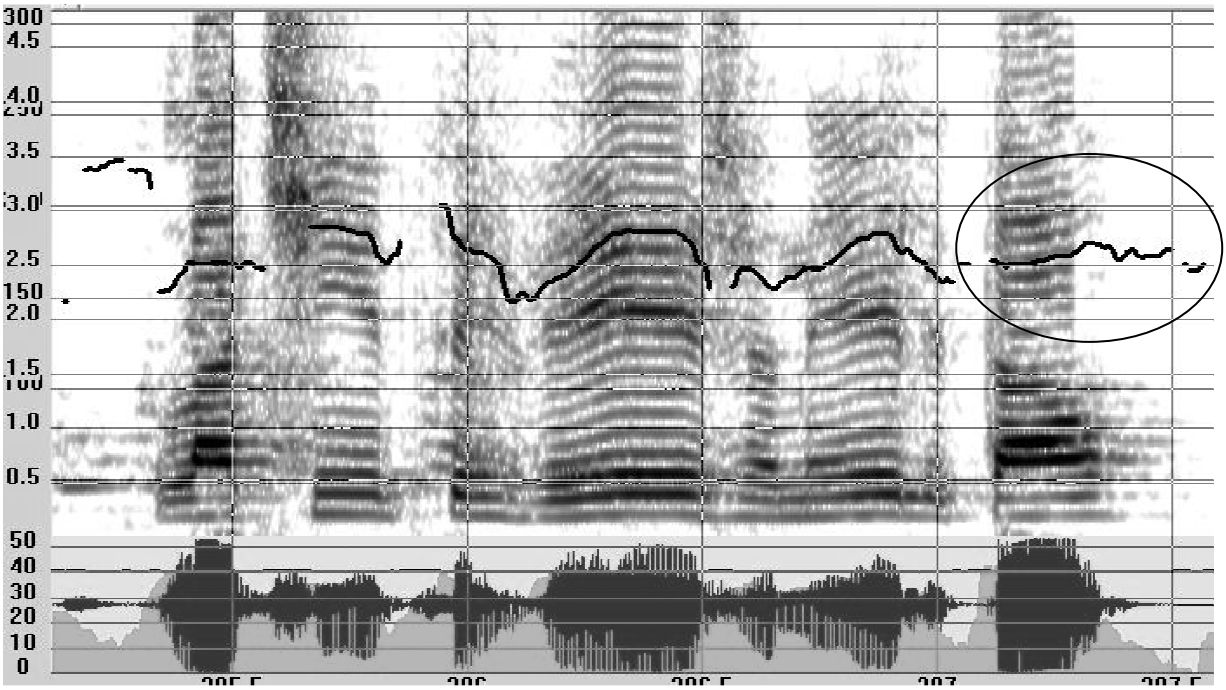


A unidade demarcada pela elipse é um COB, porque essa unidade possui uma força ilocucionária e seu sinal prosódico possui foco. Esse foi um caso facilmente identificável.

Porém, há situações como o da expressão “e tal” que comumente é empregada como APC,com função informacional de preenchimento, mas por vezes pode, realizar a força ilocucionária de um enunciado e ser uma unidade de COM. Examinemos um exemplo.

Exemplos 5.34 (bfamdl02) (CD áudio 5.34)
*BAL: [149] <latinha de> <cerveja> /=COB= nele /=COB= **e tal** //=COM=$

Figura 5.9– Representação de ‘e tal’ como COM

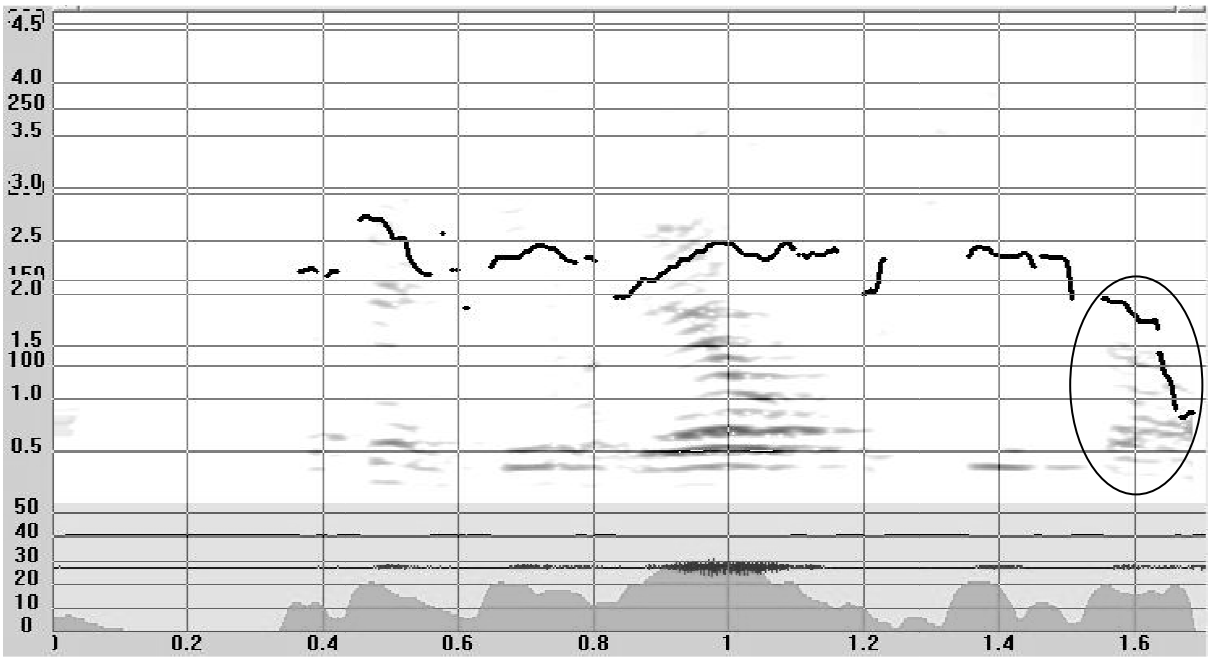


A unidade demarcada pela elipse é um COM. A Unidade de Comentário é uma unidade informacional textual e é a mais importante de todas, pois é a única necessária e suficiente para execução de um enunciado. Sua função é a de realização da força ilocucionária, ou seja, a de cumprir um ato de fala, veiculando a intenção comunicativa.

Exemplos 5.35 (bfamdl02) (CD áudio 5.35)

*BAL: [211] então ele vai em qualquer lugar //=COM=$[212] tá vendo /=CNT= tem o velcro aqui /=COM= **e tal** //=APC=$

Figura 5.10– Representação de 'e tal' como APC



Nesse caso, a unidade demarcada pela elipse não carrega a força ilocucionária do enunciado. Ela apenas funciona como preenchimento da unidade de COM e possui um perfil prosódico nivelado ou descendente, logo trata-se de uma unidade de APC.

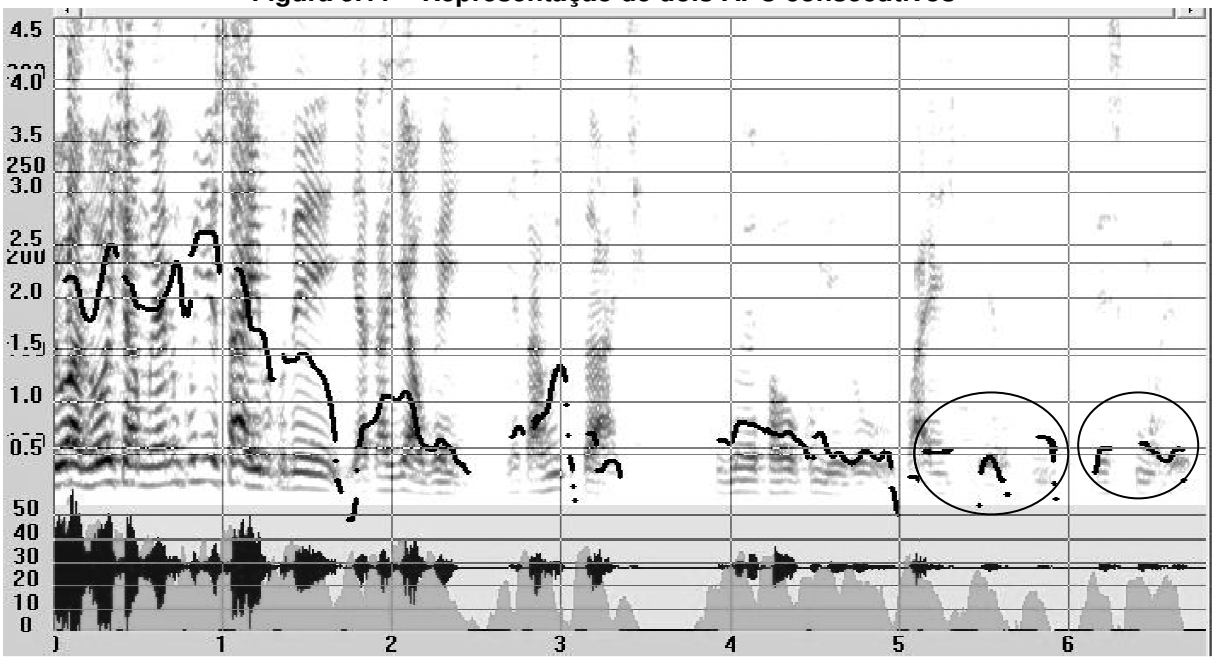
Com base nesses exemplos, podemos reafirmar que se a unidade possui autonomia pragmática, é um COM; se apenas serve como integração textual a uma unidade ilocucionária, é um APC.

5.2.3.3 Dois APC consecutivos ou um APC escansionado?

Nas análises realizadas no subcorpus, só localizamos um exemplo de enunciados que apresentam duas unidades consecutivas de APC, coisa que de fato acontece pouco, também, no italiano e que será esboçado na subseção 5.4.2. Quando isso ocorre, há um acréscimo de uma informação ao APC precedente, sem necessidade de composicionalidade sintática; isto é, cada APC possui informações independentes, podendo funcionar sem a existência do outro. Entonacionalmente, o primeiro APC costuma ser mais alto que o segundo. Infelizmente, como já mencionado na subseção 5.2.2, não foi possível calcular as medidas do segundo APC porque havia creaky voice, mas a média de F0 do COM é 107,7 Hz e a média de intensidade 60,83 dB. O primeiro APC tem média de F0 de 94,85 Hz e média de intensidade 56,65 dB

Exemplo 5.36 (bfammn01) (CD áudio 5.36)
*MAI: [67] segundo ele falou comigo /=COM= **o parente dele** **/=APC=** que &m [/2]=EMP= **desse que morreu** //**=APC**= $

Figura 5.11 – Representação de dois APC consecutivos



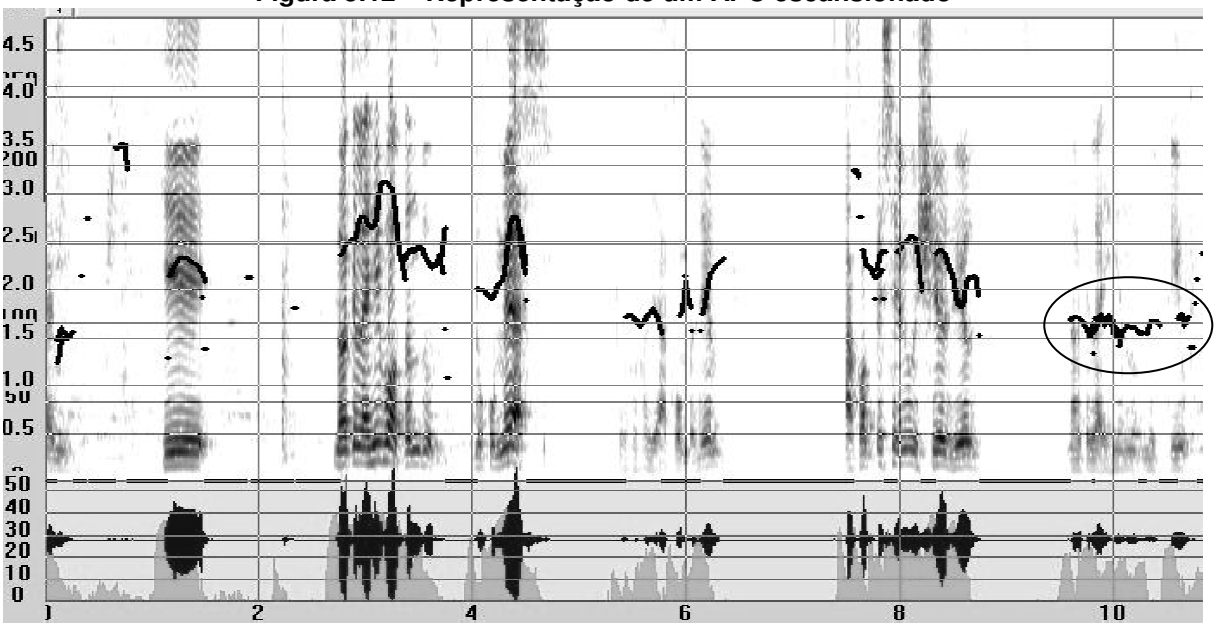
O primeiro APC está delineado por meio da 1ª elipse e tem função de informação tardia; o segundo APC tem a mesma função e está demarcado pela 2ª elipse.

Veja-se, ainda que esse exemplo não é um caso de APC escansionada, porque cada unidade funciona como APC por si só, isto é, não precisam da existência um do outro pra que façam sentido. Não se trata de COB e APC porque realmente a unidade '*o parente dele*' não tem foco e, portanto, não possui força ilocucionária. Ambas as unidades são dadas, como mostra o contexto.

Diferentemente, todavia, acontece quando temos um APC escansionado em mais unidades tonais. A principal distinção entre uma sequência de dois apêndices e um apêndice escansionado está no fato de que o primeiro apresenta um perfil prosódico concluído e o segundo não.

Exemplo 5.37 (bfammn01) (CD áudio 5.37)
*MAI [19] que /=DCT= na hora que ele lá envinha voltando do [/1]=SCA= do comércio /=TOP= que e' foi fazer a compra dele /=PAR= a cobra percebeu o cheiro dele /=COM= **na hora que ele lá envinha no [/1]=SCA= no trilho //=APC=**

Figura 5.12 – Representação de um APC escansionado



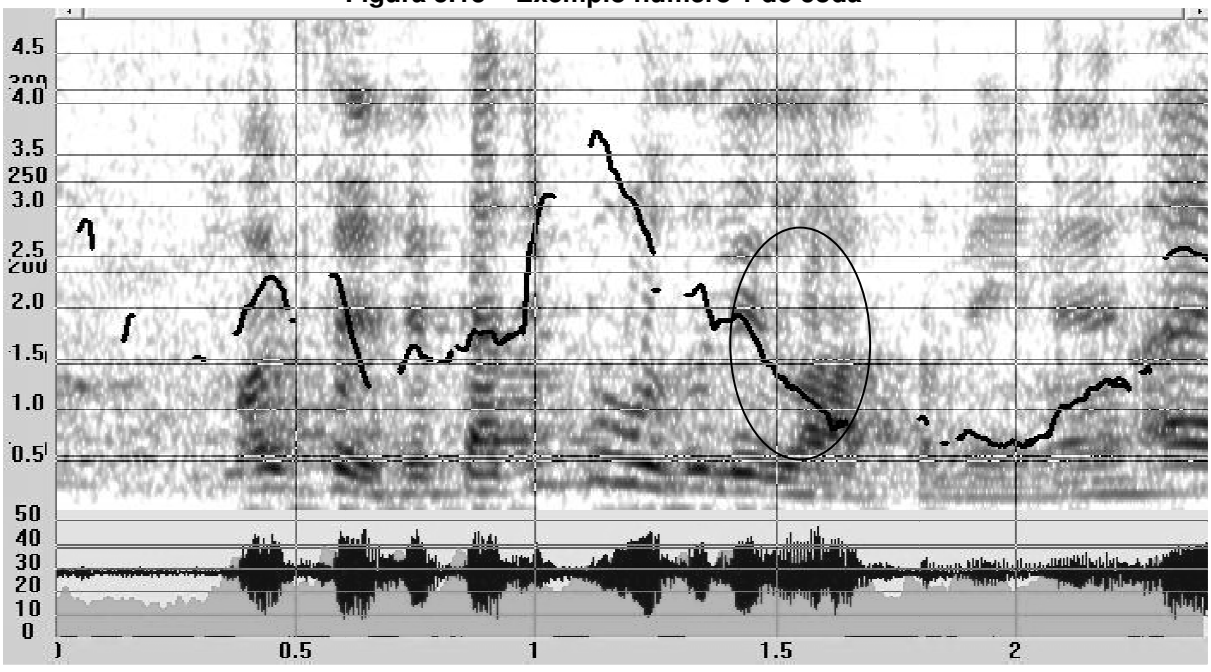
5.2.3 Coda

Outra situação que merece ser mencionada é quando há uma unidade tonal que parece ter todas as características prosódicas de um APC, mas com características informacionais distintas. Isso pode ocorrer quando temos um COM cujo foco funcional é à esquerda e cujo conteúdo locutivo se estenda por várias sílabas de coda. Essa situação faz com que seja impossível ou não natural a realização do COM, sem se produzir uma quebra entre o foco e o restante do conteúdo silábico. Assim sendo, não podemos dizer que depois da quebra haja um APC, já que essa quebra é praticamente obrigatória. Trata-se, portanto, de uma coda que produz uma unidade escansionada (SCA) à direita e não, como é frequente, à esquerda do foco funcional.

A título de exemplo, ilustramos alguns casos dessa ocorrência.

Exemplo 5.38 (bpubdl02) (CD áudio 5.38)
*EUG: [71] olha p' cê ver /=CNT= cê tá de blusa /=SCA= **lilás** /=COM=também /=APC= né //=PHA=$

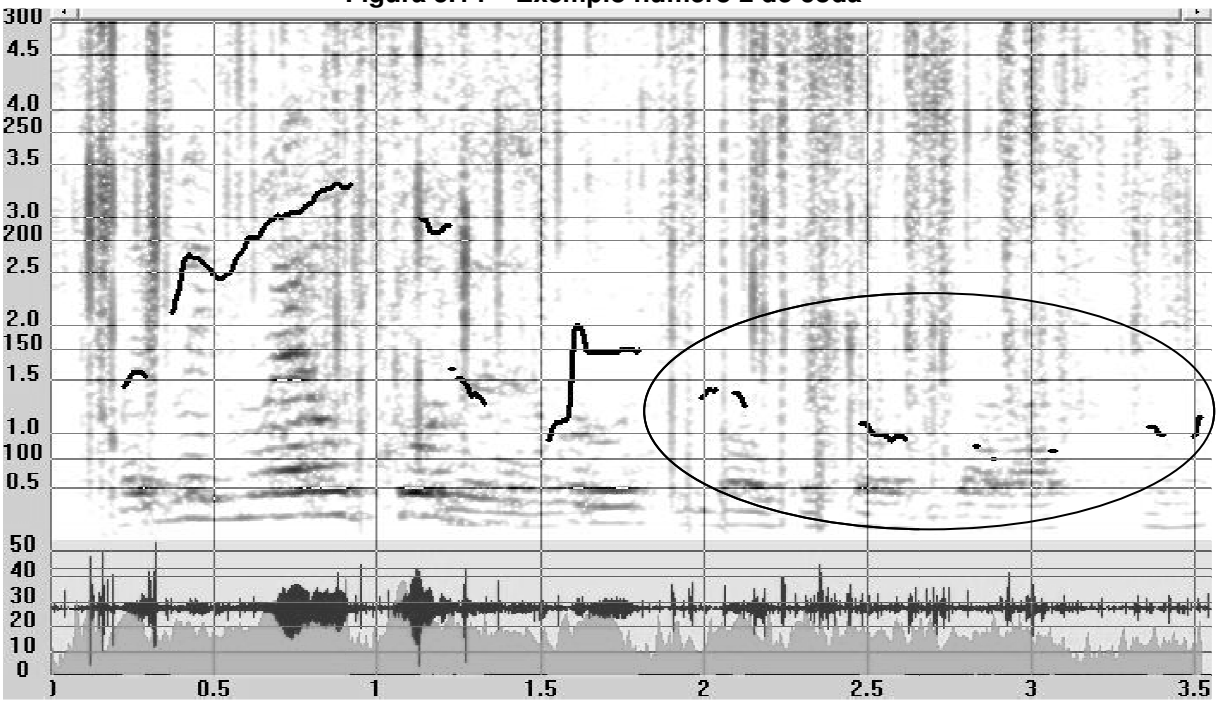
Figura 5.13 – Exemplo número 1 de coda



Exemplo 5.39 (bfamcv01) (CD áudio 5.39)

*LUI: [17] mas nũ [/2]=EMP= mas nũ [/2]=EMP= eles nũ eram todos <escrotos> /=SCA= **igual o pessoal do Galáticos não** //=COM=$

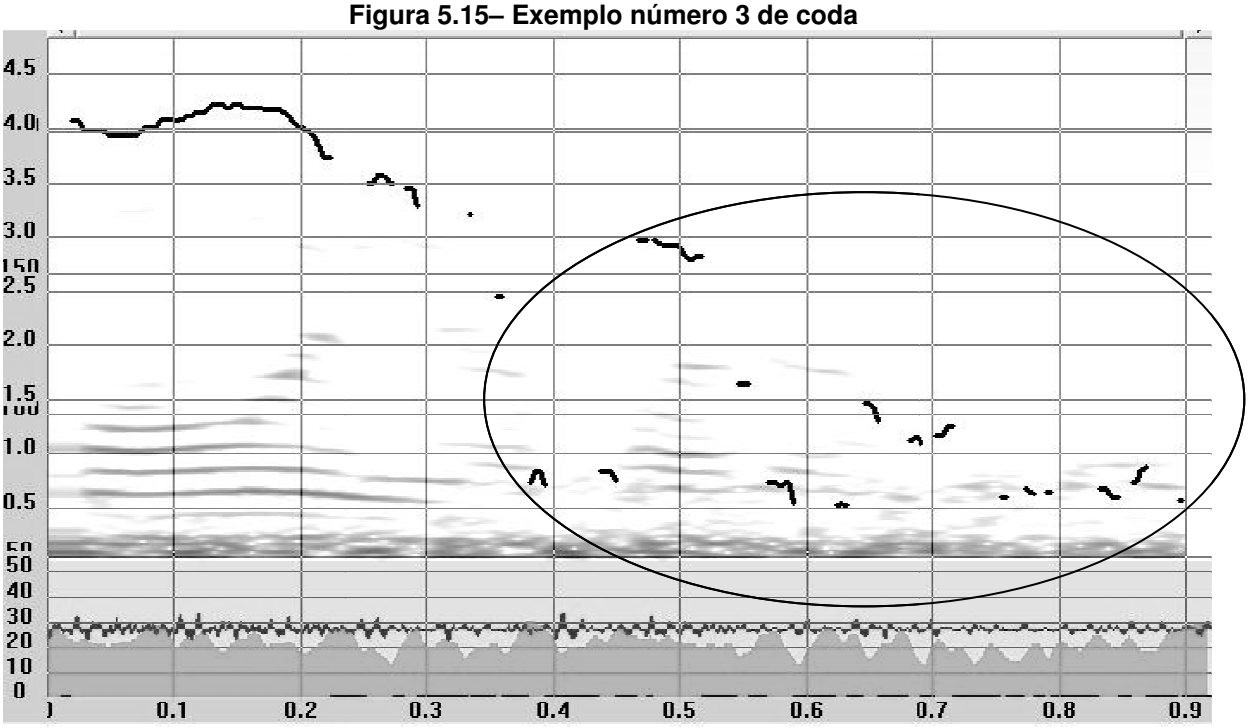
Figura 5.14 – Exemplo número 2 de coda



Exemplo 5.40 (bfamdl05) (CD áudio 5.40)

*ANE: [35] qual é /=SCA= **a paralela** //=COM=$

Figura 5.15– Exemplo número 3 de coda



5.3 Análise Morfossintática

De um ponto de vista lexical, a unidade de APC apresenta uma alta variabilidade morfossintática, conforme se verificou no capítulo 2. Ela pode ser composta por palavras isoladas, quase sempre advérbios, construções sintáticas ou orações. No caso do subcorpus brasileiro, nos detivemos em identificar apenas estruturalmente a unidade de APC e vimos despontar expressões e construções sintáticas. A tabela 5.16 esboça essas ocorrências.

Tabela 5.16 – Frequência de sintagmas sobre o total de APC

	Conversações	%	Diálogos	%	Monólogos	%	Total Parcial	%
Expressões	10	30,3	12	32,4	15	35,7	37	33
Construções Sintáticas	23	69,7	25	67,6	27	64,3	75	67
TOTAL	33	100	37	100	42	100	112	100

A tabela acima nos revela que 67% sobre o total de APC são construções sintáticas e 33% são expressões.

Passamos a esboçar detalhadamente, por meio da tabela 5.17, os correlatos morfossintáticos, explicitando o número de ocorrências por tipologia textual.

Tabela 5.17 - Frequência de expressões nas conversações, diálogos e monólogos

	ADV	%	ADJ	%	ART.	%	PRON.	%	ADV + INTERJ.	%	NUM. + ADV	%
Conversações	7	18,9	-		1		2	5,4	-		-	
Diálogos	9	24,3	1		-		-		1		1	
Monólogos	13	35,1	-		-		2	5,4	-		-	
Total	29	78,3	1	2,7	1	2,7	4	10,8	1	2,7	1	2,7

Legenda

ADV - advérbio
ADJ – adjetivo
ART - artigo
PRON. - pronome
ADV + INTERJ. – advérbio + interjeição
NUM + ADV. – numeral + advérbio

As expressões isoladas, segundo Tucci, E, são constituídas pelos advérbios, adjetivos, nomes próprios e pronomes. Em nossa amostra, todavia, expedem ADV (78,3%) ADJ (2,7%), ART (2,7%), PRON (10,8%), ADV + INTERJ. (2,7%) e NUM + ADV (2,7%) com função de APC.

Como se pode notar, os ADV são os que mais se sobressaem com função de APC em qualquer uma das tipologias, o que já era esperado porque dentre as várias funções do ADV está a de determinar um fato, ampliando a informação nele contida. Outras categorias são bem mais raras.

Analisemos, agora, os correlatos que compõem as construções sintáticas observadas na amostra com função de APC. A tabela 5.18 esboça os resultados.

Tabela 5.18 - Frequência de Construções Sintáticas nas conversações, diálogos e monólogos

	SN	%	SP	%	SV	%	SADV	%	SADV + SN	%	SADJ	%
Conversações	7	9,3	6	8	8	10,6	1	1,3	1	1,3	-	
Diálogos	7	9,3	10	13,3	7	9,3	-		-		1	1,3
Monólogos	10	13,3	9	12	8	10,6	-		-		-	
TOTAL	24	32	25	33,3	23	30,6	1	1,3	1	1,3	1	1,3

Como se pode notar, os SN, SP e SV se dividem, mais ou menos igualmente. Assim sendo, os resultados totais nos permitem afirmar que se desconsideradas as tipologias textuais, estatisticamente não há discrepância entre as classes que assumem função de APC na amostra. Basicamente o que se encontram são SN, SP e SV, seguindo as mesmas tendências gerais da língua italiana no que se refere à codificação morfossintática da unidade de APC (TUCCI, E 2006).

5.4 O mini-corpus italiano

Quando do início de nossa pesquisa sobre a unidade de APC só tínhamos disponíveis, no Italiano, os dados obtidos por meio do estudo realizado por Tucci, E (2006). Em 2011, porém, o laboratório *LABLITA*[53] criou uma ferramenta que nos permitiu visualizar os dados pesquisados por eles, sobre essa unidade textual (http://lablita.dit.unifi.it/app/dbipic/).

Embora estejam disponíveis todos os 81 textos que compõem o C-ORAL-ROM informal, optamos por identificar apenas os dados dos 20 textos do mini-corpus italiano, já que os dois subcorpora possuem a mesma arquitetura e foram compostos pelos grupos italiano e brasileiro com o intuito de estabelecer pesquisas comparativas.

Dos 20 textos do mini-corpus italiano, 14 pertencem ao contexto familiar (03 conversações, 05 diálogos e 06 monólogos) e 6 ao contexto público (2 conversações

[53] Por ora, esta ferramenta em rede, no Brasil, está acessível somente aos integrantes do C-ORAL-BRASIL. No capítulo 3 essa ferramenta foi ilustrada.

e 2 diálogos e 2 monólogos). O número de ocorrências da unidade de APC em cada tipologia textual pode ser observado na tabela 5.19, abaixo.

Os 20 textos são constituídos por um total de 5018 enunciados. Desse total, 2485 são enunciados simples e 526 enunciados complexos com unidades tonais. Desse total de enunciados complexos, apenas 243 contém pelo menos uma unidade de APC.

Tabela 5.19 - ocorrência de APC no mini-corpus italiano

TEXTOS	Nº de APC	%
ifamcv01_corte	42	43
ifamcv09	13	13
ifamcv15	20	20
ipubcv01	18	18
ipubcv05	05	5
Total Parcial	**98**	**100**
ifamdl04	19	23
ifamdl12	09	11
ifamdl15	10	12
ifamdl17	16	20
ifamdl19	06	7
ipubdl02	5	6
ipubdl05	17	21
Total Parcial	**82**	**100**
ifammn02_corte	20	32
ifammn03_corte	13	21
ifammn05	05	8
ifammn08	-	_
ifammn14	03	5
ifammn17	09	14
ipubmn01	03	5
ipubmn04	10	16
Total Parcial	**63**	**100**
TOTAL GERAL	243	100

No Italiano, como se pode notar, a incidência de APC é maior na tipologia conversação, com um percentual de 40% de ocorrências sobre o total. Em seguida estão os diálogos com 34% de ocorrências e, por último, os monólogos com 26% de ocorrências.

O percentual de enunciados simples e complexos do C-ORAL-BRASIL pode ser

comparado à seção informal do C-ORAL-ROM, de acordo com a tipologia da interação. Os textos dialógicos (conversações e diálogos) representam 2/3 do número de enunciados de cada amostra, enquanto a tipologia monológica conta com uma porcentagem de enunciados de 1/3 do total. Dado que conversações e diálogos se comportam da mesma maneira, podem ser considerados juntos na categoria "dialógico".

No italiano, língua que apresenta a maior proporção de enunciados complexos, há 69,5% de enunciados complexos nos monólogos e 30,5% de enunciados simples. Nos diálogos há 52% de enunciados simples e 48,5% de enunciados complexos.

Passamos a apresentar na tabela 5.20 o número de unidades terminadas na amostra, excluíndo-se os enunciados simples e os enunciados complexos composto por uma ou mais unidades dialógicas, seguindo o mesmo critério praticado com os dados do subcorpus do PB.

Tabela 5.20 – Características Informacionais do subcorpus do italiano

Tipologia Informacional	Conversação	Diálogo	Monólogo
Total de unidades terminadas	**1769**	**2054**	**1195**
Total de enunciados	**1481**	**1714**	**842**
Total de enunciados simples	987	1169	329
Total de unidades escansionadas simples (COM + SCA, TMT, EMP)	95	126	90
Total de enunciados complexos com unidades dialógicas	144	178	116
Total de enunciados complexos com unidades textuais	**172**	**168**	**186**
Junção de enunciados com COM + unidades textuais e dialógicas	**83**	**73**	**121**
Total de Padrão ilocucionário	**183**	**172**	**80**
Padrão ilocucionário simples	106	93	25
Padrão ilocucionário simples com unidades escansionadas	23	17	10
Padrão ilocucionário com unidades dialógicas	15	22	14
Total de padrão ilocucionário com unidades textuais	**31**	**28**	**21**
Junção de padrão ilocucionário + unidades textuais e dialógicas	**8**	**12**	**10**
Estrofes	**105**	**168**	**273**

Como se pode observar, o número de unidades terminadas no italiano é inferior (5018) ao número de unidades terminadas (5109) no PB. Entretanto, ao se compararem os enunciados complexos com unidades textuais (546) veremos que nas três tipologias a ocorrência de APC é maior do que no PB (333). O mesmo acontece com as estrofes com unidade de APC. No PB ocorrem 466 casos, enquanto no italiano, 546. Diferentemente, porém, é o comportamento dos padrões ilocucionários nas duas línguas. Enquanto no PB há 504 ocorrências dessa natureza, no italiano encontramos apenas 80.

Tendo delineado as características informacionais do subcorpus italiano, passamos a exibir as unidades terminadas com unidades textuais, padrões ilocucionários e estrofes que são aquelas onde podem haver incidência da unidade de APC.

Na tabela 5.21 será demonstrada a proporcionalidade entre os enunciados complexos da amostra e os enunciados complexos com APC.

Tabela 5.21 – Comparação entre enunciados complexos e enunciados complexos com APC, no italiano

Tipologia	Total de enunciados complexos	Enunciados complexos com APC
Conversações	255/1769 – 14,4 %	81/255 – 31,8%
Diálogos	241/2054 – 11,7%	53/241 – 21,8%
Monólogos	307/1195 - 25,7%	49/307 – 16%

Como se pode notar, no italiano, como no PB, há mais enunciados com unidades terminadas nas conversações e diálogos (26,1%) do que nos monólogos (25,7%), embora a diferença percentual não seja relevante.

Em compensação, quando comparados os resultados dos enunciados com unidades terminadas mais APC aos enunciados com unidades terminadas, observamos que diferentemente do PB, os resultados são bastante diferentes; isto é, os diálogos representam 53,6% sobre o total de enunciados complexos com APC, enquanto nos monólogos só há 16% de incidência.

Esse resultado nos permite afirmar, então, que no italiano os APC são muito mais frequentes nos diálogos do que nos monólogos, enquanto no PB ocorre exatamente o oposto.

Agora passemos a examinar os padrões ilocucionários com unidades textuais.

Tabela 5.22 – Comparação entre padrões ilocucionários com unidades textuais e padrões ilocucionários com APC, no italiano

Tipologia	Total de padrões ilocucionários com unidades textuais	Padrões ilocucionários com APC
Conversações	40/1769 – 2,3%	11/40 – 27,5%
Diálogos	35/2054 – 1,7%	8/35 – 22,9%
Monólogos	31/1195 – 2,6%	10/31 – 32,3%

Os dados da tabela acima nos permitem afirmar que há mais incidência de padrões ilocucionários com unidades textuais entre os diálogos (conversações e monólogos)

com 4% de ocorrências sobre o total de unidades terminadas, do que nos monólogos, com apenas 2,6% de frequência.

O mesmo ocorre quando da frequência de padrões ilocucionários com APC, sendo esses mais salientes nos diálogos (conversações e diálogos) com 50,4% de ocorrências e os monólogos com 32,3% de ocorrências sobre o total analisado.

Verifiquemos as estrofes. A tabela 5.23 ilustrará os dados.

Tabela 5.23 – Comparação entre estrofes e estrofes com APC, no italiano

Tipologia	Total de ESTROFES	Estrofes com APC
Conversações	105/1769 – 5,9%	4/105 – 3,8%
Diálogos	168/2054 – 8,2%	3/168 – 1,8%
Monólogos	273/1195 – 22,8%	7/273 – 2,6%

Assim como em PB, as estrofes são mais comuns nos monólogos representando 22,8% sobre o total de unidades terminadas. Os diálogos (conversações e diálogos) representam 14,4% de ocorrências sobre o total analisado.

Ao se analisarem as estrofes com APC, verifica-se que a maior incidência está nos diálogos (conversações e diálogos) com 5,6% de ocorrências sobre o total analisado. Nos monólogos aparecem apenas 2,6%. Essa diferença, todavia, pode ser explicada pela própria natureza da tipologia textual, conforme já fora mencionado quando da referência a essas unidades terminadas em PB.

5.4.1 Distribuição do APC com unidades intercaladas

Assim como no PB, o italiano também possui unidades intercaladas entre a unidade ilocucionária de COM (CMM, COB) e APC. Vejamos a tabela 5.24 que ilustra os valores encontrados.

Tabela 5.24 – Número de ocorrência dos APC com unidades intercaladas no italiano

TEXTOS	COM/ALL/APC	%	COM/PAR/APC	%	COM/PHA/APC	%
Conversações	-		-		2	
Diálogos	1		4		3	
Monólogos	-		-		2	

Abaixo, enumero alguns exemplos para ilustrar os dados da tabela acima.

Exemplo 5.41 (ifamcv01_corte,454)
*ELA,"è una ragazzina bella /=COM= eh /=PHA= questa //=APC="
(*ELA: é uma moca bonita /=COM= eh /=PHA= essa //=APC)

Exemplo 5.42 (ifamdl04)
*ART,"'un guardo /=SCA= quando parte /=SCA= un pezzettino di pelle /=SCA= che collega /=SCA= il fianco al quadrante /=COM= ma insomma /=PHA= questo poi < xxx > //=APC="
(*ART: um olhar /=SCA= quando parte /=SCA= um pedacinho de pele /=SCA= que conecta /=SCA= o lado ao conjunto /=COM= mas enfim /=PHA= isso também //=APC=)

Exemplo 5.43 (ifamdl04)
*ART,"che [/1]=EMP= che /=SCA= lucida /=COM= la pelle /=APC= < se è troppo grinzosa > /=PAR= eccetera //=APC="
(*ART: que [1]=EMP= que /=SCA= ilumina /=COM= a pele /=APC= se é muito rugosa /=PAR= etecetera //APC=)

5.4.2 Dois Apêndices

Conforme se poderá observar, a presença de dois apêndices é muito mais comum no italiano do que no PB, mesmo sendo uma ocorrência rara. Enquanto em nossa amostra só há um caso desse tipo, sobre 112 exemplos analisados, no mini-corpus italiano foram localizados seis exemplos, sobre 243 unidades terminadas analisadas.

A título de exemplo, ilustramos alguns casos.

Exemplo 5.44 (ifamcv01_corte,280)
*ELA,"dove è /=COM= questo /=APC= anche //=APC="
(*ELA: onde está /=COM= esse /=APC= também /=APC=)
Exemplo 5.45 (ifamcv01_corte,362)
*MAX,"dovevi esse' un po' ingrassata /=COM= però /=APC= qui //=APC="
(*Max: você devia ter engordado /=COM= contudo /=APC= aqui //=APC=)

Dentre todos os enunciados analisados, há um que merece ser comentado por se tratar de uma situação atípica em relação à etiquetagem, conforme se verificará abaixo.

Exemplo 5.46 (ipubdl05)
*GCM,"&he /=TMT= tu contestualizzi /=COB= < l' opera /=APC= tu &comin +=EMP="
(*GCM: é /=TMT= você contextualize /=COB= <o trabalho /=APC= você &comu+=EMP=)

Veja-se que nesse exemplo, ao etiquetar a primeira unidade, após a tomada de tempo, há um COB ao invés de um COM, como seria comum de ocorrer, porque em princípio um COB deveria vir associado a pelo menos uma unidade de COM. Porém, ao escutar o enunciado, percebeu-se que a unidade "*tu contestualizzi*", carregava um sinal de continuidade. Assim sendo, a equipe do Lablita optou em etiquetá-la como COB e não COM. O mesmo aconteceu com um único enunciado do PB. Veja:

Exemplo 5.47 (bfammn02)
*DFL: [164] e quando eles /=SCA= vinha trazer a Maria Julieta /=TOP= ele ficava uns poucos dias depois vinham buscar /=PAR= e então /=DCT= ele participava /=COB= **daquelas nossas brincadeira** /=APC=

5.5 Análise Contrastiva: APC no PB versus APC no italiano

A fim de estudar a estrutura informacional em uma perspectiva interlinguística, buscou-se analisar o comportamento da unidade de APC no PB e no italiano. Como o subcorpus brasileiro é altamente acional, para se estabelecer um parâmetro de comparação foi necessário manter a mesma proporção entre as tipologias diálogo e monólogo, e maximinizar o número de atividades realizadas pelo falante no momento da interação.

Um primeiro parâmetro de comparação entre as duas línguas a se fazer é a proporção de unidades terminadas. Vejamos a tabela 5.25.

Tabela 5.25 – Características Informacionais dos corpora de PB e Italiano

	Português Brasileiro						Italiano					
Tipologia Informacional	Cv	%	Dl		Mn		Cv		Dl		Mn	
Total de unidades terminadas	**1855**	**100**	**2304**	**100%**	**950**	**100%**	**1769**	**100%**	**2054**	**100%**	**1195**	**100%**
Total de enunciados	**1534**	**82,7%**	**1972**	**85,6%**	**633**	**66,6%**	**1481**	**83,7%**	**1714**	**83,4**	**842**	**70,5**
Total de enunciados simples	1095	71,4%	1452	73,6%	351	55,5%	987	66,6%	1169	68,2	329	39,1
Total de unidades escansionadas simples (COM + SCA, TMT, EMP)	91	5,9%	121	6,1%	63	10%	95	6,4%	126	7,4	90	10,7
Total de enunciados complexos com unidades dialógicas	196	12,8%	232	11,8%	63	10%	144	9,7%	178	10,4	116	13,8
Total de enunciados complexos com unidades textuais	**108**	**7%**	**125**	**6,3%**	**100**	**15,8%**	**172**	**11,6%**	**168**	**9,8**	**186**	**22,1**
Junção de enunciados com COM + unidades textuais e dialógicas	**44**	**4%**	**42**	**2,9%**	**56**	**16%**	**83**	**8,4%**	**73**	**6,2**	**121**	**36,8**
Total de Padrão ilocucionário	**202**	**10,9%**	**225**	**9,8%**	**77**	**8,1%**	**183**	**10,3%**	**172**	**8,4**	**80**	**6,7**
Padrão ilocucionário simples	147	72,8%	148	65,8%	34	44,2%	106	57,9%	93	54,1	25	31,3
Padrão ilocucionário simples com unidades escansionadas	13	6,4%	19	8,4%	10	13%	23	12,6%	17	9,9	10	12,5
Padrão ilocucionário com unidades dialógicas	24	11,9%	30	13,3%	8	**10,4%**	**15**	**8,2%**	**22**	**12,8**	**14**	**17,5**
Total de padrão ilocucionário com unidades textuais	**14**	**6,9%**	**20**	**8,9%**	**21**	**27,3%**	**32**	**16,9%**	**28**	**16,3**	**21**	**26,3**
Junção de padrão ilocucionário + unidades textuais e dialógicas	**4**	**2%**	**8**	**3,6%**	**4**	**5,2%**	**8**	**4,4%**	**7**	**7**	**10**	**12,5**
Estrofes	**119**	**6,4%**	**107**	**4,6%**	**240**	**25,3%**	**105**	**5,9%**	**168**	**8,2**	**273**	**22,8**

Tendo em vista a tabela acima, observamos que no PB há 82,7% de enunciados nas conversações, 85,6% nos diálogos e 66,6% nos monólogos. No italiano o percentual é parecido, à exceção dos monólogos. Nas conversações há 83,7% de enunciados, nos diálogos 83,4% e nos monólogos, 70,5%. Esses resultados nos permitem afirmar que no italiano os textos dialógicos se comportam de maneira similar aos textos do PB, havendo proporcionalidade de enunciados quando da comparação. Os textos monológicos, entretanto, apresentam medidas muito diferentes, permitindo-nos aventar a hipóstese de que as pequenas diferenças encontradas entre as duas tipologias no subcorpus do PB são devidas à presença de duas conversações em que os falantes não realizam qualquer atividade levando, portanto, algumas medidas na direção dos monólogos.

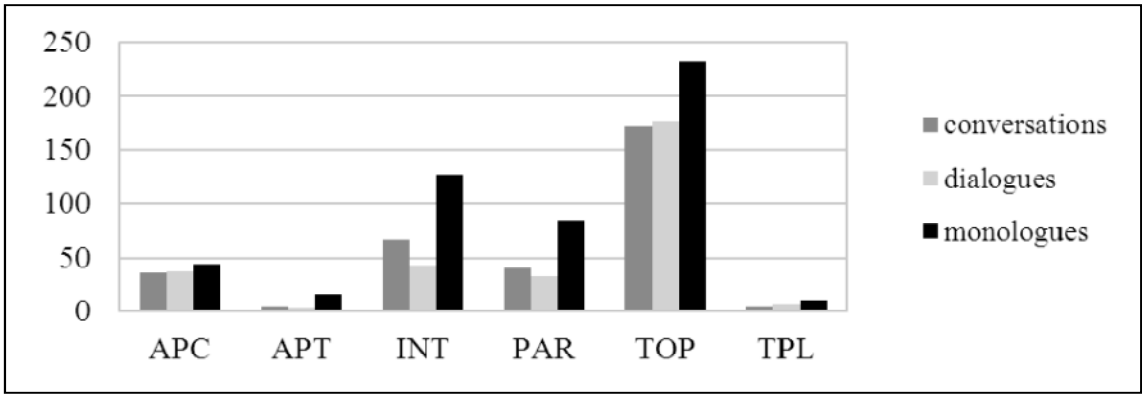
Quanto aos monólogos no italiano, a diferença fundamenta-se no fato de haver nessa tipologia tanto menos estrofes e padrões ilocucionários, quanto mais enunciados.

É interessante notar que a diferença mais significativa entre uma língua e outra está no fato de que em termos percentuais o italiano tem menos enunciados simples que o PB. Enquanto no PB há 71,4% de enunciados simples nas conversações, 73,6% nos diálogos e 55% nos monólogos, no italiano essas medidas são, respectivamente, 66,6%, 68,2% e 39,1%, o que nos conduz ao fato de que no italiano há mais enunciados complexos do que no PB. Esta hipótese é reforçada pelo fato de que as unidades textuais também são superiores no italiano, conforme se poderá verificar no gráfico 5.2.

Quanto aos enunciados complexos com unidades textuais, enquanto na amostra brasileira o percentual é de 11%, 9,2% e 31% de enunciados complexos com unidades textuais para conversação, diálogo e monólogo, no italiano, respectivamente, encontramos 20%, 16% e 58,9%. O mesmo acontece com os padrões ilocucionários. No italiano, os padrões ilocucionários são mais comuns, enquanto os padrões ilocucionários simples são mais encontrados no PB. A diferença quanto as estrofes não parece muito significativa.

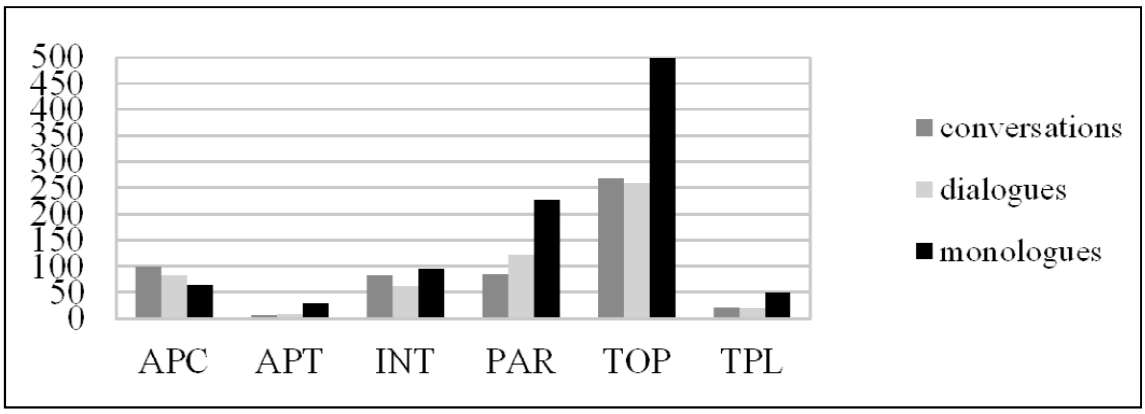
Verifiquemos, agora, a distribuição das unidades textuais, por tipologia, em ambas as línguas. O gráfico 5.1 ilustra os valores relativos ao PB e o gráfico 5.2, ao italiano.

Gráfico 5.1 – distribuição das unidades textuais, por tipologia, no PB



Fonte: Raso; Mittmann (2012 p. 167)

Gráfico 5.2 – distribuição das unidades textuais, por tipologia, no italiano



Fonte: Raso; Mittmann (2012 p. 173)

Analisando o gráfico 5.2 percebemos que no italiano, a distribuição de INT não varia muito nas três tipologias, mesmo que nos monólogos haja mais INT e nos diálogos, menos. No PB, todavia, gráfico 5.1, o número de ocorrências do INT em monólogos é muito mais elevado do que nas outras duas tipologias. Uma hipótese aventada por Raso; Mittmann (2012) refere-se ao fato de que no Brasil o uso de discursos reportados é mais frequente.

Outra diferença interessante entre os dois subcorpus diz respeito à inversão da distribuição dos APC nas três tipologias. Enquanto no PB há mais APC nos monólogos e menos nas conversações, no italiano os APC são mais comuns nas conversações e menos nos monólogos.

Tendo sido estabelecidas as diferenças entre as unidades textuais, passamos a verificar o comportamento da unidade de APC, nas duas línguas, em relação à unidade ilocucionária de COM (CMM e COB) a qual ela é integrada. Vejamos a tabela abaixo.

Tabela 5.26 – Total de unidades terminadas com APC no PB e italiano

TIPOLOGIA TEXTUAL	ENUNCIADOS COMPLEXOS COM APC		PADRÕES ILOCUCIONÁRIOS TEXTUAIS C/ APC		ESTROFES C/ APC		Total Geral (EC+PI+Ef) c/APC	
	PB	IT	PB	IT	PB	IT	PB	IT
Conversação	25/152 16,4%	81/255 31,8%	4/18 22,2%	11/40 27,5%	2/119 1,7%	4/105 3,8%	31/289 40,3%	96/399 63,1%
Diálogo	32/167 19,2%	53/241 22%	4/28 14,3%	8/35 22,9%	1/107 0,9%	3/168 1,8%	37/302 34,4%	64/449 46,7%
Monólogo	31/156 20%	49/307 16%	2/25 8%	10/31 32,3%	9/240 3,8%	7/273 2,6%	42/421 31,8%	66/611 50,9%
TOTAL	88/475 18,5%	183/803 22,8%	10/71 14%	29/106 27,4%	12/466 2,6%	14/546 2,6%	110/1012 10,9%	226/1459 15,5%

Legenda
EC = enunciados complexos
PI = padrão ilocucionário
Ef = estrofe
c/ = com

Enquanto no italiano há 243 ocorrências de APC sobre 1018 unidades terminadas com unidades textuais, no PB encontramos apenas 112 APC sobre 1012; isto é, a unidade de APC é muito mais recorrente no italiano, pois apresenta mais que o dobro (58%) dos APC encontrados no PB.

Em relação aos enunciados complexos com unidades textuais, o que se verifica é que há 3,1% a mais de ocorrência de APC, no italiano, na tipologia diálogo (conversação e diálogo) e nos monólogos apenas 0,8%. Nos padrões ilocucionários há mais 2,8% de ocorrências de APC no italiano do que no PB nos diálogos

(conversações e diálogos), e 4%, nos monólogos. E em relação às estrofes, no italiano há mais 4,2% de ocorrências de APC do que o PB nos diálogos (conversações e diálogos), e 0,7%, nos monólogos.

5.6 Resumindo

Este capítulo objetivou a análise das unidades terminadas com ocorrência de APC nas duas línguas, PB e italiano.

Expôs, também, a análise informacional, entonacional e distribucional dos dados do subcorpus brasileiro.

Por fim, estabeleceu-se uma análise contrastiva entre as duas línguas em relação ao comportamento da unidade de APC nas tipologias conversação, diálogo e monólogo.

Capítulo 6

CONCLUSÕES

Esse trabalho teve como objetivo estudar a unidade informacional de Apêndice de Comentário, tal qual descrita pela Teoria da Lingua em Ato (Cresti, 2000), em uma perspectiva baseada em *corpus*. Os aspectos enfocados nesse estudo foram: (a) a descrição da forma entonacional de APC; (b) os critérios funcional, entonacional e distribucional que regem a unidade de APC; (d) as características informacionais da unidade de APC; (e) os correlatos morfossintáticos que desempenham função de APC e (f) a análise contrastiva sobre a incidência da unidade de APC no PB e no italiano.

6.1. A definição de APC e seus critérios de ocorrência

Os dados dessa pesquisa foram extraídos do subcorpus que compõe o C-ORAL-BRASIL (Raso, Mello 2012), constituído de 20 textos etiquetados informacionalmente. Para o italiano, também foi compilado um *subcorpus* de 20 textos extraídos do C-ORAL-ROM.

No *subcorpus* de PB foram localizadas 112 unidades informacionais de APC, enquanto que no italiano o número foi de 243 ocorrências. A análise prosódica dos APC encontrados em ambas as línguas revelou a existência de apenas uma forma entonacional (Cresti, 2000; Firenzuoli, 2003) como sendo a de um perfil nivelado e descendente para essa unidade textual.

Observou-se, também, que embora seja uma unidade textual (e ocupe a quarta posição entre as demais unidades textuais no PB, com 10% de ocorrências, sendo mais frequente nos monólogos) , a unidade de APC tem como função apenas integrar a unidade de COM, já que não serve de âmbito para aplicação da força ilocucionária como a unidade de TOP, não possui função metalinguística como o

PAR, nem introduz uma metailocução como o INT, tampouco possui autonomia pragmática como a unidade de COM.

Funcionalmente deve estar posicionada após a unidade da qual faz a integração, a unidade de COM (CMM e COB) e pode exercer a função de informação tardia, repetição, retomada textual ou preenchimento (TUCCI, E 2006).

Na tipologia conversação, a informação tardia foi a classificação informacional que mais se destacou com 37,5% de ocorrências. Em seguida encontramos os preenchimentos (34,4%), as retomadas textuais (15,6%) e, por último, as repetições (12,5%). Nos diálogos, há a predominância dos preenchimentos (47,2%), as informações tardias (36,1%), repetições (11,1%) e retomadas (5,5%). Os monólogos procedem de maneira semelhante aos diálogos. Os preenchimentos destacam-se com 45,2% de ocorrências, as informações tardias com 35,7%, as repetições com 16,6% e as retomadas textuais com 2,4%.

Em âmbito geral, notou-se que as unidades de APC no PB desempenham mais a função de preenchimento, com 43% de ocorrências sobre o total de APC, seguido pela informação tardia com 36%, depois as repetições com 14% e, por último, as retomadas textuais incidindo em apenas 7% sobre o total de APC na amostra.

Entonacionalmente, são características da unidade de APC possuir as médias de F0 e intensidade inferiores às médias da unidade de COM e abaixamento do tom de voz. Nessa unidade, nenhum movimento funcional é encontrado, uma vez que ela visa à integração de novas estruturas linguísticas de modo a realizar uma expressão semântica, uma correção ou mesmo uma reestruturação de um enunciado. Do total de 112 APC localizados na amostra, 87% possuem curva analisável e apenas 13% não-analisável.

Distribucionalmente, a unidade de APC localiza-se após a unidade de COM, tendo como função informacional mais comum, nas **conversações,** a informação tardia, (37,5%), os preenchimentos (25%) e as repetições e retomadas textuais juntas (18,8%). Quanto aos **padrões ilocucionários** (CMM), a função mais recorrente é a de retomada textual (6%), repetições e preenchimentos totalizam (6%). Nas **estrofes**

(COB), a única função informacional encontrada foi a de preenchimento, com (6%) de ocorrências. Quanto aos **diálogos**, após a unidade de COM, os preenchimentos são mais recorrentes (41,7%), depois aparecem as informações tardias (33,3%), as repetições (8,3%) e as retomadas textuais (2,7%). Nos **padrões ilocucionários**, os preenchimentos se destacam com (5,5%), em seguida com o mesmo valor estão as repetições (2,7%) e informações tardias (2,7%). Nas **estrofes**, a única função exercida pelo APC é de retomada textual (2,7%). Quanto aos **monólogos**, após a unidade de COM, as informações tardias e preenchimentos possuem o mesmo valor percentual (26,2%) cada uma. Em seguida estão as repetições com (16,7%) e as retomadas textuais com apenas (2,4%). Nos **padrões ilocucionários** (CMM), a maior função desempenhada pelo APC é de preenchimento (7,1%), depois de informação tardia (2,4%). Nas **estrofes** (COB), a categoria informacional mais saliente é a de preenchimento (11,9%), seguida apenas pela informação tardia (7,1%).

Distribucionalmente, ainda nos atentamos para o fato de que entre a unidade de Comentário e a unidade de APC podem aparecer intercaladas as seguintes unidades: alocutivos, fáticos, conativos, expressivos e parentéticos; todavia, em nosso estudo, só foram veificadas as presenças das unidades de alocutivo, parentético e fático.

Observou-se, também, que há determinados contextos em que duas outras unidades podem ocupar a mesma posição do APC, com um perfil entonacional às vezes parecido, e serem confundidas com o APC. Tratam-se das unidades de PAR e COB. Para distinguir uma unidade de PAR de uma de APC, primeiramente, deve-se observar que se se retirar a unidade de APC percebe-se, frequentemente (do ponto de vista prosódico), a falta de algo para a realização completa do enunciado; o mesmo não ocorre quando da eliminação da unidade de PAR. Segundo, a unidade de PAR possui sempre valor modal ou, pelo menos, constitui uma intervenção metalinguística, cujo ponto de vista é externo àquele do resto do enunciado.

Quanto aos COB em posição de um possível APC, as principais pistas para decidir se tornam o valor cognitivo de ‘novo’ e as medidas de F0 e intensidade.

Outra situação que merece ser mencionada é a da coda. A coda ocorre quando há uma unidade tonal que parece ter todas as características prosódicas de um APC, mas com características informacionais distintas. Isso pode acontecer quando temos um COM cujo foco funcional é à esquerda e cujo conteúdo locutivo se estenda por várias sílabas de coda. Essa situação faz com que seja impossível ou pelo menos não natural a realização do COM, sem que se produza uma quebra entre o foco e o restante do conteúdo silábico; assim sendo, não há como afirmar que depois da quebra haja um APC, já que essa quebra é praticamente obrigatória; trata-se, portanto, de uma coda que produz uma unidade escansionada (SCA) à direita e não, como é frequente, à esquerda do foco funcional.

Morfossintaticamente, 67% do total de APC analisados nesse estudo são construções sintáticas e 33% são expressões, sendo os ADV os que mais se sobressaem com função de APC em qualquer uma das tipologias. Esse resultado já era esperado, porque é sabido que dentre as várias funções do ADV está a de determinar um fato, ampliando a informação nele contida, função essa desempenhada pelo APC. Outras categorias são bem mais raras, em PB.

Interessante destacar, ainda, a distinção estabelecida entre uma sequência de dois apêndices e um apêndice escansionado. Enquanto o primeiro apresenta um perfil prosódico concluído , o segundo, não.

6.2 Análise contrastiva entre APC em PB e Italiano

Os resultados desse estudo apontaram para o fato de que a unidade de APC é muito mais recorrente no italiano, pois apresenta mais que o dobro (58%) dos APC encontrados no PB. Contudo, essa diferença se dá unicamente pelo maior número de unidades terminadas com unidades textuais encontradas nessa língua. Assim sendo, podemos afirmar que a única diferença significativa entre as duas línguas está no fato de que no PB os APC são mais salientes nos monólogos, enquanto no italiano despontam nos diálogos (diálogos e conversações).

Em síntese,

- O italiano mostra uma presença muito maior de **unidades terminadas complexas,** enquanto o PB apresenta muito mais enunciados simples. Esse é um aspecto muito interessante para a comparação entre as duas línguas.
- Mesmo considerando como baseline somente os enunciados complexos, a porcentagem de APC em italiano **é quase 50% a mais do que em PB.**
- O maior número de APC se deve principalmente ao padrões ilocucionários, mas também aos enunciados. Não há diferença quanto às estrofes.
- É interessante notar que nas estrofes dos monólogos, quando a fala necessariamente se complexifica, os APC do PB superam os do italiano nos monólogos.
- No italiano, há quase 50% a mais de APC do que no PB sobre a porcentagem de unidades terminadas. A presença absoluta de APC é na realidade mais que o dobro, já que a estrutura do enunciado no italiano é muito mais complexa do que a do PB.
- DO **ponto de vista entonacional**, não aparecem diferenças entre as duas línguas.
- Do **ponto de vista informacional**, as proporções entre as várias funções nas duas línguas são perfeitamente comparáveis.
- Do **ponto de vista morfossintático**, a comparabilidade entre as duas línguas é estrita.

Assim sendo, a pesquisa

1. confirma a análise da fala espontânea do PB, com base em uma teoria elaborada a partir do italiano;
2. confirma em detalhes as características informacionais, entonacionais e morfossintáticas da unidade de APC
3. observa que essa unidade é menos presente no PB do que em italiano, e que essa menor presença se justifica pelo fato de, em geral, a estrutura da fala

em italiano aparecer mais rica de unidades textuais do que a fala em PB (à exceção do INT)

REFERÊNCIAS BIBLIOGRÁFICAS

AUDACITY TEAM. Audacity. Retrieved from http://audacity.sourceforge.net/. Acesso em janeiro, 2012.

ALVES DE DEUS. (2008). A Unidade Informacional de Tópico no Português do Brasil. Faculdade de Letras da UFMG. (Dissertação de Mestrado).

AUSTIN, J. (1962). *How to do things with words*. Oxford: Clarendon. Bernardo (a c. di), Logica deontica e semantica, Bologna, Il Mulino, pp. 147-165. Bernard Pottier (2000).Représentations mentales et catégorisations linguistiques, Louvain e Paris, Peeters, 2000.

BALLY, C. (1932). *Linguistique générale et linguistique française*. Berna: Francke Verlag.

BIBER, D., Conrad, S. ; Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.

BICK, E. (2012). A anotação gramatical do C-ORAL-BRASIL. In: RASO, T.; MELLO, H. C-ORAL-BRASIL. Corpus de referência da fala espontânea informal do português do Brasil. Belo Horizonte: UFMG.

BICK, E. (2000). The Parsing System “Palavras”: Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework. Aarhus University. Aarhus, Denmark: Aarhus University.

BICK, E. MELLO, H. ; RASO, T. (2010). Annotation parameters for a spontaneous speech corpus. IX Encontro de Linguistica de Corpus, 8-9 de outubro de 2010. Anais... . Porto Alegre.

BLANCHE-BENVENISTE, C.; BILGER, M.; ROUGET, Ch.; VAN DEN EYNDE, K.; MERTENS, P. (1990). *Le Français Parlé*: Études Grammaticales. Paris: Éditions du C.N.R.S.

BOERSMA, P.; WEENINK, D. Praat: doing phonetics by computer. Retrieved from http://www.praat.org/. Acesso em janeiro, 2012.

BUHMANN, J.; CASPERS, J.; HEUVEN, V. J., HOEKSTRA, H., MARTENS, J.P. , SWERTS, M. *et al.* (2002). Annotation of prominent words, prosodic boundaries and segmental lengthening by non-expert transcribers in the spoken Dutch corpus. In: *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC 2002).* Paris: ELRA, p. 779-785. Disponível em: < http://www.lrec-conf.org/proceedings/lrec2002/>

CASTILHO, A. T. de (Ed.). (1993). *Gramática do Português Falado*, vol. 3, Campinas: Editora da Unicamp.

CRESTI, E. (2010). The definition of Focus in Language into Act Theory. V° LABLITA workshop &II° Brazilian Seminar on Pragmatics and Prosody: Illocution, modality, attitude, information patterning and speech annotation. **Anais...** . Belo Horizonte.

CRESTI, E. (2009). La stanza: un´unità di costruzione testuale del parlato. Em: *Atti del X Congresso SILFI: sintassi storica e sincronica dell´italiano. Subordinazione, coordinazione, giustapposizione*. Firenze: Cesati.

CRESTI, E. (2008). As novidades na Teoria da Língua em Ato. Curso oferecido na PósGraduação em Estudos Lingüísticos da UFMG.

CRESTI, E. (2005a). Enunciato e frase: teoria e verifiche empiriche. In: M. Biffi, O. Calabrese, L. Salibra (Orgs.). Italia Linguistica: discorsi di scritto e di parlato. Scritti in onore di Giovanni Nencioni. Siena: Prolagon, p. 249-260.

CRESTI, E. (2005b). Per una nuova classificazione dell'illocuzione. In: E. Burr (ed.), Tradizione e innovazione - Atti del VI convergno SILFI (Duisburg 28.06/02.07 2000). Firenze: Cesati, pp. 233-246.

_____ . (2002a). L'articolazione informativa topic-comment e comment-appendice: correlati intonativi. Em: Regnicoli, A. (Org.), *La fonetica acustica come strumento di analisi della variazione linguistica in Italia. Atti delle XII Giornate del Gruppo di Fonetica Sperimentale (XII GFS)*. Roma: Il Calamo, p. 153-160.

CRESTI, E. (2002b). Illocuzione e modalità. In: Beccaria, P. – Marello, C. (eds.). La parola al testo. Scritti per Bice Mortara-Garavelli. Torino: Ed. dell'Orso, pp. 133-145.

CRESTI, E. (2000). Corpus di italiano parlato. Firenze: Accademia della Crusca, 2 voll.

CRESTI, E. (1994). La Stanza: un'unità di costruzione testuale del parlato LABLITA, Università di Firenze. In press. Ferenck Kiefer. (1994). Modality, in Ronald E. Asher (a c. di), The Encyclopedia of language and linguistics, Oxford, Pergamon Press, pp. 2515-2520.

CRESTI, E. - BACELAR do Nascimento, F. - MORENO Sandoval, A. - VERONIS, J. - MARTIN, Ph., - CHOUKRI, K. (2004). The C-ORAL-ROM CORPUS. A multilingual resource of spontaneous speech for romance languages. In: Lino, M.T. - Xavier, M.F. - Ferreira, F. – Costa, R. – Silva, R. (eds.) Proceedings of the 4th International Conference on Language Resources and Evaluation. Paris: ELRA, vol. 2, pp. 575-79.

CRESTI, E.; MONEGLIA, M. (2010). Informational patterning theory and the corpus-based description of spoken language: The compositionality issue in the topic-comment pattern. In: A. Panunzi (Ed.); Bootstrapping Information from Corpora in a Cross- Linguistic Perspective. Firenze: FUP.

CRESTI, E. ; MONEGLIA, M. (2007). C-ORAL-ROM. Comparing Romance Languages in Spontaneous Speech Corpora. In: Silva, T. C. – Mello, H. R. (eds.).

Conferências do V Congresso Internacional da Associação Brasileira de Lingüística. Belo Horizonte: UFMG.

CRESTI, E.; MONEGLIA, M. (2005). C-ORAL-ROM. Integrated Reference Corpora for Spoken Romance Languages. Amsterdam/Philadelphia: John Benjamins.

CRESTI, E.; FIRENZUOLI, V. (2002). L'articolazione informativa topic-comment e commentappendice: correlati intonativi. In: A. (a cura d) Regnicoli (Ed.); La fonetica acustica come strumento di analisi della variazione linguistica in Italia: atti delle XII giornate di Studio del Gruppo di fonetica sperimentale (A.I.A.), Macerata, Universita degli studi, 13-15 dicembre 2001. Anais... p.301. Macereta: Il calamo. Retrieved September 6, 2007, fromhttp://lablita.dit.unifi.it/preprint/preprint-02coll10.pdf.

DE CESARE; A. M. (2004). *L'avverbio* anche *e il rilievo informativo del testo*, in FERRARI (ed.), pp. 191-218.

FERRARI, A. (2003). *Le ragioni del testo*. Firenze, Accademia della Crusca.

FIRENZUOLI, V. (2003). Verso un approccio allo studio dell'intonazione a partire da corpora di parlato: esempi di profili intonativi di valore illocutivo dell'italiano. In: Maraschio, N. Atti del XXXIV Congresso internazionale di studi della SLI "Italia linguistica anno Mille – Italia linguistica anno Duemila". Roma: Bulzoni, p. 535-5.

FIRENZUOLI, V. ; SIGNORINI, S. (2003). L'unità informativa di topic: correlati intonativi. In: Atti delle giornate del gruppo di fonetica sperimentale (XIII GFS). Pisa: ETS, p. 177-184.

FLEISS, J. L. (1971). Measuring nominal scale agreement among many raters. Psychological Bulletin, v. 76, p. 378-382.

FROSALI, F. (1951). "Le unità di informazione di Ausilio dialogico: valori percentuali, caratteri intonativi, lessicali e morfo-sintattici in un corpus di italiano parlato (C-ORAL-ROM)". In: Cresti, Emanuela (org.). Prospettive nello studio del lessico italiano. Firenze University Press, 2008, pp. 417-424. Georg von Wright. (1951). Essays on Modal Logic, Amsterdam, North-Holland.

GRABE, W. (2004). Perspectives in applied linguistics: A North American view. AILA Review, 17, p. 105-132.

GONÇALVES, S. C. L. (2008). Projeto ALIP (Amostra Linguística do Interior Paulista): questões teóricas e metodológicas sobre constituição de um banco de dados de língua falada. In: TAGNIN, S. E. O.; VALE, O. A. (Orgs.). *Avanços da Linguística de Corpus no Brasil*. São Paulo: Humanitas, p. 217-245.

HALLIDAY, M. A. K. ; HASAN, R. (1989). Language, Context, and Text: Aspects of Language in a Social-semiotic Perspective. Oxford: Oxford University Press.

t´HART, J. ; COHEN, A. ; COLLIER, R. (1990). A perceptual study on intonation: an experimental approach to speech Mellody. Cambridge: Cambridge University Press.

Laboratorio Linguistico del Dipartimento di Italianistica dell‟Università di Firenze. Disponível em: <http://lablita.dit.unifi.it/>. Acesso em: 10 julho 2007.

LADEFOGED (1975). *A course in phonetics*. Orlando: Harcourt Brace. ISBN 0-15-507319-2. 2nd ed 1982, 3rd ed. 1993, 4th ed. 2001, 5th ed. Boston: Thomson/Wadsworth 2006. Japanese translation 2000.

LANDIS, J. R.; KOCH, G. G. (1977). The measurement of observer agreement for categorical data. **Biometrics**, v. 33, p. 159-174.

LOMBARDI VALLAURI, E. (2009). *La struttura informativa. Forma e funzione negli enunciati linguistici*. Roma: Carocci.

MACWHINNEY, B. J. (2000). The CHILDES Project. Tools for Analyzing Talk. Mahwah, NJ: Lawrence Erlbaum, 2 voll.

MAIA ROCHA, B. (2011). A unidade de introdutor locutivo no português do Brasil. Faculdade de Letras da UFMG. (Dissertação de Mestrado)

MARCUSCHI, L. A. (2001). Da fala para a escrita: atividades de retextualização. Sao Paulo: Cortez.

MARTIN, P. (2004). WinPitch Corpus: A text to Speech Alignment Tool for Multimodal Corpora. .Lisbon: LREC.: LREC. Retrieved from <http://lablita.dit.unifi.it/coralrom/papers/index.html>, 2004.

MELLO, H.-RASO,T. (2011). Para a transcrição da fala espontânea: o caso do C-ORAL-BRASIL. Em: *Revista Portuguesa de Humanidades*, Faculdade de Filosofia da Universidade Católica Portuguesa.

MIRANDA, I. C. C. (2001). Aspectos prosódicos da fala do idoso. Belo Horizonte: Faculdade de Letras da UFMG. (Dissertação de Mestrado)

MITTMANN, M. A unidade informacional Topico na fala espontanea: um estudo baseado em corpus. Tese (Doutorado em Linguistica) – Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte (em preparação).

MO, Y.; COLE, J.; LEE, E.-K. (2008). Naïve listeners' prominence and boundary perception. In: Proceeding of the Speech Prosody International Conference. p. 735-738, Campinas. Disponível em: http://aune.lpl.univ-aix.fr/~sprosig/sp2008/papers/id053.pdf

MONEGLIA, M. (2011). Spoken corpora and pragmaticis. Anais do congresso RBLA, Belo Horizonte, v. 11, n.2, p.479-519.

MONEGLIA, M. (2000). Specifications on the C-ORAL-ROM Corpus. http://lablita.dit.unifi.it/coralrom/papers/Specifications-CORALROM.pdf

MONEGLIA, M. C-ORAL-ROM. (2005). Un corpus di riferimento del parlato spontaneo per l'italiano e le lingue romanze. In: Korzen, J. (ed.). Lingua, cultura e intercultura. L'italiano e le altre lingue. Atti del VIII convegno SILFI (Copenhagen 22-26 July 2004). Copenhagen: Samfunzliteratur Press, pp. 229-42.

MONEGLIA, M.; CRESTI, E. (2006). C-ORAL-ROM. Prosodic boundaries for spontaneous speech analysis. In: KAWAGUCHI, Y.; ZAIMA, S.; TAKAGAKI, T. (Eds.). *Spoken Language Corpus and Linguistics Informatics.* Amsterdam: John Benjamins.

MONEGLIA, M – CRESTI, E. (1997). L´intonazione e i criteri di trascrizione Del parlato adulto e infantile. In: Bortolini, U. – Pizzuto, E. Il Progetto CHILDES Italia. Pisa: Del Cerro, pp. 57-90.

NESTERENKO, I. (2006). Prosodic boundaries in spontaneous Russian: perceptual annotation and automatic classification. In: Procedings of the International Speech Prosody Conference. Dresden. Disponível em: http://aune.lpl.univ-aix.fr/~sprosig/sp2006/contents/papers/PS5-29_0110.pdf

OLIVEIRA, C. (2009). Apêndice de Comentário e Comentários Ligados: uma distinção à luz da Teoria da Língua em Ato. Revista Eletrônica de Divulgação Científica em Língua Portuguesa, Linguística e Literatura Letra Magna. Ano 05, n.10,1º semestre (WWW.letramagna.com).

_____. (2009b).Distinção entre Apêndice de Comentário, Comentários Ligados e Inciso, três unidades informacionais, em final de enunciado, à luz da Teoria da Língua em Ato. Revista e_Hum V2. n.1.

_____. (2010). O Apêndice de Comentário: classificação informacional e morfossintática à luz da Teoria da Língua em Ato. Revista Eletrônica de Divulgação Científica em Língua Portuguesa, Linguística e Literatura Letra Magna. Ano 06, n.13, 2º semestre (WWW.letramagna.com)

PANUNZI, A.; GREGORI, L. (2011). IPIC: an XML data base for the study of Informational Patterning. Presentation at the 7th LABLITA International Workshop in Corpus Linguistics. Firenze. Disponível em: <http://lablita.dit.unifi.it/lablita_workshop/ws7>.

RASO, T.; LEITE, F. (2010). Estudo contrastivo do uso de Alocutivos em italiano, português e espanhol europeus e português brasileiro. *Domínios de Linguagem*, n.1.

RASO, T.; MELLO, H. (Orgs.) (2012). C-ORAL-BRASIL. Corpus de referência da fala espontânea informal do português do Brasil. Belo Horizonte: UFMG.

RASO, T.; MELLO, H. (2010). The C-ORAL-BRASIL corpus. Em: Moneglia, M.; Panunzi, A., (orgs.) Bootstrapping Information from Corpora in a Cross Linguistic Perspective. Firenze University Press.

RASO, T., MELLO. H., JESUS, A. U., DE DEUS, L. A. (2007). *A primeira aplicação da Teoria da Língua em Ato ao PB*. Em: *Revista de Estudos da Linguagem*.

RASO, T.; MITTMANN, M. M. (2011). La Teoria della Lingua in Atto applicata ai testi del corpus C-ORAL- BRASIL.

RASO, T.; MITTMANN, M. M. (2010) Validacao estatistica dos criterios de segmentacao da fala espontanea no corpus C-ORAL-BRASIL. **Revista de Estudos da Linguagem**, v. 17, n. 2, p.73-91. Retrieved September 23, 2010, from http://relin.letras.ufmg.br/revista/upload/17-2_04.pdf.

RASO, T. – ULISSES, A. J. (2008). *Tópico e Apêndice no português do Brasil: algumas considerações*. Em: Estudos da Linguagem, 2.

ROCHA, B N. R. de M. (2012). Características Prosódicas do Tópico em PE e o Uso do Pronome Lembrete. Faculdade de Letras da UFMG, Belo Horizonte. (Dissertação de Mestrado)

SARDINHA, T. B. MOREIRA FILHO, J. L.; ALAMBER, E. (COORD) (2011). Corpus Brasileiro. Corpus Brasileiro. Retrieved July 30, 2011, from <http://corpusbrasileiro.pucsp.br>

SARDINHA, T. B. (2005). *Lingüística de Corpus.* Manole: São Paulo. Teubert, W. (1996). Editorial. International Journal of Corpus Linguistics, Vol.1, No. 1. iii-x.

SCARANO, A. (2009). The prosodic annotation of C-ORAL-ROM and the structure of information in spoken. In: L. Mereu (Ed.); **Information structures and its interfaces**. p.51-74. Berlin and New York: Mouton de Gruyter. doi: 10.1515/9783110213973.1.51.

SEARLE, J. (1969). Speech Acts: An Essay in the Philosophy of Language. Cambridge: Cambridge University Press.

SIGNORINI, S. (2004a). Il Topic: criteri di identificazione e correlati morfosintattici in un corpus di italiano parlato. In: F. Albano Leoni ; F. Cutugno; M. Pettorino – R. Savy (a c. di) Atti del convegno nazionale ""Il parlato italiano"". Napoli: M. D'Auria Editore, p. 15-39.

TUCCI, E. (2005/2006). L'unità di appendice in un corpus di italiano parlato (C-ORAL-ROM): caracteristiche intonative, semantiche e morfo-sintattiche. Tesi de laurea triennale in italianistica. Universitá degli studi di Firenze, Facoltà di lettere e filosofia, anno academico.

ULISSES, Andréa de Jesus. (2008). A unidade de apêndice no português do Brasil. Faculdade de Letras da UFMG. (Dissertação de Mestrado).

VALE, H. P. (2010). A unidade informacional de parentético no português do Brasil: uma análise baseada em corpus. Faculdade de Letras da UFMG. (Dissertação de Mestrado).

Anexos

Anexo 1 – Análise Informacional

Texto	Numeração	Exemplo	Análise Informacional
bfamcv01	1	*LEO: 26] não /=INP_r= <a gente perdeu no campo /=COM_r= **e tal**> //=APC=$	Preenchimento
	2	*EVN: [54] <o´ /=CNT= o Arnaldinum> é caro /=CMM= e tem aqueles problemas //=CMM=$ [55] existem outros lugares /=INT= tão bons quanto o Arnaldinum /=CMM= e mais baratos //=CMM=$ [56] tem o SESC /=COB= que é bom pa caramba //=COM=$ [57] quadra é igual à do <Arnaldinum> //=COM=$ *LUI: [58] <e o> Sagrado //=COM=$ *GIL: [59] velho //=COM=$ *EVN: [60] Sagrado //=COM=$ [61] agora /=PHA= o Sagrado é mais caro /=i-COM= eu acho /=PAR= que o Arnaldinum //=COM=$ *GIL: [62] sabe que que eu penso /=COM= velho //=ALL=$ [63] <eu nũ sei> que cês acham //=COM=$ *EVN: [64] <não> //=COM=$ *GIL: [65] da gente /=TOP= pra alguns setores da organização /=TOP= chamar o pessoal /=COM= **dos outro times** //=APC=$	Informação Tardia
	3	*GIL: [39] problema é que esse último campeonato nosso /=TOP= <foi ruim> //=COM=$ *EVN: [40] é /=CMM= <pois é> //=CMM=$ [41] não tá bom //=COM=$ *LUI: [42] a gente não devia ter tirado ele do Arnaldinum //=COM=$ [43] <não &de> +=EMP=$ *EVN: [44] <uhn> //=COM=$ *LEO: [45] <gente> tirou mesmo só por causa do &s [/2]=SCA= do seu Joaquim lá //=COM=$ [46] do <doidão> lá //=COM=$ *LUI: [47] <foi> //=COM=$ *EVN: [48] ô /=EXP= mas tem outros lugares /=CMM= o negócio é que a gente não procurou //=CMM=$ [49] eu te mandei uma lista //=COM=$ [50] com uns dez lugar diferentes //=COM=$ *LUI: [51] ah /=EXP= <velho /=ALL= mas eu> gosto do Arnaldinum //=COM=$ *EVN: [52] <pra olhar> //=COM=$ *LUI: [53] eu quero fazer o próximo campeonato no Arnaldinum /=CMM= e foda-se <pro seu Joaquim> //=CMM=$ *EVN: [54] <o´ /=CNT= o Arnaldinum> é caro /=CMM= e tem aqueles problemas //=CMM=$ [55] existem outros lugares /=INT= tão bons quanto o Arnaldinum /=CMM= e mais baratos //=CMM=$ [56] tem o SESC /=COB= que é bom pa caramba //=COM=$ [57] quadra é igual à do	Retomada

		<Arnaldinum> //=COM=$ *LUI: [58] <e o> Sagrado //=COM=$ *GIL: [59] velho //=COM=$ *EVN: [60] Sagrado //=COM=$ [61] agora /=PHA= o Sagrado é mais caro /=i-COM= eu acho /=PAR= que o Arnaldinum //=COM=$ *GIL: [62] sabe que que eu penso /=COM= velho //=ALL=$ [63] <eu nũ sei> que cês acham //=COM=$ *EVN: [64] <não> //=COM=$ *GIL: [65] da gente /=TOP= pra alguns setores da organização /=TOP= chamar o pessoal /=COM= dos outro times //=APC=$ *LUI: [66] mas a <gente tenta> fazer reunião /=CMM= galera nũ <comparece> //=CMM=$ *GIL: [67] <por exemplo> +=EMP=$ [68] <não não> /=EXP= <mas> assim //=COM=$ *EVN: [69] <é /=COM= porra> //=EXP=$ *GIL: [70] agora /=TOP= que deu pau /=APT= &d +=EMP=$ [71] que eu acho que deu muito pau /=COM= **nessa taça** //=APC=$	
	4	[77] cês querem participar //=COM_r=$ [78] cês querem olhar com a gente /=COB_r= cês querem sugerir um lugar /=COB_r= <qualquer> coisa desse tipo //=COM_r=$ *EVN: [79] <é> //=COM=$ *GIL: [80] <porque> assim /=UNC= &he /=TMT= &he /=TMT= quando eu descobri lá o São Vicente /=TOP= eu adorei //=COM=$ [81] achei aquele lugar incrível //=COM=$ [82] assim /=INT= <Nossa> //=COM_r=$ *LEO: [83] <piso de madeira> //=COM=$ *GIL: [84] que [/1]=EMP= que [/1]=EMP= que [/1]=EMP= que +=EMP=$ [85] &he /=TMT= tá /=i-COM_r= &he /=TMT= chão de madeira /=COB_r= que lugar legal pra se <jogar> //=COM_r=$ *EVN: [86] <hhh é> //=COM=$ *GIL: [87] <tendeu> //=COM=$ *EVN: [88] <cê se> sente numa <prisão /=COB= e tal hhh> //=COM=$ *GIL: [89] <o trem coberto /=COB= não> +=EMP=$ [90] <assim> /=INT= &he /=TMT= &he /=TMT= quadra coberta //=COM=$ *LEO: [91] <ham ham> //=COM=$ *GIL: [92] eu pensei /=INT= poxa //=COM_r=$ [93] que ótimo //=COM_r=$ [94] a gente podia fazer a taça aqui /=COB_r= todo mundo vai adorar /=COB_r= e tal //=COM_r=$ [95] e chegou lá /=TOP= todo mundo +=EMP=$ [96] todo mundo //=COM=$ [97] nũ teve um que nũ reclamou //=COM=$ [98] então assim /=INT= eu nũ tenho parâmetro pra isso //=COM=$ [99] já descobri //=COM=$ *EVN: [100] é //=COM=$ *GIL: [101] então /=INP= <vão [/1]=EMP= vamo> <&cu [/1]=SCA= colocar pra eles /=COM= ué> //=PHA=$	Retomada

		*LEO: [102] <beleza> //=COM=$ [103] <então a gente já sabe em quem que a gente vai colocar a> culpa //=COM=$ *EVN: [104] é <hhh> //=COM=$ *LEO: [105] <hhh> Gilberto //=COM=$ [106] hhh //=COM=$ [107] você está condenado //=COM=$ *EVN: [108] hhh você está demitido //=COM=$ *LEO: [109] hhh ham ham //=COM=$ [110] tá todo mundo olhando assim //=COM=$ [111] Gilberto /=CMM= <você está /=SCA= demitido> //=CMM=$ *GIL: [112] <você está /=SCA= demitido> //=COM=$ *LEO: [113] hhh tã tã tã //=COM=$ *GIL: [114] hhh eu tenho &dire +=EMP=$ [115] e sem os seus direitos <trabalhistas hhh> //=COM=$ *EVN: [116] <é hhh> //=COM=$ [117] hhh olha o Zé Carioca ali //=COM=$ *LEO: [118] <hhh> ai /=CMM= Zé Carioca é muito bom //=CMM=$ *EVN: [119] <é> //=COM=$ *GIL: [120] <ô> //=COM=$ [121] <mas> +=EMP=$ *EVN: [122] <aí> +=EMP=$ *GIL: [123] <é> //=COM=$ [124] <&he /=TMT= então> +=EMP=$ *EVN: [125] <é> /=CMM= concordo /=CMM= **isso aí que cê falou** //=APC=$	
	5 e 6	*LUI: [136] <eu acho que /=SCA= o nome /=SCA= do> [/1]=EMP= do [/1]=EMP= do [/1]=EMP= do [/1]=EMP= do [/1]=EMP= do [/1]=EMP= da modalidade esportiva que a gente criou /=TOP= que foi Futebol Arte /=PAR= ficou /=SCA= queimado /=COB= **por causa dessa taça** /=APC= pra mim é uma <coisa que> <é [/1]=SCA= foi triste> /=COB= é triste> falar /=COB= mas /=DCT= cê fala que é Futebol Arte /=CMM= a galera começa a zoar /=CMM= j**á** //=APC=$	1= Informação Tardia 2 = Preenchimento
	7	*GIL: [142] bem /=CNT= enfim //=COM=$ [143] é //=COM=$ [144] pra próxima taça /=TOP= a gente tem que também mandar um e-mail //=COM=$ [145] &f [/1] tipo /=INT= é quase como se fosse pedindo &des +=EMP=$ [146] não pedindo desculpa /=COB= mas assim /=INT= falando /=COB= que essa [/1]=SCA= a próxima vai ser diferente /=COB= em tais e tais aspectos /=COM= **e tal** //=APC=$	Preenchimento
	8	*LUI: [148] não //=COM=$ [149] e [/1]=EMP= e [/1]=EMP= e [/1]=EMP= e principalmente /=TOP= convocando a galera /=COB= falar assim /=INT= o' /=CNT= galera /=COB= negócio seguinte /=INT= tá pensando em organizar um torneio /=COB= e a gente quer /=SCA= que vocês participem /=SCA= da organização /=COB= <mandando	Retomada

		representante> +=COB=$ *LEO: [150] <e acima de tudo /=TOP= eu acho que a gente tem> que chamar os times que /=i-COM= tipo /=PAR= o [/1]=EMP= realmente os times que /=SCA= merecem a [/1]=SCA= a nossa //=COM=$ *EVN: [151] é /=INP= a <gente tem que> <restringir também /=COM= **isso>** //=APC=$	
	9	*GIL: [183] porque /=DCT= ele ficou naquele grupo difícil /=COM= **e tal** //=APC=$	Preenchimento
	10	*GIL: [137] <é> //=COM=$ [138] todo mundo <encheu o saco> /=COM= <**por causa disso**> //=APC=$	Informação Tardia
bfamcv02	11	*RUT: [90] <morro de vergonha> /=CMM= cê nũ viu que o> da Tiane eu nem <entrei> na igreja //=CMM=$ *JAE: [91] <é> //=COM=$ *RUT: [92] eu nũ fui não /=COM= minha filha /=ALL= **lá não** //=APC=$	Informação Tardia
	12	[*RUT: [279] <eu tenho é> só vergonha /=COM= **de entrar <na igreja>** //=APC=$	Informação tardia
	13	*RUT: [203] da família &d [/3]=EMP= <da família> dele /=TOP= minha filha /=ALL= <também nũ> tomo nem conhecimento /=COM= **também** //=APC=$	Preenchimento
bfamcv03	14	*TON: [247] é ocê /=COM= Onofre /=ALL= **que tá cagando assim** //=APC=$	Informação Tardia
	15	*CAR: [269] fechar nũ dá não //=COM=$ *CEL: [270] cai //=COM=$ *TON: [271] aqui o' /=CMM= aqui o' //=CMM=$ [272] ou senão só encostar aqui também o' //=COM=$ *CEL: [273] não /=INP= mas aí vai dar [/2]=SCA= <u>vai</u> &di [/1]=SCA= <u>ser pior pa nós</u> /=COM= uai //=PHA=$ *CAR: [274] por que /=COM= **que vai ser pior** //=APC=$	Repetição literal de forma não-contígua
	16	*CAR: [174] não //=COM=$ [175] quem vai jogar agora sou eu /=COM= Toninho //=ALL=$ [176] cê nũ faz besteira não /=CMM= porque //=CMM=$ [177] <u>cê nũ entendeu</u> /=COM= <u>cê nũ</u> //=APC=$	Repetição Parcial de forma contígua
bfamcv04	17	*HEL: [173] nũ pode apontar <pra nada <u>que tem</u>> +=EMP=$ *BRU: [174] <se for> /=SCA= um passarinho /=TOP= cê nũ pode fazer /=INT= hhh //=COM=$ [175] nũ pode //=COM=$ *HEL: [176] hum hum //=COM=$ *BRU: [177] cê nũ pode fazer &a [/1] nenhum som //=COM=$ *LUC: [178] rapidim //=COM=$ *BRU: [179] e +=EMP=$ *LUC: [180] vai explicando aí //=COM=$ *BRU: [181] tá //=COM=$ [182] e /=DCT= tem outra coisa também //=COM=$ [183] se /=DCT= for uma mesa /=TOP= por exemplo /=PAR= cê não pode fazer assim //=COM=$ [184] cê não pode aprontar [/1]=SCA= apontar pra mesa //=COM=$ [185] tá	Repetição literal de forma não contígua

		//=COM=$ *HEL: [186] tá //=COM=$ [187] pra /=SCA= nenhum <objeto> /=COM= que tenha //=APC=$	
	18	*HEL: [290] <&tip> +=EMP=$ [291] hum <hum> //=COM=$ *BRU: [292] <aí> cê faz um homem /=TOP= aí a pessoa fala /=INT= homem //=COM_r=$ [293] cê faz isso /=CMM= primeira //=CMM=$ [294] aí /=PHA= depois +=EMP=$ *HEL: [295] <e pr' ocê> falar que tá certo /=COM= isso //=APC=$	Informação Tardia e Preenchimento.
	19	*HEL: [307] se alguém se /=SCA= enrola /=COM= assim //=PAR=$ [308] aí /=DCT= <tipo /=UNC= pode hhh> +=EMP=$ *LUC: [309] <se alguém se enrola hhh> //=COM=$ *BRU: [310] perde //=COM=$ *HEL: [311] não //=COM=$ [312] mas <hhh> +=EMP=$ *LUC: [313] <acaba o tempo> //=COM=$ *BRU: [314] <hhh é> //=COM=$ *HEL: [315] <hhh que saco> //=COM=$ *LUC: [316] <acaba o tempo> //=COM=$ *HEL: [317] <não> /=COB= mas aí hhh +=EMP=$ [318] <e se precisar de um esclarecimento> /=COM= <assim /=PAR= **no> meio do negócio hhh** //=APC=$	Informação Tardia
bpubcv01	20	*FLA: [7] sai do doador /=CMM= lá /=APC= e chega pra gente assim //=CMM=$	Preenchimento
	21	*BRU: [127] <aquilo ali /=TOP= acho que eu já vi na televisão> /=COM= já //=APC=$	Preenchimento
	22	*MAR: [167] <é o> mesmo processo //=COM=$ *EMM: [168] <centrifugação> //=COM=$ *MAR: [169] <você o centrifuga> /=COB= a plaqueta é mais pesada /=COB= vai pro fundo /=COB= e o plasma <xxx> //=COM=$ *FLA: [170] <que são aqueles gruminhos> //=COM=$ *BRU: [171] <ah /=CMM= tá //=CMM=$ [172] porque primeiro cê centrifuga> /=COB= **uma vez /=APC**= aí tira o +=EMP=$	Informação tardia
	23	*MAR: [163] <é /=CMM= esse vai ser> //=CMM=$ *BRU: [164] mas como é que faz separar /=SCA= o plasma da plaqueta //=COM=$ *MAR: [165] ah /=CMM= tá //=CMM=$ *FLA: [166] <o que> +=EMP=$ *MAR: [167] <é o> mesmo processo //=COM=$ *EMM: [168] <centrifugação> //=COM=$ *MAR: [169] <você o centrifuga> /=COB= a plaqueta é mais pesada /=COB= vai pro fundo /=COB= e o plasma <xxx> //=COM=$ *FLA: [170] <que são aqueles gruminhos> //=COM=$ *BRU: [171] <ah /=CMM= tá //=CMM=$ [172] porque primeiro cê centrifuga> /=COB= uma	Retomada

		vez /=APC= aí tira o +=EMP=$ *MAR: [173] <isso> //=COM=$ *FLA: [174] é //=COM=$ [175] um <conjuntinho que> +=EMP=$ *EMM: [176] <a gente separa /=COM= **isso>** //=APC=$	
	24	*BRU: [245] <yyyy> //=COM=$ *FLA: [246] <porque lá tem> /=INT= Chagas /=COB= agaivê /=COB= <agateelevê> +=COB=$ *MAR: [247] <xxx> /=UNC= <agateelevê> //=COM=$ *EMM: [248] <agateelevê> /=COB= sífilis /=COB= <&Ch [/1]=EMP= &he> +=TMT=$ *BRU: [249] <que que é agateelevê> //=COM=$ *FLA: [250] é um [/1]=SCA= uma doença sexualmente transmissível também //=COM=$ *MAR: [251] é //=COM=$ *FLA: [252] mas /=INP= mata /=CMM= **isso aí** /=APC= será //=CMM=$	Retomada
bpubcv02	25	*CAR: [24] então /=PHA= e tá ali ainda naquele [/1]=SCA= naquele fio ainda /=SCA= <ligado> /=SCA= contrário //=COM=$ *JEA: [25] <tá> //=COM=$ *OSV: [26] <tá> /=CMM= tá //=CMM=$ *JEA: [27] só <vieram fazer a leitura> //=COM=$ *OSV: [28] <mas ali nũ tem problema não> /=CMM= que ali é só extensão /=CMM= né <hhh> //=PHA=$ *CAR: [29] <n' é> não //=COM=$ [30] nũ é gato /=COM= nem **nada não** //=APC=$	Preenchimento.
	26	*CAR: [38] <entregar> pro Ibraim /=COB= **e tudo** /=APC= que <aí> //=COM=$	Preenchimento
	27	*OSV: [39] <aí tá &d> [/1]=SCA= da [/1]=EMP= do ano todo //=COM=$ *OSV: [40] <de janeiro> +=EMP=$ *CAR: [41] <e aqui /=UNC= e a> conta nũ vem pra cá mais não /=CMM= vem nada //=CMM=$ *CAR: [42] depois eles passam a entregar lá na prefeitura /=COM= né //=PHA=$ *CAR: [43] o Ibraim tem que ver <isso lá yyyy> //=COM=$ *OSV: [44] <é /=CMM= é //=CMM=$ *OSV: [45] não /=CMM= eu> falei com ele já /=COM= **pra mudar isso** //=APC=$	Informação Tardia
	28	*CAR: [212] eu nunca fui na casa lá dela não /=CMM= mas diz que é <muito ruim> /=CMM= **lá o local** //=APC=$	Repetição com variação
	29	*OSV: [249] ah /=CMM= não /=CMM= essa foi aquela última que [/1]=SCA= que o Alessandro fez /=CMM= eu pedi pra e' lá //=COM=$ *JEA: [250] <foi> //=COM=$ *OSV: [251] <foi o Alessandro> foi lá	Informação Tardia

		/=COB= foi /=SCA= acho que foi quinta-<feira> //=COM=$ *CAR: [253] <xxx> "na /=SCA= varanda /=SCA= da casa /=SCA= tem um <quarto> /=SCA= as rachaduras /=SCA= de mais ou menos /=SCA= um metro de rampa" //=UNC=$ *OSV: [254] <quarta-feira> /=COM= **que ele foi lá** //=APC=$	
	30	*OSV: [271] se for o caso só dela mesmo /=TOP= esses caso que /=SCA= não tiver jeito /=TOP= da [/1]=EMP= da gente fazer /=APT= tira ela então /=COM= né /=PHA= **no caso** //=APC=$	Preenchimento,
	31	*OSV: [32] então tem que levar isso pro Ibraim também //=COM=$ [33] apesar que o Geraldo já foi lá e já +=EMP=$ [34] lá na CEMIG /=TOP= já conversou com eles /=COM= né //=PHA=$ [35] e o <Ibraim> +=EMP=$ *CAR: [36] <tem que levar e> mostrar /=COM= **o** //=APC=$	Preenchimento
	32	*TIQ: [59] <tem nada> <não /=COM= né> //=PHA=$ *OSV: [60] <eu pedi pro> Ibraim passar em nome da <prefeitura> /=COB= pa nũ vim mais em nome do Geraldo //=COM=$ *CAR: [61] <yyyy> //=COM=$ *OSV: [62] tá dando problema /=COM= né //=PHA=$ [63] passar em nome da prefeitura /=TOP= e /=SCA= ir tudo pra lá duma vez /=COM= **em vez de vir <pra cá**> //=APC=$	Informação Tardia
	33	*OSV: [154] “em vistoria realizada no dia quatorze do sete nanananã /=SCA= no endereço acima mencionado /=SCA= constatamos uma residência /=SCA= que dista” +=EMP=$ [155] "dista" /=COM= né hhh //=PHA=$ [156] “&apro” +=EMP=$ [157] distancia /=COM= **que es queria <colocar**> //=APC=$	Informação Tardia
bfamdl01	34	*FLA: [98] o’ o detergente //=COM=$ *REN: [99] oitenta-e-cinco centavos o <Minuano> //=COM=$ *FLA: [100] <Minuano é> bom //=COM=$ *FLA: [101] <quanto [/1]=EMP= quanto é o> Ypê //=COM=$ *REN: [102] <eu também acho> //=COM=$ *REN: [103] também acho /=COM= **que é** //=APC=$	Preenchimento
	35	*FLA: [526] é /=INP= eu imaginei //=COM=$ *FLA: [527] ali arroz /=COM= **Tio João** //=APC=$	Informação Tardia
	36	*REN: [568] a culpa é sempre do arroz //=COM=$ *REN: [569] ah /=INP= também nũ tem outro /=CMM= a gente vai ter que levar esse /=CMM= **de todo jeito** //=APC=$	Preenchimento
	37	**REN: [335] um e <noventa-e-nove> //=COM=$	Informação Tardia

		*FLA: [336] <ah> /=CMM= bom //=CMM=$ *FLA: [337] o' nós trocando /=COM= **levando só dele** //=APC=$	
	38	*FLA: [356] é //=COM=$ *FLA: [357] a Bruna tava querendo acetona esses dia //=COM=$ *FLA: [358] tem que lembrar ela /=COM= **comprar acetona** //=APC=$	Informação Tardia
bfamdl02	39	*BEL: [74] cê nunca pensou em /=INT= em [/1]=EMP= em [/1]=EMP= tipo assim /=INT= junto com [/1]=SCA= com lingüística e tal /=TOP= mexer com isso /=COB= assim /=PAR= dando aula mesmo /=COM= **e tal** //=APC=$	Preenchimento
	40	*BEL: [138] <é> //=COM=$[139] e por falta de cuidado /=TOP= é foda /=COM= **também** //=APC=$	Preenchimento
	41	*BEL: [143] pois é //=COM=$[144] uma [/1]=EMP= um cara /=TOP= o primeiro cara em São Paulo +=EMP=$ *BAL: [145] que foi &pro [/1]=SCA= que <foi processado> +=COM=$ *BEL: [146] <cê viu> //=COM=$ *BAL: [147] eu vi isso hoje //=COM=$ *BEL: [148] porque chamou o cara de veado /=CMM= e jogou uma <garrafa [/1]=SCA= uma> +=EMP=$ *BAL: [149] <latinha de> <cerveja> /=COB= nele /=COB= e tal //=COM=$ *BEL: [150] <é> //=COM=$[151] foi o primeiro <cara> //=COM=$ *BAL: [152] <não /=EXP= eu> adorei /=CMM= <adorei> //=CMM=$ *BEL: [153] <não> //=COM=$[154] ô /=PHA= isso /=TOP= &i [/1]=EMP= de [/1]=EMP= de boa /=PAR= se isso +=EMP=$[155] sei lá /=INT= se tiver três casos disso /=TOP=$ *BAL: [156] na verdade /=TOP= <assim /=PHA= até viu> //=COM=$ *BEL: [157] / <o pessoal vai ser muito> <mais respeitado> //=COM=$ *BAL: [158] existem vários> /=COB= só que a maioria /=TOP= &he /=EMP= tá julgando improcedência /=COB= tal /=COB= porque /=DCT= &he /=EMP= de certa forma /=TOP= a bancada evangélica /=TOP= eles tão /=SCA= muito contra /=COM= **essa coisa** /=APC= né //=PHA=$	Retomada Textual
	42	*BEL: [176] &he /=TMT= se ocê tá +=EMP=$[177] eles tavam num posto de gasolina /=COM= eu acho //=PAR=$[178] aqueles postinhos que ficam bebendo e tal //=COM=$[179] o cara /=i-COB= tipo /=PAR= falou de forma /=SCA= ofensiva /=COB= ainda [/1]=EMP= e teve agressão <ainda> //=COM=$ *BAL: [180] <ham ham> //=COM=$ *BEL: [181] então é completamente diferente /=COM= **assim** //=APC=$[	Preenchimento

	43	*BAL: [211] então ele vai em qualquer lugar //=COM=$[212] tá vendo /=CNT= tem o velcro aqui /=COM= **e tal** //=APC=$	Preenchimento
	44	*BEL: [253] <é> //=COM=$[254] não /=INP= e ainda cheguei falando com a minha mãe /=COB= toda empolgada /=INT= mãe /=ALL_r= minha mala /=TOP_r= tá super bem arrumada //=COM_r=$[255] tá até pesando menos /=COM_r= **e tal** //=APC=$[	Preenchimento
bfamdl03	45	*LUZ: [1] porque /=DCT= eu só soube que eu nũ +=EMP=$ [2] eu tive certeza absoluta que eu nũ era daqui quando eu saí //=COM=$ [3] que eu senti que então /=SCA= eu tava no meu lugar /=COM= né //=PHA=$ [4] porque eu [/1]=SCA= eu me senti /=SCA= respirando /=i-COB= né /=PHA= adequada /=COM= né /=PHA= **no lugar** //=APC=$ [5]	Repetição com variação
	46	*LAU: [27] eu vi no +=EMP=$ [28] mandaram um e-mail pra mim //=COM=$ [29] tem [/1]=EMP= vai ter concurso pra seis +=EMP=$ [30] es vão admitir seis professores na /=SCA= Escola de Belas Artes /=COM= **da uefeemegê** //=APC=$	Informação Tardia
	47	*LAU: [72] <mas +=EMP=$ [73] não> //=COM=$ [74] es tavam falando /=COB= mas /=DCT= <mas eu acho que> +=EMP=$ *LUZ: [75] não /=INP= es tavam falando <que é meio> certo que é dele /=COM= <assim> //=PHA=$ *LAU: [76] <é> /=INP= mas pode ser que seja da +=EMP=$ *LUZ: [77] mas tem o Ronan /=SCA= também /=COM= né /=PHA= <**aqui>** //=APC=$	Preenchimento
	48	*LUZ: [70] Marco Paulo //=COM=$ [71] Marco <Paulo> //=COM=$ *LAU: [72] <mas +=EMP=$ [73] não> //=COM=$ [74] es tavam falando /=COB= mas /=DCT= <mas eu acho que> +=EMP=$ *LUZ: [75] não /=INP= es tavam falando <que é meio> certo que é dele /=COM= <assim> //=PHA=$ *LAU: [76] <é> /=INP= mas pode ser que seja da +=EMP=$ *LUZ: [77] mas tem o Ronan /=SCA= também /=COM= né /=PHA= <aqui> //=APC=$ *LAU: [78] <não> /=PHA= tem o &Ro> +=EMP=$ *LUZ: [79] <então /=INP= assim> /=PAR= a [/1]=EMP= essa aí <vai ser muito &di> +=EMP=$ *LAU: [80] <tem o Ronan> /=COB= e tem a /=SCA= <Giovana> //=COM=$ *LUZ: [81] e tem a Giovana //=COM=$ [82] <então> vai ser muito disputada /=COM= **essa aí** //=APC=$	Repetição literal de forma não-contígua
	49	*LAU: [127] <tem que fazer /=COM= né> //=PHA=$ [128] é //=COM=$ [129] eu sei	Preenchimento

		//=COM=$ *LUZ: [130] mas é pra mestre ou pa doutor /=COM= **essa aí** //=APC=$	
	50	*LUZ: [189] <ham ham> //=COM=$ [190] &t [/1]=EMP= tá //=COM=$ [191] tá certo //=COM=$ [192] porque eu acho que no mesmo concurso /=TOP= cê nũ pode fazer duas //=COM=$ [193] quando é o mesmo /=COM= **coisa** //=APC=$	Preenchimento
	51	*LUZ: [313] <tava> //=COM=$ [314] tava úmida /=COM= &he /=TMT= **a coisa** //=APC=$ [315] agora não //=COM=$ [316] <ressecou> mesmo //=COM=$	Preenchimento
bfamdl04	52	*KAT: [1] que sio' arrumou no seu cabelo //=COM=$ [2] que espetou de novo //=COM=$ *SIL: [3] nada //=COM=$ *KAT: [4] do lado aí //=COM=$ [5] tá caindo //=COM=$ *SIL: [6] capaz //=COM=$ *KAT: [7] ahn //=COM=$ *SIL: [8] capaz /=COM= **que tá** //=APC=$	Preenchimento
	53	*KAT: [11] pois é //=COM=$ [12] depois que a siora fez aquela /=SCA= hidratação com aloe que ele ficou assim //=COM=$ [13] espetado pos lados //=COM=$ [14] antes de passar aquele creme ele nũ tava assim /=COM= **espetado pos lados não** //=APC=$	Repetição parcial de forma não-contígua
	54	*KAT: [11] pois é //=COM=$ [12] depois que a siora fez aquela /=SCA= hidratação com aloe que ele ficou assim //=COM=$ [13] espetado pos lados //=COM=$ [14] antes de passar aquele creme ele nũ tava assim /=COM= espetado pos lados não //=APC=$ *SIL: [15] tava não //=COM=$ [16] pode ser o creme /=COM= **que nũ deu certo com ele** //=APC=$ [	Informação tardia
	55	*SIL: [94] conta lá /=i-COB= por favor /=PAR= pra mim /=COB= se tem <isso> //=COM=$ *TOM: [95] <cê sabe> qual que é o símbolo do Napoli /=COM= né //=PHA=$ [96] o asno //=COM=$ *SIL: [97] eu achei que era menos //=COM=$ [98] chei que devia ser doze /=CMM= seis //=CMM=$ *KAT: [99] o quê //=COM=$ *SIL: [100] copos //=COM=$ [101] copos de Urano /=COM= **que tem aí** //=APC=$	Informação Tardia.
	56	*SIL: [135] isso é cuidado //=COM=$ [136] nũ é igualzim de casa de pobre /=COB= que tudo que tem põe pra fora não //=COM=$ [137] tendeu //=COM=$ [138] cuidado //=COM=$ [139] gualzim lá em casa //=COM=$ [140] um /=TOP= usa o copo /=CMM= deixa lá /=CMM= **arrupiado** //=APC=$	Informação Tardia
	57	*SIL: [206] leva não //=COM=$ [207] mas os trem tá aumentando demais //=COM=$ [208]	Informação Tardia

		o prano de saúde ia aumentar [/2]=SCA= vai aumentar de novo //=COM=$ [209] foi por isso que es te deram só seis meses /=COM= na [/1]=EMP= **naquela folheta** //=APC=$	
bfamdl05	58	*ANE: [113] ah /=EXP= mas é porque aqui ea deve ser outro nome /=COM= **então** //=APC=$	Preenchimento
	59	*ENC: [200] mas ele nũ tá ativo ainda não /=CMM= ele só tá instalado /=CMM= **só** //=APC=$	Preenchimento
	60	*ANE: [207] <só tem esse> apartamento pra vender //=COM=$ *CES: [208] obrigado> //=COM=$ *ENC: [209] eu creio que sim //=COM=$ [210] que +=EMP=$ *ANE: [211] tá //=COM=$ *ENC: [212] tá //=COM=$ *ANE: [213] esquecemo de perguntar <pa ele a> previsão /=COM= né //=PHA=$ *CES: [214] <bom> //=COM=$ [215] é //=COM=$ [216] deixa eu <perguntar pra ele /=COB= a> previsão //=COM=$ *ANE: [217] <pergunta lá> //=COM=$ *CES: [218] eu acho que a previsão dele tá pra entregar em dezembro //=COM=$ [219] a previsão de <u>entrega do prédio</u> //=COM=$ *ENC: [220] no começo yyyy //=COM=$ *CES: [221] ahn //=COM=$ *ENC: [222] no começo yyyy //=COM=$ *CES: [223] então /=INP= até final de novembro //=COM=$ *ENC: [224] isso //=COM=$ *ANE: [225] &he /=TMT= essa localização aqui /=TOP= eu nũ acho ruim //=COM=$ *RAQ: [226] eu gostei também foi por causa do construído já ao redor /=COM= <né> //=PHA=$ *ANE: [227] <é> //=COM=$ *RAQ: [228] <é sim> //=COM=$ *ANE: [229] <tá tudo> <construído> //=COM=$ *CES: [230] <Anete> //=COM=$ [231] <ele> [/1]=EMP= e' disse que é &d +=EMP=$ [232] final de novembro //=COM=$ [233] <é /=INP= porque agora é rápido> /=COM= isso <aqui> //=APC=$	Retomada ou preenchimento
	61	*ANE: [268] é aqui //=COM=$ *CES: [269] segundo //=COM=$ [270] acho que esse aqui é o &segun +=EMP=$ [271] terceiro //=COM=$ [272] prontinho //=COM=$ [273] trezentos-e-um /=COB= trezentos-e-dois /=COB= trezentos-e-três tá pra cá //=COM=$ [274] ah /=EXP= tá aberto //=COM=$ [275] tem uma varandinha /=COB= cozinha estilo americano //=COM=$ *ANE: [276] esse mesmo /=COM= **que eu vim <ver>** //=APC=$	Informação Tardia
	62	*CES: [405] <aqui> /=TOP= ele é maior /=COM= <**do que aqui>** //=APC=$	Informação Tardia

	63	*CES: [82] nũ pode subir /=CMM= é contramão //=CMM=$ *ANE: [83] uhn //=COM=$ *CES: [84] e essa é a que nós subimos //=COM=$ *ANE: [85] andei isso aqui tudo /=COM= **com o Tiago** //=APC=	Informação Tardia
bpubdl01	64	*PAU: [45] porque +=EMP=$ [46] ah /=EXP= eu tenho uma aqui //=COM=$ [47] pequenininha //=COM=$ [48] então aqui /=TOP= quarenta centímetros +=COB=$ *ROG: [49] que ali e' ficou com quarenta-e-cinco //=COM=$ *PAU: [50] aonde /=CMM= que <ficou /=CMM= com quarenta-e-cinco> //=APC=$	Repetição literal de forma não-contígua
bpubdl02	65	*JAN: [43] <e uma> dessa /=COM= aqui /=PAR= **também <o'>** //=APC=$	Preenchimento
	66	*EUG: [70] mas ele nũ deixa de ser legal também //=COM=$ [71] olha p' cê ver /=CNT= cê tá de blusa /=SCA= lilás /=COM= **também** /=APC= né //=PHA=$	Preenchimento
	67	*EUG: [43] <eu &v> +=EMP=$ *EUG: [44] <é uma> gracinha /=COM= né //=PHA=$ *EUG: [45] <porque> essa borracha parece uma cortiça /=COM= né //=PHA=$ *JAN: [46] <é> //=COM=$ *EUG: [47] <minha esposa> pegou uma dessa //=COM=$ *JAN: [48] <hum hum> //=COM=$ *EUG: [49] ela é muito legal //=COM=$ *EUG: [50] eu pego p' cê ver /=COM= **dessa aqui** //=APC=$	Informação Tardia
	68	*EUG: [84] o modelo tradicional eu vou ver se eu acho /=COM= **dele lá** //=APC=$	Preenchimento
	69	*EUG: [61] eu peguei esse modelo primeiro pr' ocê aqui /=COB= esse modelo de /=SCA= borracha /=PAR= que enquanto cê distrai /=TOP= com esse modelo aqui /=PAR= eu vou ver se eu acho /=SCA= os outros que cê me pediu //=COM=$ *JAN: [62] ahn //=COM=$ *JAN: [63] tá //=COM=$ *EUG: [64] tá //=COM=$ *JAN: [65] e desse /=TOP= é dessa cor mesmo que eu quero //=COM=$ *EUG: [66] mas é a cor mais bonitinha /=TOP= é essa <mesmo> //=COM=$ *JAN: [67] <é> //=COM=$ *JAN: [68] porque senão levar uma roxa //=COM=$ *JAN: [69] eu nũ sei que eu hhh faço com ela //=COM=$ *EUG: [70] mas ele nũ deixa de ser legal também //=COM=$ *EUG: [71] olha p' cê ver /=CNT= cê tá de blusa /=SCA= lilás /=SCA= também /=COM= né //=PHA=$ *JAN: [72] é //=COM=$	Retomada Textual

		*EUG: [73] hoje em dia +=TOP=$ *EUG: [74] gozado que ela tem a tira lilás e ficou discreta //=COM=$ *EUG: [75] porque a [/1]=SCA= a palmilha é toda em &cor [/2]=SCA= parecendo <cortiça> //=COM=$ *JAN: [76] <hum hum> //=COM=$ *JAN: [77] hum hum //=COM=$ *EUG: [78] enquanto <cê> &dis [/1] experimenta essa aí /=TOP= eu vou pegar as outras aqui //=COM=$ *JAN: [79] <tá> //=COM=$ *EUG: [80] dá licença //=COM=$ *JAN: [81] Nossa /=EXP= essa é ótima //=COM=$ *EUG: [82] olha /=CNT= eu peguei as três caixas que eu achei trinta-e-nove lá /=CMM= mas /=SCA= todas três são desse modelo de cruzar //=CMM=$ *JAN: [83] uhn //=COM=$ *EUG: [84] o modelo tradicional eu vou ver se eu acho /=COM= dele lá //=APC=$ *JAN: [85] nts //=COM=$ *JAN: [86] essa aqui é bonitim //=COM=$ *EUG: [87] tem nenhum par mais /=COM= **daquele modelo** //=APC=$	
	70	*EUG: [155] era [/1]=EMP= era a rasteira mais cara que tinha aqui //=COM=$ *EUG: [156] pôs na liquidação por cinqüenta-e-nove /=COB= quando tava na liquidação //=COM=$ *EUG: [157] aqui o' /=CNT= essa aqui que eu achei dela pretinha //=COM=$ *EUG: [158] achei Picadilly /=COM= **trinta-e-nove lá dentro** //=APC=$	Informação Tardia
bfammn01	71	*MAI: [4] ele nũ é muito parente chegado não /=COB= mas &t [/1]=SCA= deve ser /=SCA= primo [/1]=EMP= primo quarto /=COM= por aí /=PAR= **deve ser** //=APC=$ [5]	Repetição literal de forma contígua
	72	*MAI: [9] e quando chegou lá /=TOP= &he /=TMT= montou uma casinha pa ele /=COB= pa família dele /=COB= e tal /=COB= e e' vinha na cidade pa compar alguma coisa &dif [/1]=SCA= diferente /=COM= **que nũ era da roça** /=APC= né //=PHA=$	Informação Tardia
	73	*MAI: [12] ela é [/2]=EMP= &che [/1]=EMP= ea chega a medir /=i-COM= assim /=PAR= mais +=EMP=$ [13] o diâmetro dea deve dar uns [/1]=SCA= uns quarenta cinqüenta centímetro de [/1]=SCA= de &s [/2]=EMP= de grossura /=COM= **o diâmetro dela** //=APC=$	Repetição literal de forma contígua
	74	*MAI: [21] que /=DCT= na hora que ele lá envinha voltando do [/1]=SCA= do comércio /=TOP= que e' foi fazer a compra dele /=PAR= a cobra percebeu o cheiro dele /=COM= na hora que ele lá envinha no [/1]=SCA= **no trilho** //=APC= $	Repetição Parcial de forma contígua
	75	*MAI: [20] é um trilho dentro da mata	Repetição com

		/=COB= a mata virgem //=COM=$ [21] n' é matinha igual essas capoeirinha aqui não /=COB= é mata mesmo /=COB= de /=SCA= madeira /=SCA= da grossura que /=SCA= quato homem nũ abarca um pau //=COM=$ [22] uma /=SCA= tora /=COM= **da madeira** //=APC=$	variação
	76	*MAI:e' [/1]=EMP= aí ele /=TOP= pr' ea não [/1]=SCA= não avançar ele de uma vez /=TOP= ele tirou um chapéu que ele tinha lá //=COM=$ [31] chapéu /=SCA= que es lá no interior fala chapéu de barreta //=COM=$ [32] tipo &ba [/2]=EMP= chama barreta /=COB= o /=SCA= material /=COM= **daquele chapéu** //=APC=$	Repetição Parcial de forma não-contígua
	77	*MAI: [34] e ea veio /=COB= ea saiu no [/1]=SCA= no [/1]=SCA= no trilho /=COB= e rasgou o chapéu todo /=COB= nũ sendo ele ea [/1]=TOP= ea &rap [/2]=SCA= continuou atrás dele /=COM= **quebrando** //=APC=$	Informação Tardia
	78	*MAI: [34] e ea veio /=COB= ea saiu no [/1]=SCA= no [/1]=SCA= no trilho /=COB= e rasgou o chapéu todo /=COB= nũ sendo ele ea [/1]=TOP= ea &rap [/2]=SCA= continuou atrás dele /=COM= quebrando //=APC=$ [35] ea nũ andava no [/1]=SCA= no limpo não /=CMM= era só dentro do mato //=CMM=$ [36] igual animal /=SCA= forte mesmo /=COM= a quebração que ea tava fazendo no meio do mato //=APC=$	Informação Tardia
	79 e 80	*MAI: [67] segundo ele falou comigo /=COM= o /=SCA= **parente dele** /=APC= que &m [/2]=EMP= **desse que morreu** //=APC= $	1= Informação Tardia 2= Informação Tardia
	81	*MAI: [65] &he /=TMT= no outro dia que /=SCA= o sol tava [/1]=SCA= já tinha saído e tava tudo bem /=TOP= ea pegou e &f [/2]=SCA= e foi atrás de [/1]=SCA= de alguém /=COM= **pa vim matar a cobra** //=APC=$	Informação Tardia
	82	*MAI: [71] certamente ea tava querendo fazer o seguinte /=INT= eu /=TOP= eu matei esse /=CMM= eu vou matar o resto tudo /=CMM= da [/1]=EMP= &he /=TMT= **dentro da casa** //=APC=$	Informação Tardia
	83	*MAI: [73] ea pegou e +=EMP=$ *MAI: [74] &he /=TMT= a dona foi atrás dele [/1]=SCA= de alguém /=COB= as pessoa veio com arma /=SCA= forte /=COB= espingarda /=SCA= de cartucho /=COB= e teve que atirar no corpo da [/1]=SCA= do cara que ea tava enrolada no corpo pa poder matar ela //=COM=$ *MAI: [75] diz que [/2]=EMP= diz que era enorme //=COM=$ *MAI: [76] só até aí que eu sei o caso /=COM= **que ele me contou** //=APC=$	Informação tardia
	84	*MAI:[73] ele matou /=COB= e essa família saiu desse lugar /=COM= **não ficou mais**	Informação Tardia

		//=APC=$	
	85	*MAI: [79] é lugar /=SCA= perigoso mesmo o norte de Minas /=COM= **lá** //=APC=$	Preenchimento
	86	*MAI: [85] na Amazonas tem muito dessa cobra aí //=COM=$[86] só ni lugar /=SCA= que tem /1 =SCA= tem mata mata muito /=SCA= forte /=COB= mata /=SCA= perigosa /=COB= que tem <esses tipo> de cobra /=COM= né //=PHA=$ *DUD: [87] <uhn> //=COM=$ *MAI: [88] no norte de Mina /=TOP= tinha esse /2 =SCA= antigamente /=PAR= tinha esse tipo de cobra tudo /=COM= né //=PHA=$[89] talvez agora já acabou /=COB= porque já desmataram muito /=COM= né //=PHA=$[90] nũ existe muita mata mais igual /1 =SCA= igual era /=COB= mas tem muita mata forte aí /=COB= nũ é igual essas /1 =SCA= essas capoeirim /=COB= cê fala assim /=INT= ah /=EXP_r= isso aqui é uma mata //=COM_r=$[91] isso aqui /=TOP= em vista de lá /=TOP= é /1 =EMP= é um cerrado /=COM= aqueas mata lá //=APC=$	Retomada Textual (com reformulação)
	87	*MAI: [68] nũ sei /=COB= né /=PHA= a imaginação hhh dum [/1]=SCA= dum animal /=COM= **o que que pode ser** /=APC= né //=PHA=$	Informação Tardia
bfammn02	88	*LUC: [47] &he /=TMT= essa [/1]=EMP= &he /=TMT= a [/1]=EMP= a [/1]=EMP= a história //=COM=$ *LUC: [48] desde o começo /=COM= assim /=PAR= e tudo o mais //=APC=$	Preenchimento
	89	*DFL: [92] <e isso> /=TOP= assim /=PAR= era muito importante /=COM= pra ela //=APC=$	Informação Tardia
	90	*DFL: [125] papai /=INT= não /=CMM_r= é &mov [/1]=SCA= poesia moderna /=CMM_r= e meu pai [/3]=EMP= e tudo mais /=COB_r= então /=INP= o tio Carlos tinha um certo dever de gratidão /=SCA= com o papai /=COB= porque /=DCT= o papai é que /=SCA= foi <estudar direito no Rio> +=COB=$ *LUC: [126] hum hum //=COM=$ *LUC: [127] <ele publicou a primeira> /=CMM= nũ é //=CMM=$ *LUC: [128] é /=PHA= ele <publicou a primeira> <poesia> //=COM=$ *DFL: [129] <é /=UNC= &pap> +=EMP=$ *DFL: [130] é /=UNC= &p +=EMP=$ *DFL: [131] é> //=COM=$ *DFL: [132] papai fundou um <jornalzinho lá> na &i [/6]=SCA= ele fundou um jornalzinho lá na Itabira /=COB= e tudo mais /=COB= mas /=INP= de &po [/2]=EMP= de pouca [/1]=SCA= poucos números //=COM=$ *LUC: [133] <hum hum> //=COM=$	Informação Tardia

		*LUC: [134] hum hum //=COM=$ *DFL: [135] e então /=DCT= &he /=TMT= e /=SCA= papai foi estudar no Rio /=COB= e mandava /=COB= pra ele /=APC=	
	91	*DFL: [154] e ela nũ ficava na casa dos avós /=SCA= velhos /=COB= né /=PHA= ela ia <lá pra> casa /=COB= que tinha aquela meninada //=COM=$ *LUC: [155] <claro> //=COM=$ *LUC: [156] <hum hum> //=COM=$ *DFL: [157] <e nós> /=TOP= exatamente da mesma idade //=COM=$ *DFL: [158] eu sou vinte-e-sete de &fever ea é [/4]=SCA= de fevereiro /=CMM= e ela é logo do princípio de março //=CMM=$ *LUC: [159] hum hum //=COM=$ *DFL: [160] então /=TOP= &he /=TMT= tinha uma certa hora /=TOP= que a Maria Julieta ia de manhã lá pra casa /=COB= almoçava lá em casa /=COB= e ficava //=COM=$ *DFL: [161] na hora do banho ela ia pra casa da outra avó //=COM=$ *LUC: [162] hum hum //=COM=$ *DFL: [163] e eu tomava meu banho / e ia pra casa da outra avó também //=COM=$ *DFL: [164] e quando eles /=SCA= vinha trazer a Maria Julieta /=TOP= ele ficava uns poucos dias depois vinham buscar /=PAR= e então /=DCT= ele participava /=COB= **daquelas nossas brincadeira** /=APC=	Informação Tardia
	92	*DFL: [165] e às vezes ele tava lendo jornal /=TOP= de vez em quando ele [/1]=SCA= ele interrompia /=COB= ele tava sempre atento às nossas brincadeira /=PAR= e' falava /=INT= o que é o que é //=COM_r=$ [166] a gente <assustava /=COB= e tinha que +=EMP=$ [167] às vezes +=EMP=$ [168] se sabia /=CMM= falava /=CMM= <mas se nũ sabia /=CMM= né> //=PHA=$ *LUC: [169] <uma charada /=COB= ali> /=APC= <na hora> //=COM=$	Informação Tardia
	93	*DFL: [173] <eu até já> falei isso na [/1]=SCA= &n [/1]=EMP= ni entrevista /=COB= **e tudo mais** /=APC= às vezes ele punha uma=SCA= uma toalha na cabeça /=COB= ou então ele vestia uma blusa ao contrário /=COB= e fingia de cachorrinho /=COB= e levantava a perna pra fazer xixi //=COM=$	Preenchimento
bfammn03	94	*ALO: [37] sio' /=SCA= precisa de ir lá p' siora despedir //=COM_r=$ [38] do pai lá /=COM_r= **e tal** //=APC_r=$	Preenchimento
	95	*ALO: [90] no Natal sessenta-e-dois /=TOP= a dona Maricota me fez uma compra /=SCA= lá na minha loja //=COM=$ *ALO: [91] eu tava liquidando /=COM= **até a loja lá** //=APC=$	Repetição Parcial de forma não-contígua
	96	*ALO: [107] ela me fez uma compra lá /=COB_r= comprou presente pr' ocês todo	Informação Tardia

		/=COB_r= pos menino /=COB_r= roupa /=COB_r= e tal /=COB_r= e /=DCT_r= ea nũ me pagou /=COB_r= também nũ cobrei /=COB_r= e já perdoei isso há muito tempo /=COB_r= nem converso mais não //=COM_r=$ [108] ah não /=COB_r= porque /=DCT_r= ela [/1]=EMP= a minha irmã /=SCA= quer pagar //=COM_r=$ [109] todo jeito /=COB= porque /=DCT= acho que teve uma /=SCA= conversa com ela há &m +=EMP=$ [110] ea já tinha morrido /=COB= nũ sei há quantos ano //=COM=$ *ALO: [111] e /=DCT_r= ela pediu pa pagar /=COM_r= **essa conta** //=APC_r=$	
	97	*ALO: [29] aí /=PHA= pôs lá um [/1]=SCA= uns tamboretes /=COB= pôs as vela /=COB= **e tal** /=APC=	Preenchimento
bfammn04	98	*REG: [48] porque o Gabriel ficou assim /=INT= super preocupado /=COB_r= **tal** /=APC_r= enquanto ele nũ me viu ele nũ <ficou tranqüilo> /=COM= sabe //=EXP=$	Preenchimento
	99	*REG: [54] mas aí ele viu /=COB= viu o João /=COB= me viu /=COB= **tal** /=APC= beleza //=COM=$	Preenchimento
	100	*REG: =$ [65] arrepiava /=COB= na hora que falava //=COM $ [66] como que pode /=COM_r= **nũ sei o quê** //=APC_r=$	Preenchimento
	101	*REG: [79] eu sei que eu morri de dor /=COB= assim /=PAR= depois que passa /=COB= né /=PHA= a gente quer tomar remédio /=COB= **tal** /=APC= eu falei assim /=INT= ô menino //=COM_r=$	Preenchimento
	102	*REG: [85] a médica nũ deixou nada prescrito não /=COM_r= **tal** //=APC_r=$	Preenchimento
	103	*REG: [87] falei assim /=INT= pois é /=CMM_r= providencia /=CMM_r= **tal** //=APC=$	Preenchimento
	104	*REG: [122] dor é coisa do passado /=COM_r= **nũ sei o quê** //=APC_r=$ [123] beleza //=COM=$	Preenchimento
	105	*REG: [150] e ocê /=TOP_r= nessa tranqüilidade /=COB_r= ainda falando de comida /=COB_r= de nũ sei o quê /=COM_r= **tal** //=APC_r=$	Preenchimento
bfammn05	106	*CAR: [32]queria uma criança que nũ me desse trabalho /=COM= **e tudo** //=APC=$	Preenchimento
	107	*CAR: [138] e meu marido tem uma &o [/1]=SCA= tem uma história também /=COB= que eu tenho certeza /=INT= se ele tivesse aqui /=PAR= ele ia te contar <u>essa história</u> //=COM=$ *MAR: [139] como é que é /=COM= **essa história** //=APC=$	Repetição literal de forma não-contígua
bfammn06		Não há APC	
bpubmn01	108	*SHE: [114] então é aquele entra-e-sai //=COM=$ *SHE: [115] então o' /=INP= dou aula pra quem /=SCA= quer assistir //=COM=$ *LUA: [116] mas isso é que idade //=COM=$	Preenchimento

		*SHE: [117] primeiro ano /=COB= né /=PHA=$ *LUA: [118] uhn //=COM=$ *SHE: [119] / de primeiro a terceiro //=COM=$ *SHE: [120] é Ensino Médio //=COM=$ *LUA: [121] Nossa //=COM=$ *SHE: [122] é //=COM=$ *SHE: [123] é complicado //=COM=$ *SHE: [124] e sem material /=COM= **também** /=APC= né //=PHA=$	
	109	*SHE: [47] não tenho tempo de fazer aquele planejamento /=COB= igual eu gostaria de fazer /=COM= **e tal** //=APC=$	Preenchimento
	110	*SHE: [49] mas uns dois /=SCA= uns dois três dia assim /=TOP= e &e=SCA= eu penso /=i-COB= né /=PHA= &he /=TMT= tipo um [/1]=EMP= faço em tópicos /=COM= **lá** //=APC=$	Preenchimento
	111	*SHE: [87] então /=INP= na escola que eu tenho a coordenadora /=TOP= eu vejo que /=SCA= a parte de /=INT= conversar com o aluno /=COB= né /=PHA= em relação à disciplina /=COB= problema familiar /=COB= &he /=TMT= problema de material /=COB= o aluno que tá desinteressado /=COB= **tudo** /=APC=	Preenchimento
	112	*SHE: [112] ah /=EXP_r= nũ sai não /=COB_r= **e tal** /=APC_r= mas que que é que tá acontecendo /=COB_r= não /=COB= tem um problema sempre pra resolver //=COM=$	Preenchimento

Anexo 2 – Análise Acústica

Texto	Numeração do exemplo	Média F0 – COM (HZ)	Média F0 – APC (HZ)	Média Int. – COM (dB)	Média Int. – APC (dB)
bfamcv01	1	189,9 HZ	131,5 HZ	71,07 dB	70,85 dB
	2	189,9	100	55,21	52,56
	3	270,4	237,7	65,28	56,54
	4	114,2	107,1	56,68	51,41
	5 e 6	COM1= 142Hz COM2= 132,7 Hz	APC1= 120,1 Hz APC2= 114,8 HZ	COM1= 53,18 dB COM2= 57,73dB	APC1= 49,47 dB APC2= 48,59 dB
	7	140,1	113,4	62,79 dB	55,65 dB
	8	205,9	240,5	72,05	71,23
	9	152,2	50	52,22	51,06
	10	142,6	128,5	62,41	61,64
bfamcv02	11	352	184,6	54,98	53,97
	12	222,3	172,5	63,24	54,14
	13	248,5	173,2	60,71	48,91
bfamcv03	14	216,8	144,3	65,68	60,46
	15	121,8	77,68	37,9	41,69
	16	146,6	90,9	46,34	34,22
bfamcv04	17	236	209,3	61,55	50,46
	18	289,7	203,8	62,74	59,37
	19	293,5	267,3	73,35	69,75
bpubcv01	20	167,9	214,9	62,9	67,75
	21	160,6	130,8	66,67	63,46
	22	162,2	170,5	69,93	70,76
	23	113,7	133,9	68,46	67,58
	24	239,95	50	63,91	59,39
bpubcv02	25	148	93,26	63,21	59,94
	26	130,4	118	69,98	58,92
	27	193,3	138,8	77,61	67,17
	28	126	72,45	64,17	54,89
	29	109,7	99,97	67,86	58,83
	30	147,8	111,4	71,08	58,98
	31	170,5	153,9	72,02	65,35
	32	146,3	128,5	69,62	68,16
	33	162,5	124,2	73,77	71,01
bfamdl01	34	262,9	241,4	68,01	64,78
	35	247,7	202,5	69,28	67,2
	36	281,9	126,5	79,67	71,65
	37	259,8	197,1	70,86	65,82
	38	308,8	230,7	73,89	75,1
bfamdl02	39	205,8	185,5	56,56	48,8
	40	224,5	150,8	52,67	44,46
	41	216,4	153	66,94	59,64
	42	300,4	83,89	53,13	37,33
	43	161,15	131,2	64,87	61,99
	44	246	215,2	71,5	70,63
bfamdl03	45	177,2	144,2	61,05	62,06
	46	138,3	?	58,9	64,12
	47	282,4	191,2	63,8	62,16
	48	221,9	198,8	59,94	55,77
	49	228,6	95,81	63,03	61,43
	50	83,77	71,89	56,13	51,98
	51	75,81	66,79	63,1	60,37

bfamdl04	52	255,8	165,3	61,17	50,39
	53	256,9	196,8	64,44	61,52
	54	216,1	132,3	58,89	53,35
	55	253,1	200,8	74,74	67,31
	56	235,7	173,3	59,37	52,48
	57	234,6	128,6	58,8	52,88
bfamdl05	58	75,22	52,1	62,95	57,88
	59	175,2	116,6	48,44	44,02
	60	240,3	138,3	61,15	50,71
	61	258,3	141,9	59,95	58,69
	62	140,1	84,23	54,22	57,24
	63	60,46	55,19	68,35	34,25
bpubdl01	64	102,6	97,71	60,84	61,78
bpubdl02	65	273,2	157,3	76,76	73,33
	66	187	109,6	69,87	63,21
	67	135,3	115,8	63,42	63,09
	68	186,7	184,1	69,1	64,7
	69	154,1	133,6	62,44	60,52
	70	165,5	139,2	75,2	74,11
bfammn01	71	137,1	57,88	92,42	49,18
	72	163,5	133,3	59,55	58,03
	73	179,7	100,6	55,34	44,8
	74	250	57,63	122,9	56,91
	75	97,53	90,5	47,38	38,54
	76	152,1	226,1	53,88	43,92
	77	134,7	108,4	67,06	58,43
	78	186,9	183,4	62,99	50,3
	79 e 80	107,7	APC1= 94,85 APC2 = ?	60,83	APC1= 56,65 APC2 = ?
	81	168,9	159,2	61,55	53,8
	82	164,1	112,2	52,25	47,37
	83	179,4	109,9	61,7	56,06
	84	209	180,3	61,08	49,65
	85	129,9	94,66	56,83	41
	86	227,1	294	45,07	46
	87	100,2	92,51	56,64	48,58
bfammn02	88	123	102,5	53,61	49,75
	89	212,1	93,63	48,38	44,95
	90	247,4	120,5	56,34	55,9
	91	250,8	175,9	50,84	56,5
	92	157,15	143,4	67,25	66,59
	93	227,5	174,5	55,64	52,79
bfammn03	94	108,5	72,82	62,76	49,24
	95	139,8	88,73	59,69	57,36
	96	121,2	85,77	60,79	51,42
	97	137,8	129,19	57,29	52,36
bfammn04	98	275,6	244,7	57,2	56,49
	99	212,6	205,5	58,28	57,49
	100	449,89	346,6	58,47	55,98
	101	283,56	265,4	61,73	54,41
	102	253,4	288,1	56,73	48,28
	103	374,2	223,2	58,27	52,37
	104	269,1	268,8	61,62	54,93
	105	241,8	137,3	55,73	48,97

bfammn05	106	221,1	194,7	60,49	53,2
	107	232,6	221,6	42,82	39,14
bfammn06	-				
bpubmn01	108	254,8	132,8	66,8	63,18
	109	226,4	224,5	67,70	57,24
	110	293,4	155,9	70,24	56,84
	111	207,8	199,40	67,47	53,66
	112	309,2	288,8	74,47	71,9

Anexo 3 – Análise Morfossintática

Texto	Numeração	Exemplo	Análise Morfossintática
bfamcv01	1	*LEO: 26] não /=INP_r= <a gente perdeu no campo /=COM_r= e tal> //=APC=$	ADV
	2	*GIL: [65] da gente /=TOP= pra alguns setores da organização /=TOP= chamar o pessoal /=COM= dos outro times //=APC=$	SP
	3	*GIL: [70] agora /=TOP= que deu pau /=APT= &d +=EMP=$ [71] que eu acho que deu muito pau /=COM= nessa taça //=APC=$	SP
	4	*EVN: [125] <é> /=CMM= concordo /=CMM= isso aí que cê falou //=APC=$	SN
	5 e 6	*LUI: [136] <eu acho que /=SCA= o nome /=SCA= do> [/1]=EMP= do [/1]=EMP= do [/1]=EMP= do [/1]=EMP= do [/1]=EMP= do [/1]=EMP= da modalidade esportiva que a gente criou /=TOP= que foi Futebol Arte /=PAR= ficou /=SCA= queimado /=COB= por causa dessa taça /=APC= pra mim é uma <coisa que> <é [/1]=SCA= foi triste> /=COB= é triste> falar /=COB= mas /=DCT= cê fala que é Futebol Arte /=CMM= a galera começa a zoar /=CMM= já //=APC=$	1 = SP 2 = ADV
	7	*GIL: [146] não pedindo desculpa /=COB= mas assim /=INT= falando /=COB= que essa [/1]=SCA= a próxima vai ser diferente /=COB= em tais e tais aspectos /=COM= e tal //=APC=$	ADV
	8	*EVN: [151] é /=INP= a <gente tem que> <restringir também /=COM= isso> //=APC=$	SN
	9	*GIL: [183] porque /=DCT= ele ficou naquele grupo difícil /=COM= e tal //=APC=$	ADV
	10	*GIL: [137] <é> //=COM=$ [138] todo mundo <encheu o saco> /=COM= <por causa disso> //=APC=$	SP
bfamcv02	11	*RUT: [92] eu nũ fui não /=COM= minha filha /=ALL= lá não //=APC=$	SADV
	12	*RUT: [279] <eu tenho é> só vergonha /=COM= de entrar <na igreja> //=APC=$	SV
	13	*RUT: [203] da família &d [/3]=EMP= <da família> dele /=TOP= minha filha /=ALL= <também nũ> tomo nem conhecimento /=COM= também //=APC=$	ADV
bfamcv03	14	*TON: [247] é ocê /=COM= Onofre /=ALL= que tá cagando assim //=APC=$ [248] n' é o senhor	SV
	15	*CAR: [274] por que /=COM= que vai ser pior //=APC=$	SV
	16	*CAR: [177] cê nũ entendeu /=COM= cê nũ //=APC=$	SN
bfamcv04	17	*HEL: [186] tá //=COM=$ [187] pra	SV

		/=SCA= nenhum <objeto> /=COM= que tenha //=APC=$	
	18	*HEL: [295] <e pr' ocê> falar que tá certo /=COM= isso //=APC=$	SN
	19	*HEL: [317] <não> /=COB= mas aí hhh +=EMP=$ [318] <e se precisar de um esclarecimento> /=COM= <assim /=PAR= no> meio do negócio hhh //=APC=$	SP
bpubcv01	20	*FLA: [7] sai do doador /=CMM= lá /=APC= e chega pra gente assim //=CMM=$	ADV
	21	*BRU: [127] <aquilo ali /=TOP= acho que eu já vi na televisão> /=COM= já //=APC=$	ADV
	22	*BRU: [171] <ah /=CMM= tá //=CMM=$ [172] porque primeiro cê centrifuga> /=COB= uma vez /=APC= aí tira o +=EMP=$	SN
	23	*EMM: [176] <a gente separa /=COM= isso> //=APC=$	SN
	24	*FLA: [252] mas /=INP= mata /=CMM= isso aí /=APC= será //=CMM=$	SN
bpubcv02	25	*CAR: [29] <n' é> não //=COM=$ [30] nũ é gato /=COM= nem nada não //=APC=$	PRONOME
	26	*CAR: [38] <entregar> pro Ibraim /=COB= e tudo /=APC= que <aí> //=COM=$	PRONOME
	27	*OSV: [45] não /=CMM= eu> falei com ele já /=COM= pra mudar isso //=APC=$	SV
	28	*CAR: [212] eu nunca fui na casa lá dela não /=CMM= mas diz que é <muito ruim> /=CMM= lá o local //=APC=$	SADV + SN
	29	*OSV: [253] <quarta-feira> /=COM= que ele foi lá //=APC=$	SV
	30	*OSV: [271] se for o caso só dela mesmo /=TOP= esses caso que /=SCA= não tiver jeito /=TOP= da [/1]=EMP= da gente fazer /=APT= tira ela então /=COM= né /=PHA= no caso //=APC=$	SP
	31	*CAR: [36] <tem que levar e> mostrar /=COM= o //=APC=$	ARTIGO
	32	*OSV: [63] passar em nome da prefeitura /=TOP= e /=SCA= ir tudo pra lá duma vez /=COM= em vez de vir <pra cá> //=APC=$	SV
	33	*OSV: [157] distancia /=COM= que es queria <colocar> //=APC=$	SV
bfamdl01	34	*REN: [103] também acho /=COM= que é //=APC=$	SV
	35	*FLA: [527] ali arroz /=COM= Tio João //=APC=$	SN
	36	*REN: [569] ah /=INP= também nũ tem outro /=CMM= a gente vai ter que levar esse /=CMM= de todo jeito //=APC=$	SP
	37	*FLA: [337] o' nós trocando /=COM= levando só dele //=APC=$	SV
	38	*FLA: [358] tem que lembrar ela /=COM= comprar acetona //=APC=$	SV
bfamdl02	39	*BEL: [74] cê nunca pensou em /=INT=	ADV

		em [/1]=EMP= em [/1]=EMP= tipo assim /=INT= junto com [/1]=SCA= com lingüística e tal /=TOP= mexer com isso /=COB= assim /=PAR= dando aula mesmo /=COM= e tal //=APC=$	
	40	*BEL: [138] <é> //=COM=$[139] e por falta de cuidado /=TOP= é foda /=COM= também //=APC=$	ADV
	41	*BAL: [158] existem vários> /=COB= só que a maioria /=TOP= &he /=EMP= tá julgando improcedência /=COB= tal /=COB= porque /=DCT= &he /=EMP= de certa forma /=TOP= a bancada evangélica /=TOP= eles tão /=SCA= muito contra /=COM= essa coisa /=APC= né //=PHA=$	SN
	42	*BEL: [181] então é completamente diferente /=COM= assim //=APC=$[	ADV
	43	*BAL: [211] então ele vai em qualquer lugar //=COM=$[212] tá vendo /=CNT= tem o velcro aqui /=COM= e tal //=APC=$	ADV
	44	*BEL: [255] tá até pesando menos /=COM_r= e tal //=APC=$	ADV
bfamdl03	45	*LUZ: [4] porque eu [/1]=SCA= eu me senti /=SCA= respirando /=i-COB= né /=PHA= adequada /=COM= né /=PHA= no lugar //=APC=$ [5] <falei> /=INT= Nossa //=COM_r=$	SP
	46	*LAU: [30] es vão admitir seis professores na /=SCA= Escola de Belas Artes /=COM= da uefeemegê //=APC=$	SP
	47	*LUZ: [77] mas tem o Ronan /=SCA= também /=COM= né /=PHA= <aqui> //=APC=$	ADV
	48	*LUZ: [81] e tem a Giovana //=COM=$ [82] <então> vai ser muito disputada /=COM= essa aí //=APC=$	SN
	49	*LUZ: [130] mas é pra mestre ou pa doutor /=COM= essa aí //=APC=$	SN
	50	*LUZ: [193] quando é o mesmo /=COM= coisa //=APC=$	SN
	51	*LUZ: [314] tava úmida /=COM= &he /=TMT= a coisa //=APC=$	SN
bfamdl04	52	*SIL: [8] capaz /=COM= que tá //=APC=$	SV
	53	*KAT: [14] antes de passar aquele creme ele nũ tava assim /=COM= espetado pos lados não //=APC=$	SADJ
	54	*SIL: [16] pode ser o creme /=COM= que nũ deu certo com ele //=APC=$ [	SV
	55	*SIL: [101] copos de Urano /=COM= que tem aí //=APC=$	SV
	56	*SIL [140] um /=TOP= usa o copo /=CMM= deixa lá /=CMM= arrupiado //=APC=$	ADJ
	57	*SIL: [209] foi por isso que es te deram só seis meses /=COM= na [/1]=EMP= naquela folheta //=APC=$	SP

bfamdl05	58	*ANE: [113] ah /=EXP= mas é porque aqui ea deve ser outro nome /=COM= então //=APC=$ [114] <nũ sei> //=COM=$	ADV
	59	*ENC: [200] mas ele nũ tá ativo ainda não /=CMM= ele só tá instalado /=CMM= só //=APC=$	ADV
	60	*CES: [233] <é /=INP= porque agora é rápido> /=COM= isso <aqui> //=APC=$	SN
	61	*ANE: [276] esse mesmo /=COM= que eu vim <ver> //=APC=$	SV
	62	*CES: [405] <aqui> /=TOP= ele é maior /=COM= <do que aqui> //=APC=$	SP
	63	*ANE: [85] andei isso aqui tudo /=COM= com o Tiago //=APC=	SP
bpubdl01	64	*PAU: [50] aonde /=CMM= que <ficou /=CMM= com quarenta-e-cinco> //=APC=$	SP
bpubdl02	65	*JAN: [43] <e uma> dessa /=COM= aqui /=PAR= também <o'> //=APC=$	ADV + INTERJEIÇÃO
	66	*EUG: [70] mas ele nũ deixa de ser legal também //=COM=$ [71] olha p' cê ver /=CNT= cê tá de blusa /=SCA= lilás /=COM= também /=APC= né //=PHA=$	ADV
	67	*EUG: [50] eu pego p' cê ver /=COM= dessa aqui //=APC=$	SP
	68	*EUG: [84] o modelo tradicional eu vou ver se eu acho /=COM= dele lá //=APC=$	SP
	69	*EUG: [87] tem nenhum par mais /=COM= daquele modelo //=APC=$	SP
	70	[158] achei Picadilly /=COM= trinta-e-nove lá dentro //=APC=$	NUMERAL + ADV
bfammn01	71	*MAI: [4] ele nũ é muito parente chegado não /=COB= mas &t [/1]=SCA= deve ser /=SCA= primo [/1]=EMP= primo quarto /=COM= por aí /=PAR= deve ser //=APC=$ [5]	SV
	72	*MAI: [9] e quando chegou lá /=TOP= &he /=TMT= montou uma casinha pa ele /=COB= pa família dele /=COB= e tal /=COB= e e' vinha na cidade pa compar alguma coisa &dif [/1]=SCA= diferente /=COM= que nũ era da roça /=APC= né //=PHA=$	SV
	73	*MAI: [13] o diâmetro dea deve dar uns [/1]=SCA= uns quarenta cinqüenta centímetro de [/1]=SCA= de &s [/2]=EMP= de grossura /=COM= o diâmetro dela //=APC=$	SN
	74	*MAI: [21] que /=DCT= na hora que ele lá envinha voltando do [/1]=SCA= do comércio /=TOP= que e' foi fazer a compra dele /=PAR= a cobra percebeu o cheiro dele /=COM= na hora que ele lá envinha no [/1]=SCA= no trilho //=APC= $	SP
	75	*MAI: [22] uma /=SCA= tora /=COM= da madeira //=APC=$	SP
	76	*MAI: [32] tipo &ba [/2]=EMP= chama barreta /=COB= o /=SCA= material /=COM= daquele chapéu //=APC=$	SP

	77	*MAI: [34] e ea veio /=COB= ea saiu no [/1]=SCA= no [/1]=SCA= no trilho /=COB= e rasgou o chapéu todo /=COB= nũ sendo ele ea [/1]=TOP= ea &rap [/2]=SCA= continuou atrás dele /=COM= quebrando //=APC=$	SV
	78	*MAI: [36] igual animal /=SCA= forte mesmo /=COM= a quebração que ea tava fazendo no meio do mato //=APC=$	SN
	79 e 80	*MAI: [67] segundo ele falou comigo /=COM= o /=SCA= parente dele /=APC= que &m [/2]=EMP= desse que morreu //=APC= $	1=SN 2= SP
	81	*MAI: [65] &he /=TMT= no outro dia que /=SCA= o sol tava [/1]=SCA= já tinha saído e tava tudo bem /=TOP= ea pegou e &f [/2]=SCA= e foi atrás de [/1]=SCA= de alguém /=COM= pa vim matar a cobra //=APC=$	SV
	82	*MAI: [71] certamente ea tava querendo fazer o seguinte /=INT= eu /=TOP= eu matei esse /=CMM= eu vou matar o resto tudo /=CMM= da [/1]=EMP= &he /=TMT= dentro da casa //=APC=$	SP
	83	*MAI: [76] só até aí que eu sei o caso /=COM= que ele me contou //=APC=$	SV
	84	*MAI:[73] ele matou /=COB= e essa família saiu desse lugar /=COM= não ficou mais //=APC=$	SV
	85	*MAI: [79] é lugar /=SCA= perigoso mesmo o norte de Minas /=COM= lá //=APC=$	ADV
	86	*MAI: [96] isso aqui /=TOP= em vista de lá /=TOP= é [/1]=EMP= é um cerrado /=COM= aqueas mata lá //=APC=$	SN
	87	*MAI: [68] nũ sei /=COB= né /=PHA= a imaginação hhh dum [/1]=SCA= dum animal /=COM= o que que pode ser /=APC= né //=PHA=$	SN
bfammn02	88	*LUC: [48] desde o começo /=COM= assim /=PAR= e tudo o mais //=APC=$	SN
	89	*DFL: [92] <e isso> /=TOP= assim /=PAR= era muito importante /=COM= pra ela //=APC=$	SP
	90	*DFL: [135] e então /=DCT= &he /=TMT= e /=SCA= papai foi estudar no Rio /=COB= e mandava /=COB= pra ele /=APC=	SP
	91	*DFL: [164] e quando eles /=SCA= vinha trazer a Maria Julieta /=TOP= ele ficava uns poucos dias depois vinham buscar /=PAR= e então /=DCT= ele participava /=COB= daquelas nossas brincadeira /=APC=	SP
	92	*LUC: [169] <uma charada /=COB= ali> /=APC= <na hora> //=COM=$	SP
	93	*DFL: [173] <eu até já> falei isso na [/1]=SCA= &n [/1]=EMP= ni entrevista	SN

		/=COB= e tudo mais /=APC= às vezes ele punha uma=SCA= uma toalha na cabeça /=COB= ou então ele vestia uma blusa ao contrário /=COB= e fingia de cachorrinho /=COB= e levantava a perna pra fazer xixi //=COM=$	
bfammn03	94	*ALO: [38] do pai lá /=COM_r= e tal //=APC_r=$	ADV
	95	*ALO: [91] eu tava liquidando /=COM= até a loja lá //=APC=$	SN
	96	*ALO: [111] e /=DCT_r= ela pediu pa pagar /=COM_r= essa conta //=APC_r=$	SN
	97	*ALO: [29] aí /=PHA= pôs lá um [/1]=SCA= uns tamboretes /=COB= pôs as vela /=COB= e tal /=APC=	ADV
bfammn04	98	*REG: [48] porque o Gabriel ficou assim /=INT= super preocupado /=COB_r= tal /=APC_r= enquanto ele nũ me viu ele nũ <ficou tranqüilo> /=COM= sabe //=EXP=$	ADV
	99	*REG: [54] mas aí ele viu /=COB= viu o João /=COB= me viu /=COB= tal /=APC= beleza //=COM=$	ADV
	100	*REG: [66] como que pode /=COM_r= nũ sei o quê //=APC_r=$	SV
	101	*REG: [79] eu sei que eu morri de dor /=COB= assim /=PAR= depois que passa /=COB= né /=PHA= a gente quer tomar remédio /=COB= tal /=APC= eu falei assim /=INT= ô menino //=COM_r=$	ADV
	102	*REG: [85] a médica nũ deixou nada prescrito não /=COM_r= tal //=APC_r=$	ADV
	103	*REG: [87] falei assim /=INT= pois é /=CMM_r= providencia /=CMM_r= tal //=APC=$	ADV
	104	*REG: [122] dor é coisa do passado /=COM_r= nũ sei o quê //=APC_r=$ [123] beleza //=COM=$	SV
	105	*REG: [150] e ocê /=TOP_r= nessa tranqüilidade /=COB_r= ainda falando de comida /=COB_r= de nũ sei o quê /=COM_r= tal //=APC_r=$	ADV
bfammn05	106	*CAR: queria uma criança que nũ me desse trabalho /=COM= e tudo //=APC=$	PRONOME
	107	*MAR: [139] como é que é /=COM= essa história //=APC=$	SN
bfammn06		Não há APC	
bpubmn01	108	*SHE: [124] e sem material /=COM= também /=APC= né //=PHA=$	ADV
	109	*SHE: [47] não tenho tempo de fazer aquele planejamento /=COB= igual eu gostaria de fazer /=COM= e tal //=APC=$	ADV
	110	*SHE: [49] mas uns dois /=SCA= uns dois três dia assim /=TOP= e &e=SCA= eu penso /=i-COB= né /=PHA= &he /=TMT= tipo um [/1]=EMP= faço em tópicos /=COM= lá //=APC=$	ADV
	111	*SHE: [87] então /=INP= na escola que eu tenho a coordenadora /=TOP= eu vejo que /=SCA= a parte de /=INT= conversar	PRONOME

		com o aluno /=COB= né /=PHA= em relação à disciplina /=COB= problema familiar /=COB= &he /=TMT= problema de material /=COB= o aluno que tá desinteressado /=COB= tudo /=APC=	
	112	*SHE: [112] ah /=EXP_r= nũ sai não /=COB_r= e tal /=APC_r= mas que que é que tá acontecendo /=COB_r= não /=COB= tem um problema sempre pra resolver //=COM=$	ADV

Anexo 4 – Total de APC no PB

Texto	Nome do texto	Nome em inglês	Total de APC
bfamcv01	Campeonato de Futebol	Soccer championship	10
bfamcv02	Preparativos para casamento	Marriage in the family	03
bfamcv03	Sinuca	Snooker	03
bfamcv04	Imagem e ação	Pictionary	03
bpubcv01	Banco de Sangue	Blood donation	05
bpubcv02	Reunião do PTB	Party meeting	09
		TOTAL	33
bfamdl01	Supermercado	At the supermarket	05
bfamdl02	Guardando equipamento	Packing the equipment	06
bfamdl03	Conversa no carro	A chat in the car	07
bfamdl04	Arrumação de cozinha	Kitchen cleaning	06
bfamdl05	Procura de apartamento	Broker	06
bpubdl01	Obra no sítio	Country house work	01
bpubdl02	Compra de sandália	Buying sandals	06
		TOTAL	37
bfammn01	A cobra	The snake	17
bfammn02	Drummond	Drummond	06
bfammn03	Causos Aloísio	Bad debt	04
bfammn04	Parto no carro	After delivery	08
bfammn05	Adoção	Adoption	02
bfammn06	Gerente de vendas	Salesman	0
bpubmn01	Educonle	Foreign language at public school	05
		TOTAL	42
TOTAL GERAL			112