

UNIVERSIDADE FEDERAL DE MINAS GERAIS

LUCIANA BEATRIZ BASTOS ÁVILA

**Modalidade em perspectiva:
estudo baseado em corpus oral do Português Brasileiro**

BELO HORIZONTE

2014

LUCIANA BEATRIZ BASTOS ÁVILA

**Modalidade em perspectiva:
estudo baseado em corpus oral do Português Brasileiro**

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

Área de Concentração: Linguística Teórica e Descritiva

Linha de Pesquisa: Estudos Linguísticos baseados em Corpora

Orientador: Prof^a. Dr^a. Heliana Ribeiro de Mello

Belo Horizonte

Faculdade de Letras da UFMG

2014

Ficha catalográfica elaborada pelos Bibliotecários da Biblioteca FALE/UFMG

A958m Ávila, Luciana Beatriz Bastos.
Modalidade em perspectiva [manuscrito] : estudo baseado em *corpus* oral do português brasileiro / Luciana Beatriz Bastos Ávila. – 2014.
253 f., enc. : il., grafs (color), tabs (p&b) + 1 CD-ROM.
Orientadora: Heliana Ribeiro de Mello.
Área de concentração: Linguística Teórica e Descritiva.
Linha de Pesquisa: Estudos Lingüísticos Baseados Em Corpora.
Tese (doutorado) – Universidade Federal de Minas Gerais, Faculdade de Letras.
Inclui CD-ROM com arquivos de áudio, guia do usuário MMAX2, demonstração de arquivo anotado.
Bibliografia: f. 206-218.
Anexo: f. 219-253.

1. Modalidade (Linguística) – Teses. 2. Linguística de corpus – Teses. 3. Língua portuguesa – Português falado – Brasil – Teses. 4. Atos de fala (Linguística) – Teses. 5. Linguística – Processamento de dados – Teses. I. Mello, Heliana Ribeiro de. II. Universidade Federal de Minas Gerais. Faculdade de Letras. III. Título.

CDD: 418



FOLHA DE APROVAÇÃO

Modalidade em perspectiva: estudo baseado em corpus oral do português brasileiro

LUCIANA BEATRIZ BASTOS AVILA

Tese submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em ESTUDOS LINGÜÍSTICOS, como requisito para obtenção do grau de Doutor em ESTUDOS LINGÜÍSTICOS, área de concentração LINGÜÍSTICA TEÓRICA E DESCRITIVA, linha de pesquisa Linha G-Estudos Inter-Relação Linguagem, Cognição e

Aprovada em 30 de abril de 2014, pela banca constituída pelos membros:

Prof(a). Heliana Ribeiro de Mello - Orientador
UFMG

Prof(a). Giulia Bossaglia
UFMG

Prof(a). Ricardo Augusto de Souza
UFMG

Prof(a). Lilian Vieira Ferrari
UFRJ

Prof(a). Maria Cláudia de Freitas
PUC-RJ

Belo Horizonte, 30 de abril de 2014.

*à memória do meu pai, José Walter,
que me ensinou a assimetria do tempo-espaço.*

AGRADECIMENTOS

Não se engana quem pensa que um doutorado é um osso duro de roer. No entanto, seria ainda mais duro se não estivéssemos cercados de um sem número de pessoas que nos apoiam, encorajam, dão colo, provocam, motivam.

À Prof. Heliana Mello, minha orientadora e amiga, parceira e cúmplice nessa jornada, que me guiou pela mão para a terceira margem e se empenhou de forma delicada, gentil e generosa para eu me tornar uma doutora. Por sua contribuição intelectual e por sua compreensão, paciência e carinho desde sempre.

À Amália Mendes, que me acolheu em Lisboa em meu período de Estágio Doutoral e não mediu esforços para que me integrasse à vida de trabalho no Centro de Linguística da Universidade de Lisboa. Pela nossa história de afeto e “muita amizade”. Torcendo por parcerias futuras.

Ao Prof. Tommaso Raso e Prof^ª Lílian Ferrari, pelos valiosos comentários em minha sessão de qualificação.

Aos membros da Banca Examinadora, por aceitarem prontamente o convite para leitura e discussão deste trabalho.

Aos colegas do LEEL, grandes companheiros: Priscila, Adriana e Raíssa, com quem dividi a peleja para aprender sobre a modalidade; Luís, pelas ótimas trocas; Bruno Rocha, Elisa, Helô, Maryualê, Andressa, Carol, Bruna e Bruno Alberto, pelos bons momentos passados juntos. Ao Pedro Perini, com quem venho construindo um produtivo diálogo sobre linguística e uma bela amizade. Ao Fred, por nossa história comprida de carinho.

Aos colegas do CLUL, Liliana, Raíssa, Aida, Sandra Antunes, Sandra Pereira, Fernando; aos professores Luísa Alice, Gabriela e João, pelos almoços, cafés da tarde, solzinho na esplanada e toblerones compartilhados. Ao Sérgio, da informática, pelo suporte técnico no meu momento de desespero, quando o programa não rodava. Meu agradecimento especial ao Agostinho, que já no primeiro dia de trabalho me ensinou, pacientemente, a operar o *software* de anotação e me provocou um lindo *insight* para a solução de uma questão que me perseguia.

À minha mãe Celeste, a quem não cabe adjetivos. Ela é superlativa.

À minha família, porque sim. À Clara, grande companheira de viagem, e a todos os sobrinhos, plenos de sonhos e esperanças.

À Fernanda, que estava lá na hora certa.

Ao Nando e à Angélica, amigos obrigatórios nos meus agradecimentos, na alegria e na tristeza, na saúde e na doença. Meu amor sempre.

Ao Francisco e à Antonieta, pela amizade, carinho, almoços de domingo, sabores maravilhosos e um bom vinho.

Ao Eduardo, amigo queridíssimo, pelo apoio incondicional e por tornar a vida mais fácil todos os dias. Café preto sem açúcar, fruta picada e CBN.

À Isabel, Mariana e Laurent, pelas portas e braços sempre abertos, do lado de cá e de lá do Atlântico.

À Chris, que foi chegando devagarinho e veio pra ficar. Risos, lágrimas, conversas matinais.

À Corita, Evandro e Fernandinho, reencontro maravilhoso, pelo carinho e a sensação de abraço apertado em Viçosa.

Aos amigos de sempre, espalhados pelo mundo, Viçosa, Juiz de Fora, Rio, Belo Horizonte, Lisboa, Ouro Preto, São Paulo, Campinas, Niterói, Nicosia, Madrid, Paris, Genève, Basel, Berna, Berlim... Vocês formam o mapa do meu afeto.

Ao Nuno, tão diferente e tão surpreendente, pelo apoio, paciência, carinho, sol&mar, comida preparada. Você segurou a minha onda no momento mais barra pesada da vida.

Ao Francisco Carlos e à Miúcha, que abriram suas casas para me acolherem. Aos companheiros de casa em Belo Horizonte e Lisboa, por me ensinarem a enxergar o outro.

Às funcionárias que cuidaram de meu pai e cuidam de minha mãe, a ajuda de Vocês nos cerca de tranquilidade.

Ao Thiago e ao Ronaldo, que além de amigos maravilhosos, me ajudaram com *abstracts*, *papers*, “uma tradução rapidinha”.

Ao Prof. Kleos, pela ajuda com as imagens e a mágica da formatação.

Aos meus colegas de departamento.

Às funcionárias da PPG, Suely e Margarida, pelo apoio na burocracia.

Aos funcionários do PosLin, prestativos e solícitos a nossas demandas.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), pela bolsa de estudos concedida para Estágio Doutoral em Portugal, no âmbito do Programa de Doutorado Sanduíche no Exterior (PDSE), e à UFV, pela bolsa de estudos no Brasil.

A todos os músicos de samba, chorinho, jazz, fado, soul, reggae, rock & blues. A vida é mais leve assim.

A todas as pessoas que passaram pela minha vida, que não estão aqui nomeadas, mas que certamente deram cor e graça a esta minha aventura.



“É preciso que as pessoas entrem e saiam.

Que vivam por toda a parte.”

(Herberto Helder)

RESUMO

Este estudo trata da descrição da modalidade em um *corpus* de fala espontânea da variedade mineira do português brasileiro. Modalidade, em uma definição mais enxuta, é a avaliação de um sujeito conceptualizador, que relativiza o material locutório que enuncia, em termos do grau de certeza, possibilidade, necessidade, capacidade e volição. No entanto, não existe consenso sobre a definição precisa desta categoria, por diferentes razões: (a) a modalidade é campo de estudo para a lógica e a linguística, que aplicam para este fim metodologias diversas; (b) a categoria se diferencia e está interrelacionada com outras como as de tempo, aspecto, modo; (c) esta noção se confunde e se sobrepõe às de ilocução, atitude e emoção. Com base nos princípios da Teoria da Língua em Ato, na metodologia da Linguística de Corpus e Linguística Computacional, o trabalho tem como objetivo descrever o comportamento de diferentes índices marcadores de modalidade e propor um esquema de anotação da modalidade para um *minicorpus* do C-ORAL-BRASIL. O C-ORAL-BRASIL é um *corpus* de fala espontânea do português brasileiro, o quinto braço do C-ORAL-ROM, um *corpus* comparável representativo das quatro principais línguas românicas europeias (italiano, espanhol, francês e português), segmentado prosodicamente em enunciados e suas subunidades informacionais. O enunciado é a unidade mínima de referência pragmaticamente interpretável. Para fins desta pesquisa, utilizo uma amostra da parte informal do *corpus* C-ORAL-BRASIL, um *minicorpus* composto de 20 textos de três tipologias interacionais, divididos em privados e públicos: 7 monólogos — 6 privados e 1 público —; 7 diálogos — 5 privados e 2 públicos — e 6 conversações — 4 privadas e 2 públicas. Para o tratamento nesta tese, foram selecionados 1.197 índices modais, distribuídos em 1.046 enunciados, codificados e posteriormente anotados semanticamente através do *software* MMAX2. Após a análise quantitativa e qualitativa, podemos concluir que (i) a categoria da modalidade se comporta de forma distinta da escrita na fala espontânea; (ii) os índices modais são amplamente utilizados em situações dialógicas, sejam públicas ou privadas; (iii) apenas algumas unidades informacionais podem ser modalizadas: Comentário, Parentético, Tópico, Introdutor Locutivo; (iv) a unidade de Comentário é a mais modalizada e não há restrição de realização de itens nesta UI; (v) a modalidade epistêmica é a mais frequente entre os tipos de significado; (vi) os verbos são a estratégia preferencial para marcação de modalidade.

Palavras-chave: modalidade; *corpus* oral; fala espontânea; anotação semântica; português brasileiro

ABSTRACT

This study deals with the description of modality in a *corpus* of spontaneous Brazilian Portuguese spoken language, more specifically, the Minas Gerais variety. Modality, in a more concise definition, is the evaluation of a conceptualizer, which relativizes the uttered locutive material in terms of degree of certainty, possibility, necessity, capability, and volition. However, there is no consensus on a precise definition of such category, explained by a number of reasons: (a) modality is a field of study for logics and linguistics that utilizes various methodologies; (b) the categories is both differentiated and interrelated to others, such as time, aspect, mode; (c) this notion is mistaken and overlaps those of illocution, attitude, and emotion. Based on the principles of the Speech Act Theory, in Corpus and Computational Linguistics, this work aims at describing the behavior of different modality markers and proposing a modality annotation scheme for a C-ORAL-BRASIL *minicorpus*. C-ORAL-BRASIL is a *corpus* of Brazilian Portuguese spoken language, the fifth branch of C-ORAL-ROM, a comparable *corpus*, representative of the four main European Romance Languages (Italian, Spanish, French, and Portuguese), prosodically segmented in utterances and their informational sub-units. The utterance is the minimal reference unit pragmatically interpretable. To better fit the purposes of this research, I utilized the informal part of the C-ORAL-BRASIL *corpus*, a *minicorpus* composed of 20 texts of three interactional typologies, divided into private and public: 7 monologues – 6 private and 1 public –; 7 dialogues – 5 private and 2 public –; and 6 conversations – 4 private and 2 public ones. For the treatment in this thesis, 1.197 modal indexes, distributed in 1.046 utterances, codified and, later on, semantically annotated through the MMAX2 *software* were utilized. After the quantitative and qualitative analysis, we can conclude that (i) the category of modality behaves distinctively in written and spontaneous oral language; (ii) the modal indexes are widely utilized in dialogic situations, be them public or private; (iii) only a few informational units can be modalized: Comment, Parenthetical, Topic, Locutive Introducer; (iv) the Comment unit is the most modalized and there is no realization restriction for this IU; (v) the epistemic modality is the most frequent one among the types of meaning; (vi) verbs are the preferred strategy for modality marking.

Keywords: modality; oral *corpus*; spontaneous speech; semantic annotation, Brazilian Portuguese.

LISTA DE FIGURAS

Figura 2.1 – Quadrado modal	25
Figura 2.2 – Mapa semântico da modalidade. Fonte: van der Auwera e Plungian (1998, p. 111)	44
Figura 2.3 - <i>Continuum</i> modalidade epistêmica/evidencialidade	58
Figura 2.4 – Projeção do ato comunicativo	66
Figura 4.1 – Frequência de enunciados modalizados no <i>minicorpus</i>	102
Figura 4.2 – Frequência de enunciados por tipologia e contexto interacional	104
Figura 4.3 – Frequência de índices modais por tipologia e contexto interacional	105
Figura 4.4 – Distribuição das estratégias modalizadoras na amostra (%)	107
Figura 4.5 – Frequência absoluta das unidades informacionais em relação aos índices modalizados	108
Figura 4.6 – Frequência relativa das unidades informacionais no <i>minicorpus</i>	109
Figura 4.7 – Distribuição dos valores modais (%)	110
Figura 4.8 – Distribuição absoluta de unidades informacionais por valor modal	112
Figura 4.9 – Frequência relativa de unidades informacionais por valor modal	113
Figura 4.10 – Frequência dos verbos modalizadores	116
Figura 4.11 – Ocorrência dos principais verbos modais / tipo de modalidade	117
Figura 4.12 – Frequência relativa dos valores modais dos principais verbos modalizadores	118
Figura 4.13 – Distribuição e frequência dos verbos epistêmicos na amostra	124
Figura 4.14 – Distribuição e frequência dos epistêmicos em unidades informacionais	127
Figura 5.1 – Tela principal do software MMAX2	155
Figura 5.2 – Criação de <i>markable</i>	158
Figura 5.3 – Estabelecer <i>links</i> entre <i>markables</i>	159
Figura 5.4 – <i>Links</i> entre <i>markables</i> estabelecido	159
Figura 5.5 – <i>Modal-class</i> : número do <i>set</i> modal	160
Figura 5.6 – Remover o <i>link</i>	161
Figura 5.7 – Apagar um <i>markable</i>	161
Figura 5.8 – Adicionar uma sequência ao <i>markable</i>	162
Figura 5.9 – Remover sequência de um <i>markable</i>	162
Figura 5.10 – Caixa de anotação de valores e atributos	163
Figura 5.11 – Tela principal de anotação	163
Figura 5.12 – Painel de controle	164
Figura 5.13 – Tela de validação da anotação	164

LISTA DE QUADROS E TABELAS

Quadro 2.1 – Correspondência da modalidade epistêmica entre a tradição lógica e a funcional	32
Quadro 2.2 – Termos gramaticais para a MOF e suas funções	42
Tabela 3.1 – Características dos textos do <i>minicorpus</i>	91
Tabela 3.2 – Informações sobre os metadados	94
Tabela 3.3 – Informações sobre lemas e lexemas	95
Tabela 3.4 – Etiquetagem de unidades informacionais	95
Tabela 3.5 – Etiquetagem de valores e subvalores modais	96
Tabela 3.6 – Etiquetagem de <i>Part of speech</i> (PoS) e índices morfológicos	97
Tabela 3.7 – Organização e codificação de dados	98
Tabela 3.8 – Exemplo de organização dos dados das construções condicionais	99
Tabela 4.1 – Frequência de enunciados modalizados x contexto e tipologia interacional	103
Tabela 4.2 – Frequência de índices modais por contexto e tipologia interacional	104
Tabela 4.3 – Estratégias modalizadoras	106
Tabela 4.4 – Frequência relativa das unidades informacionais na amostra	108
Tabela 4.5 – Tipos de modalidade	109
Tabela 4.6 – Estratégias modalizadoras e tipos de modalidade	110
Tabela 4.7 – Distribuição dos valores modais nas unidades informacionais	112
Tabela 4.8 – Distribuição dos verbos modalizadores no <i>minicorpus</i>	115
Tabela 4.9 – Distribuição e frequência dos principais verbos modalizadores	118
Tabela 4.10 – Distribuição de <i>tokens</i> de verbos epistêmicos por tipologia interacional	124
Tabela 4.11 – Distribuição dos verbos epistêmicos em unidades informativas	126
Tabela 4.12 - Número de ocorrências dos advérbios modais	130
Tabela 4.13 - Distribuição dos advérbios x tipologia interacional	131
Tabela 4.14 – Distribuição dos advérbios modais nas unidades informacionais	131
Tabela 4.15 – Distribuição de <i>tokens</i> de condicionais por tipologia interacional	136
Tabela 4.16 – Frequência dos padrões sintáticos de condicionais	136
Tabela 4.17 – Frequência das condicionais quanto à estrutura informacional	137
Tabela 4.18 – Relação do padrão sintático de condicionais e estrutura informacional	137
Tabela 5.1 – Valores e subvalores para o esquema do português europeu	150
Tabela 5.2 – Elementos identificados e/ou anotados em diferentes projetos de anotação semântica	152
Tabela 5.3 – Correspondência entre valores na tradição linguística e nos esquemas de anotação	153
Tabela 5.4 – Valores e subvalores modais e suas definições	173
Tabela 5.5 – <i>Sources</i> correspondentes a cada valor e subvalor modais	182
Tabela 5.6 – Frequência dos valores modais na amostra	198

LISTA DE ABREVIATURAS

ALL - Alocutivo	INT - Introdutor Locutivo
APC - Apêndice de Comentário	IPO - Instituto de Pesquisa Perceptual
APT - Apêndice de Tópico	LABLITA - Laboratório Linguístico do
b - Brasileiro (Português)	Departamento de Italianística
fam - Familiar	LEEL - Laboratório de Estudos Empíricos e
pub - Público	Experimentais da Linguagem
cv - Conversação	M - Masculino
dl - Diálogo	PAR - Parentético
mn - Monólogo	PB - Português do Brasil
CMM - Comentário Múltiplo	PE - Português Europeu
CNT - Conativo	PLN - Processamento de Linguagem Natural
COB - Comentário Ligado	PHA - Fático
COM - Comentário	PRL - Lista de Parentéticos
DCT - Conector Discursivo	SCA - Unidade informacional escandida
EMP - Unidade informacionalmente vazia	TLA - Teoria da Língua em Ato
EXP - Expressivo	TMT - Tomada de Tempo
F - Feminino	TOP – Tópico
F ₀ - Frequência fundamental	TPL - Lista de Tópicos
Hz - Hertz	UI - Unidade de informação
i- - Unidade informacional interrompida	UNC - Unidade informacional irreconhecível
INP - Incipitário	XML - Linguagem de marcação estendida

SUMÁRIO

Capítulo 1 INTRODUÇÃO	17
1.1 Justificativa	18
1.2 Quadro hipotético	20
1.2.1 Hipóteses	20
1.2.2 Objetivos	20
1.3 Organização do texto	21
PARTE I – DISCUSSÃO TEÓRICA	23
Capítulo 2 MODALIDADE: estado da arte	24
2.1 Breve percurso histórico: a modalidade na lógica	24
2.2 A modalidade é? Visões sobre o tema	27
2.2.1 A abordagem formalista – a semântica dos mundos possíveis	28
2.2.2 <i>Realis</i> x <i>Irrealis</i>	31
2.2.3 Subjetividade	33
2.3 Tipos de significados modais	39
2.3.1 As visões não-tradicionais: tipologias alternativas	40
2.3.2 A visão tradicional	45
2.3.2.1 Modalidade epistêmica: a linguagem das possibilidades	47
2.3.2 Modalidade deôntica	50
2.3.3 Modalidade dinâmica	53
2.4 Modalidade epistêmica e evidencialidade: o limite (?) entre categorias	55
2.5 O escopo da modalidade e as fronteiras entre algumas categorias	58
2.5.1 Modalidade e negação	59
2.5.2 Modalidade e modo	61
2.5.3 Modalidade, ilocução e atitude	63
2.6 Modalidade na fala espontânea	67
PARTE II – PROCEDIMENTOS METODOLÓGICOS E ANÁLISE DOS DADOS	74
Capítulo 3 PROCEDIMENTOS METODOLÓGICOS	75
3.1 A Teoria da Língua em Ato	75
3.1.1 O enunciado	76
3.1.2 Unidades Informacionais	79

3.1.2.1 Comentário (COM)	79
3.1.2.2 Tópico	80
3.1.2.3 Apêndice de Comentário (APC)	81
3.1.2.4 Apêndice de Tópico (APT)	81
3.1.1.5 Parentético (PAR)	82
3.1.1.6 Introdutor locutivo (INT)	83
3.2.2 Unidades Dialógicas	83
3.2.2.1 Incipitário (INP)	83
3.2.2.2 Conativo (CNT)	84
3.2.2.3 Fático (PHA)	84
3.2.2.4 Alocutivo (ALL)	84
3.2.2.5 Expressivo (EXP)	85
3.2.2.6 Conector Discursivo (DCT)	85
3.2.3 A perda do isomorfismo	85
3.2.3.1 Unidade de escansão (SCA)	86
3.2.3.2 Comentários Múltiplos (CMM)	86
3.2.3.3 Estrofes e Comentários Ligados (COB)	87
3.2 O corpus de fala espontânea: o C-ORAL-BRASIL	88
3.3 Coleta e organização dos dados	89
3.3.1 O <i>minicorpus</i> do C-ORAL-BRASIL I	89
3.3.2 A busca e organização dos índices	92
3.4 Análise dos dados	99
Capítulo 4 A MODALIDADE NO MINICORPUS DO C-ORAL-BRASIL I:	101
resultados e análise	
4.1 Frequência de enunciados modalizados na amostra	101
4.2 Enunciados e tipologia interacional	103
4.3 Frequência dos índices morfolexicais de modalidade	105
4.3.1 Distribuição dos índices em relação à unidade informacional	107
4.4 Frequência da modalidade quanto aos tipos modais	109
4.4.1 Frequência de valor modal por índice lexical	110
4.4.2 Relação entre o valor modal e as unidades informacionais	111
4.5 Análise de índices marcadores de modalidade	114
4.5.1 Os verbos modalizadores	114

4.5.1.1	Frequência dos verbos modalizadores	114
4.5.2	Os verbos epistêmicos	117
4.5.2.1	Os números para os verbos epistêmicos	122
4.5.2.2	Padrões sintáticos, semântica e questões pragmáticas	125
4.5.3	Advérbios e adjetivos modais	129
4.5.3.1	Frequência de advérbios e locuções adverbiais modais	129
4.5.3.2	Distribuição em unidades informacionais	131
4.5.4	As construções condicionais: <i>se p, então q</i>	135
4.5.4.1	Os números para as construções condicionais de forma canônica	135
	PARTE III – ANOTAÇÃO SEMÂNTICA DA MODALIDADE	141
	Capítulo 5 O PROJETO MASS (MODAL ANNOTATION IN SPONTANEOUS SPEECH): anotação semântica da modalidade no C-ORAL-BRASIL	142
5.1	Trabalhos anteriores em anotação semântica	143
5.2	Panorama geral (ou quem identifica e/ou anota o quê)	152
5.3	Proposta para anotação da modalidade no C-ORAL-BRASIL: o projeto MASS (Modal Annotation in Spontaneous Speech)	154
5.3.1	Metodologia	154
5.3.1.1	O software de anotação MMAX2	154
5.3.1.2	A preparação dos textos	156
5.3.1.3	Definição e formalização da tarefa de anotar	158
5.3.1.4	O esquema de anotação em versão XML	165
5.3.2	Escolha dos valores modais	168
5.3.3	<i>Triggers, Sources e Targets</i> : elementos a serem anotados	174
5.3.3.1	<i>Triggers</i>	174
5.3.3.2	<i>Sources</i>	180
5.3.3.2.1	<i>Source of the event mention</i>	180
5.3.3.2.2	<i>Source of the modality</i>	181
5.3.3.3	<i>Targets</i>	185
5.3.3.4	Identificação de <i>targets</i> , de acordo com o valor modal	193
5.3.4	Polaridade	197
5.4	Resultados	198
5.4.1	Acordo entre anotadores	199
	Capítulo 6 Considerações finais	201

REFERÊNCIAS BIBLIOGRÁFICAS

206

ANEXOS

Anexo 1 MASS 1.0 - Manual de anotação

Anexo 2 Guia de uso MMAX2 (apenas na versão eletrônica)

Anexo 3 Demonstração de arquivo anotado (apenas na versão eletrônica)

Capítulo1

INTRODUÇÃO

O presente trabalho se insere em um projeto sobre a modalidade, desenvolvido desde o ano de 2008, na Universidade Federal de Minas Gerais, denominado “Modalidade na fala espontânea do Português Brasileiro: um estudo de corpus”. Este projeto tem como objetivo investigar a modalidade em um corpus oral de fala espontânea do português brasileiro, por diferentes motivos: (1) pela dificuldade de compreensão dessa categoria gramatical; (2) pelo número ainda tímido de estudos sobre o tema no português brasileiro; e (3) pela necessidade de “descrições e estudos pontuais” da manifestação oral do português brasileiro.

Algumas das decisões teórico-metodológicas aqui adotadas, portanto, são fruto das discussões realizadas pelo grupo de pesquisa “Interfaces linguagem, cognição e cultura – INCOGNITO”, certificado pelo CNPq e liderado pelos professores Heliana Mello e Tommaso Raso. Assim, parti de estudos preliminares, empreendidos dentro do projeto C-ORAL-BRASIL, que levantaram as principais estratégias modalizadoras na fala espontânea do PB (MELLO, CARVALHO, CORTES, 2010; MELLO, RAMOS, AVILA, 2011; CORTES; MELLO, 2013; MELLO; CAETANO, 2012).

O trabalho que ora apresento pretende ser um passo à frente no que diz respeito à descrição do comportamento de diferentes índices marcadores do fenômeno semântico da modalidade em um *corpus* oral da variante mineira do português brasileiro.

Desde o século XVII, segundo Arendt (1958/2010, p. 310), “todos os grandes autores, cientistas e filósofos [...] declaravam ver coisas jamais antes vistas e ter pensamentos jamais antes pensados”. Como proposta de tese de doutoramento, pretendo tratar de um tema que já foi bastante pensado e discutido desde a Antiguidade Clássica e por autores de diferentes persuasões: a categoria semântica da **modalidade**. No entanto, tal como os grandes nomes da história, revolucionários ou não, sofremos insistentemente do “pathós da novidade” e, portanto, arrisco pensar a modalidade, diferentemente da concepção clássica¹, como uma categoria menos estática e como uma noção que se realiza também em um contexto interacional. Logo, esse estudo pode ser realizado tão somente pela análise da língua em uso e, no caso da pesquisa em tela, através de um estudo empírico.

¹ Trato da concepção clássica da modalidade, e de outras definições da categoria, no capítulo 2.

Com base nos princípios da Teoria da Língua em Ato (CRESTI, 2000) e na metodologia da Linguística de Corpus, a pesquisa que proponho desenvolver tem como objetivo descrever (e tentar explicar) o comportamento de diferentes índices marcadores de modalidade — lexicais ou gramaticais —, como, por exemplo, verbos modais, advérbios ou locuções modalizadoras, verbos de crença, verbos de volição, construções adjetivas modalizadoras, construções condicionais e o futuro.

A pesquisa terá como foco a diamesia oral da variante brasileira do português, a partir da análise (qualitativa e quantitativa) da parte informal de um *corpus* oral representativo da fala espontânea,² o C-ORAL-BRASIL (cf. RASO; MELLO, 2012), a ser descrito mais adiante. Como exemplos desses índices, temos:

(1.1) *RUT: [178] <mas eles> também **deve** ter condições / uai //

(1.2) *JAE: [88] mas **é lógico** que ea vai pôr ocês / uai //

1.3) *CAR: [12] e /=TXC= e eu **achei** que **ia** me **dar** trabalho // =COM=\$

(1.4) [123] o Zé Carlos **vai dar** a televisão // =COM=\$

(1.5) *LUC: [361] <isso **pode** ser /=SCA= **potencialmente**> divertido /=COM=\$

1.1 Justificativa:

Um estudo sobre a modalidade se justifica por diferentes razões. Primeiro por ocupar um lugar de destaque nos estudos da linguística nas últimas décadas e ser um conceito ainda controverso (e complexo), que se pode sobrepor a outros como os de atitude, ilocução ou modo. Mello e Raso (2011) propõem uma discussão interessante sobre as categorias de atitude, ilocução e modalidade, a fim de facilitar a identificação de traços que possam caracterizar esses diferentes conceitos. De acordo com os autores, essas três definições muitas vezes se sobrepõem, o que leva a problemas metodológicos cruciais para a análise prosódica, uma vez que a prosódia codifica coisas diferentes ao mesmo tempo. Os autores, na tentativa de separar os domínios de aplicação de cada um desses conceitos, argumentam que eles

² Para fala espontânea, acompanho a proposta de Cresti e Scarano (1998, p. 5) que entendem o espontâneo “como o cumprimento de atos linguísticos, nem programados, nem programáveis, porque produzidos durante e no desenvolvimento de uma interação, sempre nova e imprevisível, entre locutores”. As autoras destacam o caráter dialógico da fala considerada espontânea, fundamentadas na Teoria dos Atos de Fala, que relaciona o ato de falar com o cumprimento de diversas ações linguísticas (AUSTIN, 1962). Tais ações linguísticas, em si, finitas e circunscritas, não são o produto de um locutor ativo e isolado, mas sim se relacionam a um locutor engajado em uma situação dinâmica e interativa com vários interlocutores. Segundo as autoras, o princípio que rege os textos da fala espontânea repousa sobre um fundamento ilocutório, ausente na escrita, que contém uma articulação informacional específica.

devem ser estabelecidos como instâncias de diferentes fenômenos os quais são aplicados a diferentes níveis do ato comunicativo, e podem, a princípio, ser composicionais.

Em segundo lugar, entendo que partir de *corpora* orais de fala espontânea para a análise do fenômeno pode nos levar a caminhos de investigação com foco em “contextos específicos de uso de expressões modais e como eles correspondem a necessidades específicas de comunicação” (CORNILLIE; PIETRANDREA, 2012) e, por consequência, à importância dos movimentos linguísticos dos participantes na negociação de sentidos no curso de uma interação. De fato, se o objetivo que aqui se delineia é trabalhar com a descrição de construções da língua, os estudos de *corpora*, em termos metodológicos, vêm contribuir para a manipulação de um grande número de dados representativos do uso, com o recurso de ferramentas computacionais sofisticadas. Dar conta da distribuição das expressões de uma língua se constitui como uma tarefa hercúlea. A pesquisa de *corpus*, dessa forma, pode determinar a frequência com a qual uma determinada expressão (ou construção) pode ser usada em diferentes contextos. Gries destaca, em artigo sobre a polissemia do verbo *to run*, como “os métodos quantitativos da linguística de corpus podem fornecer evidência empírica, sugerindo respostas para alguns problemas notoriamente difíceis da linguística cognitiva” (GRIES, 2006, p. 57).

Por último, mas não menos importante, o tema da modalidade cobre diferentes construtos de abordagens baseadas no uso ainda não combinados. Os estudos linguísticos no Brasil sobre a modalidade ainda carecem de análises que levem em conta as expressões em contexto, com dados da fala espontânea, organizados sistematicamente, ainda que haja projetos que explorem o estudo da língua falada (como o NURC e seus desdobramentos em publicações).

Como consequência, o trabalho que ora se apresenta pode fomentar a discussão sobre semelhanças e diferenças em relação a variedades da língua portuguesa e em relação a outras línguas de base românica (o italiano, o francês e o espanhol). Desse modo, a pesquisa se mostra relevante na medida em que a análise pode colaborar na reavaliação de procedimentos teórico-metodológicos nesse campo e contribuir para uma melhor compreensão da relação da cognição humana e o comportamento social.

Para além da própria importância do esforço descritivo da modalidade na fala, a tarefa de se propor um esquema de anotação semântica da modalidade, para seu reconhecimento automático, se constitui como campo ainda inexplorado no português brasileiro tanto para dados escritos quanto dados orais. Um projeto de anotação para o PB, inspirado em outros desenvolvidos para a língua inglesa e o português europeu (PE), pretende, portanto, contribuir

para um número de aplicações para o Processamento de Linguagem Natural (PLN) e, também, para o próprio estudo linguístico da modalidade.

1.2 Quadro hipotético:

1.2.1 Hipóteses:

(a) Há uma relação entre a tipologia da modalidade, os perfis linguísticos e a estrutura informacional dos enunciados do Português Brasileiro;

(b) As construções que carregam índices de modalidade cumprem diferentes funções, a depender do contexto de uso;

(c) O uso de índices de modalidade depende da tipologia do texto, do gênero discursivo, do propósito comunicativo e, além disso, da relação entre os participantes da interação.

1.2.2 Objetivos:

Como objetivos, pretendemos:

Objetivo geral:

- reavaliar, a partir de um estudo de um corpus de fala espontânea, a definição de modalidade, no sentido de delimitar a fronteira entre semântica e pragmática, e, a partir disso, definir e descrever os marcadores modais;

Objetivos específicos:

- explorar o comportamento de diferentes índices marcadores de modalidade, — lexicais ou gramaticais —, em um corpus oral da variante brasileira do português.
- sugerir uma correlação entre a tipologia da modalidade do Português Brasileiro, com base nas unidades informacionais que compõem seus enunciados;

- investigar os contextos de uso dos diferentes índices de modalidade.
- elaborar um esquema de anotação da modalidade, a ser aplicado em uma amostra de um *corpus* representativo da fala espontânea do português brasileiro.

1.3 Organização do texto:

Este trabalho está organizado em três grandes eixos: uma parte teórica, que trata da definição de modalidade e da sua tipologia; os procedimentos metodológicos e a descrição dos itens modais, de base empírica, em uma amostra de *corpus* oral do PB, e a última parte em que apresento uma proposta de esquema de anotação semântica da modalidade. A narrativa, assim, se constrói desde uma discussão teórica, passa pela descrição do fenômeno, a partir de consistente base empírica, e termina com sua aplicação para fins de PLN, com a elaboração do esquema de anotação.

O capítulo 2 traz uma discussão do conceito de modalidade corrente na literatura sobre o tema; os tipos de significados modais; as fronteiras desta categoria semântica com as de negação, modo, ilocução e atitude; a fronteira entre modalidade e evidencialidade; a modalidade na fala espontânea e a minha proposta de definição que leva em conta a modalidade ancorada em uma situação comunicativa.

Em seguida, no capítulo 3, são detalhados os procedimentos metodológicos, é introduzido o quadro teórico da Teoria da Língua em Ato, o projeto C-ORAL-BRASIL e o *minicorpus* utilizado nesta pesquisa.

O capítulo 4 apresenta a descrição da modalidade na amostra de 20 textos da parte informal do *corpus* C-ORAL-BRASIL. Este *minicorpus* é representativo e está definido segundo critérios de alta qualidade. Após serem discutidos os resultados para a distribuição e frequência dos índices no que respeita a tipologia interacional, as unidades informacionais, e os valores modais, empreendo uma análise de três estratégias modalizadoras lexicais – verbos modalizadores, verbos epistêmicos e advérbios – e uma gramatical – as construções condicionais.

O capítulo 5, que constitui a terceira parte da tese, introduz o projeto MASS – Modal Annotation of Spontaneous Speech – que se trata de um esquema de anotação da modalidade para a fala espontânea do português brasileiro. Parte desta pesquisa foi desenvolvida no

Centro de Linguística da Universidade de Lisboa, sob orientação da Dr^a. Amália Mendes, dentro do Programa de Doutorado Sanduíche no Exterior da Capes (PDSE/Capes).³

Por fim, no capítulo 6, apresento as considerações finais, no sentido de apontar os objetivos alcançados e sinalizar para perspectivas futuras de trabalho.

Em resumo:

Esta introdução pretendeu explicitar o objeto de estudo, os problemas a serem enfrentados em relação à categoria semântica da modalidade, a justificativa da pesquisa, o quadro hipotético, os objetivos a serem perseguidos e a estrutura do trabalho. No capítulo seguinte, pretendo situar o debate acerca do que vem sendo dito sobre modalidade, o consenso entre as definições e os problemas da conceituação e da tipologia aplicadas aos dados da fala. Por fim, avanço para uma proposta de reformulação do conceito.

³ Processo nº BEX 9537/12-0.

PARTE I – DISCUSSÃO TEÓRICA

Capítulo 2

MODALIDADE: estado da arte

Este capítulo pretende elucidar o caminho percorrido pelo conceito de modalidade, desde as abordagens lógicas, passando pela sua apropriação pelos semanticistas formais, pelas concepções correntes na literatura linguística, sejam formalistas ou funcionalistas, para chegar ao que nos interessa como debate, a saber, a modalidade aplicada a dados de fala espontânea. O objetivo é assumir um conceito que pressuponha a divisão entre semântica e pragmática no que diz respeito a essa categoria, atribuindo à semântica aquilo que está relacionado à conceptualização e ao significado do material locutório e, à pragmática, os aspectos referentes à estrutura informacional, à força ilocucionária, ao conjunto de informações contextuais e às estratégias adotadas pelos interlocutores no curso da interação.

2.1 Breve percurso histórico: a modalidade na lógica

A noção de modalidade vem sendo debatida, ao longo da história, por filósofos, lógicos, linguistas. Autores de diferentes persuasões tratam o tema de maneira diversa, fundados, principalmente, na ideia comum de que a modalidade está comprometida com a noção de verdade e com a opinião do falante sobre o que enuncia uma determinada proposição. No entanto, a tarefa de definir o que é essa categoria é difícil, por uma série de motivos elencados por Mello, Carvalho e Cortes (2010, p. 108):

(a) o de tal categoria haver sofrido, em sua tradição de estudo, uma grande influência da Lógica, havendo assim uma mescla entre estudos filosóficos e matemáticos e aqueles da linguagem natural, o que acarreta uma mistura metodológica nem sempre produtiva para o estudo da língua em uso (cf. verificação de valores de verdade); (b) o de essa categoria existir em interrelação, no sistema gramatical de uma língua, com vários fenômenos gramaticais, como *tempo*, *aspecto* e *modo*, *prosódia*, *organização da informação*, dentre outros; e (c) o fato de o próprio conceito de modalidade confundir-se com aqueles de atitude, ilocução e emoção.

Ou como coloca Bybee et al. (1994, p. 176), “pode ser impossível chegar a uma caracterização sucinta do domínio nocional da modalidade.”⁴

⁴ Tradução do original: “it may be impossible to come up with a succinct characterization of the notional domain of modality.” (BYBEE et al., 1994, p. 176).

Segundo a tradição aristotélica, uma proposição é uma sentença declarativa que contém uma verdade ou uma falsidade, não podendo ocorrer as duas simultaneamente.

Dessa forma, Aristóteles dedica um capítulo do seu *De Interpretatione* à sua teoria da modalidade e propõe uma análise das declarações modais, ou seja, da relação das proposições “que são possíveis e aquelas que negam que seja possível, entre aquelas que declaram que é suscetível de se produzir e aquelas que o negam, e entre aquelas que tratam do impossível e do necessário” (ARISTOTE, 2007, p. 309).⁵ Em dois capítulos, Aristóteles trata das relações entre as declarações que dizem respeito ao possível, ao impossível, ao contingente, ao necessário e suas respectivas negações (Capítulo 12, ‘As declarações modais e suas contradições’) e, na sequência, analisa as consequências das declarações modais (Capítulo 13, ‘A consecução lógica entre declarações modais’).

A possibilidade e a necessidade estão relacionadas por uma dupla negação: ‘É possível que p = Não é necessário que não- p ’ e ‘É necessário que p = Não é possível que não- p ’. Estas duas são as noções centrais da modalidade, mas não refletem a sua totalidade, pois há uma gama de significados modais que circulam na periferia, formando o quadrado modal – necessidade/não-necessidade/impossibilidade/possibilidade), como ilustrado na Figura 2.1 abaixo:

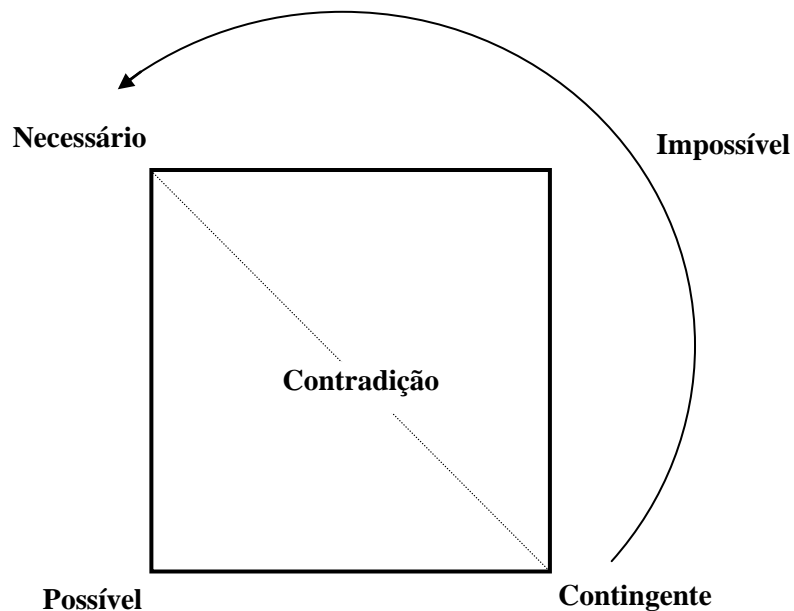


Figura 2.1 - Quadrado modal

⁵ No original francês: “que c’est possible et celles qui nient que ce soit possible, entre celles déclarant que c’est susceptible de se produire et celles qui le nient, et entre celles qui traitent de l’impossible et du nécessaire” (ARISTOTE, 2007, p. 309).

O quadrado modal pode dar conta, por exemplo, de explicar que “Ele pode subir 3 degraus de cada vez” e “Não é impossível pra ele subir 3 degraus de cada vez” se relacionam semanticamente, mas não são sinônimas.

Como extensão da lógica formal, as lógicas modais identificam uma proposição “com o conjunto de mundos possíveis nos quais ela é verdadeira, ou com uma função de mundos possíveis em valores de verdade” (HAAK, 2002, p. 116). Incorporam operadores que modulam a verdade ou a falsidade, para representar argumentos ligados à ideia de necessidade e possibilidade. Segundo Machado e Cunha (2005, p. 75), “um *modal* é uma expressão (necessariamente, possivelmente) que é usada para qualificar a verdade de um julgamento. Dois operadores modais são utilizados, portanto: “é possível” e “é necessário”. O primeiro significa que uma sentença é verdadeira em algum mundo possível; já o segundo significa que a sentença é verdadeira em todos os mundos possíveis. Considere a proposição em (2.1) abaixo⁶ e suas modificações pelos operadores em (2.1a) e (2.1b):

(2.1) João é solteiro.

(2.1a) Possivelmente João é solteiro.

(2.1b) Com certeza João é solteiro.

A sentença em (2.1), que representaria o “real”, pode ser tomada como a modalidade zero. Em (2.1a), o operador “possivelmente” indica que há um mundo possível em que João é solteiro. Já em (2.1b), afirma-se que em todos os mundos possíveis, a proposição “João é solteiro” é verdadeira.

Haak (2002, p. 229) nos esclarece sobre esse caráter necessário e contingente:

Há uma longa tradição filosófica de distinguir entre verdades *necessárias* e verdades *contingentes*. A discussão é frequentemente explicada da seguinte maneira: uma verdade necessária é uma verdade que não poderia ser de outra forma, uma verdade contingente, uma que poderia; ou, a negação de uma verdade necessária é impossível ou contraditória, a negação de uma verdade contingente é possível e consistente; ou, uma verdade necessária é verdadeira em todos os mundos possíveis, uma verdade contingente é verdadeira no mundo real, mas não em todos os mundos possíveis.

As lógicas modais sofreram críticas principalmente porque sua interpretação parece cheia de dificuldades. Segundo Kiefer (2009, p. 179), “os tratamentos lógicos da modalidade

⁶ Este exemplo foi adaptado de Lyons (1977, p. 165). O exemplo original é “John is a bachelor”.

são restritos às propriedades das proposições e excluem todos os traços não-proposicionais das sentenças.”⁷

Estas ideias já tinham surgido, mesmo que de forma incipiente, nos *Primeiros Analíticos*, em que Aristóteles anuncia “Temos que dizer primeiro que se quando A é, é necessário que B seja, então se A é possível é também necessário que B seja possível” (citado por MACHADO; CUNHA, 2005, p. 76). O tema continuou a ser desenvolvido em outros âmbitos, sobretudo na linguística, da qual passo a tratar agora.

2.2 A modalidade é? Visões sobre o tema:

As lógicas modais e a linguística utilizam os mesmos termos para tratar da modalidade. Entretanto, a lógica a divide em objetiva e subjetiva e se ocupa, fundamentalmente, do caráter objetivo da modalidade. Já na tradição linguística, a leitura subjetiva da modalidade parece ser a preferencial, ou melhor dito, a modalidade na linguagem é essencialmente subjetiva, uma vez que está relacionada com as atitudes dos falantes.⁸

De fato, um dos primeiros estudos linguísticos sobre a modalidade, o de Jespersen, em 1924, nos oferece a definição “expressar certas atitudes da mente do falante em relação ao conteúdo das sentenças”⁹ e aponta para uma diferenciação de categorias modais, dividindo-as a partir da presença ou ausência de “um elemento volitivo”.¹⁰ A primeira corresponderia ao obrigativo, jussivo e permissivo; a última, ao necessitativo, potencial, hipotético, dubitativo.

Há, no entanto, além da abordagem relacionada à subjetividade, à qual esta tese vai se alinhar, outras abordagens para o estudo da modalidade: (a) uma, mais formal, sob a perspectiva de uma semântica de mundos possíveis; e (b) outra que propõe a distinção entre *realis* e *irrealis*.

Nas seções seguintes, apresento estas três teorias semânticas que definem de diferentes perspectivas o fenômeno da modalidade.

⁷ No original: “Logical treatments of modality are restricted to properties of propositions and exclude all nonpropositional features of sentences.” (KIEFER, 2009, p. 179).

⁸ Esta formulação vai ser problematizada logo à frente.

⁹ “express[ing] certain attitudes of the mind of the speaker towards the content of the sentences.” (JESPERSEN, 1924, p. 313).

¹⁰ “element of will” (JESPERSEN, 1924/1992, p. 320-1).

2.2.1 A abordagem formalista – a semântica dos mundos possíveis:

De uma forma ou de outra, observamos que as definições propostas por linguistas sobre a modalidade estão comprometidas com o valor de verdade de um determinado conteúdo proposicional ou também são analisadas sob a perspectiva de uma semântica de mundos possíveis (KRATZER, 1981).

Esta noção foi introduzida na filosofia lógica na década de 1960 por Hintikka (1962) e Kripke (1963), com raízes em Leibniz, para quem as pessoas viviam em um mundo que é apenas um entre infinitas possibilidades de mundos criados por Deus. Para fins deste trabalho, não ampliarei a discussão da lógica modal, e passo a apresentar a proposta de Kratzer, em que os mundos possíveis servem a modelar determinadas relações semânticas entre expressões linguísticas. Nas palavras de Rubinstein (2012, p. 191), “[...] os mundos possíveis representam todas as maneiras possíveis de as coisas poderem ser (sem o compromisso de estas possibilidades realmente existirem ou não no mundo em que vivemos [...])”.¹¹

Nesta perspectiva, a modalidade seria “o fenômeno linguístico por meio do qual a gramática permite que se diga coisas sobre, ou na base de, situações que não necessitam ser reais.” (PORTNER, 2009, p. 1).¹² Os modais seriam pistas linguísticas que lançariam luz aos conjuntos contextualmente relevantes de mundos possíveis.

Duas das ideias centrais da abordagem de Kratzer (1981, 1991, 2012), tomada como uma “encarnação influente”¹³ da lógica modal, são: (i) os modais relativos não são ambíguos: a distinção do valor de um determinado índice modal é dada pelo contexto ou pelo que ela denomina *background* conversacional. Os modais de possibilidade corresponderiam a quantificadores existenciais, e os de necessidade a quantificadores universais, estas forças modais que incidem sobre a base são definidas pelo item lexical; e (ii) ordenamento de mundos: as expressões modais são quantificações sobre mundos possíveis, no entanto, estes mundos possíveis são tomados como um conjunto ordenado (ou parcialmente ordenado) de mundos. Este conjunto é criado através da interação de dois *backgrounds* conversacionais.

A base modal, formalmente, é uma função de mundos possíveis para um conjunto de proposições, que servem para diferenciar as leituras dos significados modais. A cada situação de uso, a base modal é o conjunto dos mundos compatíveis com o conhecimento, a crença, os

¹¹ Tradução para: “[...] possible worlds represent all the possible ways things could be (without committing to whether or not these possibilities actually exist alongside the one world we live in [...])” (RUBINSTEIN, 2012, p. 191).

¹² No original: “the linguistic phenomenon whereby grammar allows one to say things about, or on the basis of, situations which need not be real.” (PORTNER, 2009, p. 1).

¹³ “influential incarnation” (von FINTEL, 2006, p. 3).

desejos do falante, as regras aplicáveis, e podem variar de mundo para mundo, de acordo com a sua avaliação.

Para o segundo parâmetro, o subconjunto relevante de mundos possíveis para qualquer modal é identificado, primeiro, pela descoberta de todos os mundos nos quais todas as proposições na base modal são verdadeiras e, em seguida, é feita a hierarquização (ou o escalonamento) destes mundos, de acordo com o quão próximo estão do ideal determinado pela fonte de ordenação. Nas palavras de Hara (2006, p. 9), a fonte de ordenação “força uma ordenação particular entre os mundos epistêmicos da base modal em termos de sua acessibilidade.”¹⁴

Para von Fintel (2006, p. 1), discípulo de Kratzer, a modalidade é a categoria de significado linguístico que tem a ver com a expressão de possibilidade e necessidade. Uma sentença modalizada estabelece uma proposição subjacente e prejacente (análise no nível proposicional) no espaço das possibilidades. Por exemplo:

(2.4) Sandy might be home.

Sandy pode estar em casa.

diz que há uma possibilidade de Sandy estar em casa.

(2.5) Sandy must be home.

Sandy deve estar em casa.

diz que entre todas as possibilidades, Sandy está em casa.

von Fintel (2006) argumenta que a modalidade, assim como a temporalidade, está no centro da propriedade de deslocamento, que permite à linguagem falar de coisas para além do aqui-agora. Além desta capacidade de projeção (prospectiva ou retrospectiva), Fintel aponta para os diferentes tipos de significados dos itens modais¹⁵ e nos fornece exemplos com a forma verbal ‘*to have to*’, apresentados abaixo:

(2.6) It has to be raining.

Tem que estar chovendo.

[depois de observar pessoas entrando com sombrinhas; modalidade epistêmica]

¹⁴ No original: “[The ordering source] forces a particular ordering among epistemic worlds of the modal base in terms of their accessibility.” (HARA, 2006, p. 9)

¹⁵ No caso da língua portuguesa, comportam mais de um sentido os verbos ‘poder’, ‘dever’ e ‘ter que’, que carregam tanto o valor epistêmico quanto o raiz/deôntico e dinâmico. Mais adiante, discuto este ponto.

(2.7) Visitors have to leave by six pm.

Os visitantes têm que sair por volta de seis da tarde.

[regulamento de hospital; deôntico]

(2.8) You have to go to bed in ten minutes.

Você tem que ir pra cama em dez minutos.

[um pai severo; bulomaica ou bulética]

(2.9) I have to sneeze.

Tenho que espirrar.

[dado o estado atual do nariz de alguém; circunstancial]

(2.10) To get home in time, you have to take a taxi.

Para chegar em casa no horário, Você tem que pegar um táxi.

[modalidade teleológica]

Os exemplos entre (2.6) e (2.10) poderiam ser tomados como ocorrências do tipo deôntico. No entanto, guardam entre si diferenças sutis. Em (2.7), o tipo bulético ou bulomaico indica a vontade ou intenção do falante. O valor circunstancial, em (2.8) expressa o que é possível ou necessário de acordo com determinadas circunstâncias. A modalidade teleológica é orientada para um alvo, isto é, é o domínio do necessário e do possível com a finalidade de alcançar um determinado objetivo (von FINTEL; IATRIDOU, 2005; HACQUARD, 2006), no caso em (2.9), o objetivo seria “chegar em casa no horário”.

Introduzi aqui algumas das ideias principais acerca da abordagem formal da modalidade, desenvolvida desde a década de 1960 e cuja contribuição de Kratzer é fundamental, como o fato de as expressões modais serem usadas para descrever maneiras alternativas de como pode ser o mundo e incidirem sobre um argumento proposicional, conhecido por “prejacente”. São quantificadores sobre mundos possíveis e esta quantificação, em qualquer mundo avaliativo, está contextualmente localizada e é determinada por relações de acessibilidade. Foram apresentados também os dois parâmetros principais desta abordagem, quais sejam, a base modal (epistêmica, deôntica, bulomaica, circunstancial, teleológica), a ordenação de mundos e a fonte de ordenação, definidos em um *background* conversacional e a partir da interação entre eles.

2.2.2 *Realis* x *Irrealis*:

Como dito, há na literatura sobre modalidade abordagens que privilegiam em suas análises uma divisão binária entre modal e não-modal, e relacionam essa distinção ao contraste factual / não-factual, ou *realis* / *irrealis* (CHUNG; TIMBERLAKE, 1985; GIVÓN, 1995, KIEFER, 1994; MITHUN, 1999). Para Mithun (1999, p. 173), “o *realis* retrata situações como presentes, ou tendo ocorrido ou atualmente ocorrendo, reconhecíveis por percepção direta. O *irrealis* retrata situações como puramente dentro domínio do pensamento, reconhecível apenas através da imaginação”.¹⁶ Em outras palavras, esta distinção separa o mundo entre situações ou eventos reais ou irrealis.

Ferdinand de Haan (2005, p. 41-45) argumenta que a distinção *realis-irrealis* talvez não seja suficiente para dar conta de uma tipologia da modalidade e uma definição interlinguística desta categoria, uma vez que o termo *irrealis* é vago e pode se referir a diferentes circunstâncias, além de o conteúdo semântico de morfemas *irrealis* variar de língua para língua, mesmo as línguas intimamente relacionadas.

A categoria de futuro é apresentada por de Haan como uma que pode ilustrar esta inconsistência, apesar de ser uma categoria que pode ser tomada como prototipicamente *irrealis*, já que indica eventos que ainda não ocorreram e, portanto, são irrealis. Nas línguas Amele e Muyuw (ambas línguas da Papua Nova Guiné, mas de famílias diferentes), o futuro é uma categoria *irrealis*, mas não é o caso da língua Caddo (CHAFE, 1995, p. 358), exemplificada abaixo.

(2.11) cíbáw-/a/ / ci-yi=bahw- /a/
 1SG.AG.REAL-see-FUT
 ‘Irei olhá-lo.’

No exemplo em (2.11), o morfema de futuro - / a / não ocorre com o prefixo de *irrealis* t’a-/t’i-, mas com o prefixo de *realis* ci-.

Mithun (1995, p. 378–80) pontua que há línguas como o Pomo Central, falado na Califórnia, em que o futuro pode ser usado tanto como *realis* ou *irrealis*, a depender do julgamento do falante sobre a probabilidade de um evento realmente acontecer.

¹⁶ No original: “the *realis* portrays situations as actualized, as having occurred or actually occurring, knowable through direct perception. The *irrealis* portrays situations as purely within the realm of thought, knowable only through imagination”. (MITHUN, 1999, p. 173).

Givón (1995) trata esta distinção a partir da afirmação de que a modalidade é um domínio funcional complexo que abarca subdomínios semânticos e pragmáticos, difíceis de serem separados, uma vez que agrupam construções que servem tanto a construções primariamente semânticas quanto primariamente pragmáticas. Seu objetivo é delinear alguns princípios, coerentes, pelos quais se pode predizer uma variação de ambientes gramaticais em que o modo subjuntivo provavelmente se gramaticaliza e podem representar um subconjunto de *irrealis*. O *irrealis* é tomado como uma categoria comunicativo-cognitiva (funcional) e tipológica-gramatical (formal).

Abaixo, um quadro que representa a modalidade epistêmica na tradição lógica e seu equivalente comunicativo:

tradição lógica	equivalente comunicativo
a. verdade necessária	pressuposição
b. verdade factual	asserção <i>realis</i>
c. verdade possível	asserção <i>irrealis</i>
d. não-verdade	asserção-NEG

Quadro 2.1 – Correspondência da modalidade epistêmica entre a tradição lógica e a funcional

A tradição lógica trata a modalidade desconectada de seu contexto comunicativo natural. Nas palavras de Givón (1995, p. 114), “a interpretação comunicativo-pragmática das quatro modalidades, [...], as reellenca em termos dos estados epistêmicos e objetivos comunicativos dos dois participantes na negociação comunicativa — falante e ouvinte”.¹⁷ A redefinição comunicativa da modalidade epistêmica fica assim estabelecida:

- (a) **pressuposição**: assume-se uma pressuposição como verdadeira, por definição, por acordo *a priori*, por convenção genérica culturalmente compartilhada, por ser óbvio para todos presentes numa situação de fala, ou por ter sido enunciada pelo falante e não contestada pelo ouvinte.

¹⁷ Tradução minha para: “the communicative-pragmatic interpretation of the four modalities, [...], recasts them in terms of the epistemic states and communicative goals of the two participants of the communicative transaction — speaker and hearer.” (GIVÓN, 1995, p. 114).

- (b) **asserção *realis***: a proposição é **fortemente afirmada** como verdadeira; mas a contestação do ouvinte é considerada apropriada, apesar de o falante ter evidência ou outras bases para defender sua firme crença.
- (c) **asserção *irrealis***: a proposição é **fracamente afirmada** por ser tanto possível, provável ou incerta (submodos epistêmicos), ou necessária, desejada ou indesejada (submodos avaliativo-deôntico). Mas o falante não está pronto para confirmar a afirmação com evidência ou outras bases; a contestação do ouvinte é prontamente considerada, esperada ou até mesmo solicitada.
- (d) **asserção-NEG**: a proposição é **fortemente afirmada** como falsa, mais comumente em contradição às crenças explícitas ou assumidas do ouvinte; a contestação do ouvinte é antecipada e o falante tem evidência ou outras bases para confirmar sua firme crença.

Mudar o foco das noções de *realis* e *irrealis* para uma ótica cognitiva e comunicativa implica, cognitivamente, deslocar de questões de verdade lógica para questões de certeza subjetiva, e, comunicativamente, deslocar do sentido (semântico) da orientação pelo falante para o sentido (pragmático) interativo socialmente negociado, envolvendo tanto o falante quanto o ouvinte.

Independentemente da perspectiva que é tomada esta distinção, parece difícil tomar a oposição *realis* x *irrealis* e, especificamente o *irrealis*, como uma noção para a definição de modalidade, ainda que em algumas línguas esteja gramaticalmente marcada e que na literatura sobre crioulos e pidgins esta seja a forma padrão para descrever as distinções modais (de HAAN, 2005, p. 45).

Passo, portanto, para a terceira abordagem sobre o fenômeno, a modalidade como uma categoria relacionada à subjetividade.

2.2.3 Subjetividade:

Lyons (1977, p. 797-ss), em sua tentativa de síntese sobre a modalidade, que guia muitos outros autores em termos de relacionar a categoria com a expressão da atitude de um falante, anuncia, ao tratar dos advérbios modais, que a modalidade representa “a opinião ou a atitude do falante em relação à proposição que a sentença expressa ou a situação que a

proposição descreve” (LYONS, 1977, p. 452).¹⁸ Kiefer (1994, p. 2516) se alinha a ele, e afirma que a modalidade é “a atitude volitiva, emotiva e cognitiva do falante em relação a um estado-de-coisas”¹⁹ e também Nitta (2000, p. 81) que define o fenômeno “como a expressão do ponto de vista do falante sobre os conteúdos da sentença e a atitude comunicativa do falante”.²⁰

Palmer (1986), igualmente, sugere uma definição para modalidade baseada na subjetividade: “Modalidade na linguagem está [...] relacionada às características subjetivas de um enunciado [...] [e] poderia ser definida como ‘**a gramaticalização das atitudes e opiniões (subjetivas) de um falante**’.”²¹ (PALMER, 1986, p. 16, grifos meus). O autor toma a modalidade desde uma perspectiva tipológica e a divide em dois tipos: proposicional e de evento.

A primeira é definida como o julgamento, pelo falante, do valor de verdade ou estatuto factual de uma proposição, e é subdividida em epistêmica e evidencial. A epistêmica expressa o julgamento sobre o estatuto factual da proposição e pode ser especulativa (expressa incerteza); dedutiva (indica uma inferência a partir de uma evidência observável); assumptiva (indica inferência a partir do que é geralmente conhecido); evidencial (os falantes indicam a evidência que têm sobre o estatuto factual da proposição); reportada; e sensória. O segundo tipo é a atitude do falante em relação a um potencial evento futuro que ainda não ocorreu. É também subdividida em deôntica, relacionada à obrigação ou à permissão, e dinâmica, relativa à habilidade ou volição.

Na mesma clave, Bybee, Perkins e Pagliuca (1994, p. 178, grifo meu) conceituam a modalidade como a “gramatização das atitudes e opiniões (**subjetivas**) do falante”²² e propõem uma divisão de modalidade em quatro tipos: epistêmica, orientada-para-o-agente (i.e. deôntica), orientada-para-o-falante e subordinada²³. Elas sublinham a importância do estudo diacrônico do desenvolvimento de elementos modais, a fim de compreender a variação dos significados modais em uma língua.

¹⁸ No original: “the speaker’s opinion or attitude towards the proposition that the sentence expresses or the situation that the proposition describes.” (LYONS, 1977, p. 452).

¹⁹ Tradução minha para: “the speaker’s cognitive, emotive or volitive attitude towards a state of affairs” (KIEFER, 1994, p. 2516).

²⁰ Tradução de: “as expressing the speaker’s point of view on the sentence contents and the speaker’s communicative attitude.” (NITTA, 2000, p. 81).

²¹ “Modality in language is [...] concerned with subjective characteristics of an utterance [...] [and] could be defined as ‘the grammaticalization of speaker’s (subjective) attitudes and opinions.’” (PALMER, 1986, p. 16).

²² Tradução minha para: “grammaticization of speaker’s (subjective) attitudes and opinions” (BYBEE et al., 1994, p. 178).

²³ Esta tipologia será apresentada na seção 2.3.1.

A dimensão subjetiva da modalidade é tratada de diferentes formas na literatura sobre a modalidade epistêmica. A oposição entre o traço objetivo e subjetivo da modalidade é apresentada em Lyons (1977) que define que a modalidade epistêmica objetiva expressa a chance objetivamente mensurável de que um estado de coisas seja verdade ou não e, por sua vez, um enunciado epistêmico subjetivo indica uma suposição subjetiva em relação à sua verdade. A definição de Lyons de subjetividade x objetividade recai sobre diferenças na qualidade da evidência que leva ao julgamento (NUYTS, 2005, p. 14). Lyons (1977, p. 797) toma em consideração os seguintes exemplos em (2.12) e (2.13):

(2.12) Alfred may be unmarried.

Alfred pode ser descasado.

Em uma interpretação subjetiva do exemplo em (2.12), o falante, baseado em suas próprias suposições, subjetivamente se compromete com a possibilidade de Alfred ser descasado. Em uma leitura objetiva, o falante tem a informação, por exemplo, do número de pessoas casadas em uma determinada comunidade e, baseado neste fato, constrói a possibilidade de Alfred ser descasado.

(2.13) Alfred must be unmarried.

Alfred deve ser descasado.

Em (2.13), uma leitura subjetiva da ocorrência seria mais natural, segundo Lyons, uma vez que o falante subjetivamente qualifica seu comprometimento com a factualidade da proposição. Já uma interpretação objetiva nasceria, por exemplo, se se considerasse a porcentagem de pessoas descasadas numa determinada comunidade e que fosse identificado o estado civil de todas estas pessoas solteiras, exceto o de Alfred. O falante, portanto, se compromete com a informação dada ao seu interlocutor, e que a proposição *Alfred deve ser descasado* é de certa maneira verdadeira.

Segundo Nuyts (2001), alguns autores consideram estas características como diferentes significados da modalidade (epistêmico, para a subjetividade, e deôntico, para a objetividade; c.f. HENGEVELD, 1988), e outros problematizam o tema, como Coates (1983) que reivindica que tanto o valor epistêmico quanto o valor deôntico podem ter usos subjetivos e objetivos.

Narrog (2005, p. 169-170) aponta que há dois problemas relacionados a esta questão. O primeiro seria a dificuldade em definir o que é “subjetivo” ou “a atitude do falante”, porque ambos os conceitos levam uma carga de vaguidade. O segundo ponto seria definir a fronteira entre a “subjatividade” e a “objetividade”. Na verdade, Narrog argumenta que não é possível alcançar um limite para o que é subjetivo ou objetivo, uma vez que a subjatividade de um falante pode estar expressa em diferentes elementos de uma sentença, seja na escolha do vocabulário, na perspectiva adotada para conceptualizar a situação, etc.

Pelo viés cognitivista do fenômeno, também é considerada a dimensão subjetiva da modalidade. Mortelmans (2007, p. 869-870) concorda que a noção de modalidade está longe de ser definida de forma mais ou menos unificada e destaca que os trabalhos na moldura cognitiva focam, em sua maioria, os verbos modais, mas há publicações que contemplam outras expressões de modalidade como os advérbios, adjetivos modais e predicados mentais (NUYTS, 1994, 2001, 2002); os verbos semi-auxiliares (CORNILLIE, 2007); e os marcadores evidenciais (MATLOCK, 1989; LEE, 1993).

A ideia de força, como pontua Mortelmans, tem papel fundamental na forma como a modalidade é conceptualizada pelos cognitivistas. A Hipótese da Dinâmica de Forças de Talmy (1988, 2000) se caracteriza como um esquema imagético básico de causa como imposição de força e suspensão de barreiras. Expressa “como as entidades interagem no que diz respeito à força. Incluído aqui está a imposição da força, resistência a tal força, barreira à expressão de tal força, a superação de tal resistência, a suspensão de tal barreira, e coisas do gênero.” (TALMY, 1988, p. 49).²⁴

Este modelo é uma generalização da noção de causação, em que processos são conceptualizados como resultado de uma interação de forças e relações causais agindo de diversas maneiras sobre os participantes de um evento (cf. CROFT; CRUSE, 2004, p. 66). Talmy (1988, 2000) importa da fisiologia a terminologia que emprega: a entidade de força focal é chamada Agonista e o elemento de força oposto a ela é o Antagonista. Tento ilustrar este ponto com uma ocorrência em nossa amostra²⁵:

(2.14) * MAI: [47] aí **ele jogou** [/1] **jogou a capa**_{E1} / **ea** / **estraçalhou toda a capa**_{E2} /
a' o tamanho do dente // (*bfamnn01*)

²⁴ Tradução para: “how entities interact with respect to force. Included here is the exertion of force, resistance to such a force, the overcoming of such a resistance, blockage of the expression of force, removal of such blockage, and the like.” (TALMY, 1988, p. 49).

²⁵ A notação utilizada nestes exemplos será explicitada na Parte II deste trabalho, quando introduzo os pressupostos da Teoria da Língua em Ato. No entanto, adianto aqui que uma barra simples (“/”) corresponde a uma quebra no enunciado percebida como não-terminal e uma barra dupla (“//”) como uma quebra percebida como terminal.

O exemplo em (2.14) traz dois tipos de eventos causativos: o Evento 1 (E₁) “**ele jogou [1] jogou a capa /**” e o Evento 2 (E₂) “**ea / estraçalhou toda a capa**”. Em E₁, o Antagonista (aquele que aplica a força, o causador “ele”) exerce uma força sobre o Agonista (aquele em que a força é aplicada, o causado “a capa”), que realiza um movimento. A situação em E₂ é diferente: o Antagonista (“ea” = a cobra) impõe uma força que impede o movimento do causado “a capa”, que resulta na sua destruição.

Dentre os trabalhos nessa linha, sobressai-se o de Sweetser (1990), que argumenta que há uma tendência a usar um vocabulário de um domínio externo (sociofísico) para um domínio interno (emocional e psicológico). Diacronicamente, o sistema metafórico guia numerosas mudanças semânticas. Sincronicamente, é representado por várias palavras polissêmicas e extensões abstratas de sentido do vocabulário do mundo físico.

A modalidade, assim, é uma categoria que está em ambiguidade sincrônica entre mundo externo e mundo interno, quer dizer, há uma ambiguidade de expressões modais entre sentidos raiz²⁶ ou deônticos e sentidos epistêmicos.

Os significados que denotam obrigação, permissão ou habilidade, a autora denomina-os “raiz”. Como exemplo, temos:

(2.15) John *must* be home by ten; Mother won't let him stay out any later.

John tem que estar em casa às dez; a Mãe não deixará que ele fique fora até mais tarde.

Epistêmicos são os significados que denotam necessidade, probabilidade ou possibilidade. Como atestado abaixo:

(2.16) John *must* be home already; I see his coat.

John já deve estar em casa; estou vendo seu casaco.

A proposta, portanto, é de que significados “raiz” são extensíveis ao domínio epistêmico justamente porque usamos, geralmente, a linguagem do mundo externo para aplicar ao mundo mental interno, o qual é metaforicamente estruturado como paralelo àquele mundo externo. Sweetser argumenta que os verbos modais não têm sentidos separados não-

²⁶ Sweetser prefere usar o termo raiz. Não só porque é um termo mais amplo, mas também porque, como veremos a seguir, ela argumenta que a modalidade epistêmica está “enraizada” na modalidade sociofísica. Segundo ela, o termo deôntico pode ser tomado, algumas vezes, meramente como uma obrigação social ou moral.

relacionados, mas sim mostram uma extensão do sentido-raiz para o domínio epistêmico — uma extensão fortemente motivada pelo sistema linguístico circundante.

A autora considera que um dos obstáculos para a evolução de uma compreensão unificada de modalidade (que já sabemos ser um conceito controverso) decorre do fato de que análises semânticas da modalidade raiz não foram sistematicamente relacionadas à necessidade e à probabilidade lógicas.

O objetivo de Sweetser é demonstrar que uma análise em termos de dinâmica de forças para os modais raiz é possível e extensível ao domínio epistêmico. A análise da linguista avança em relação ao modelo de Talmy (1981, 1988), no sentido de considerar que a modalidade se refere a forças e barreiras intencionais e diretas. Tomando como exemplo o modal *may*, temos que: no domínio sociofísico, o verbo representa, em termos de dinâmica de forças, uma ausência de barreira potencial, e encontra, no domínio epistêmico, um sentido correspondente, quer dizer, uma barreira ausente no processo de raciocínio do falante a partir das premissas disponíveis para a conclusão expressa na sentença qualificada pelo verbo.

Ex.: permissão (*may, let, allow*): instância de retirar uma barreira potencialmente presente.

(2.17) The crack in the stone let the water flow through. (barreira física)
A fenda na pedra deixou a água escorrer.

(2.18) I begged Mary to let me have another cookie. (barreira “social”)
Implorei à Maria para me deixar comer outro cookie.

Gonzalv ez-Garcia (2000), sob uma perspectiva discursiva, ou seja, uma perspectiva mais din amica (e sistem atica) que pressup oe a “rela ao entre modalidade, o contexto de enuncia ao e a fun ao interpessoal da polidez [nos termos de Brown e Levinson, 1987]”²⁷ (GONZALV ES-GARCIA, 2000, p.127), prop oe que a modalidade seja caracterizada como express ao do envolvimento do falante com o conte udo proposicional de um dado enunciado (ag encia ou subjetividade) e que pode ser tomada como ramificadora, atrav es de toda a arquitetura l exico-gramatical da linguagem.

O autor argumenta que muitos dos significados normalmente atribu idos a verbos modais individuais s ao, na verdade, derivados tanto do ambiente sentencial do verbo quanto

²⁷ No original: “[...] relationship between modality, the context of utterance and the interpersonal function of politeness.” (GONZALV ES-GARCIA, 2000, p.127).

de algum contexto de enunciação mais amplo. Assim, o significado modal pode ser tomado como emergente da interação de duas camadas de significado estreitamente conectadas: uma, abarcando o significado linguístico do verbo modal herdado em conjunto com outros itens modais vizinhos; e outra, que diz respeito aos princípios ligados a polidez e estratégias de proteção de face.

Para ilustrar seu ponto, o autor lança mão de exemplos com “*might*” selecionados do *International Corpus of English (ICE-GB)* e de um excerto do “*Four Weddings and a Funeral*” (“Quatro casamentos e um funeral”, no título brasileiro). Abaixo, transcrevo o primeiro exemplo apresentado pelo autor e sua respectiva análise (GONZALVÉS-GARCIA, 2000, p.128-129):

(2.19) “Maybe I might miss something out because we are all human and I’ve overlooked it and you’ve remembered it” (ICE-GB Corpus, S2A-061-36).

Talvez eu pudesse deixar escapar algo porque somos todos humanos e eu o tenha esquecido e você se lembrado. (ICE-GB Corpus, S2A-061-36).

Gonzalvés-Garcia considera este uso do modal “*might*” (“pudesse”) como “possibilidade remota”, com a leitura reforçada pelo advérbio epistêmico “*maybe*” (“talvez”). Entretanto, outras formas linguísticas (“*are*”, “*have overlooked*” e “*have remembered*”) parecem indicar, em contraste, uma leitura mais factual do modal e, além disso, o uso da forma “*might*” parece ser motivado pela necessidade do falante de salvar a sua face e despertar a simpatia da audiência.

Em resumo, o autor praticamente assume uma visão pragmática da modalidade e este estudo apresenta as fronteiras (tênuas) entre semântica e pragmática.

2.3 Tipos de significados modais:

A tradição lógica estabelece o estudo das modalidades alética, epistêmica e deôntica, relacionadas, respectivamente, com as noções de verdade/falsidade, conhecimento e conduta. Seguindo a tradição, as noções de necessidade, possibilidade, probabilidade e obrigação são tomadas em consideração para definição dos significados modais.

No entanto, assim como não há qualquer unanimidade na definição da noção da modalidade, também não há, entre os pesquisadores, um consenso de quais categorias poderiam estar englobadas sob o rótulo de “modal”. Na literatura linguística tradicional sobre a modalidade destacam-se três valores semânticos: o epistêmico, o deôntico e o dinâmico,

mas muitos estudos organizam de formas distintas as dimensões relacionadas à categoria, a depender da perspectiva adotada para sua definição. Entre as diferentes abordagens, mantém-se a oposição entre os significados epistêmicos e não-epistêmicos, havendo uma maior diferenciação no que diz respeito aos últimos.

Nas subseções abaixo, apresento, primeiramente, as visões não-tradicionais da tipologia dos índices modais e, em seguida, a classificação tradicional, sendo que, nesta última, destaco as modalidades epistêmica, deontica e dinâmica, que tomo como ponto de partida para a descrição e análises que serão empreendidas no escopo deste trabalho.

2.3.1 As visões não-tradicionais: tipologias alternativas:

Uma das classificações mais conhecidas na tipologização dos modais é a de Bybee e colaboradoras (BYBEE, 1985; BYBEE et al., 1994; BYBEE; FLEISCHMAN, 1995). O objetivo do trabalho de Bybee e colegas (1985, 1994) é propor caminhos para o desenvolvimento das noções de modo e modalidade, buscando determinar como e por que determinados significados gramaticais surgem neste domínio. As autoras distinguem quatro tipos de modalidade: (i) modalidade orientada-para-o-agente (MOA); (ii) modalidade orientada-para-o-falante (MOF); (iii) modalidade epistêmica (ME); e (iv) modalidade subordinada.

A modalidade orientada-para-o-agente “compreende todos os significados modais que indicam a existência de condições sobre um agente com relação à conclusão de uma ação expressa no predicado principal, por exemplo, obrigação, desejo, habilidade, permissão e possibilidade raiz” (BYBEE; FLEISCHMAN, 1995, p. 6).²⁸

As noções semânticas específicas deste tipo são:

(a) obrigação: existência de condições sociais externas que levam um agente a concluir a ação predicada, expressa como uma obrigação forte ou fraca. Por exemplo:

(2.20) All students **must** obtain the consent of the Dean of the faculty concerned before entering for examination.

Todos os estudantes devem obter o consentimento do Diretor da Faculdade referida antes de entrar no exame.

²⁸ Tradução do original: “[...]“encompasses all modal meanings that predicate conditions on an agent with regard to the completion of an action referred to by the main predicate, e.g., obligation, desire, ability, permission and root possibility” (BYBEE; FLEISCHMAN, 1995, p. 6).

(2.21) I just insisted very firmly on calling her Miss Tillman, but one **should** really call her President.

Insisti muito firmemente em chamá-la de Srta. Tillman, mas deve-se realmente chamá-la de Presidente.

Em (2.20), tem-se caracterizada uma obrigação mais fraca. Já em (2.21), temos um exemplo de obrigação forte.

(b) necessidade: existência de condições físicas que levam um agente a completar a ação predicada.

(2.22) I **need** to hear a good loud alarm in the mornings to wake up.

Eu preciso ouvir um despertador bem alto pelas manhãs para acordar.

(c) habilidade: existência de condições internas de habilidade no agente, com relação à ação predicada.

(2.23) I can only type very slowly as I am a beginner.

Como sou um iniciante, eu só consigo digitar muito devagar.

(d) desejo: existência de condições internas de volição no agente, com relação à ação predicada.

(2.24) Juan Ortiz called to them loudly in the Indian tongue, bidding them come forth if they **would** (= wanted to) save their lives.

Juan Ortiz chamou-os bem alto na língua indiana, ordenando-lhes que voltassem, se eles quisessem salvar suas vidas.

A modalidade orientada-para-o-falante abrange “marcadores de diretivos, tais como imperativos, optativos e permissivos, que representam atos de fala através dos quais um falante tenta provocar o endereçado a agir” (BYBEE; FLEISCHMAN, 1995, p. 6).²⁹ Este tipo não indica a existência de condições sobre o agente, mas permite ao falante impor tais condições sobre o interlocutor. O quadro abaixo mostra, em resumo, os termos gramaticais usados para a MOF e suas respectivas funções:

²⁹ No original: “[m]arkers of directives, such as imperatives, optatives or permissives, which represent speech acts through which a speaker attempts to move an addressee to action” (BYBEE; FLEISCHMAN, 1995, p. 6).

Termos gramaticais	Função
Imperativo	Comando direto a uma segunda pessoa
Proibitivo	Comando negativo
Optativo	Desejo ou expectativa do falante expresso na oração principal
Hortativo (exortativo)	O falante encoraja ou incita alguém a agir
Admoestativo	O falante emite um aviso
Permissivo	O falante concede permissão

Quadro 2.2 – Termos gramaticais para a MOF e suas funções

Bybee et al. (1994) consideram os modos acima elencados como parte da modalidade deôntica. No entanto, é possível questionar a validade deste tipo ser uma categoria modal, se se exclui o “modo” do domínio da modalidade.

As autoras ainda distinguem a categoria de modalidade subordinada, que se trata de estabelecer uma categoria formal para dar conta das modalidades em cláusulas subordinadas.

Uma segunda tipologia considerada alternativa dentro dos estudos da modalidade é a proposta por van der Auwera e Plungian (1998). A modalidade, para os autores, estaria relacionada com domínios semânticos em que possibilidade e necessidade são as variantes paradigmáticas (van der AUWERA; PLUNGIAN, 1998, p. 80). São considerados quatro domínios: modalidade interna ao participante, modalidade externa ao participante,

A modalidade interna ao participante se refere ao tipo de possibilidade ou necessidade interna a um participante engajado no estado de coisas. Veja os exemplos:

(2.25a) Boris can get by with sleeping five hours a night.

Boris pode passar bem dormindo cinco horas por noite.

(2.25b) Boris needs to sleep ten hours every night for him to function properly.

Boris precisa dormir dez horas toda noite para funcionar adequadamente.

O exemplo (2.25a), em termos de possibilidade, mostra uma capacidade/habilidade do participante ‘Boris’, e o exemplo em (2.25b) é um caso de uma necessidade interna do participante.

O segundo domínio de contraste entre possibilidade e necessidade é a modalidade externa ao participante que se refere a circunstâncias externas ao participante engajado no estado de coisas, o que torna este estado de coisas possível ou necessário. Como atestado nos exemplos (2.26a) e (2.26b):

(2.26a) To get to the Station, you can take bus 66.

Para chegar à estação, você pode pegar o ônibus 66.

(2.26b) To get to the Station, you have to take bus 66.

Para chegar à estação, você tem que pegar o ônibus 66.

Em (2.26a), o ônibus 66 é uma entre outras possibilidades para se chegar à estação e em (2.26b), pegar o ônibus 66 constitui-se como uma necessidade externa, já que é o único meio de transporte para se chegar à estação.

O terceiro domínio é a modalidade deôntica, caracterizado como um subdomínio da modalidade externa ao participante. Este tipo identifica as circunstâncias favorecedoras externas ao participante, como uma pessoa(s), normalmente o falante, e/ou alguma norma social ou ética, que permite ou obriga o participante a se engajar no estado de coisas. Em (2.27a), temos um caso de possibilidade deôntica: uma pessoa ou norma permite e torna possível a partida de John. A necessidade deôntica está expressa em (2.27b), em que uma autoridade ou uma norma impõe a partida de João como necessária e obrigatória.

(2.27a) John may leave now.

John pode partir agora.

(2.27b) John must leave now.

John deve partir agora.

O quarto e último domínio é a modalidade epistêmica, que se refere a um julgamento de uma proposição pelo falante, que pode ser considerada incerta ou provável.

(2.28a) John may have arrived.

John pode ter chegado.

(2.28b) John must have arrived.

John deve ter chegado.

(2.28a) ilustra uma possibilidade baseada em julgamentos anteriores; o falante trata a chegada de John como incerta. Em (2.28b), o falante se refere à chegada de John como provável. Dado um conjunto de evidências anteriores, o falante é levado a crer que necessariamente John chegou.

Assim, os autores, a partir dos passos de gramaticalização de Bybee et al. (1994), desenvolveram um mapa semântico da representação da modalidade. Como mostra a Figura 2.2 abaixo:

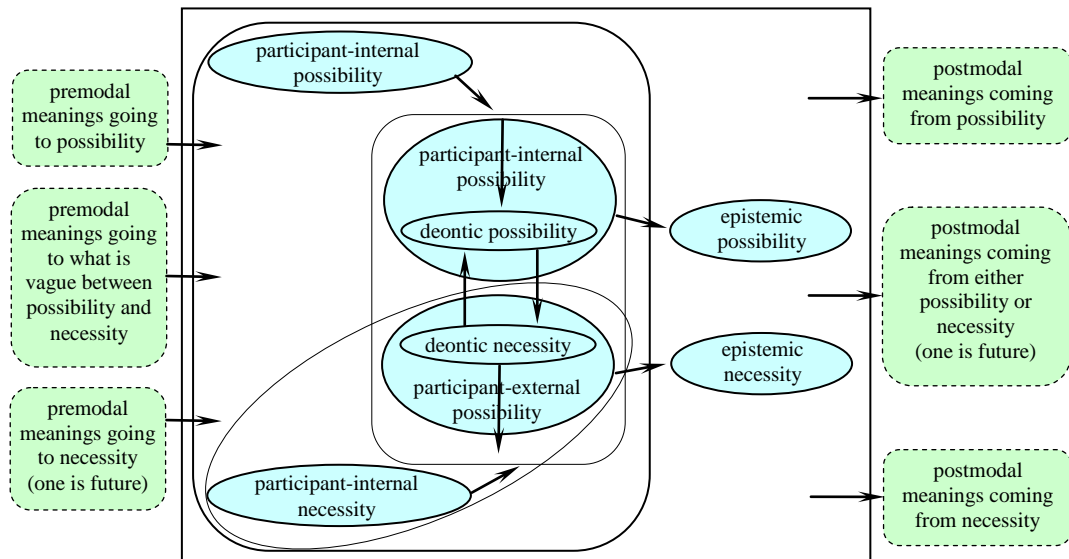


Figura 2.2 – Mapa semântico da modalidade. Fonte: van der Auwera e Plungian (1998, p. 111)

Neste mapa estão representados três domínios, da esquerda para a direita: pré-modal, modal e pós-modal. Os domínios pré e pós-modais são representados abstratamente e os passos de gramaticalização são representados por setas.

Por fim, a modalidade raiz abarca os significados não-epistêmicos em oposição à modalidade epistêmica. Como visto, ela pode estar relacionada com o significado deontico (TALMY, 1988; SWEETSER, 1990) e também incluir parte do tipo dinâmico (COATES, 1983).

Como dito anteriormente, levarei em conta neste trabalho uma classificação tradicional na literatura linguística, considerando os tipos mais centrais desde o ponto de vista do uso das línguas naturais. Utilizo os significados epistêmico (relacionado ao conhecimento e crença), deontico (relacionado a obrigação e permissão) e dinâmico (relacionado a capacidade, habilidade física e volição). Na próxima seção, apresento alguns sentidos modais numa visão tradicional, além de descrever mais detidamente as categorias básicas aplicadas nesta pesquisa nas seções 2.3.2.1, 2.3.2.2 e 2.3.2.3.

2.3.2 A visão tradicional:

Em uma visão tradicional, encontra-se normalmente, além das três categorias básicas – epistêmica, deôntica e dinâmica – as categorias alética, bulomaica (ou bulética), circunstancial e teleológica.

A modalidade alética, por definição, refere-se à verdade contingente, possível ou necessária de uma proposição. von Wright (1951), em seu “*Um ensaio em lógica modal*”, afirma que há quatro tipos de modos: o alético, o epistêmico, o deôntico e o existencial. Os modos aléticos são os modos da verdade, e a preocupação central para as análises da lógica modal. De acordo com von Wright (1951, p. 8), “[a]s modalidades aléticas são ditas *de dicto* quando são sobre o modo ou forma na qual uma proposição é ou não é verdadeira. As modalidades são usadas *de dicto* em frases como ‘É necessário que...’ ou ‘É impossível que...’, etc.”³⁰

Nas abordagens da semântica formal (LYONS, 1977), o valor alético se diferencia do valor epistêmico, uma vez que o último está relacionado aos modos de conhecimento. Esta diferença se aproxima daquela já discutida anteriormente do caráter objetivo e subjetivo da modalidade epistêmica.

Segundo Palmer (1986, p. 11), “não há qualquer distinção entre [...] o que é logicamente verdadeiro e o que o falante acredita ser verdade, efetivamente, e “não há qualquer distinção gramatical formal em inglês e, talvez, também em qualquer outra língua, entre a modalidade alética e a epistêmica”.³¹ Vejamos o exemplo:

- (2.29) Uma refinaria representa investimento de muitos milhões. Além do mais, o refino do petróleo **pode** ser feito em qualquer lugar do país. (CP, 19Or:Br:Intrv: Com)

Na ocorrência em (2.29), coletada no *Corpus* do Português (DAVIES; FERREIRA, 2006), o falante afirma uma possibilidade física ou natural, compatível com o que está estabelecido em um determinado universo.

³⁰ Tradução minha para: “The alethic modalities are said to be *de dicto* when they are about the mode or way in which a proposition is or is not true. The modalities are used *de dicto* in phrases such as “it is necessary that ...”, “it is impossible that ...”, etc.” (von WRIGHT, 1951, p. 8).

³¹ No original: “there is no distinction between [...] what is logically true and what the speaker believes, as a matter of fact, to be true” and “there is no formal grammatical distinction in English, and, perhaps, in no other language either, between alethic and epistemic modality.” (PALMER, 1986, p. 11).

Entretanto, alinho-me a Palmer no sentido de que compreendo que em um determinado enunciado em que um conhecimento é apresentado como dado e não especulativo (o que seria, em lógica modal, reconhecido como uma verdade alética), considero que seja um conhecimento epistêmico ou ainda como uma possibilidade dinâmica. Portanto, a diferença colocada entre verdade no mundo e verdade na mente de um determinado indivíduo perde o sentido. Portner (2009, p. 135) propõe, inclusive, o termo “modalidade factual” para abarcar os tipos epistêmico, alético e metafísico.

Conforme Portner (2009, p. 10), há uma dificuldade em se classificar precisamente o uso alético de um modal e que este tipo pode ser considerado o mais básico, a partir do qual outros podem ser definidos.

A modalidade bulomaica indica o grau de aprovação ou desaprovação do falante (ou outra pessoa) em relação a um determinado estado de coisas. Este tipo, assim como o epistêmico e o deontico (que podem ser considerados mais fortes ou mais fracos), podem ser analisados de forma escalar, com polos positivo e negativo.

O tipo circunstancial se refere ao que é possível ou necessário, de acordo com determinadas circunstâncias, por exemplo:

(2.30) **É possível** instalar uma internet a rádio na zona rural de Viçosa.

A sentença em (2.30) expressa a informação de que há as condições necessárias que viabilizam a instalação de um serviço de internet na zona rural de uma cidade, não importa se a instalação vai ocorrer ou não.

Segundo Mello et al. (2009, p. 119), alguns autores não consideram este tipo de forma autônoma, concebendo-os como enunciados modalizados ou aleticamente ou dinamicamente.

A modalidade teleológica expressa o que é necessário para se atingir um determinado objetivo.

(2.31) Para entrar no Ministério Público, **tenho que** estudar com disciplina.

Em (2.31), em que o objetivo é ser funcionário do Ministério Público e, para alcançar tal objetivo, é necessário um estudo disciplinado.

2.3.2.1 Modalidade epistêmica: a linguagem das possibilidades

Lyons (1977, p. 797-800) afirma que as interpretações epistêmicas de índices modais são frequentemente mais subjetivas, e que essas interpretações são mais básicas que as objetivas. Kiefer confirma este caráter subjetivo da modalidade. De acordo com ele, “[a] modalidade epistêmica subjetiva expressa diferentes graus de comprometimento com a factualidade”³² (KIEFER, 2009, p. 183), e completa dizendo que, dessa forma, se relaciona com a categoria de evidencialidade.³³

Bybee et al. (1994, p. 179) propõem que a modalidade epistêmica marca “a extensão até a qual o falante está comprometido com a verdade da proposição”³⁴ ou, como define Palmer (1986, p. 51), o “grau de comprometimento do falante com a verdade do que ele diz”.³⁵

No mesmo sentido, van der Auwera e Plungian (1998) argumentam que uma proposição é avaliada como certa ou provável em relação a um conjunto de julgamentos. Coates (1983, p. 41), por sua vez, sustenta que a modalidade epistêmica sempre indica “a confiança (ou falta de confiança) do falante na verdade da proposição expressa”³⁶ e, dessa forma, varia entre o limite da confiança e da dúvida.

Nuyts (2001) nos oferece a seguinte definição:

A modalidade epistêmica é [...] uma avaliação das chances que um certo estado-de-coisas hipotético tomado em consideração (ou algum aspecto dele) vai ocorrer, está ocorrendo ou ocorreu em um mundo possível, o qual funciona como o universo de interpretação para a avaliação do processo e o qual, no caso default, é o mundo real (ou antes, a interpretação do avaliador desse mundo [...])³⁷ (NUYTS, 2001, p. 21)

A avaliação, nesse caso, iria da absoluta certeza de que um estado-de-coisas é real à absoluta certeza de que não é real. Entre os dois extremos estaria um *continuum* que inclui a possibilidade e a probabilidade. Como os exemplos extraídos de nossa amostra indicam:

³² No original: “Subjective epistemic modality thus expresses different degrees of commitment to factuality and it thus relates to evidentiality.” (KIEFER, 2009, p. 183).

³³ Trato da relação semântica entre modalidade epistêmica e evidencialidade mais adiante, na seção 2.4.

³⁴ “the extent to which the speaker is committed to the truth of the proposition” (BYBEE, PERKINS PAGLIUCA, 1994, p. 179)

³⁵ “the degree of commitment by the speaker to the truth of what he says”. (PALMER, 1986, p. 51).

³⁶ “[speaker’s] confidence (or lack of confidence) in the truth of the proposition expressed”. (COATES, 1983, p. 41).

³⁷ “Epistemic modality is [...] an evaluation of the chances that a certain hypothetical state of affairs under consideration (or some aspect of it) will occur, is occurring, or has occurred in a possible world which serves as the universe of interpretation for the evaluation process, and which, in the default case, is the real world (or rather the evaluator’s interpretation of it [...])” (NUYTS, 2001, p. 21).

- (2.32) [141] os cara que são bem mais / boleiros / eles / **com <certeza>** vão saber alguma coisa // (*bfamcv01*)
- (2.33) *LUI: [7] **com certeza** es nũ vão participar / uai> // (*bfamcv01*)
- (2.34) [198] minha mãe / **sem chance** // (*bfammn04*)
- (2.35) *LUA: [95] <e> assim / se cê tivesse uma coordenação pedagógica melhor / **talvez** / adiantasse um pouco assim / ou / pelo menos ajudasse a organizar / né // (*bpubmn01*)

O exemplo em (2.32) representa a certeza que um estado-de-coisas vai ocorrer (“saber alguma coisa”), ao contrário de (2.33) e (2.34) que apontam para a certeza de que não vão ocorrer. Já em (2.35) temos um exemplo da nuance de possibilidade que Nuyts fala de um estado-de-coisas ocorrer ou não, dadas determinadas circunstâncias, no caso a hipótese de “uma coordenação pedagógica melhor”.

Sweetser (1990) considera a modalidade epistêmica como extensão da modalidade raiz. A partir da análise da modalidade raiz em termos de forças e barreiras sociofísicas, ela explora a transferência dessa visão para a modalidade epistêmica, para atingir uma análise unificada sobre modalidade.

A autora se alinha a Palmer (1986), no sentido de argumentar que os verbos modais em inglês nos seus sentidos epistêmicos expressam o julgamento do falante. Assim, pretende encontrar uma conexão semântica motivada entre o domínio epistêmico de raciocínio e julgamento e o domínio da modalidade sociofísica.

Por exemplo, se o *may* sociofísico representa, em termos de dinâmica de forças, uma ausência de barreira potencial, o *may* epistêmico tem um sentido paralelo no domínio do raciocínio, isto é, há uma barreira ausente no processo de raciocínio do falante, a partir das premissas disponíveis para a conclusão expressa na sentença qualificada pelo verbo. Segundo Sweetser, nossa experiência sobre os domínios físico, social e do raciocínio compartilham alguma estrutura, que nos permite um mapeamento metafórico bem sucedido entre os aspectos relevantes dos três domínios.

Os fatores pragmáticos (como as pistas contextuais e a prosódia) vão influenciar para o ouvinte (ou para o indivíduo endereçado) compreender se um modal corresponde a um domínio ou outro. Veja os exemplos com os verbos ‘dever’ e ‘poder’ no português brasileiro:

(2.36) **dever epistêmico**(2.36a) **necessidade**

*RUT: [178] <mas eles> também deve ter condições /=COM= uai /=PHA=\$
(*bfamcv02*)

(2.36b) **probabilidade**

o diâmetro dea **deve** dar uns [/1] uns quarenta a cinquenta centímetro de [/1]
de &s [/2] de grossura / o diâmetro dela // (*bfammn01*)

(2.37) **poder epistêmico**(2.37a) **possibilidade**

*BAL: [229] então isso aqui a gente **pode** <inventar de colocar em qualquer
lugar> // =COM=\$ (*bfamd102*)

(2.37b) **impossibilidade**

*MAI: ele nũ [/2] ele nũ [/2] ele nũ <**pode** nem> falar p' ocê <que é verdade
// (*bfammn01*)

Em recente trabalho sobre o significado epistêmico, Boye (2012) propõe, igualmente, que a modalidade epistêmica está definida em termos da noção do grau de certeza, do grau de comprometimento, o que ele vai chamar de “apoio epistêmico”³⁸. Esta categoria, assim como a de evidencialidade, estaria subordinada a outra categoria descritiva, a epistemicidade, que seria uma generalização das noções de justificativa epistêmica e apoio epistêmico, ao que os filósofos chamam de “apoio justificativo”.³⁹

Kärkkäinen (2003) destaca que, ainda que as definições propostas para a modalidade epistêmica façam referência à verdade da proposição, os *types* relacionados a ela são de possibilidade, probabilidade e certeza (inferida).

No PB, a modalidade epistêmica pode ser expressa por elementos de natureza lexical como verbos modais, advérbios modais, verbos de atitude proposicional, construções adjetivas, expressões modais, e de natureza gramatical como o futuro perifrástico e as condicionais. Encerro a seção com mais alguns exemplos de nossa amostra:

(2.38) *TER: [21] ô Jael // [22] mas / gente velha / já prometeu o [/1] os presente / <já
/ **pode**> garantir que ganhou // (*bfamcv02*)

³⁸ “epistemic support” (BOYE, 2012, p. 2).

³⁹ “justificatory support” (BOYE, 2012, p. 2-3).

- (2.39) [171] não / trinta reais / aí eu &j [2]=SCA= eu [1] eu fico imaginando que e' fica pensando assim / Nossa Sio' / **às vezes** lá em casa tá precisando de fazer uma compra e tudo / né // (*bpubmn01*)
- (2.40) [72] quela [1] quela aula igual se dá no EDUCONLE que cê **acha** que é &o [1] é tudo de bom / né // (*bpubmn01*)
- (2.41) // **foi verdade** <mesmo> // (*bfammn01*)
- (2.42) *PAU: [153] porque **é capaz** d'eu subir uma parede lá // (*bpubdl01*)
- (2.43) [59] porque até aí / essas criança até dez anos / o [1] o mundo que nós tamo vivendos hoje / com dez ano já **dá pa** ver que dá trabalho // (*bfammn05*)
- (2.44) [44] es **vão pegar** os caras que / tipo / tavam reclamando e tal / es vão pegar [3] es **vão pegar tentar fazer** um negócio desse / e eu aposto que cê **vai ver** os caras que já conhecem a gente há mais tempo / tipo José [1] Zé Mourinho falando assim / não / o / campeonato d'ocês é bem melhor // (*bfamcv01*)
- (2.45) *BRU: [268] <e **se** for uma palavra composta / cê faz assim> // (*bfamcv04*)

Em (2.38) temos uma ocorrência com o verbo modal “poder” com valor epistêmico, que denota possibilidade. (2.39) traz um exemplo com a locução adverbial “às vezes”, normalmente utilizada no sentido temporal de “em algumas ocasiões ou circunstâncias”, mas que, na diatopia mineira, toma o sentido modal de possibilidade, sendo sinônima de “talvez”. A ocorrência em (2.40) traz um verbo de atitude proposicional (ou ainda verbo epistêmico ou verbo de crença) “acha”, que indica o grau de certeza sobre o que está enunciado. As construções adjetivas estão representadas nos exemplos (2.41) e (2.42), com as expressões “foi verdade” e “é capaz”. No primeiro caso, temos um sentido de certeza e, no segundo, de possibilidade. O uso modal do verbo “dar”, exemplificado em (2.43), é ainda pouco reconhecido na literatura (cf. SALOMÃO, 2008), no entanto, é recorrente na fala, em sentidos possibilitativos. Os marcadores de modalidade de caráter gramatical estão ilustrados em (2.44) e (2.45): o primeiro exemplo traz uma série de usos de futuro perifrástico que indicam o grau de certeza ou comprometimento em relação à realização de um evento futuro; no segundo exemplo, temos uma construção condicional na forma canônica “se *p*, então *q*”, em que o conectivo “se” é o marcador modal.

2.3.2 Modalidade deôntica:

Segundo a tradição lógica, o valor deôntico diz respeito às noções da obrigação e da permissão. Este valor está ligado a princípios morais ou legais e de conduta social.

Lyons (1977, p. 792-793) afirma que a mesma distinção aplicada à modalidade epistêmica entre leitura subjetiva e objetiva pode ser estendida à modalidade deôntica, como exemplifica a partir da proposição “*Alfred is a bachelor*” (“Alfred é celibatário”):

(2.46) Alfred is obliged to be unmarried.

Alfred é obrigado a ser descasado.

(2.47) “I (hereby) oblige Alfred to be unmarried”.

“Eu obrigo Alfred ser descasado”.

Em (2.46) teríamos uma interpretação deôntica mais objetiva e, em (2.32), uma mais subjetiva. Mais além, no exemplo (2.47) temos um performativo. Tento resgatar em nossa amostra um exemplo que, em minha opinião, é ambíguo no que diz respeito a uma leitura subjetiva ou objetiva como coloca Lyons:

(2.48) [132] po' parar o carro hhh // (*bfamd105*)



Se levarmos em consideração a ocorrência acima, com o verbo modal “poder” deôntico, em sua forma aferética ‘*po*’, desconectada de seu contexto de uso e da forma como foi enunciada, não haveria dúvida de que as seguintes interpretações seriam possíveis:

(2.48a) É permitido parar o carro.

(2.49b) Eu te permito parar o carro.

No entanto, não é o caso quando a escutamos [🔊]. O contexto permite a interpretação de “é o caso de” ou “é oportuno” e percebemos uma ilocução do tipo representativo do subtipo “conclusão”. Não estou colocando em dúvida o caráter deôntico da ocorrência, mas gostaria de pensar como (e se) a força ilocucionária, em sobreposição à modalidade⁴⁰, atenuaria ou reforçaria a ideia de permissão.

Kiefer (2009) argumenta que se deve fazer uma distinção entre o estabelecimento de uma necessidade deôntica ou de uma possibilidade deôntica (entre proposições deônticas) e a emissão de uma ordem (impor uma obrigação) ou dar permissão. As primeiras, segundo ele, pertencem à semântica (e se proposicional, são descritas em termos de lógica formal), as últimas são parte da teoria dos atos de fala e devem, assim, ser tratadas na pragmática.

⁴⁰ Sobre a distinção entre modalidade e ilocução, ver seção 2.4.3.

A literatura identifica a modalidade deôntica com os atos de fala diretivos (e com o modo imperativo, cf. BYBEE et al., 1994), uma vez que a obrigação e permissão derivariam de uma fonte de autoridade. As evidências em *corpora* de fala espontânea, com arquivos de áudio disponíveis para checar os atos ilocucionários, apontam para outro caminho. Os exemplos de Tucci (2007, p. 62)⁴¹, do *corpus* de referência do italiano, problematizam a ideia de comando nos usos de *dovere* e *potere*, em enunciados assertivos, que expressam necessidade contingente, em (2.50); impossibilidade condicionada, em (2.51); e esclarecimento de uma regra operativa, em (2.52):

(2.50) *LIA: si *doveva* rigovernare //^{COM} [ifamcv02]

(2.51) *GAL: ‘un *potevi* fa’ cip //^{COM} [ifamcv14]

(2.52) *MAM: ma te la *devi* fare a lei / <e lei la fa a te> //^{COM} [ifamcv25]

Ou como atestado na amostra do *corpus* de fala espontânea do PB, em que temos diferentes classes de atos de fala a partir de Moneglia (2011, p. 490)⁴², representativos, diretivos e expressivos:

(2.53) **dever > necessidade / obrigação:**

(2.53a) *LUI: [213] eu acho que a gente **deve** chamar os <times> legais // (bfamcv01)

% ill.: asserção forte



(2.54) **ter que > necessidade / obrigação**

(2.54a) [211] então **tem que** ser eu mesma // (bfamcv02)

% ill.: expressão de obviedade



(2.55) **poder**

(2.55a) **permissão**

[205] a porta tá fechada mas cê **pode** abrir // (bfamdl05)

% ill.: autorização



⁴¹ Os exemplos em (2.50), (2.51), (2.52) correspondem, respectivamente aos exemplos 20, 21 e 22 de Tucci (2007, p. 62), transcritos aqui *ipsis litteris*.

⁴² A tarefa de anotação ilocucionária ainda está por ser realizada no âmbito do projeto C-ORAL-BRASIL. Os exemplos elencados foram anotados com a colaboração do aluno de doutorado Bruno Rocha, exclusivamente para ilustrar nosso argumento.

(2.55b) **proibição**

*BRU: [177] cê nũ **pode** fazer &a [/1] nenhum som // (bfamcv04)

% ill.: ordem

(2.55c) **restrição**

[181] porque eu acho que no mesmo concurso / cê nũ **pode** fazer duas // (bfamd103)

% ill.: asserção



No *subcorpus* analisado, foram identificadas expressões de modalidade deôntica com os *types* ‘poder’ (em grande número, com significado de permissão), ‘ter que’ (a ampla maioria, com significado de obrigação) e ‘dever’ (em número reduzido)⁴³.

2.3.3 Modalidade dinâmica:

A modalidade dinâmica, de acordo com Palmer (1990) e Papafragou (2000), inclui uma habilidade e a intenção (vontade), isto é, a expressão da capacidade do sujeito. Salomão (1990) dá os seguintes exemplos com o verbo ‘poder’:

(2.56) Ele **pode correr** 4 km sem cansar (habilidade física)

(2.57) Ele **pode citar** Virgílio de cabeça (habilidade mental)

Os exemplos de Salomão se sustentam dentro da definição de Palmer e Papafragou, no entanto, se realizada uma busca minuciosa em *corpus*, dificilmente esses usos ocorreriam no português brasileiro ou mesmo na língua falada. Em uma busca no *Corpus* do Português, por exemplo, com a expressão “pode nadar”, encontrei apenas duas ocorrências no português europeu:

(2.58) A maioria dos caranguejos não **pode nadar**, podendo rastejar lentamente para a frente, e mais frequentemente, movendo-se para o lado, especialmente quando rastejam a grande velocidade. (CP, 19Ac: Pt: Enc)

(2.59) **Pode nadar**, mergulhar ou andar junto ao fundo, usando a pressão da água sobre as asas e cauda para se manter em baixo, enquanto procura larvas de insectos e outros pequenos animais.

⁴³ Na análise dos verbos modais, tratarei da substituição do uso do ‘dever’ deôntico por ‘ter que’ no PB.

Os dois exemplos pertencem ao mesmo gênero, verbete de enciclopédia. Pelo que foi discutido anteriormente, estas são ocorrências que ficam na fronteira entre a possibilidade alética, o conhecimento epistêmico e a habilidade dinâmica, uma vez que o verbo modal “poder” carrega o sentido de um conhecimento estabelecido em um determinado universo, não se constituindo como uma especulação, mas igualmente carrega o sentido da habilidade dos seres em “nadar” (no caso, também “rastejar”, “mergulhar” e “andar junto ao fundo”).

Ainda poderia colocar em discussão, o estatuto do verbo “saber” como um modal que expressa conhecimento epistêmico ou habilidade do sujeito em realizar uma determinada atividade em contextos como em (2.60) e (2.61):

- (2.60) Se a pessoa **sabe fazer feijão**, então ela deve ensinar isso, porque tem muita gente que não **sabe fazer feijão**.
- (2.61) É um talento, o homem. **Sabe falar** e **sabe escrever**. Se ele quiser escrever uma carta para você chorar, você chora. Se quiser escrever uma carta para você rir, você ri.

Palmer (1990, p. 36) aponta que a modalidade dinâmica diz respeito à habilidade ou à volição do sujeito da sentença, não do falante, e, dessa forma, não é subjetiva como outras modalidades.

No PB, a modalidade dinâmica se diferencia igualmente da modalidade epistêmica e os verbos mais frequentes são “conseguir” e “aguentar”, para sentidos de capacidade, e “querer” e a forma “gostaria”, no sentido de volição. Na amostra, encontramos, ainda que em número bastante reduzido, ocorrências do tipo dinâmico, como ilustro abaixo:

- (2.62) *CEL: [85] e' nũ <**consegue** fazer isso nunca tá meio de hhh> [/3] <tá de ladinho> // (bfamcv03)
- (2.63) [184] o papai nũ **güenta** carregar mais // (bfammn05)
- (2.64) [50] e eu &a [/3] e [/1] e vou **conseguir** / sabe / vencer essas dificuldades aí // (bpubmn01)
- (2.65) [168] tá pensando em organizar um torneio / e a gente **quer** / que vocês participem / da organização / <mandando representante> // (bfamcv01)

Os exemplos em (2.62) e (2.63) apontam para a habilidade física do sujeito, já o em (2.64) indica habilidade sócio-afetiva ou disposição interna do sujeito. Em (2.65), finalmente, temos um exemplo com o verbo “querer”, no sentido volitivo.

Há quem argumente (LARREYA, 2009) que a volição está incluída no significado deôntico de um enunciado como “*John must leave*” (“John deve sair”) em dois sentidos: no primeiro, a obrigação expressa no enunciado incluiria uma vontade, uma vez que há uma exigência imposta, seja por princípio moral ou por uma pessoa, que explicita um desejo, no caso, de que John saia; no segundo, o uso de “*must*” carrega em seu significado a implicação de que, se John obedece à injunção contida no enunciado, isso caracterizaria um ato voluntário por parte do agente ‘John’. A primeira seria uma vontade externa e a segunda uma vontade interna.

Sobre a compulsão interna ou intenção, o que me faz aninhar a volição à categoria dinâmica e não à deôntica, Larreya (2009) a considera como uma vontade não-factual, em exemplos como “*I want to go*” (“Eu quero ir”), em que o evento pode ou não se realizar. Seria uma vontade que não depende de uma força externa, mas unicamente da disposição interna do sujeito.

2.4 Modalidade epistêmica e evidencialidade: o limite (?) entre categorias:

Os estudos sobre evidencialidade partem da análise de línguas em que há um sistema evidencial gramaticalizado, como as línguas nativas norte-americanas (CHAFE, 1995; MITHUN, 1995), o tibetano (VOKURKOVA, 2011; MELAC, 2012) ou o cuzco quéchua (FALLER, 2002). Apesar de não ser uma categoria gramatical em inglês ou português, ela pode ser expressa por índices lexicais, como advérbios ou verbos como parecer e dizer (na expressão “diz que”).

Justamente por não ser uma categoria com marcação gramatical e por alguns dos itens lexicais poderem igualmente ser índices que expressam modalidade, a fronteira entre evidencialidade e epistemicidade é difusa e a relação entre as duas noções ainda ocupa o debate entre linguistas (CHAFE, 1986; FITNEVA, 2001; NUYTS, 2001; PLUNGIAN, 2001, entre outros). Dendale e Tasmowski (2001, p. 341-342) levantam os problemas de conceituação e apontam três caminhos encontrados em estudos recentes: “*disjunção* (em que

são conceptualmente distinguidas), *inclusão* (em que uma está incluída no escopo semântico da outra) e *sobreposição* (em que elas parcialmente se interseccionam).⁴⁴

Palmer (2001), por exemplo, organiza a modalidade epistêmica e a evidencial subordinadas a uma categoria hierarquicamente superior da modalidade proposicional. A mesma decisão de inclusão da evidencialidade como uma categoria modal aparece em Bybee (1985), Bybee et al. (1994) e Frawley (1992). Na outra ponta, a da exclusão, estão os estudos de de Haan (1999, 2005).

Como visto anteriormente, a modalidade epistêmica expressa os diferentes graus de comprometimento em relação à validade do material enunciado. Já os evidenciais são marcadores que indicam a fonte e a confiabilidade do conhecimento do falante. As duas categorias lidam com a evidência, no entanto, a noção de comprometimento, no domínio da modalidade epistêmica, está relacionada com a postura epistêmica do falante, isto é, com a probabilidade ou certeza que atribui àquilo que está dizendo. Já no domínio evidencial, esta noção está relacionada ao grau de confiança de onde parte a evidência.

Este grau de confiança, segundo Cornillie (2009, p. 58), pode ser explicado em termos de *status* da evidência compartilhado ou não-compartilhado. Como são várias as fontes de evidência (o falante; o falante e outros participantes da comunicação; ou exclusivamente os outros), a confiabilidade pode variar, mas não envolve uma avaliação ou probabilidade epistêmica. Nas palavras de de Haan (2005, p. 379), “a evidencialidade *assevera* a evidência, enquanto a modalidade epistêmica *avalia* a evidência”⁴⁵. Na mesma linha, Pietrandrea (2005, p. 33) afirma que “enquanto a evidencialidade qualifica a fonte que justifica a asserção de uma proposição, a modalidade qualifica a crença genuína do falante sobre a verdade da proposição”.⁴⁶

A evidencialidade, pois, lida com a fonte da evidência que um falante tem sobre uma dada informação. Os marcadores de evidência expressam se um falante foi testemunha direta do evento ou atividade que descreve ou se esta informação chega de outra fonte. Dessa forma, a categoria pode ser dividida em duas subcategorias: evidência direta e evidência indireta (o ouvi-dizer-que, a citação e a inferência).

⁴⁴ Tradução minha para: “disjunction (where they are conceptually distinguished from each other), inclusion (where one is regarded as falling within the semantic scope of the other), and overlap (where they partly intersect)”.

⁴⁵ Tradução para: “[e]videntiality asserts the evidence, while epistemic modality evaluates the evidence” (de HAAN, 2005, p. 379).

⁴⁶ No original: “[w]hile evidentiality qualifies the source that justifies the assertion of a proposition, modality qualifies the genuine belief of the speaker about the truth of the proposition.” (PIETRANDEA, 2005, p. 33).

Em nossa amostra, levanto alguns exemplos para ilustrar a questão e sugerir, de alguma forma, que as categorias, de fato, se distinguem. Vou partir de três exemplos para evidenciar meu ponto.

(2.66) *LUI: [236] **eu acho que** a gente deve chamar os <times> legais // (*bfamcv01*)

(2.67) [65] se o brasileiro ã lê os manuais / hhh no mercado de reposição / &auto [1] de autopeça / **eles acham que abrir uma empresa é comprar um produto por um real /=COB= na base cem / e vender por dois acha que tá ganhando o &do [2] o dobro** // (*bfammm06*)

(2.68) *ROG: [187] dá // [188] tem muita pedra ali / uai // [189] lá embaixo ainda tem pedra / perto da [2] perto da garagem / lá tem // [190] assim vai ficar uma pracinha boa aqui //

*PAU: [191] e **a Isa tava achando que** ela ali ia ficar pequena // [192] falei assim / depois de feito é que a gente vê o tamanho / né // (*bpubdl01*)

No exemplo em (2.66), temos um julgamento epistêmico de um conceptualizador⁴⁷ primário (o falante), o mais frequente em nossa amostra. Em (2.67), há a ocorrência de uma avaliação por parte do falante de um julgamento epistêmico de um conceptualizador em terceira pessoa, o que poderia ser textualizado como “eu acho que eles acham”. O uso de um pronome genérico como “eles” pode indicar essa adesão a uma avaliação que parte do senso comum. Por fim, em (2.68), o falante reporta um julgamento epistêmico de uma terceira pessoa, porque partilha dessa informação (ele/ela acha, porque compartilhamos esse conhecimento).

Parece que há em (2.67) e (2.68) dois níveis de interpretação: o primeiro nível estaria relacionado ao falante e à confiabilidade/fonte da informação, que seria evidencial, uma vez que é uma informação/avaliação filtrada pelos olhos do falante, no primeiro caso, e, no segundo exemplo, há uma avaliação reportada pelo falante; o outro nível estaria relacionado a um conceptualizador em terceira pessoa e à sua (suposta) avaliação, portanto, uma interpretação epistêmica.

O seguinte *continuum* representa os graus de comprometimento/afastamento do falante, conceptualizador primeiro, no que diz respeito ao material enunciado:

⁴⁷ O conceptualizador é a entidade que constrói um ponto de vista em uma determinada cena. Esta noção está ligada à de *construal* que é a nossa capacidade de conceber e retratar a mesma situação de maneiras diferentes.

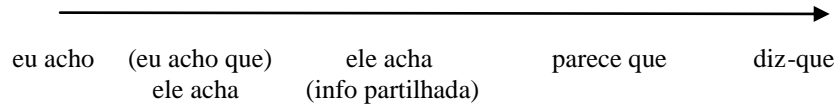


Figura 2.3 - *Continuum* modalidade epistêmica/evidencialidade

Nos exemplos, observam-se três diferentes graus de participação do enunciador na constituição da avaliação sobre o que está sendo dito, o que me leva a tentar reformular o conceito de modalidade e assumir a noção de conceptualizador⁴⁸, nos termos langackerianos, para explicar este fenômeno semântico.

2.5 O escopo da modalidade e as fronteiras entre algumas categorias

Nas seções anteriores foram apresentadas diferentes abordagens e tipologias no que diz respeito à modalidade. As definições desta categoria podem ser mais amplas, como a de Palmer (2001) que diz que a modalidade se relaciona com o status da proposição que descreve o evento, ou mais restritas como a de van der Auwera e Plungian (1998), para quem a modalidade envolve os domínios de possibilidade e necessidade como variantes paradigmáticas. A primeira definição não deixa claro o que inclui ou exclui a modalidade; na segunda, identificam-se termos reconhecidos como “modais”, mas exclui, por exemplo, habilidade e volição.

Bally (1942, p. 3) apresenta uma definição enunciativa da modalidade como o *modus* do *dictum*, em suas palavras: “a forma linguística [*Modus* ou *dictum*] de um julgamento intelectual, de um julgamento afetivo ou de uma vontade que um sujeito pensante enuncia a propósito de uma percepção ou de uma representação de sua alma”.⁴⁹ Segundo ele, uma frase explícita é composta por duas diferentes partes: uma, o *dictum*, é a representação percebida pelos sentidos, pela memória e a imaginação; a outra, o *modus*, é uma operação psíquica de um sujeito pensante.

Ao *modus* é conferida mais importância que ao *dictum*. O autor oferece exemplos para o que considera as operações do pensamento⁵⁰ sobre o que se enuncia, que indicam um entendimento, um sentimento e uma vontade: “alguém crê que vai chover” ou “não crê” ou “duvida”, “alguém se alegra porque chove” ou “alguém lamenta que chove”, “alguém deseja

⁴⁸ Desenvolvo este ponto na seção 2.5.

⁴⁹ No original: “la forme linguistique d’un jugement intellectuel, d’un jugement affectif ou d’une volonté qu’un sujet pensant énonce à propos d’une perception ou d’une représentation de son esprit” (BALLY, 1942, p. 3).

⁵⁰ Pensar, para Bally (1932/1965, p. 35), seria “reagir a uma representação, constatando-a, apreciando-a ou desejando-a”.

que chova” ou “que não chova”.

Esta parece ser, entretanto, uma definição muito ampla — apesar de indiscutivelmente nos inspirar a considerar a dimensão do enunciado — e seguirei na tentativa de estreitá-la e responder à pergunta: que tipo de modificação do *dictum* deve ser considerada sob o escopo da modalidade ou, em outras palavras, como definir o que compreende o *modus*? Como definir: (a) de que modificação se trata; (b) o que está incluído no escopo da modalidade; (c) quais os domínios semânticos envolvidos; (d) como estabelecer fronteiras entre as diferentes categorias.

A partir de agora, portanto, argumentarei no sentido de diferenciar as categorias que podem modificar o que é dito e, em particular, apontar as nuances entre negação, modo, ilocução e atitude.

A pergunta que emerge, ao me defrontar com dados orais, em situações reais de uso da língua é: “por que seria insuficiente esse conceito de modalidade fundado na lógica aristotélica?” Dito de outra forma: a definição de modalidade e o seu campo de aplicação até agora utilizados pela semântica, como herança da lógica formal, para a análise de sentenças isoladas servem completamente ao estudo desse fenômeno na fala e na dinâmica de uma interação?

2.5.1 Modalidade e negação:

A categoria da negação vem sendo estudada, em uma perspectiva filosófica, desde Aristóteles. A pergunta que se persegue responder, em termos de valor de verdade, é: como o valor de verdade de uma sentença muda, se for adicionado um elemento negativo? Na lógica, a negação é um operador que inverte as condições de verdade de um determinado conteúdo proposicional.

Lawler (2007) coloca que a negação é um fenômeno linguístico, cognitivo e intelectual e que é fundamental para todo o pensamento humano. Nas palavras de Kiefer (2009, p. 193), “a negação de um evento teria que ser caracterizada em termos da ocorrência do evento ser feita relativamente à sua não-ocorrência”.⁵¹

Nas línguas naturais a negação funciona como um operador associado a quantificadores e modais, sendo que a relação entre estas categorias se dá de forma complexa. Vejamos alguns casos ambíguos:

⁵¹ [...] the negation of an event would have to be characterized in terms of the occurrence of the event being made relative to its nonoccurrence [...]” (KIEFER, 2009, p. 193).

(2.69) Ele não pode ir ao jogo.

(2.69a) Não é possível ele ir ao jogo.

(2.69b) Não é permitido ele ir ao jogo.

Destaco de nossa amostra algumas ocorrências em que temos o operador de negação associado a um índice modal. A partir deles, discuto a relação entre dois índices modais (um deles modificado pela negação).

(2.70) *GIL: [2] <ô / mas> / voltando à questão / falando em [/2] e também falando em povo mascarado / esse povo do Galáticos é muito palha / eu acho que es nũ deviam mais participar / e <tal> // (*bfamcv01*)

(2.71) *PAU: [247] ele não pode chegar no final da semana e / desmanchar tudo o que foi feito durante a semana / né //
 *ROG: [248] pode não // [249] mas tem gente que desmancha / sô Paulo // (*bpubdl01*)

(2.72) [185] cê não pode aprontar [/1] apontar pra mesa // (*bfamcv04*)

Os verbos “dever” e “poder”, modificados pelo operador de negação, nos casos acima, expressam interdição/impedimento, restrição e proibição, respectivamente, portanto, estão inscritos no significado deôntico da modalidade. Quero destacar o exemplo (2.70) em que temos dois índices modais na mesma unidade informacional: um verbo epistêmico ‘achar’ e o verbo modal ‘dever’ deôntico. Neste caso, vai atuar o princípio da composicionalidade: o valor epistêmico ordena o valor deôntico. Em outras palavras, o falante, a partir de seu conhecimento de mundo e de suas crenças, emite sua opinião que envolve uma avaliação relacionada à interdição.

Apresento mais dois exemplos do *subcorpus* em estudo, para discutir o escopo da negação:

(2.73) *CES: [391] acho que nũ tem ninguém aí não // (*bfamdl03*)

(2.74) *LUI: [232] <eu nũ acho que a gente [/6] eu nũ acho que a gente deve chamar só veteranos não> // (*bfamcv01*)

Na ocorrência em (2.73), o verbo de atitude proposicional ‘achar’ não está sob o escopo da negação. A avaliação aqui apresentada é de que o falante considera que o material

locutório da encaixada não ocorre. Já em (2.74) o verbo é posto no escopo da negação: o falante enuncia uma sua opinião negativa.

No PB, há três tipos de negação: a simples preposta, a simples posposta e a dupla negação. Neste trabalho, investigarei ainda a relação de modalidade e negação para fins de anotação semântica. Este ponto vai ser desenvolvido na Parte III desta tese.

2.5.2 Modalidade e modo:

O modo é uma categoria morfológica e se refere a um conjunto de tipos de enunciados que indica a maneira de o verbo expressar um estado de coisas. Esta categoria é apresentada em sua relação com a modalidade ou como a expressão gramaticalizada da modalidade, por exemplo, o modo imperativo se relaciona com a modalidade deôntica e os outros modos com a modalidade epistêmica (OLIVEIRA, 2003, p. 254).

Bybee et al. (1994), como já visto, incluem o modo no domínio da modalidade, enquanto van der Auwera e Plungian (1998), no domínio das ilocuções. Como Green (2009) que afirma que “o modo com o conteúdo subdetermina a força. Por outro lado, é uma hipótese plausível que o modo gramatical é um dos dispositivos que usamos, em conjunto com as pistas contextuais, entonação e outros, para indicar a força com a qual estamos expressando um dado conteúdo”.⁵²

Palmer (1986) define o “modo” como uma categoria morfossintática de natureza verbal, no mesmo nível de tempo e aspecto, sendo caracterizado, em relação à modalidade, como um recurso linguístico menos prototípico. Lyons (1977, p. 848) também caracteriza o modo como uma categoria gramatical e, portanto, diferente da modalidade, que estaria associada ao domínio semântico. O que é corroborado por Narrog (2005, p. 167) que afirma que o termo modo “se refere a formas linguísticas específicas ou paradigmas de formas, tipicamente na inflexão verbal, cuja função primeira é expressar modalidade”.⁵³

Halliday (1970) distingue modalidade e modo, relacionando a primeira à expressão de necessidade e possibilidade e, o segundo, à força ilocucionária. O modo estaria ligado a noções de realidade, existência, factualidade, e a modalidade a noções de possibilidade, probabilidade, necessidade e volição.

⁵² No original: “Mood together with content underdetermine force. On the other hand, it is a plausible hypothesis that grammatical mood is one of the devices we use, together with contextual clues, intonation and so on, to indicate the force with which we are expressing a given content.” (GREEN, 2009).

⁵³ No original: “[...] refers to specific linguistic forms or paradigms of forms, typically in verb inflection, whose primary function is to express modality”. (NARROG, 2005, p. 167)

Na tradição gramatical do português do Brasil, o modo verbal é classificado em três tipos: indicativo (relacionado ao factual; o subjuntivo (relacionado ao hipotético e contrafactual); e o imperativo (relacionado a comandos, ordens, pedidos).

O modo indicativo, de maneira mais geral, é não-marcado para a modalidade, com exceção para os usos modais do futuro do presente e futuro do pretérito, conforme os exemplos abaixo:

(2.75) [33] eu vou olhar pr' ocê se eu tenho trinta-e-nove / dessa / dessa // (*bpubdl02*)

(2.76) [37] aí ea falou assim / cê ia ter neném aonde // (*bfammn04*)

Já o subjuntivo, segundo Palmer (1986, p. 39), é um marcador genérico de modalidade. Morfossintaticamente, ocorre em orações encaixadas (PALMER, 1986, p. 22) e, com frequência, é uma marca redundante de um elemento modal presente na oração principal. No *minicorpus* analisado, encontramos algumas dessas ocorrências:⁵⁴

(2.77) *REN: [550] espero que eu não tinha [/1] tenha perdido // (*bfamdl01*)

(2.78) [39] queria uma criança que nũ me desse trabalho / e tudo // (*bfammn05*)

Podem aparecer também em ocorrências, como as em (2.79) e (2.80), com valor de hipótese:

(2.79) [148] porque esse aqui é quando for desenhar // (*bfamcv04*)

(2.80) [239] pra que Deus venha &ri [/1] contribuir pa sua sorte // (*bfamdl04*)

O modo imperativo, por sua vez, poderia estar ligado à modalidade deôntica, como reconhece Portner (2007, p. 399), “[a] ideia de que os imperativos devam ter alguma coisa a ver com a interpretação de modais deônticos é extremamente intuitiva”.⁵⁵ De fato, uma enunciação imperativa é uma tentativa de impor a um indivíduo uma obrigação. No entanto, a opção neste trabalho, dados seus pressupostos, é de que os imperativos estão ligados a atos de fala do tipo diretivos e não à expressão da modalidade.

⁵⁴ Vale lembrar que no PB, em contextos anteriormente tidos como tipicamente de uso do subjuntivo, observa-se uma variação com o emprego de formas do indicativo. No entanto, não está no escopo deste trabalho a discussão desta flutuação. Para esta variação no espanhol, ver o trabalho de Mejías-Bikandi (1996), apoiado na Teoria dos Espaços Mentais de Fauconnier (1994). Nesta moldura teórica, o modo é visto como um mecanismo que regula o compartilhamento de informações entre espaços mentais.

⁵⁵ Tradução para: “The idea that imperatives should have something to do with the interpretation of deontic modals is extremely intuitive.” (PORTNER, 2007, p. 339)

Na próxima seção, trato da modalidade e sua relação com a ilocução e a atitude, no sentido de demonstrar que são categorias diferentes, porém que se interrelacionam.

2.5.3 Modalidade, ilocução e atitude:

O conceito de ilocução ou ato ilocucionário foi primeiramente formulado por Austin (1962) em uma série de conferências em que sustentava que “*falar alguma coisa é fazer alguma coisa*”.⁵⁶ Austin distingue no ato linguístico três níveis: o locutório, o ilocutório e o perlocutório.

Este conceito é retomado mais tarde por Searle (1969) que considera que um ato ilocucionário é a unidade mínima da comunicação linguística humana. O filósofo revê a taxinomia austiniana e propõe cinco categorias básicas para como usamos a linguagem (SEARLE, 1979/2002, p. 19-31): (a) assertivos (dizer às pessoas como as coisas são); (b) diretivos (levar as pessoas a fazer coisas); (c) compromissivos (nos comprometer a fazer coisas); (d) expressivos (expressar nossos sentimentos); e (e) declarações (provocar mudanças no mundo através das emissões linguísticas).

Na mesma linha, Cresti (2001) reconhece a divisão em três níveis do ato linguístico e afirma que a ilocução coocorre com o ato locutório e produz um efeito de caráter afetivo-pulsional. Principalmente a partir de critérios prosódicos, ela propõe cinco classes de atos ilocucionários:

- (a) **Asserção:** uma disposição, atitude ou estado de espírito de autoconfiança por parte do falante, com base nas próprias realizações de pensamento, que permite a manifestação de julgamentos, descobertas, avaliações, representações, como novos objetos no mundo.
- (b) **Direção:** uma disposição, atitude ou estado de espírito de levar em consideração as habilidades, possibilidades, disponibilidade do ouvinte, enquanto se espera transformar o mundo através de ações, informações, movimentos, ou que o ouvinte transforme a si mesmo com relação a seu horizonte de atenção, seu conhecimento, sua habilidade, seu ponto de vista.

⁵⁶ “[...] to say something is to do something [...]”. (AUSTIN, 1962, p. 12)

- (c) **Expressão:** uma disposição, atitude ou estado de espírito de manifestação “estética” de estados mentais, sentimentos, emoções, crenças, esperando que o ouvinte se torne consciente disso e compartilhe disso.
- (d) **Rito:** um comportamento externo de execução de tarefas lingüísticas que têm valor e efeito legal e social, e que pode ser realizado com o mínimo de participação afetiva, apenas suficiente para a ativação fisiológica da fala.
- (e) **Recusa:** uma disposição, atitude ou estado de espírito de liberdade e separação por parte do falante com relação ao seu interlocutor, que permite um confronto completo com esse último.

Cresti assume que a ilocução não se opõe à modalidade, mas que as duas pertencem a dois níveis diferentes. Tucci (2011), em análise de *corpus* de fala espontânea do italiano, afirma que há duas razões principais para se separar as duas noções: a primeira, de caráter formal, consiste em assumir que, por definição, cada enunciado corresponde a uma e somente uma força ilocucionária, enquanto que mais de um valor modal pode coexistir num enunciado; a segunda, de um ponto de vista mais qualitativo, consiste em atestar que não há uma correspondência entre o valor modal e a força ilocucionária, apesar de os índices modais contribuírem para a interpretação ilocucionária de um enunciado.

Outros autores, igualmente, assumem que a modalidade se circunscreve perfeitamente no nível da semântica e não da pragmática (FRAWLEY, 1992; KIEFER, 1997; NARROG, 2005). Kiefer (1997, p. 247) afirma que “os atos de asseverar, permitir, ordenar ou prometer são estranhos à modalidade. Não há como elaborar uma questão significativa sobre como um ato afeta a modalidade. Isso vale para a ilocução assim como para a perlocução”.⁵⁷

Desde um ponto de vista pragmático coerente não se pode considerar a modalidade como a mesma coisa que ilocução. Mello e Raso (2011, p. 5-ss.), em estudo experimental, fazem uma distinção entre as categorias de modalidade, ilocução e atitude.⁵⁸

Os autores pleiteiam que a modalidade pertence ao domínio da semântica, no qual a postura do falante em relação à expressão do material locutório se manifesta, de maneira que a mesma ilocução pode ser modalizada de diferentes formas, sem afetar o nível ilocucionário. A ilocução, por sua vez, pertence ao domínio pragmático, no qual a postura do falante em relação ao seu interlocutor se manifesta.

⁵⁷ No original: “[t]he act of asserting, permitting, ordering or promising is alien to modality. There is no way to ask a meaningful question about how such an act affects modality. This holds for illocution as well as for perlocution.” (KIEFER, 1997, p. 247).

⁵⁸ A distinção entre modalidade e atitude está na próxima seção 2.4.4.

Em relação à atitude, uma das dificuldades para se distinguir modalidade e esta categoria reside justamente na própria polissemia do termo atitude, que pode ser definida como uma posição/postura assumida, ou como emoção (MOZZICONACCI, 2001), ou como comportamento/maneira de agir em relação a uma outra pessoa. Nos trabalhos de anotação semântica de opinião e sentimentos (cf. WIEBE et al., 2005), os conceitos de modalidade e atitude também se confundem.

Mello e Raso (2011) constatam que a categoria de atitude, comparada às de modalidade e ilocução, ainda é pouco discutida na literatura. Eles apresentam alguns estudos como o de Local (2005) que menciona a importância da categoria para os estudos de entonação, ou a definição de Moraes et al. (2010) de que “a atitude prosódica geralmente se refere à expressão de afetos sociais, voluntariamente controladas pelo falante”.⁵⁹ Ainda segundo Moraes et al. (2010), existem algumas atitudes que afetam o conteúdo proposicional de um enunciado, como a ironia, a incredulidade, a surpresa, e outras que estão ligadas a relações sociais estabelecidas entre os participantes de uma interação, como a polidez, a autoridade, a arrogância.

Assim, é reconhecida a dimensão sociointeracional da atitude, isto é, o falante revela seu humor enquanto realiza uma ilocução específica (com uma modalidade específica). A mesma ação pode ser realizada de diferentes maneiras, por exemplo: triste, contente, desafiadora, envolvente, surpresa, irritada etc.

Como experimento para a distinção das três categorias, os pesquisadores consideraram três ilocuições diferentes, pertencentes à classe dos diretivos (cf. CRESTI, 2000; MONEGLIA, 2011), a saber, sugestão/recomendação, convite e pergunta, para materiais locutórios com diferentes modalidades: o primeiro, “Vem pro Brasil”; o segundo, “Pode vir pro Brasil”; e o terceiro, “Tem que vir pro Brasil”. Foram consideradas também duas atitudes: comprometida e irritada.

Comparando a mesma ilocução produzida com diferentes atitudes, observou-se que enquanto a ilocução tem um núcleo, a atitude está distribuída por toda a unidade tonal. Dessa forma, para se analisar a ilocução, é necessário olhar a posição do núcleo, a forma do núcleo e o alinhamento do núcleo, isto é, a relação do núcleo com o conteúdo segmental.

Também se observou que, se se mantêm a ilocução e a atitude e se varia a modalidade, esta última não vai afetar os parâmetros prosódicos. O único aspecto que parece mudar é a parte de preparação da curva, devido a uma mudança no material segmental.

⁵⁹ No original: “prosodic attitude generally refers to the expression of social affects, voluntarily controlled by the speaker.” (MORAES et al., p. 1).

Outro aspecto observado por Mello e Raso (2011, p. 15) foi que, a depender da ilocução, um mesmo índice modal pode ser interpretado com diferentes significados. Por exemplo, para o material locutório “eu devo passar na casa dele”, se enunciado como uma asserção, recebe uma leitura epistêmica; enunciado como uma pergunta, uma interpretação deôntica.

Em resumo, conforme o experimento de Mello e Raso (2011, p. 16), as categorias da modalidade, ilocução e atitude se diferenciam e, ao mesmo tempo, se interrelacionam. Nas suas palavras, “uma atitude específica ‘prefere’ algumas ilocuições e uma ilocução específica ‘prefere’ algumas modalidades, ao ponto de que o mesmo item lexical vai receber uma interpretação preferencial devido à ilocução em que figura”.⁶⁰

Um esquema possível para esta relação seria o seguinte:

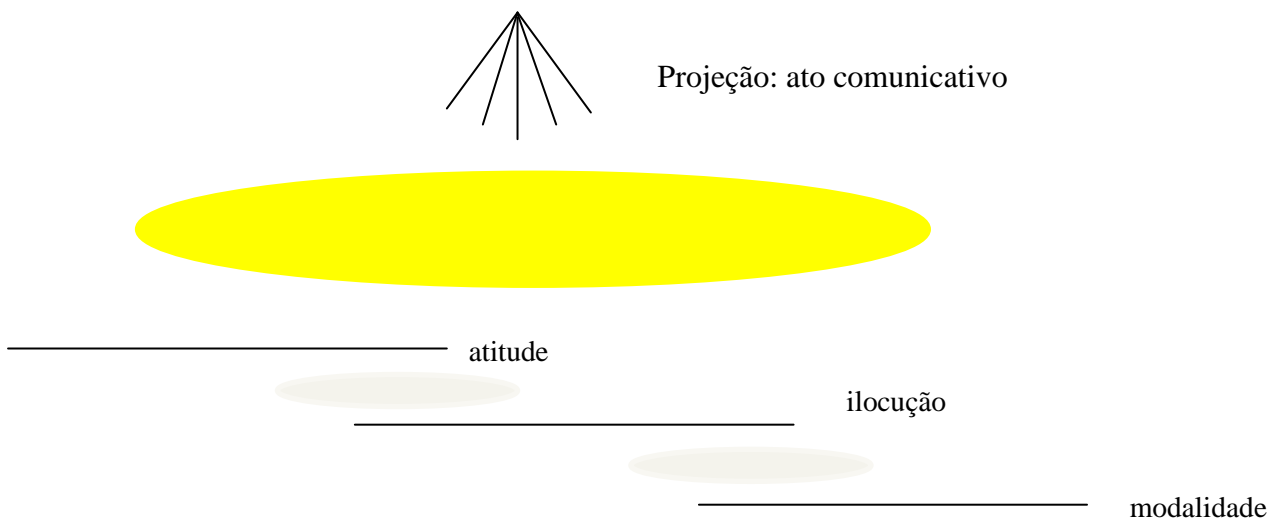


Figura 2.4 – Projeção do ato comunicativo

A figura é um esforço para representar a hierarquia no ato comunicativo: o campo atitudinal pode projetar uma espécie de zona de sombra na camada da ilocução, que, por sua vez, projeta uma zona de sombra na camada da modalidade. A modalidade seria o “*Modus do Dictum*”, a ilocução o “*Actum do Dictum*”; e a atitude o “*Modus do Actum*”.

Parto, portanto, para tentar a formulação de uma definição mais apropriada para que abarque, igualmente, a modalidade na fala espontânea, a partir de dados empíricos.

⁶⁰ “a specific attitude ‘prefers’ some illocution and a specific illocution ‘prefers’ some modalities, to the point that the same lexical index will receive a preferred interpretation due to the illocution it figures in.” (MELLO; RASO, 2011, p. 16).

2.6 Modalidade na fala espontânea

Tomarei, para início de conversa (e porque, como explicitado nos objetivos, pretendo reavaliar a noção de modalidade), a posição de Cresti (2002) e Tucci (2007, 2008, 2009) que assumem que a modalidade é inerente a qualquer enunciado, com ou sem a presença de índices modais. Para elas, a modalidade está sob o domínio do nível semântico, não do pragmático, uma vez que se constitui como um processo cognitivo, não uma ação.

Como a fala não é organizada em proposições lógicas, mas em unidades pragmáticas, a incidência do valor modal veiculado por um índice, de acordo com Tucci (2007), é uma propriedade não do enunciado, mas da unidade informacional. Há três casos a serem considerados:

(i) se o enunciado é simples, quer dizer, se contém apenas a unidade de Comentário, a modalidade pertence simultaneamente ao enunciado inteiro e à unidade informativa;

(2.81) [31] três é o carinho no um ã dá // (*bfamcv03*)

(ii) se dentro de uma mesma unidade informacional temos dois índices modais, vale o *princípio de composicionalidade*, em que um índice domina o outro e determina a modalidade;

(2.82) *EVN: [148] **acho** <que a gente> **tem** / **que** olhar direito // (*bfamcv01*)

(iii) se um enunciado possui duas unidades informacionais, se aplica um *princípio de não-composicionalidade*, e o domínio da modalidade, portanto, recai sobre a unidade informacional.

(2.83) [397] **vai arrumar** lá o dia todo /=TOP= e de lá que ea **vai po** [/1] po coisa
//=COM=\$ (*bfamcv02*)

Dentre os valores modais, três primordiais são utilizados por Cresti e Tucci: o alético,⁶¹ o deôntico e o epistêmico. A modalidade alética se restringe à avaliação que o

⁶¹ A lógica modal diz respeito quase exclusivamente à modalidade alética, isto é, àquela relacionada com a verdade necessária ou contingente das proposições. Na linguística, este tipo de modalidade desempenha um papel marginal, por esta razão não será levada em conta neste trabalho, como explicitado na seção 2.3.1.

falante faz do estado-de-coisas segundo noções lógicas de verdade, e à apresentação do estado-de-coisas como verdadeiro, como nos seguintes exemplos:⁶²

(2.84) Um cisne **pode** ser negro.

Um leopardo **deve** ser malhado.

Como as noções de capacidade e habilidade estão ligadas à verdade do mundo, são encaixadas nessa categoria.⁶³

A modalidade deôntica está relacionada às noções de permissão e de obrigação, com as quais o falante avalia um estado-de-coisas referente a alguma ação, como nos exemplos seguintes:

(2.85) Não se **pode** abandonar a pessoa sozinha.

Marco **deve** partir hoje. (~ *ter que*)

A expressão de desejo, em alguns casos, também se insere nessa categoria, uma vez que o falante posiciona-se perante uma ação.

A modalidade epistêmica manifesta-se sempre que o falante imprime ao estado de coisas uma avaliação baseada em seus conhecimentos e crenças sobre o mundo e denota graus de possibilidade e de probabilidade, com relação aos estado-de-coisas:

(2.86) Júlio **pode** ter partido.

Devem ser os sete.

No que se refere ao discurso oral, o comprometimento com o conteúdo proposicional não se aplica, uma vez que o escopo da modalidade não pode estar limitado a proposições. O conteúdo linguístico é organizado em enunciados,⁶⁴ revelando um determinado padrão informacional (CRESTI, 2000). A visão de modalidade tomada pelas autoras italianas é a

⁶² Os exemplos que ilustram as categorias de modalidade são de Tucci (2007). Tradução minha.

⁶³ Os sentidos de capacidade, habilidade e volição são tratados de forma separada neste trabalho, como visto na seção 2.3.3.

⁶⁴ Na moldura da “Teoria da Língua em Ato” (CRESTI, 2000), como veremos em detalhes mais adiante, o enunciado é a unidade de referência da língua falada e é definido como qualquer expressão linguística interpretável pragmaticamente, ligada a: (a) uma condição semântica de plena significação da expressão em questão e (b) uma realização entoada segundo um padrão melódico de valor ilocutório.

seguinte: “expressão da avaliação do falante sobre o material locutório” e seu escopo é a unidade informacional.

Relembrando as perguntas que me motivaram até aqui, a saber: (a) de que modificação se trata a modalidade; (b) o que está incluído no seu escopo; (c) quais os domínios semânticos envolvidos; (d) como estabelecer fronteiras entre as diferentes categorias, creio que é possível ir um pouco além e entender a modalidade como uma nuance/relativização do material locutório enunciado, não necessariamente uma proposição, ligada à dimensão subjetiva e que inclui os domínios da possibilidade, da necessidade, da habilidade e da volição. Diferencia-se de negação, modo, ilocução e atitude, entretanto está estreitamente relacionada com estas noções.

Na minha formulação, a partir do que foi discutido anteriormente, a modalidade pode ser tomada como **o julgamento ou a avaliação de um sujeito conceptualizador, que relativiza o que está sendo enunciado, em termos de grau de certeza e das noções de possibilidade, necessidade, habilidade e volição. Ela está ancorada em uma situação comunicativa e vai cumprir diferentes funções pragmático-discursivas no curso da interação.**

Esta definição proposta se aproxima e se afasta do conceito corrente na literatura “a opinião ou atitude do falante em relação à proposição ou à situação contida na proposição” (LYONS, 1977, p. 452-3) por alguns motivos, os quais elenco a seguir.

(a) opinião, atitude > julgamento, avaliação que relativiza

Como vimos, a categoria da atitude está relacionada a um nível sociointeracional, no qual o falante demonstra seu humor enquanto realiza uma ilocução específica e com uma modalidade específica. Usar este rótulo, portanto, pode levar a interpretações equivocadas sobre a categoria da modalidade.

A modalidade é uma modificação que cria uma nuance ou uma relativização do que está sendo enunciado. Não pode ser confundida com qualquer modificação, nem com qualquer função modalizadora, nem com um quantificador (e.g. ‘aproximadamente’, ‘cerca de’,...), nem com um angulador (e.g. ‘tipo’, ‘uma espécie de’,...), é uma modificação que pressupõe o grau de comprometimento do sujeito da modalidade e/ou os domínios da possibilidade, necessidade, habilidade e volição.

(b) falante > sujeito conceptualizador

Utilizo uma noção semântico-cognitiva, o conceptualizador, nos termos de Langacker (1987, 1991, 2001, 2007, 2008), em lugar do “falante”, uma vez que a primeira noção é mais abrangente do que a última, pois abarca além do falante, o endereçado ou um terceiro indivíduo. O conceptualizador seria o próprio falante ou uma terceira pessoa para cujo ponto de vista o falante se projeta. Observe os exemplos abaixo:

(2.87) [117] eu falei / não / eu / **devo** ter a conta ainda / porque eu nã + (*bfammn03*)

(2.88) *LUC: [361] <isso **pode** ser / **potencialmente**> divertido / (*bfamcv04*)

(2.89) *DFL: [58] aquilo que o / o professor **achava** mais importante // (*bfammn02*)

(2.90) *KAT: [43] então ela acha que é a meia que tá melhorando // (*bfamd104*)

(2.91) [48] cê nã consegue / cem por cento de nada na sua vida / então / acho que / tô
[1] tô caminhando // (*bpubmn01*)

Em (2.87) e (2.88), temos um conceptualizador em primeira pessoa, que coincide com o falante, e enuncia a sua avaliação sobre o material locutório. Em (2.87), o enunciador se desloca para um momento anterior e reporta o que enunciou àquela altura, marcada linguisticamente pela presença do verbo *dicendi* ‘falar’. A ocorrência em (2.88) apresenta, inclusive, uma redundância da possibilidade epistêmica com o uso do verbo modal ‘poder’ e do advérbio modal ‘potencialmente’ (ou, conforme Halliday, 1970 e Lunguinho, 2010, o fenômeno da concordância modal⁶⁵). No exemplo em (2.89), o falante enuncia a avaliação de um conceptualizador em terceira pessoa. O falante em (2.90) pede a confirmação da avaliação de um outro conceptualizador (“ela”). Na ocorrência do monólogo público em (2.91), a falante no papel de professora se projeta para um ponto de vista impessoal, representado pelo dêitico “cê”, o que aponta para um grau de afastamento em relação ao que está sendo enunciado.

Segundo Langacker (2008), a conceptualização, metaforicamente, é a forma como vemos uma cena. Em suas palavras: “Um arranjo de ponto de vista é a relação global entre os “observadores” e a situação sendo “observada” [...] os observadores são os conceptualizadores que apreendem os significados de expressões linguísticas: o falante e o

⁶⁵ A concordância modal se caracteriza se caracteriza pela presença sintática de dois itens modais que são interpretados como se houvesse apenas um. (HALLIDAY, 1970; LUNGUINHO, 2010).

ouvinte.”⁶⁶ (LANGACKER, 2008, p. 73).

O arranjo de ponto de vista, portanto, seria uma espécie de ajuste ou afinação entre quem vê e o que vê. Este ajuste é tomado como default quando os interlocutores estão juntos em um local fixo, a partir do qual observam e descrevem o que acontece no mundo em volta deles e é “uma parte essencial para o substrato conceptual que sustenta o significado de uma expressão e molda a sua forma”.⁶⁷ As perspectivas default tendem a ser praticamente invisíveis e se explicitam quando se considera expressões que cumprem uma ação, como uma pergunta ou ordem.

O que a modalidade parece sugerir é que se trata da modificação avaliadora de uma perspectiva default, ou seja, aquela perspectiva explícita, marcada semanticamente. Por exemplo, em “O livro tá sobre a mesa”, estamos diante de uma perspectiva default. Já em “Eu acho que o livro tá sobre a mesa”, temos o grau de certeza modalizador como acréscimo ao ponto de vantagem do arranjo default.

Cresti (no prelo, p. 13), corroborando esta posição, afirma que “[c]ada cena é uma função de uma perspectiva: o ponto de vista do falante ou do endereçado ou de alguém externo, e pode ser marcada socialmente e/ou estilisticamente [...]”⁶⁸. Uma cena se caracteriza por coordenadas nos eixos temporal e espacial, em um determinado universo possível.

(c) proposição > material locutório enunciado

Como estamos tratando da modalidade na fala, com um entendimento particular do que é esta fala espontânea (CRESTI; SCARANO, 1998) e dentro do quadro teórico da Língua em Ato (CRESTI, 2000), há que se deslocar o alvo de incidência da proposição para o material locutório enunciado, no escopo de uma unidade informacional (cf. TUCCI, 2007).

Recuperando, resumidamente, a formulação de Tucci, a modalidade não é uma propriedade do enunciado, mas da unidade informacional. Há três casos a serem levados em conta: (a) se o enunciado é simples, isto é, se contém apenas a unidade de Comentário, a modalidade pertence simultaneamente a todo o enunciado e à unidade informacional; (b) se

⁶⁶ No original: “A viewing arrangement is the overall relationship between the “viewers” and the situation being “viewed”. [For our purposes], the viewers are conceptualizers who apprehend the meanings of linguistic expressions: the speaker and the hearer.” (LANGACKER, 2008, p. 73).

⁶⁷ Tradução para: “an essential part of the conceptual substrate that supports an expression’s meaning and shapes its form”.

⁶⁸ No original: “Every scene is a function of a perspective: the point of view of the speaker or the addressee or someone external, and can be marked socially and/or stylistically [...]” (CRESTI, no prelo, p. 13).

dentro de uma mesma unidade informacional temos dois índices modais, vale o *princípio de composicionalidade*, em que um índice domina o outro hierarquicamente e determina a modalidade; e (c) se um enunciado possui duas unidades informacionais, aplica-se o *princípio de não-composicionalidade*, e o domínio da modalidade, portanto, recai sobre a unidade informacional.

Alguns dados da amostra, no entanto, me levam a considerar que a modalidade tem uma incidência local, no âmbito da unidade informacional, mas também é uma propriedade dinâmica, que atua igualmente, em uma cadeia referencial, no domínio textual.

Segundo Cresti (no prelo, p. 12), cada *chunk* linguístico, de um ponto de vista semântico, corresponde a uma cena (cf. BARWISE; PERRY, 1981; FAUCONNIER, 1984) e, para permitir o desenvolvimento de uma função textual, as expressões participantes devem estar na mesma cena.

Segundo Simon-Venderbergen e Aijmer (2007), os marcadores modais podem evoluir de uma função puramente qualificacional para funções textuais e retóricas. Podem ser marcadores de confirmação, adversidade, continuidade e causalidade. Vejamos este excerto de uma conversação familiar, do tipo argumentativo predominantemente, cujo tema é um campeonato de futebol:

(2.92) Excerto 1:

*LEO: [1] o Juninho <foi> //

*GIL: [2] <ô / mas> / voltando à questão / falando em e também falando em povo mascarado / **esse povo do Galáticos é muito palha / eu acho que es nũ deviam mais participar / e <tal> //**

*LUI: [3] <não> //

*LEO: [4] <não> //

*LUI: [5] <**eu acho não**> //

*LEO: [6] <**com certeza**> //

*LUI: [7] <**com certeza es nũ vão participar / uai**> //

Destaco do excerto acima os enunciados [2], [5], [6] e [7]. No enunciado [2], o participante *GIL expõe sua opinião sobre o time Galáticos (“é muito palha”) e faz também um julgamento sobre a participação desta equipe no campeonato que estão organizando (“eu

acho que es nũ deviam mais participar / e <tal>”). Os participantes *LUI e *LEO, por sua vez, fazem eco a esse julgamento, utilizando operadores modais de confirmação, como “acho não” e “com certeza”.

Os enunciados são diferentes, com suas respectivas modalidades, no entanto, o que parece é que podemos recuperar nos domínios locais elementos referenciais pertencentes a uma determinada cena anterior, em uma coordenação mútua de sistemas cognitivos, quer dizer, os sistemas cognitivos dos participantes de uma interação estão alinhados, o que Verhagen (2005) chama “intersubjetividade”. Segundo ele, a língua se presta não só a objetivos de informação, mas principalmente de argumentação. Assim, desenvolve o conceito de argumentatividade, ou seja, a capacidade das pessoas de ações cognitivamente coordenadas no fluxo, para induzir o interlocutor a realizar inferências e criar um discurso coerente.

Neste capítulo, apresentei o estado da arte dos estudos de modalidade, as tipologias tradicionais e as visões alternativas. Ainda, defini os valores modais que serão utilizados na pesquisa e, a partir de uma perspectiva subjetiva, tentei uma formulação para o conceito de modalidade que desse conta de explicar a que tipo de modificação me refiro quando utilizo os termos “modalidade”, “modal” e “modalizador”, e a que domínios semânticos ela estaria ligada.

A partir de agora, desde um ponto de vista empírico, introduzo os procedimentos metodológicos e os resultados da descrição do fenômeno da modalidade em um *corpus* de fala espontânea do português brasileiro.

PARTE II – PROCEDIMENTOS METODOLÓGICOS E ANÁLISE DOS DADOS

Capítulo 3

PROCEDIMENTOS METODOLÓGICOS

No capítulo anterior me ocupei em apresentar alguns estudos sobre a categoria de modalidade, destacados de uma vasta literatura sobre o tema, sejam eles vinculados a uma abordagem mais objetiva ou subjetiva, e também descrevi os tipos de significados modais da tradição linguística e de abordagens alternativas, e aqueles que utilizo na pesquisa. Ainda aponte a relação da modalidade com outras categorias a ela associadas como a negação, o modo, a ilocução e a atitude. Por fim, apresentei minha própria proposta de definição que se conforma com o estudo da modalidade na fala, que se diferencia, em larga medida, do estudo da categoria na escrita.

Neste capítulo descrevo os procedimentos metodológicos utilizados para empreender a pesquisa. Em primeiro lugar, introduzo os pressupostos da Teoria da Língua em Ato. Em seguida, apresento o projeto C-ORAL-BRASIL e a amostra de estudo, um *minicorpus* da parte informal deste *corpus* (RASO; MELLO, 2012). A seguir, apresento as outras etapas do trabalho, para fim de coleta, preparação, organização e análise dos dados.

3.1 A Teoria da Língua em Ato:

A Teoria da Língua em Ato para a análise da estrutura informacional, proposta por Cresti (2000), tem como hipótese básica, a partir de Austin (1962), que é possível dizer que há uma correspondência entre unidade de ação (os atos de fala) e unidade linguística (os enunciados), através de padrões entoacionais. Nas palavras da própria autora: “assumindo a teoria dos atos linguísticos de Austin (1962) como quadro teórico de referência, podemos descrever a verbalização oral como uma forma particular de comportamento humano, do qual cada instância é um ato linguístico, constituído da ativação simultânea de três diversos tipos de ato (*locutório, ilocutório e perlocutório*).”⁶⁹ (CRESTI, 2000, p. 42).

⁶⁹ No original: “assumendo la teoria degli atti linguistici di Austin (1962) come quadro teórico di riferimento, possiamo descrivere La verbalizzazioneorale come una particolare forma di comportamento umano, di cui ogni istanza è un atto linguístico, costituito dall’attivazione simultanea di ter diversi tipi di atti (*locutivo, ilocutivo e perlocutivo*).” (CRESTI, 2000, p. 42)

Essa teoria nasce no domínio da oralidade, especificamente da observação e análise da fala espontânea⁷⁰, quer dizer, de um texto falado, concebido ao mesmo tempo em que é executado e que não realiza nenhum texto anterior — escrito lido ou script (fala roteirizada).

De uma perspectiva comunicativa, a fala espontânea é um comportamento dinâmico de interação entre interlocutores, uma cadeia de ação-reação. Os participantes de uma troca conversacional realizam ações dirigidas a um ou mais interlocutores, o que chamamos de ilocuções. Uma ilocução se define a partir de um conteúdo afetivo, base de todas as relações entre pessoas.⁷¹

Cumprir uma ilocução seria realizar um ato ao dizer algo. Um ato de fala se realiza simultaneamente em três dimensões:

- a) ato locutório: parte linguística, ato de dizer algo.
- b) ato ilocutório: intenção comunicativa, ato ao dizer algo.
- c) ato perlocutório: é a produção de um efeito sobre o interlocutor.

A contraparte linguística de um ato de fala é um **enunciado**, que carrega sempre uma intenção comunicativa e é **a unidade de referência da diamesia falada**. Diferencia-se de uma frase por ser um elemento compreendido **em autonomia e analisável pragmaticamente**. A **prosódia** é o componente suprasegmental que nos permite identificar esses enunciados.

3.1.1 O enunciado:

A análise da fala diferencia-se da análise da escrita, uma vez que, segundo Moneglia (2011, p. 481), “a linguagem escrita pode ser propriamente segmentada de acordo com princípios sintáticos e semânticos”, enquanto “a identificação de unidades de referência em um corpus de fala pode dificilmente ser feita através dos mesmos dispositivos sintáticos e semânticos”.⁷² A primazia da escrita sobre a fala nos estudos linguísticos, aliás, leva-nos a

⁷⁰ A base para este estudo foi o LABLITA Corpus, principalmente o *subcorpus* LABLITA Corpus of Adult Spontaneous Spoken Italian. Disponível em: <http://lablita.dit.unifi.it/corpora/descriptions/lablita/>. Último acesso: 08 jan. 2014.

⁷¹ De acordo com Cresti (2000, p. 84 e ss.), um afeto é definido como “uma pulsão e representação de um esquema acional em relação ao interlocutor”, que permite uma ação concreta no mundo.

⁷² Tradução minha para: “[...] written language can be properly parsed according to syntactic and semantic principles. (...) the identification of the units of reference in a spoken corpus can hardly be identified through the same syntactic and semantic devices (Blanche-Benveniste 1997; Biber et al. 1999; Cresti 2000; Miller & Weinert 1998; Izre'el 2005).” (MONEGLIA, 2011, p. 481).

cometer o equívoco de negligenciar esta diamesia e tratá-la como um texto escrito, já que o que é estudado, normalmente, é a transcrição da fala, segmentada de acordo com pausas sintáticas o que corresponde, graficamente, ao uso da vírgula como pontuação (MELLO; RASO, 2013). No entanto, ancorado em consistente base empírica, Moneglia (2011) afirma que os dados de *corpora* orais apontam para a insuficiência dos mesmos critérios sintáticos aplicados à escrita para análise dos eventos de fala, uma vez que muitos destes eventos não apresentam sequer um verbo.

Ainda, de acordo com Mello e Raso (2013, p. 101), para o estudo efetivo da fala, há que se considerar os níveis hierárquicos definidos na expressão comunicativa da fala, o que pressupõe: “i. que inicialmente se individualize a unidade ilocucionária; ii. que dentro dessa unidade seja estabelecida a sua estrutura informacional; e iii. que somente dentro da unidade informacional seja possível uma análise sintática”.

Como mencionado acima, o enunciado é a entidade linguística que, do ponto de vista pragmático é autônoma, o que significa que não se define nem pela completude semântica, nem pela expressão da predicação. O critério de identificação de um enunciado parte de características entoacionais, dado que a força ilocucionária de um enunciado é veiculada através dela. A correspondência entre ilocução e um dado contorno prosódico é denominada “critério ilocutivo” (CRESTI, 2000). As funções linguísticas da entoação seriam, segundo Cresti (2000): (a) indicar o tipo de ilocução realizada no ato de fala; (b) delimitar os enunciados no contínuo fônico; (c) segmentar o enunciado em unidades menores; (d) assinalar o tipo informacional de cada unidade dentro do enunciado.

Assim, a prosódia pode ser considerada a interface formal entre os atos executados simultaneamente em um ato de fala⁷³. Podemos identificar como parâmetros prosódicos os seguintes elementos: movimento de F_0 (hz); intensidade (db); duração (s); alinhamento; velocidade de fala e ritmo. É importante destacar que o tipo de ilocução não está subordinado ao conteúdo locutório.

No que diz respeito à segunda função da entoação, esta delimitação dos enunciados é realizada através da percepção das variações prosódicas relevantes por um falante “competente”. As variações que resultam na segmentação do enunciado em unidades discretas são chamadas de quebras prosódicas, divididas em:

⁷³ Adiante veremos que a prosódia é apenas um dos critérios para identificação da unidade entoacional, ainda contribuem para esta definição do tipo informacional critérios funcionais e distribucionais.

(a) terminais: quebras que são percebidas como conclusivas e, portanto, indicam a completude prosódica do enunciado. Graficamente são representadas por duas barras (“//”);

(b) não-terminais: percebidas como quebras não-conclusivas. São mais fracas e cumprem a função de delimitar as unidades internas ao enunciado. Graficamente são representadas por uma barra (“/”).

As quebras prosódicas são utilizadas na transcrição e segmentação de textos orais (MONEGLIA; CRESTI, 1997). Como exemplos dessas quebras, terminal e não-terminal, respectivamente, temos:

(3.1) *REN [116] eu gosto de maçã //



(3.2) *REN: [145] desinfetante /a gente precisa //



Os enunciados podem ter um padrão simples ou complexo. No padrão simples, o enunciado é composto apenas pelo núcleo do ato de fala e é executado em uma única unidade tonal, necessária e suficiente porque carrega a força ilocucionária, denominada **Comentário**; no padrão complexo, o enunciado corresponde ao núcleo mais um conjunto de unidades com distintas funções informacionais, é executado em duas ou mais unidades tonais. Uma dessas unidades é obrigatoriamente o Comentário, as outras unidades cumprem funções informacionais textuais ou dialógicas.

As unidades tonais são marcadas pelas quebras prosódicas (terminais ou não-terminais) e, em princípio, realizam uma unidade informacional. A relação entre as unidades de um enunciado é de natureza funcional e apenas dentro de uma mesma unidade podem ser estabelecidas relações de ordem sintática.

Os critérios de individualização das unidades, aplicados simultaneamente, são de caráter: (a) funcional: a função desempenhada pela unidade no padrão informacional; (b) entoacional: perfil prosódico⁷⁴, tipo e direção do movimento de F₀, velocidade de elocução e intensidade, presença ou não de foco entoacional; e (c) distribucional: a posição da unidade

⁷⁴ O perfil prosódico é identificado com base na fonologia entoacional desenvolvida pelo grupo IPO (Instituut voor Perceptie Onderzoek, em holandês) ou Instituto para Pesquisa em Percepção (HART; COLLIER; COHEN, 1990). São definidas três classes de configuração: *root*, *prefix*, *suffix*, que se combinam em um contorno prosódico.

opcional em relação à unidade nuclear (o Comentário), que tem distribuição livre dentro do enunciado.

Há **unidades com função textual**, que compõem e agem diretamente sobre o texto do enunciado, e **unidades com funções dialógicas**, também chamadas *auxílios dialógicos*, que são instrumentos para regular o diálogo e a interação, agem sobre a situação comunicativa e/ou o interlocutor.

As unidades textuais são o Comentário, que, em um padrão complexo, pode ser realizado na forma de Comentários Múltiplos ou Comentários Ligados; o Apêndice de Comentário; o Tópico, o Apêndice de Tópico; o Parentético; e o Introdutor Locutivo. As unidades dialógicas são: o Incipitário; o Conativo; o Fático; o Alocutivo; o Expressivo; e o Conector Discursivo. Há outras unidades que não possuem valor informacional e sinalizam disfluências da fala ou escansão: a Unidade de Escansão, a Tomada de Tempo, e a Unidade Vazia.

A partir de agora, apresento a definição de cada uma destas unidades, tornadas discretas através dos critérios mencionados anteriormente, funcional, entoacional e distribucional.

3.1.2 Unidades Informacionais:

3.1.2.1 Comentário (COM):

A unidade de Comentário (COM) é a unidade nuclear primária de um padrão informacional e tem a função de **veicular a força ilocucionária** do enunciado. É a única que pode figurar sozinha no enunciado, uma vez que representa a unidade informacional necessária e suficiente para a sua realização e interpretabilidade.

Em termos de propriedades prosódicas, esta unidade corresponde ao perfil do tipo raiz e seu foco varia de acordo com os tipos de locução (recusa, asserção, direção, expressão, rito) e condições contextuais.

Quanto ao critério distribucional, o Comentário tem distribuição livre e estabelece restrições nas propriedades distribucionais de outras (possíveis) unidades informacionais, isto é, as demais unidades se organizam em função dele. Esta unidade pode ser segmentada por Unidades de Escansão.

Vejamos os exemplos:

(3.3) *BRU: [360] nũ tem nada que pode ser aproveitado // =COM=\$ (bpubcv01)



(3.4) *CAR: se for mais gente numa situação pior / =TOP= nós vamo ter que pensar
// =COM=\$ (bpubcv02)



O enunciado em (3.3) é do tipo simples, porque possui apenas a unidade de Comentário, suficiente e necessária. Em (3.4) temos um exemplo de enunciado do tipo complexo, com duas unidades informacionais: o Tópico e o Comentário.

3.1.2.2 Tópico (TOP):

O Tópico (TOP) é uma das principais unidades informacionais e tem como função **delimitar semanticamente o campo de aplicação da força ilocucionária**. Esta unidade seleciona o domínio de relevância em relação ao qual o ato de fala deve ser interpretado.

Prosodicamente, a unidade é marcada por um perfil do tipo prefixo e apresenta foco entoacional à direita. Assim como o Comentário, a unidade informacional de Tópico também tem um foco, o que permite distingui-la de outras unidades que antecedem o Comentário. No português foram identificadas quatro formas entoacionais diferentes para o Tópico (cf. ROCHA, 2012; MITTMAN, 2012).

Em termos de distribuição, antecede o Comentário, ainda que não seja adjacente a ele. A unidade também pode ser segmentada em Unidades de Escansão.

Um Tópico pode ser recursivo e também pode ser formado por subpadrões de unidades informacionais, normalmente do tipo referencial, o que se constitui como uma Lista de Tópicos. Todos os Tópicos de uma lista, juntos, fornecem um único domínio de aplicação da força do Comentário e todos eles pertencem ao mesmo domínio ontológico ou, no mínimo, eles devem ser semanticamente coerentes.

Como exemplos:

(3.5) *FLA: aí o caderno / =TOP= é um negócio meio atrasado // =COM= assim / =
PHA=\$ (bpubcv01)



(3.6) *FLA: o [/1]=EMP= o ceagaeme / =TPL(1)= o plasma / =TPL(2)= e a plaqueta
/ =TPL(3)= a gente armazena ate o dia seguinte // =COM= (bpubcv01)



Em (3.5), temos um SN que realiza um Tópico simples e a ocorrência em (3.6) representa uma Lista de Tópicos, um dos tipos de Tópico complexo.

3.1.2.3 Apêndice de Comentário (APC):

O Apêndice de Comentário (APC) integra textualmente o Comentário e é dependente desta unidade. Varia informacionalmente através de: (a) repetição: repetição ou eco (parcial ou não) de uma informação já conhecida ou uma paráfrase de um material locutivo no mesmo enunciado, ou em um enunciado diferente do mesmo falante ou mesmo do interlocutor; (b) preenchedor: não apresenta repetição de um conteúdo anterior, e se constitui como uma expressão formulaica que não apresenta quase nenhuma informação; e (c) informação retardada: o conteúdo semântico adiciona informação para fins sociais, quer dizer, oferece informação mais específica ou alterações que facilitam a compreensão do enunciado para o interlocutor.

Apresenta perfil entoacional do tipo sufixo, nivelado ou descendente, sem foco. Está distribuído à direita do Comentário e, em alguns casos, pode estar separado do Comentário por uma unidade de Fático ou Parentético. No entanto, nunca pode ser interrompido por um Parentético e, algumas vezes, é seguido por um Fático ou Conativo.

(3.7) *CAR: queria uma criança que ã me desse trabalho /=COM= e tudo // =APC=\$
(bfammn05)

(3.8) *CAR: [177] cê ã entendeu /=COM= cê ã / =APC=\$ (bfamcv03)

(3.9) *FLA: [358] tem que lembrar ela /=COM= comprar acetona // =APC=\$
(bfamd101)

3.1.2.4 Apêndice de Tópico (APT):

O Apêndice de Tópico (APT) integra o texto do Tópico e se constitui informacionalmente por integrações lexicais ou alterações e muito raramente por repetições exclusivamente do material locutório do Tópico.

Esta é uma unidade de tipo sufixo, com perfil que pode ser nivelado, descendente ou mesmo reproduzir a curva de F₀ do Tópico a que se refere. Diferentemente do Tópico, não possui foco. Distribui-se à direita da unidade de Tópico e o segue, podendo estar intercalado

por uma unidade dialógica (Fático ou Alocutivo, no PB). Não pode ser interrompido por uma unidade de Parentético.

(3.10) *BAL: um cuidado / =TOP= que cês têm que tomar / =APT= < Bel > +
=ALL=\$ (bfamd102)



3.1.1.5 Parentético (PAR):

A unidade de Parentético (PAR) expressa uma integração metalinguística do enunciado, em alguns casos com função modalizadora, apresentando um ponto de vista externo àquele do Comentário. A função metalinguística encontra sua relevância em relação a todo o enunciado ou pode se referir a uma unidade textual para trás ou para frente, na maioria das vezes um Comentário ou um Tópico, e mais raramente a outro Parentético ou uma palavra específica dentro de qualquer unidade informacional. Informacionalmente, cinco tipos de integração metalinguística foram encontrados: avaliação modal simples, adição de informação (com avaliação positiva ou negativa), comentários sobre a atividade não-linguística do falante simultâneo ao enunciado, instruções para o interlocutor que diz respeito à atividade dialógica, glosa terminológica, glosa locutiva de discurso reportado ou exemplificação.

O perfil prosódico do tipo parêntese é caracterizado por um abaixamento da F₀ e velocidade de execução mais alta que o resto do enunciado. Pode ocorrer em qualquer posição no enunciado, exceto na inicial (VALE, 2010).

(3.11) *BAL: [30] porque /=DCT= < se eu for > empregado / =TOP= por exemplo /
=PAR= alguém vê que eu sou muito foda / =TOP= < medo > de perder /
=TOP= < o posto > < deles / =APT= es vão [2] =EMP= es vão > me dizar /
=COM= < né > // =PHA=\$ (bfamd102)



Foi identificada, ainda, no *minicorpus* do C-ORAL-BRASIL, a unidade de Lista de Parentéticos:

(3.12) *LUZ: [68] aquele dia a Lilisa tava discutindo com a Deise /=COB= ela ja tava
[3]=EMP= já [1]=EMP= a Deise /=SCA= já &f [1]=SCA= tinha falado de
quem que era /=PAR= eu achei ate que ia ser do Ronan /=COB= mas /=DCT=
elas tava falando que nao /=COB= falando de um outro ai /=COM= nũ sei se é
Marco Túlio /=PRL= sei lá quem /=PRL= nũ sei quem // =PRL=
(bfamd103)



3.1.1.6 Introdutor locutivo (INT):

O Introdutor Locutivo (INT) sinaliza a suspensão pragmática do *hic et nunc* e introduz uma metailocução, por exemplo, um discurso reportado, uma exemplificação e um pensamento falado, um elenco e também um parentético. O perfil do tipo introdutor apresenta alta velocidade de execução e F_0 mais baixa do que a do Comentário subsequente. Sua posição é anterior ao Comentário que introduz.

(3.13) *FLA: [324] como diz você / =INT= não precisa // =COM_r=\$ (bfamd101)



3.2.2 Unidades Dialógicas:

As unidades dialógicas compartilham algumas propriedades que as distinguem das unidades textuais como, por exemplo, desenvolver uma função que não contribui para a construção do conteúdo semântico do enunciado e, por outro lado, contribuir para o desempenho feliz do ato de fala no contexto comunicativo e podem ocorrer em discurso reportado. Além disso, em sua maioria, são realizadas com apenas uma palavra ou expressão, não podem ser escandidas, não carregam foco, não podem ser modalizadas, possuem restrições de seleção lexical e não contribuem para a composicionalidade sintático-semântica do enunciado.

3.2.2.1 Incipitário (INP):

O Incipitário (INP) é uma unidade dialógica que marca a abertura do turno dialógico e regula o fluxo da interação. Ele abre o canal comunicativo e indica um valor de contraste ou uma oposição ao enunciado anterior.

Apresenta perfil prosódico de auxílio com F_0 alta, caracterizada por um movimento de rápida subida seguida de descida. Distribucionalmente, ocorre em início de enunciado ou turno e frequentemente é seguido de uma unidade de Fático.

(3.14) bom / =INP= ali também pode fazer o seguinte o' // =COM=\$ (bpubd101)



3.2.2.2 Conativo (CNT):

A unidade de Conativo (CNT) tem a função de provocar o interlocutor a se engajar na interação ou interromper o seu comportamento não-cooperativo. O perfil prosódico se caracteriza por intensidade alta, perfil descendente e duração curta. Na maioria das vezes ocupa posição inicial ou final, e algumas vezes posição intermediária. Há a possibilidade, ainda que pouco frequente, de ocorrer mais de uma unidade de CNT no enunciado.

(3.15) <pera aí / = CNT= pera pera aí> / =CNT= só um segundo / =CNT= só vão esperar o [1] / =SCA= a ampulheta / = COB= aí < desnegoçar > / = COM=\$ (bfamcv04)



3.2.2.3 Fático (PHA):

A unidade dialógica de Fático (PHA) tem a função de controlar o canal comunicativo, de modo a assegurar a manutenção de sua abertura. Ele estimula o ouvinte a manter a coesão social necessária numa troca conversacional ou tenta se certificar de que o enunciado foi recebido. Em posição final no enunciado, serve para marcar o acordo com o interlocutor e pode ser empregado para pedir uma confirmação. Por último, funciona como uma tomada de tempo para uma melhor programação do enunciado.

Apresenta perfil entoacional nivelado e duração muito curta. Sua posição no enunciado é livre.

(3.16) eu contei o caso < pra ele / =COM= né > // =PHA= (bfammn01)



3.2.2.4 Alocutivo (ALL):

O Alocutivo (ALL) opera no controle da comunicação, especifica a quem a mensagem é dirigida e mantém a atenção do interlocutor, desempenhando também uma função de coesão social. Tem duração curta, baixa intensidade e movimento descendente. Pode ocorrer em qualquer posição dentro do enunciado.

(3.17) < nã tem perigo não / =COM= sô > // =ALL= (bpubcv02)



3.2.2.5 Expressivo (EXP):

O Expressivo (EXP) é suporte emocional do ato ilocutório cumprido pelo Comentário, assinala o compartilhamento de uma identidade de grupo com o interlocutor. Apresenta F₀ modulada ascendente, com aumento da velocidade em relação à média do enunciado. Pode ocorrer em qualquer posição em relação ao comentário.

(3.18) é um [2] =EMP= é um pacote lá com a Oi /=COM= uai //EXP= (bpubcv01)



3.2.2.6 Conector Discursivo (DCT):

O Conector Discursivo (DCT) tem como função marcar a continuidade do discurso, indicando ao interlocutor que o processo de construção textual vai prosseguir. Ocorre sempre em início de enunciado ou de um subpadrão de estrofe. É uma unidade com duração longa, perfil nivelado ou modulado, baixa velocidade e intensidade alta.

O DCT funciona de modo oposto ao Incipitário. Enquanto o último marca descontinuidade no discurso, o primeiro marca a continuidade. O INP muitas vezes é usado para tomar o turno de outro falante, enquanto o DCT funciona mais internamente ao turno de um mesmo falante.

(3.19) então / =DCT= dia de sexta e sábado ele nũ trabalha // =COM=\$ (bpubdl01)



3.2.3 A perda do isomorfismo:

De acordo com Cresti (2000), há um princípio de isomorfismo entre enunciado e ilocução e entre unidade tonal e unidade informacional. Em alguns casos, no entanto, este princípio é quebrado, como nos seguintes: (i) quando uma unidade informacional se realizada em mais de uma unidade entoacional, fenômeno denominado escansão; (ii) quando mais de uma ilocução é realizada com determinado padrão dentro de uma única entidade linguística concluída; e (iii) quando varias ilocuições fracas e homogêneas são realizadas processualmente, em uma situação pouco acional, ao que é chamado de Estrofe.

3.2.3.1 Unidade de escansão (SCA):

O fenômeno da escansão ocorre apenas em unidades informacionais textuais. Estas unidades podem ser realizadas em mais de uma unidade entoacional. Assim, a Unidade de Escansão (SCA) não apresenta função informacional, mas é parte de uma unidade informacional maior. A SCA indica disfluências na fala, que podem ocorrer por dificuldades relativas à execução do programa melódico de uma unidade longa ou à expressividade da unidade informacional. Não possui foco. Uma característica deste fenômeno é a composicionalidade sintática entre as unidades tonais escandidas.

(3.20) com medo / =INT= que se ea entrasse dentro de casa / =TOP= **ea ia matar /**
=SCA= os filho / =SCA= com ea e tudo // =COM=\$ (bfammn01)



3.2.3.2 Comentários Múltiplos (CMM):

Os Comentários Múltiplos (CMM) se constituem como dois ou mais comentários dentro do mesmo enunciado, ou seja, separados por quebra não-terminal, e estruturados em um padrão retórico. Estes padrões podem ser de tipos variados e, até agora, foram identificados as seguintes possibilidades, de uma lista aberta: elenco; reforço; alternativa/pedido de confirmação; comparação; relação necessária; chamamento; clímax; e consequência.

(3.21) *SIL: **ou é vinho bom caro /**=CMM= **ou é cerveja //**=CMM= (bfamd104)



(3.22) **ô Heliana /** =CMM= **o vinho tava bom /** =CMM= (bfamd104)



Os Comentários Múltiplos são contíguos e como propriedades prosódicas apresentam o formato da unidade tonal é do tipo raiz + raiz (+ raiz), formando um padrão prosódico, e seus perfil prosódico e núcleo variam de acordo com o tipo ilocucionário.

3.2.3.3 Estrofes e Comentários Ligados (COB):

Uma Estrofe é uma entidade linguística concluída que não corresponde ao cumprimento de um ato de fala, mas sim a uma atividade de fala genérica. As Estrofes são unidades maiores que desenvolvem duas ou mais ilocuções antes da quebra prosódica, percebida como tendo valor terminal. Contrariamente aos Comentários Múltiplos, não são unidades programadas, mas uma sequência processual de Comentários Ligados (COB); nelas o princípio ilocucionário se enfraquece em favor de uma dimensão menos pragmática e mais textual.

Nos Comentários Múltiplos há um programa que requer mais de um Comentário para ser realizado; no caso das Estrofes, à medida que ocorre a enunciação outros Comentários vão sendo acrescentados, porque o falante os percebe como importantes no momento, sem que se tenha programado anteriormente.

Cada Comentário Ligado (COB) é caracterizado por uma força ilocucionária fraca e devem ser homogêneos. Aparentemente, as estrofes cumprem somente atos com baixa afetividade (descrição, narração, pergunta retórica, instrução, conselho, expressão de estado psicológico).

O objetivo primário da Estrofe parece ser a expressão de um pensamento em andamento que induz a realização de um texto oral. O texto da Estrofe é organizado como uma sequência de focos semânticos, cada um dos quais pode registrar um valor modal diferente. A homogeneidade ilocucionária dos COB, junto com a liberdade na variação modal, implica que a Estrofe é concebida através de uma relação atitudinal única com o destinatário.

Em síntese, a Estrofe se realiza quando a fala tende para uma dimensão menos pragmática e mais semântica. O texto, como conteúdo semântico, se torna mais evidente, e o ato linguístico, como operação pragmática, se enfraquece. As estrofes em princípio podem ser encontradas em todos os tipos de texto, mas são muito mais presentes no monológico e no contexto público. Podemos ver, portanto, a estrofe como a dimensão da acionalidade enfraquecida em favor de uma elaboração semântico-textual.

- (3.23) nos temos vinte-e-cinco funcionários /=COB= dentro de Minas Gerais
/=COB=atuando /=COB= com a base nossa aqui em [/1]=SCA= na capital
/=COB= e hoje nós tamos /=SCA= numa media de &future [/1]=SCA=
 faturamento de um-milhão-e-meio a um-milhão-e-setecentos-mil reais /=SCA=
 mês //COM= (bfammn06)



3.2 O corpus de fala espontânea: o C-ORAL-BRASIL:

O C-ORAL-BRASIL, corpus de fala espontânea do português brasileiro, é um projeto em desenvolvimento na Universidade Federal de Minas Gerais, especificamente no Laboratório de Estudos Experimentais e Empíricos da Linguagem (LEEL), coordenado pelos professores Tommaso Raso e Heliana Mello.

O *corpus* foi originalmente estruturado para o estudo da estrutura informacional do português brasileiro e suas ilocuções, com base nos pressupostos da Teoria do Padrão Informacional (CRESTI, 2000) e é construído aos moldes do C-ORAL-ROM (CRESTI; MONEGLIA, 2005), projeto que reúne os *corpora* comparáveis das quatro principais línguas românicas européias: italiano, francês, espanhol e português. Na arquitetura desses *corpora*, privilegia-se uma maior variação diafásica, a fim de se obter uma diversidade de ilocuções e de estruturas informacionais da fala.

O *corpus* se constitui de textos de fala espontânea, isto é, textos produzidos ao mesmo tempo em que são executados, e sua realização não pressupõe um texto anterior (escrito ou roteirizado). Está dividido em uma metade formal (em fase de coleta, transcrição e revisão) e em uma metade informal (concluída e já publicada, RASO; MELLO, 2012). A parte informal é composta de 139 textos (de aproximadamente 1500 palavras) e 208.130 palavras, em um total de 21h8min de gravação, organizada em duas seções: privado/familiar e público. Do total de textos, 80% pertencem ao domínio familiar (105 gravações), e 20% ao domínio público (34 gravações), distribuídos igualmente em três diferentes tipologias interacionais: 1/3 de monólogos, 1/3 de diálogos e 1/3 de conversações.

A classificação da tipologia interacional não está relacionada somente ao número de participantes na cena comunicativa, mas à dinâmica da situação e ao peso relativo de cada participante na produção linguística resultante. Além de se privilegiar a diversidade de situações, o corpus é balanceado em termos de variação diastrática. As gravações são realizadas em contexto natural enquanto os falantes realizam atividades.

Como nos *corpora* do C-ORAL-ROM, os textos são transcritos no formato CHILDES-CLAN, e procede-se à segmentação da fala em enunciados e unidades tonais e à anotação da estrutura informacional.

Escolheu-se representar de maneira sistemática uma única diatopia, no caso a do estado de Minas Gerais, e em particular a área urbana da capital, Belo Horizonte. Tal escolha se deve à limitada dimensão do *corpus*, que não poderia ser representativo de mais diatopias. Assim, pelo menos 50% dos falantes são mineiros. Tradicionalmente a fala mineira é dividida

em pelo menos três grandes áreas: Sul e Triângulo (falar paulista), Norte (falar baiano) e região constituída pela Zona da Mata, Zona Metalúrgica, Vertentes e Belo Horizonte e arredores (falar mineiro). Estas regiões estão representadas, mas a grande área metropolitana de Belo Horizonte tem a maior representação.

O *corpus* busca representar a variação diastrática, especialmente importante na fala brasileira, mas também, nesse caso, não se pretende alcançar uma representatividade sistemática. Há, entretanto, um esforço no sentido de que várias faixas socioculturais estejam presentes: em todas as ramificações do corpus constam interações com falantes de várias diastratias, interagindo tanto com falantes da mesma faixa quanto com falantes de faixas diferentes, o que se mostrará produtivo para a análise da modalidade.

Nesse aspecto, com relação aos critérios adotados para o C-ORAL-ROM, optou-se por adequar a indicação das faixas de escolaridade a uma distribuição que melhor refletisse a realidade brasileira, levando-se em conta as significativas diferenças em relação à realidade europeia.

3.3 Coleta e organização dos dados

3.3.1 O *minicorpus* do C-ORAL-BRASIL I:

Para fins desta pesquisa, utilizarei uma amostra da parte informal do corpus C-ORAL-BRASIL. Dentre os 139 textos, com média de 1500 palavras cada um, num total de 208.130 palavras, será analisado um *minicorpus*, composto de 20 textos de três tipologias interacionais, divididos em privados e públicos: 7 monólogos — 5 privados e 2 públicos —; 7 diálogos — 5 privados e 2 públicos — e 6 conversações — 4 privadas e 2 públicas.

Este *minicorpus* busca preservar a mesma estrutura da parte informal do C-ORAL-BRASIL e representa uma ampla variação diastrática e diafásica. Os seguintes critérios “de máxima qualidade” foram utilizados para a escolha dos 20 textos da amostra (RASO; MELLO, 2009, p. 32):

- (i) Representatividade de ramificação (monólogos, diálogos, conversações; privados e públicos);
- (ii) Maior variação de atividade, maior número possível de situações comunicativas diferentes;

- (iii) Alta qualidade acústica, de acordo com a qualidade do espectrograma; pouco ou nenhum ruído de fundo; pouco ou nenhum retorno do sinal; clareza da voz; ganho bom; cálculo confiável da F_0 ; baixa porcentagem de sobreposição de vozes;
- (iv) Diversidade de locutores, representatividade equilibrada entre vozes masculinas e femininas e de faixas etárias. Um informante não aparece mais de uma vez, a não ser que seja como interlocutor em atividade monológica.
- (v) Não marcação no que diz respeito à diastratia, a média dos informantes não é de diastratia nem muito alta nem muito baixa.
- (vi) Interesse do conteúdo: aumenta a atenção de transcritores e segmentadores; aumenta o nível de informatividade, pois garante uma fala mais espontânea.

Abaixo segue a Tabela 3.1, que descreve as características dos textos do *minicorpus*:⁷⁵

Características dos textos do subcorpus do C-ORAL-BRASIL etiquetado informacionalmente					
Texto	Situação	Informantes		Duração H:m:s	Nº de palavras
		M	F		
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
bfamdl02	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 história da família ao neto	1	1	00:07:23	1677
bfammn03	Narrativa de “causos” divertidos para a família*	3	3	00:07:08	1206

⁷⁵ Fonte: Raso e Mittman (2012, p. 220-221).

Características dos textos do subcorpus do C-ORAL-BRASIL etiquetado informacionalmente					
Texto	Situação	Informantes		Duração H:m:s	Nº de palavras
		M	F		
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
Total		28	27	03:58:36	31318
* presença de pequenas intervenções de um ou mais informantes que estavam fora da situação					

Tabela 3.1 – Características dos textos do *minicorpus*

Os nomes dos arquivos são construídos da seguinte forma: a primeira letra representa a língua do *corpus* (“b”); o grupo de três letras seguintes, o contexto da interação (“fam”, para contextos privados e “pub”, para contextos públicos); as duas letras finais correspondem à tipologia interacional (“mn”, para monólogos, “dl”, para diálogos, e “cv”, para conversações); os numerais em sequência correspondem à sua identificação na seção a que pertencem.

A tabela aponta que o *minicorpus* é composto de um total de 31.318 palavras em 3h 58min e 36s de gravação, totalizando de 55 informantes (28, do sexo masculino e 27, do feminino). As conversações têm 1h 7min e 28s de gravação, com 24 informantes (15 homens e 9 mulheres); os diálogos, 1h 45min e 28s de gravação, com 14 informantes (6 homens e 8 mulheres); e os monólogos, 1h 5min e 40s de duração e 17 informantes (7 homens e 10 mulheres).

Com relação ao número de informantes nos monólogos, ainda que haja mais de um participante na situação comunicativa, apenas um dos participantes é o informante central para a contagem do número de palavras do texto. Os eventuais interlocutores são os endereçados e contribuem tão-somente com pequenas interferências, normalmente para manter o fluxo discursivo.

Os números para o contexto e a tipologia interacional são os seguintes: 74% da amostra pertencem ao contexto privado/familiar, e os arquivos em contexto público representam 26% do total de palavras da amostra. Assim como em toda a parte informal do *corpus*, a divisão do *minicorpus* é balanceada: 31% são conversações; 36%, diálogos; e 33%, monólogos.

Deve-se observar que, neste *subcorpus*, a representatividade para o contexto público não é suficiente para efeitos de comparação com as situações familiares. O número de

palavras neste contexto equivale a 1/3 das palavras das conversações, 1/4 das dos diálogos, e 1/6 das palavras produzidas em monólogos.

3.3.2 A busca e organização dos índices:

A busca pelos índices modais no *minicorpus* foi feita manualmente por três anotadores. Os anotadores, duas estudantes de graduação e uma de pós-graduação, cumpriram duas tarefas, a partir do que já havia sido investigado na literatura e discutido sobre o conceito de modalidade e a tipologia: a primeira, buscar os índices nos textos, e a segunda tarefa a de classificá-los de acordo com os tipos epistêmicos, deônticos e dinâmicos.

Ainda que um objetivo futuro seja a comparação com dados de modalidade para a fala espontânea do italiano, que utilizam os valores alético, epistêmico e deôntico, decidiu-se não individualizar a modalidade alética⁷⁶ e proceder à classificação apenas dos dois últimos tipos. A partir da observação direta dos dados e nas discussões do grupo sobre os significados modais a serem aplicados, optou-se por uma abordagem subjetiva da modalidade. De fato, alguns estudiosos (von WRIGHT, 1951; LYONS, 1977⁷⁷, CRESTI, 2000; TUCCI, 2007) reconhecem o tipo alético e consideram-no em sua tipologização, porém, como Portner (2009, p. 10), argumenta: “[p]ode ser difícil determinar precisamente quais os usos de modais supostamente contados como aléticos, como em *é necessariamente verdade que...*, a leitura alética tende a se destacar. A motivação para a concepção estrita pode ser a assunção de que a modalidade alética é o tipo mais básico de modalidade, em termos do qual as outras variedades podem ser definidas, mais do que um tipo entre muitos.”⁷⁸, o que corrobora nossa posição.

A cada texto anotado, os resultados eram discutidos em reunião com a coordenadora do projeto. A busca e a classificação eram checadas e validadas e, apesar de não computado estatisticamente, foi apontado alto grau de acordo entre os anotadores. À medida que o processo de codificação avançava, foi decidido que o valor dinâmico seria incluído na tipologia, para cobrir os casos de habilidade/capacidade e os de volição/intenção.

⁷⁶ No entanto, é reconhecida a importância da modalidade alética para o estudo das ontologias, por exemplo.

⁷⁷ Segundo Portner (2009, p. 122-123), “A visão de Lyons, também, parece ser de que a modalidade na linguagem natural é simplesmente uma versão extremamente objetiva da modalidade epistêmica”.

⁷⁸ No original: “It can be hard to determine precisely which uses of modals are supposed to count as alethic, but if you put in the word true, as in *It is necessarily true that . . .*, the alethic reading tends to stand out. The motivation for the narrow conception may be an assumption that alethic modality is the most basic type of modality, in terms of which the other varieties may be defined, rather than just one type among many.” (PORTNER, 2009, p. 10).

Para os casos de ambiguidade dos verbos modais como “poder” e “dever” que permitem leituras tanto epistêmicas quanto deônticas (e para o “poder” também dinâmicas), o desempate foi feito pela análise do contexto e ainda pela utilização do software WinPitch (MARTIN, 2004) na oitava, uma vez que a análise perceptual e acústica do sinal sonoro permitem a desambiguação do valor semântico de uma mesma forma lexical.

O WinPitch Corpus é uma ferramenta fundamental para a análise de textos falados, porque permite o alinhamento preciso das informações linguísticas contidas num texto com seu correspondente sinal acústico. Assim, permite a identificação não só de sílabas como de enunciados inteiros, inclusive dos segmentos de unidades informacionais. Além do alinhamento transcrição-fala, o software possui outros recursos para a análise acústica como espectrograma, identificação da frequência fundamental, possibilidade de síntese da curva prosódica, entre outros. A análise dos arquivos de som se mostra fundamental para reconhecer, classificar e analisar os índices de modalidade, quando sugerem mais de um sentido.

Para o tratamento nesta tese, foram selecionados apenas os enunciados com um índice modal explícito e, ao final, destacamos 1.197 índices modais, distribuídos em 1.046 enunciados, excluídos os enunciados interrompidos. Após a seleção, os dados foram organizados em uma tabela com todas as ocorrências por arquivo de texto. Os enunciados foram codificados da seguinte forma:

- (a) Metadados, informações sobre cada arquivo: o nome de cada um; a tipologia interativa (monólogos, diálogos e conversações) e os participantes. Estes dados foram colhidos nos cabeçalhos de cada arquivo que fornecem os dados estatísticos sobre as categorias de sexo, idade, nível escolar, ocupação e papel comunicativo exercido pelos falantes, assim como informações relativas às atividades exercidas durante a gravação, os assuntos e os locais de gravação. Esse cabeçalho traz também comentários que podem conter observações julgadas importantes sobre a natureza situacional, para a compreensão do texto como um todo ou de parte(s) dele e observações linguísticas não acolhidas nos critérios de transcrição, mas consideradas relevantes em cada caso específico.

A estes metadados fornecidos pelo *corpus*, foram acrescentados a tipologia textual⁷⁹, o gênero textual e o assunto de cada interação, a fim de verificar a relação entre as escolhas dos itens modais e os tipos de texto em que estão inseridos. Estas variáveis não são parâmetros controlados pelo C-ORAL-BRASIL, nem pelo C-ORAL-ROM. Abaixo, um instantâneo das informações compiladas e codificadas sobre os metadados:

Arquivo	Tipologia interacional	Tipo textual	Gênero	Assunto	Informantes	S
<i>bfamcv01</i>	conversa�o privada	argumentativo	avalia�o/planejamento	campeonato de futebol	GIL	M
<i>bfamcv01</i>	conversa�o privada	argumentativo	avalia�o/planejamento	campeonato de futebol	GIL	M
<i>bfamcv01</i>	conversa�o privada	argumentativo	avalia�o/planejamento	campeonato de futebol	LUI	M
<i>bfamcv01</i>	conversa�o privada	argumentativo	avalia�o/planejamento	campeonato de futebol	LUI	M
<i>bfamcv01</i>	conversa�o privada	argumentativo	avalia�o/planejamento	campeonato de futebol	LUI	M
<i>bfamcv02</i>	conversa�o privada	narrativo	bate-papo	preparativos de casamento	RUT	F
<i>bfamcv02</i>	conversa�o privada	narrativo	bate-papo	preparativos de casamento	RUT	F
<i>bfamcv02</i>	conversa�o privada	narrativo	bate-papo	preparativos de casamento	TER	F
<i>bfamcv03</i>	conversa�o privada	narrativo	bate-papo	variados/jogo de sinuca	CEL	M
<i>bfamcv04</i>	conversa�o privada	descritivo	instru�o de jogo	Imagem e A�o	CEL	M
<i>bfamdI01</i>	di�logo privado	narrativo	sele�o de alimentos	compras do m�s	REN1	F
<i>bfamdI02</i>	di�logo privado	narrativo/descritivo	bate-papo/instru�o de uso de equipamento	variados/ organiza�o de equipamento	BEL	F

Tabela 3.2 – Informa es sobre os metadados

(b) Lemas e lexemas: identifica o do item modal de cada enunciado e suas respectivas formas. A Tabela 3.3 indica a forma de codifica o:

⁷⁹ A fim de classificar a tipologia textual, utilizo as categorias de Dolz e Schneuwly (1996/2004), divididas em NARRAR, RELATAR, EXPOR, ARGUMENTAR e DESCRIVER.

Lema	Lexema
achar	acho
dever	deviam
achar	acho
com certeza	com certeza
com certeza	com certeza
ir	vão
mesmo	mesmo
querer	queria
dar	dá
dever	deve
querer	querendo
esperar	espero

Tabela 3.3 – Informações sobre lemas e lexemas

- (c) Organização informativa: unidades informacionais em que se encontram o índice modal (Comentário, Comentários Múltiplos, Comentários Ligados, Tópico, Lista de Tópico, Parentético, Lista de Parentéticos, Introdutor Locutivo, Apêndice de Comentário e Apêndice de Tópico)⁸⁰. Abaixo segue a tabela com as etiquetas utilizadas:

Unidade informacional	Etiqueta
Comentário	COM
Comentários Múltiplos	CMM
Comentários Ligados	COB
Tópico	TOP
Lista de Tópicos	TPL
Parentético	PAR
Lista de Parentéticos	PRL
Introdutor Locutivo	INT
Apêndice	APC
Apêndice de Tópico	APT

Tabela 3.4 – Etiquetagem de unidades informacionais

⁸⁰ Na etiquetagem informacional é utilizada a etiqueta CMB, para Comentários Múltiplos Ligados. No entanto, para os propósitos de meu trabalho assumo os CMBs como COBs, pela opção de agrupar categorias semelhantes funcionalmente.

- (d) Valor modal correspondente: epistêmico, deôntico e dinâmico e seus respectivos subvalores modais. Para os subvalores, a nomenclatura está em inglês, tendo em vista que nos projetos de anotação da modalidade a língua inglesa é a língua ‘franca’.

O valor epistêmico: o significado epistêmico está relacionado com o grau de certeza de um conceptualizador sobre o material locutório enunciado e também se refere às noções de possibilidade, probabilidade e necessidade. Por exemplo:

O valor deôntico: normalmente está associado às convenções morais e sociais e, portanto, normalmente é usado para indicar obrigação, permissão ou proibição.

O valor dinâmico: relaciona-se à capacidade, habilidade e volição/intenção de um conceptualizador

A Tabela 3.3 indica os códigos utilizados:

Valores	Subvalores	Código
Epistêmicos		epi
	Conhecimento	knowledge
	Crença	belief
	Possibilidade	possibility
	Probabilidade	probability
	Necessidade	necessity
	Verificação	verification
Deônticos		deo
	Obrigação	obligation
	Permissão	permission
	Proibição	prohibition
	Necessidade	necessity
Dinâmicos		dyn
	Habilidade	ability
	Volição	volition

Tabela 3.5 – Etiquetagem de valores e subvalores modais

(e) *Part of speech* (PoS) do índice modal: verbos modalizadores⁸¹, verbos epistêmicos, verbos de compreensão e percepção, verbos de atitude proposicional, advérbio modal e locuções adverbiais, adjetivos e construções adjetivas, expressões modais, construções de futuro (futuro perifrástico, vamos+infinitivo, futuro do pretérito e imperfeito com função de retrospecto). A Tabela 3.4 indica as etiquetas:

PoS	Etiqueta
Verbos modalizadores	Vm
Verbos epistêmicos	Vepi
Verbos de compreensão/percepção	Vcomp
Verbos de atitude proposicional	Vap
Advérbios e locuções adverbiais	Adv
Adjetivos e construções adjetivas	Adj
Expressões modais	Em
Morfologia	
Futuro perifrástico	futperif
vamos+infinitivo	let's
Futuro do pretérito	futpret
Imperfeito	imperf

Tabela 3.6 – Etiquetagem de *Part of speech* (PoS) e índices morfológicos

(f) Composicionalidade: se há apenas um índice modal em cada enunciado (“s”, para simples); ou se há dois ou mais índices na mesma unidade informacional e mesmo enunciado (“me”, para mesmo enunciado); ou se há no mesmo enunciado e em unidades informacionais diferentes (“me/ud”, para mesmo enunciado e unidades diferentes).

A Tabela 3.7 traz um instantâneo da maneira como os dados referentes a enunciados, padrão informacional, composicionalidade, valor modal, PoS e subvalores modais estão organizados e codificados:

⁸¹ Sob a etiqueta verbos modalizadores, inclui não apenas os verbos “poder”, “dever” e “ter que”, mas todos aqueles que expressam modalidade epistêmica, deontica e dinâmica, mesmo que, a rigor, nem todos sejam enquadrados na categoria “verbos modais”.

Enunciado	UI/Padrão	comp	Valor modal	PoS	Subvalor modal
*GIL: [2] <ô /CNT= mas> /DCT= voltando à questão /COB= falando em [2]=EMP= e também falando em povo mascarado /COB= esse povo do Galáticos é muito palha /COB= eu acho que es nũ deviam mais participar /COM= e <tal> /UNC=\$	COM	me	epi	vepi	belief
*LUI: [5] eu acho não //COM=\$	COM	me	epi	vm	necessity
*LEO: [6] com certeza //COM=\$	COM	s	epi	vepi	belief
*LUI: [7] com certeza es nũ vão participar /COM= uai> //PHA=\$	COM	me	epi	adv	belief
No' /EXP= o Galáticos é mesmo /INT= todo mundo é <babaca> //COM=\$	COM	me	epi	adv	prediction
*EVN: [27] <eu queria ver a comunidade dele> lá /COB= ver que es falam /SCA= entre si /COB= assim /PAR= só pra eles mesmos //COM=\$	INT	s	epi	adv	belief
[35] mas é fechado /COB= nũ dá pra ir //COM=\$	COM	s	dyn	vm	volition
*GIL: [36] né / es deve meter o pau // \$	COM	s	epi	vm	possibility
[32] e &pr [2]=EMP= e principalmente /TOP= es tão querendo fazer /INT= campeonato /CMB= sem a gente /CMB= também /CMB=	COM	s	epi	vm	probability
então assim /INT= espero que /SCA= isso nũ seja /SCA= coisa pros times que jogam com a gente deixar de jogar com a gente //COM= \$	INT	s	dyn	vm	volition
*LEO: [35] como o Chub fala /INT= es vão pegar os caras que /i-COB= tipo /PAR= tavam reclamando /COB= e tal /COB= es vão pegar /COB= es vão pegar tentar fazer um negócio desse /COB= e eu aposto que cê vai ver os caras que já conhecem a gente há mais tempo /COB= tipo José [1]=SCA= Zé Mourinho /PAR= falando assim /INT= não /CMB_r= o /SCA= campeonato d' ocês é bem melhor //COM_r=\$	COM	s	dyn	vm	volition
	COB	me/ud	epi	futperif	prediction
	COB	me/ud	epi	futperif	prediction
	COB	me/ud	epi	futperif	prediction

Tabela 3.7 – Organização e codificação de dados

Em fase posterior, foram acrescentados, apenas por um anotador, dados referentes aos componentes utilizados para a anotação semântica da modalidade, como o *target*, a unidade informacional que contém o *target*, a polaridade e a *source of the modality* ou *holder* (BAKER et al., 2010). A metodologia utilizada na anotação dos dados será explicitada no capítulo 5 da Parte III deste trabalho, uma vez que, para tal, classifico apenas os índices lexicais, utilizo uma classificação mais refinada e um software específico para tarefas de anotação, o MMAX2 (MULLER; STRUBE, 2006).

Da tabela ainda consta um campo para observações sobre sutilezas de cada ocorrência, como, por exemplo, relações entre morfossintaxe e semântica, estratégias de polidez envolvidas, padrões modais etc.

As construções condicionais foram tabuladas em planilha específica, devido à particularidade dos dados a serem analisados. Além dos metadados, lemas e lexemas, e valor modal, foram acrescentados o padrão de organização informativa da construção, por exemplo, TOP/COM, e a distribuição da prótase e apódose correspondente ao padrão de cada enunciado.

Abaixo, uma amostra da tabela com as condicionais e suas variáveis:

Texto	Falante	Construção	Est. Info.	Estr. Sint.
bfamcv02	TER	[21] mas /=INP= gente velha /=TOP= já prometeu o [/1]=SCA= os presente /=TOP= <já /=SCA= pode> garantir que ganhou //=-COM=\$	TOP-COM	Pro-apo
bfamcv02	RUT	[23] se nũ morrer antes <deles> //=-COM=\$	COM	pro
bfamcv02	JAE	[27] <se for o> filho e ganhou /=TOP= <é questão da mãe> //=-COM=\$	TOP-COM	Pro-apo

Tabela 3.8 – Exemplo de organização dos dados das construções condicionais

3.4 Análise dos dados:

A partir da organização dos dados, foi realizada a descrição da modalidade no *minicorpus* etiquetado informacionalmente do C-ORAL-BRASIL. Foram analisadas as frequências de enunciados modalizados encontrados na amostra, bem como a sua distribuição por tipologia interacional. Além disso, foram descritas a frequência dos índices morfolexicais, inclusive em relação à unidade informacional de que faziam parte, e a frequência dos tipos de modalidade, em relação aos índices e a cada unidade informacional.

Em seguida, foi feita a análise qualitativa e quantitativa para cada um dos índices modais encontrados, em termos de: ambientes sintáticos, relação entre estrutura sintática e estrutura informacional, distribuição nos contextos e tipologias interacionais, tipologia textual

(narrativo, relato, expositivo, argumentativo, descritivo). Além disso, foram analisadas as variáveis: tipo de modalidade e seus subvalores, padrão de estrutura informacional (qual unidade informacional contém o marcador modal), e a composicionalidade (enunciado simples, dois índices na mesma unidade informacional, dois ou mais índices em unidades informacionais diferentes, mas no mesmo enunciado).

Neste capítulo, tratei dos procedimentos metodológicos que guiaram esta pesquisa. Na primeira seção, expus os pressupostos da Teoria da Língua em Ato, teoria esta *corpus-driven* e *corpus-based*, que define o enunciado como sua unidade mínima de referência pragmaticamente analisável. Elenquei e defini as unidades informacionais, textuais e dialógicas, e os padrões complexos de organização informacional, a saber, as Unidades de escansão, Comentários Múltiplos, Estrofes e Comentários Ligados. Em seguida, na segunda seção, apresentei a arquitetura do C-ORAL-BRASIL, *corpus* de fala espontânea do português brasileiro, especificamente em sua variedade mineira. Da parte informal deste *corpus* foi recortado um *minicorpus*, a partir de critérios de máxima qualidade, com 20 textos, privados/familiares e públicos, que tenta preservar a arquitetura do *corpus*. Descrevo as variáveis utilizadas para a análise dos dados (metadados, lemas e lexemas, estrutura informacional, composicionalidade, valores e subvalores modais e *PoS*) e, por último, explícito o tipo de análise a ser empreendida.

No próximo capítulo, sigo com a descrição da modalidade no C-ORAL-BRASIL, com a apresentação e discussão dos dados e posterior análise qualitativa de alguns dos índices lexicais que expressam o fenômeno em tela.

Capítulo 4

A MODALIDADE NO MINICORPUS DO C-ORAL-BRASIL I: resultados e análise

No capítulo anterior apresentei a moldura teórica da Teoria da Língua em Ato, o projeto C-ORAL-BRASIL e os procedimentos metodológicos utilizados para a análise da modalidade no *minicorpus*, extraído da parte informal do C-ORAL-BRASIL. No presente capítulo, descrevo os resultados para a modalidade na amostra de 20 textos deste *minicorpus* e analiso o comportamento de alguns dos índices lexicais que expressam modalidade, a saber, verbos modalizadores, verbos epistêmicos, advérbios e as construções condicionais do tipo “se *p*, então *q*”.

4.1 Frequência de enunciados modalizados na amostra

O *minicorpus* de análise possui um total de 31.465 palavras e 5.484 enunciados. Foi realizada uma busca manual em todos os enunciados para a extração dos enunciados que continham pelo menos um marcador de modalidade dentre os já elencados na seção de procedimentos metodológicos. Foram encontrados 1.197 índices marcadores de modalidade, sendo 109 referentes a construções condicionais⁸² e 1.088 referentes aos outros itens modais, distribuídos em 1.044 enunciados, o que corresponde a **19,03%** do total de todos os enunciados da amostra. Como mostra o gráfico abaixo:

⁸² Conforme destacado no capítulo de procedimentos metodológicos, as construções condicionais recebem tratamento diferenciado, porque “assumem uma estrutura textual, que difere das estratégias lexicais, e possuem uma distribuição prosódico-pragmática peculiar.” (CORTES; MELLO, 2013, p. 1).

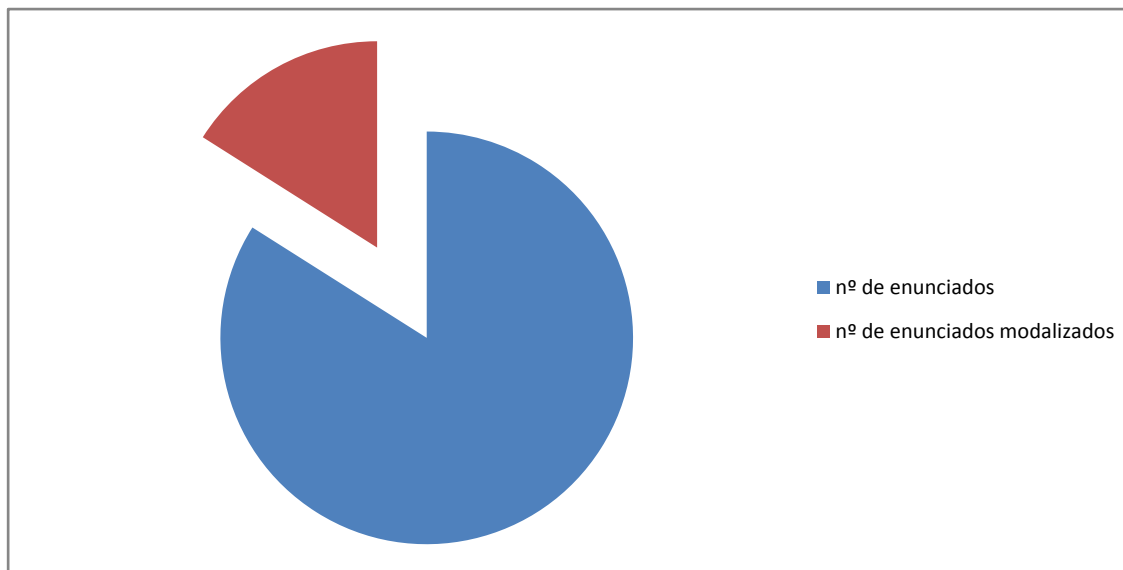


Figura 4.1 – Frequência de enunciados modalizados no *minicorpus*

Em termos de composicionalidade⁸³, os índices estão assim organizados: dos 1.088 marcadores, 825 aparecem em enunciados com apenas um índice modal; 202 estão combinados (dois ou mais índices) em uma mesma unidade informacional; e 61 deles correspondem a mais de um índice modal em um mesmo enunciado, mas em unidades informacionais diferentes. A combinação de um ou mais índices em uma mesma unidade informacional exige que seja analisado o valor global do enunciado, respeitando a hierarquia de valores de cada um dos modais.

Na amostra, há uma série de valores modais combinados (epi-epi; epi-deo; epi-dyn; epi-epi-epi; epi-deo-epi; epi-epi-dyn; deo-deo-; deo-epi; dyn-epi etc.), sendo que o valor global predominante nestes enunciados é o epistêmico. Vejamos alguns exemplos:

(4.1) *SIL: [92] <**sinceramente**> que eu nũ **sei** //COM=\$ (bfamd104)

(4.2) *HEL: [97] o que que <**pode**> <que que nũ **pode**> //COM=\$ (bfamcv04)

(4.3) *JOR: [79] eles **queriam** me dar um dinheiro por fora mas /=SCA= a gente também **tem que** ter o nossos momento de descanso //COM=\$ (bfammn06)

(4.4) *EVN: [169] vão> vão /=CNT= <eu **acho** que **tem que** ser esses> //COM=\$ (bfamcv01)

⁸³ Apresento a composicionalidade dos 1.088 marcadores de modalidade, excluídas as construções condicionais que, como já mencionado, estão organizadas de maneira particular.

Em (4.1), temos dois índices modais epistêmicos “sinceramente” e “sei”, com subvalores de crença e conhecimento, e o valor global do enunciado é epistêmico. A ocorrência em (4.2) traz um enunciado com valor global deôntico: há uma combinação de dois itens modais deônticos, com significado de permissão e impedimento. O exemplo (4.3) combina um índice dinâmico “queriam” e um deôntico “tem que”, dada a hierarquia, o valor global é dinâmico. Por último, em (4.4), a unidade informacional de Comentário contém dois marcadores modais, um epistêmico de crença “acho” e uma obrigação deôntica “tem que”, e seu valor global é epistêmico.

4.2 Enunciados e tipologia interacional

Na amostra da parte informal do *corpus*, a extração automática dos enunciados nos dá um resultado de 5.484 enunciados simples e complexos, distribuídos da seguinte forma: 4.126 enunciados privados (1.404 conversações, 1.870 diálogos e 852 monólogos) e 1358 enunciados públicos (635 conversações, 581 diálogos e 142 monólogos). A Tabela 4.2 mostra os números para a distribuição de enunciados em termos de tipologia interacional e contexto familiar ou público. Destacam-se o número total de enunciados por tipologia e contexto e a frequência deles comparados aos números totais do *minicorpus*:

	Monólogos	Diálogos	Conversações	Total modal	Total amostra	Freq. (%)
Privados	174	332	291	797	4126	19,3
Públicos	46	111	90	247	1358	18,1
Total modal	220	443	381	1044	5484	19,03
Total amostra	994	2451	2039	5484		
Freq (%)	22,1	18	18,6			

Tabela 4.1 – Frequência de enunciados modalizados x tipologia e contexto interacional

O gráfico abaixo dá a perspectiva da distribuição:

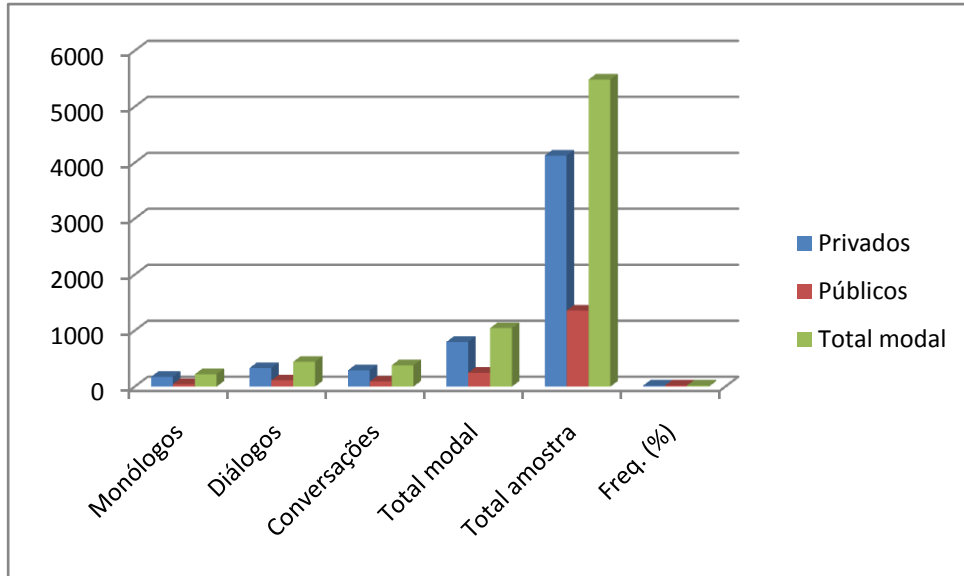


Figura 4.2 – Frequência de enunciados por tipologia e contexto interacional

Observa-se que a tipologia interacional tem influência sobre o percentual de enunciados modalizados: comparados aos diálogos e conversações (18,1% e 18,6%, respectivamente), a tipologia monológica possui mais enunciados modalizados (22,2%).

Na Tabela 4.2, estão indicados o número de índice modais encontrados em cada contexto e tipologia interacional e sua frequência em relação ao número total de índices modalizados:

	Monólogos	Diálogos	Conversações	Total modal	Freq (%)
Privados	207	365	334	906	75,6
Públicos	64	130	97	291	24,3
Total modal	271	495	431	1197	
Freq (%)	22,6	41,3	35,9		

Tabela 4.2 – Frequência de índices modais por contexto e tipologia interacional

Abaixo, o gráfico para ilustrar essa distribuição:

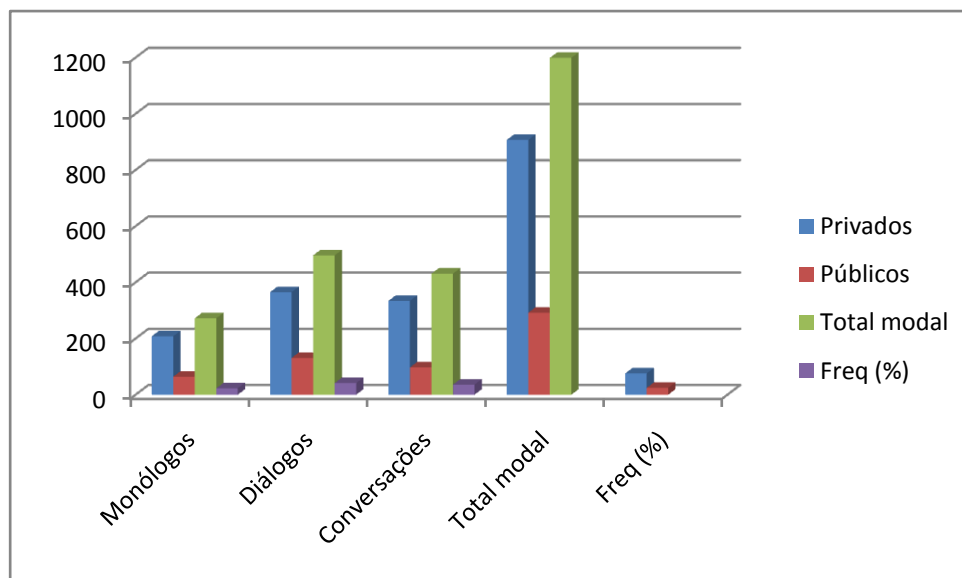


Figura 4.3 – Frequência de índices modais por tipologia e contexto interacional

A grande diferença entre a frequência dos índices em contextos privados e público deve-se ao número bastante superior de textos privados (15) em relação ao público (5). Esta comparação, portanto, é meramente ilustrativa. No que diz respeito à tipologia interacional, os índices estão assim distribuídos: para os monólogos, 22,6% do total de índices; para os diálogos, 41,3%; e para as conversações, 35,9%.

Assim, vemos que ao serem apresentados apenas os números de índices modais, deslocados do número de enunciados e de uma base de referência de toda a amostra, inverte-se o peso das tipologias: os monólogos apresentam menos modalizações em comparação com as trocas dialógicas (diálogos e conversações).

4.3 Frequência dos índices morfolexicais de modalidade

Foram encontradas 1.197 ocorrências de itens lexicais e gramaticais que expressam modalidade, divididos em verbos modais, verbos epistêmicos e de atitude proposicional, advérbios e locuções adverbiais, adjetivos e construções adjetivas, expressões modais, futuro perifrástico, futuro do pretérito, imperfeito (com valor de futuro do pretérito) e construções condicionais. A Tabela 4.3 abaixo mostra os números para cada estratégia modalizadora:

Estratégias modalizadoras	nº de enunciados
verbos modalizadores	414
verbos epistêmicos / atitude proposicional	223
advérbios e locuções adverbiais	92
adjetivos e construções adjetivas	22
expressões modais	30
futuro perifrástico	243
vamos+infinitivo (~let's)	18
futuro do pretérito	37
imperfeito (~futuro do pretérito)	9
construções condicionais	109

Tabela 4.3 – Estratégias modalizadoras

O número de ocorrência de índices lexicais é de 783 e corresponde a 65,4% da amostra, enquanto o de índices gramaticais é de 416 e corresponde a 34,6%. O número de índices gramaticais somados é idêntico ao número de verbos modalizadores.

A estratégia modalizadora empregada mais frequentemente são os verbos modalizadores (aqui, como já explicitado, incluem todos os verbos que expressam modalidade epistêmica, deôntica e dinâmica), correspondendo a 34,6% do total de ocorrências. As duas outras estratégias mais frequentes são o futuro perifrástico (20,2%) e os verbos epistêmicos e de atitude proposicional (18,5%). A Figura 4.4 demonstra a participação de cada uma das estratégias utilizadas na amostra:

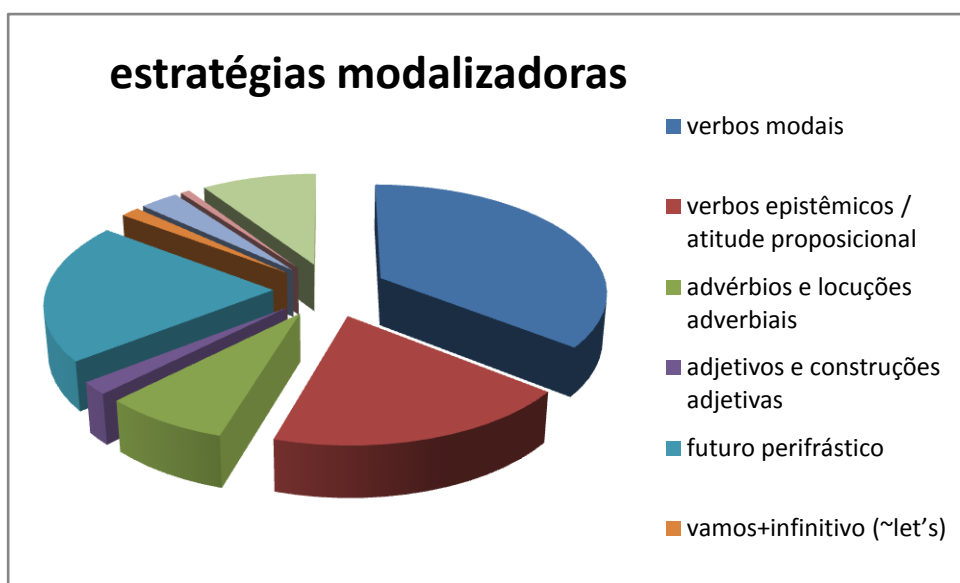


Figura 4.4 – Distribuição das estratégias modalizadoras na amostra (%)

Importante destacar que na língua portuguesa o futuro do pretérito, apesar de ser uma forma do modo indicativo, representando um futuro retrospectivo, funciona como uma forma condicional, exatamente como em outras línguas românicas como o francês e o italiano. Dessa forma, as construções condicionais (do tipo “se p , então q ”) e o futuro do pretérito (inclusive na forma temporal do imperfeito) consistem em 12,9% do total de enunciados modalizados.

4.3.1 Distribuição dos índices em relação à unidade informacional

Nesta seção, apresento a distribuição dos marcadores de modalidade em relação à articulação informacional do enunciado.

Segundo Tucci (2007), apenas algumas unidades podem conter um índice modal: o Comentário, o Tópico, o Parentético e o Introdutor Locutivo. No *minicorpus* em análise, além destas unidades informacionais centrais já apontadas nos dados do italiano, ainda aparecem como modalizados as unidade de Lista de Tópicos, Lista de Parentéticos, Comentários Múltiplos e Comentários Ligados. Destaco que, apesar de termos quatro ocorrências de índices na unidade de Apêndice de Comentário e uma na de Apêndice de Tópico, os números foram computados na unidade imediatamente anterior que integra, na medida em que se constituem como ecos ou informação atrasada de itens contidos no Comentário ou Tópico. A Tabela 4.4, a seguir, mostra a frequência relativa das unidades informacionais centrais na amostra:

Unidades textuais	Amostra	# índices modais	%
Comentário	5484	1027	18,7
Tópico	503	22	4,3
Parentético	124	17	13,7
Introdutor Locutivo	223	22	10,3

Tabela 4.4 – Frequência relativa das unidades informacionais na amostra⁸⁴

Do total de 1.088 ocorrências de índices modais, 1027 delas se realizam em unidade de Comentário, seja o Comentário-ele mesmo (820 ocorrências) ou unidades de Comentário Múltiplo (94 ocorrências), Comentários Ligados (109 ocorrências). As três outras unidades que carregam o marcador modal — Tópico, Parentético, Introdutor Locutivo — participam de maneira semelhante na caracterização da modalidade na amostra, em números absolutos, com 22, 17 e 22 ocorrências, respectivamente. Os gráficos em 4.5 e 4.6 apontam o percentual absoluto e relativo de unidades modalizadas, respectivamente:

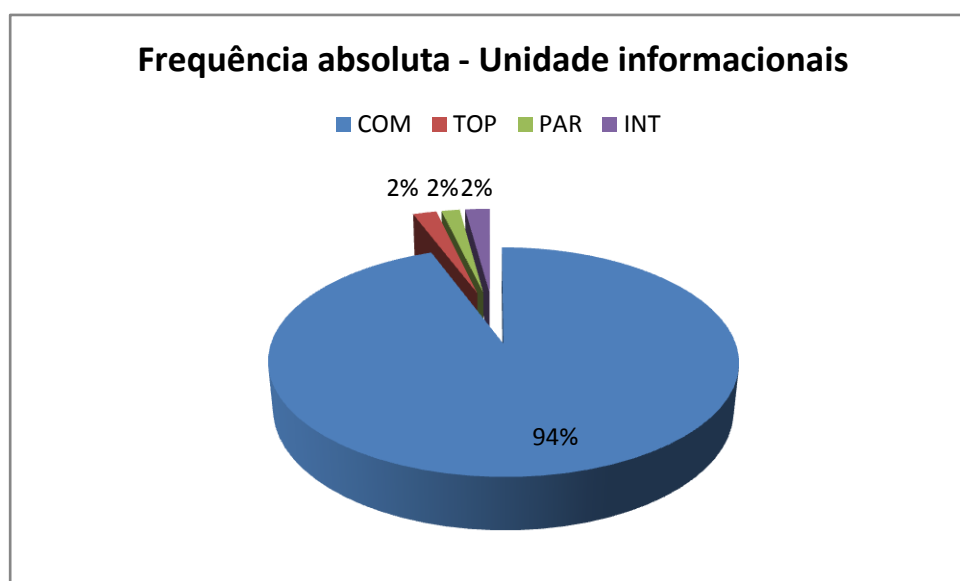


Figura 4.5 – Frequência absoluta das unidades informacionais em relação aos índices modalizados

⁸⁴ A tabela foi montada a partir da base IPIC, plataforma de busca que fornece os dados comparativos do C-ORAL-ROM italiano e seu *minicorpus* e o *minicorpus* (parte informal) do C-ORAL-BRASIL, em relação a dados gerais, tipos interacionais, contextos comunicativos, e unidades informacionais. Disponível em: <http://lablita.dit.unifi.it/ipic/>. Último acesso em: 16 mar. 2014. Trago apenas os números que me interessam nesta análise. Para fins de análise, nos números para Comentário, estão incluídos enunciados e estrofes, uma vez que os números da unidade COM contêm os CMMs e os COBs. Computo igualmente para o Tópico e o Parentético os números da Lista de Tópicos (3) e Lista de Parentéticos (3), respectivamente.

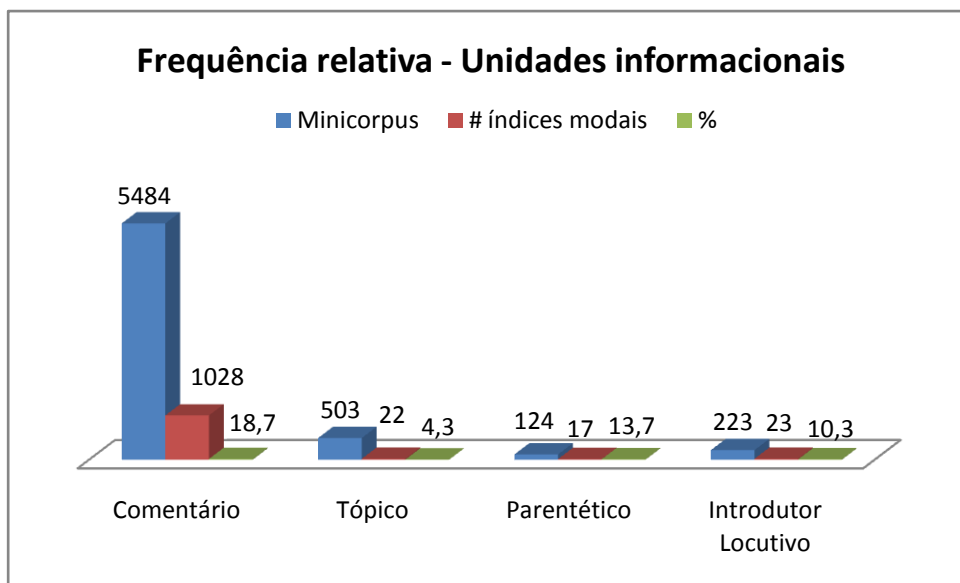


Figura 4.6 – Frequência relativa das unidades informacionais no *minicorpus*

Comparados os dois gráficos, notamos que o Comentário, tanto em termos absolutos quanto em termos relativos, é a unidade mais frequentemente modalizada. As outras unidades, em números absolutos, são modalizadas, em média, em 2% dos enunciados, já em termos relativos, o quadro se modifica: o Parentético figura como a segunda unidade mais modalizada (13,7%), seguida do Introdutor Locutivo (10,3%) e, em frequência bem menor, do Tópico (4,3%).

4.4 Frequência da modalidade quanto aos tipos modais:

Trato agora da distribuição dos valores modais (epistêmicos, deônticos e dinâmicos) em relação ao número de enunciados, à estrutura informacional e à ocorrência das estratégias modalizadoras. A distribuição dos valores modais na nossa amostra é a seguinte:

tipo de modalidade	nº de enunciados
epistêmica	904
deôntica	190
dinâmica	103

Tabela 4.5 – Tipos de modalidade

De um total de 1.197 índices que expressam a modalidade, o epistêmico é o valor modal mais amplamente utilizado em 76,6% das ocorrências. As modalidades deontica e dinâmica são realizadas em percentual bem menor, com 14,9% e 8,4% das ocorrências. Como mostra a figura a seguir:

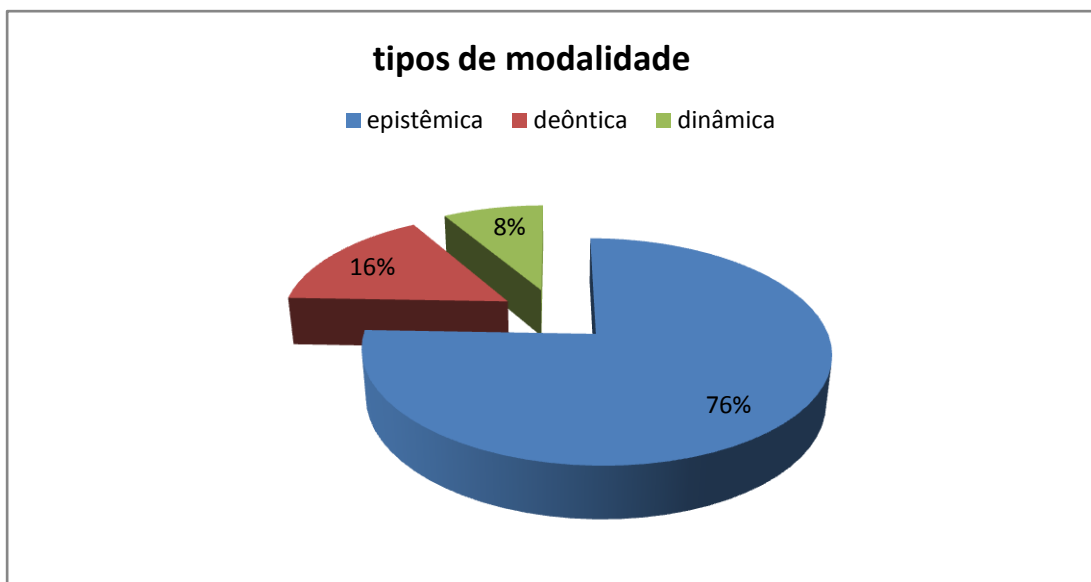


Figura 4.7 – Distribuição dos valores modais (%)

4.4.1 Frequência de valor modal por índice lexical:

No que diz respeito à frequência dos valores modais por índice lexical, temos:

Estratégias modalizadoras	epi	deo	dyn
verbos modalizadores	128	188	85
verbos epistêmicos	210	0	0
verbos de atitude proposicional	26	0	0
advérbios e locuções adverbiais	94	0	0
adjetivos e construções adjetivas	19	1	0
expressões modais	28	1	1
futuro perifrástico	244	0	0
vamos + infinitivo (~ construção <i>let</i>)	0	0	17
futuro do pretérito	37	0	0
imperfeito	9	0	0
condicionais	109	0	0

Tabela 4.6 – Estratégias modalizadoras e tipos de modalidade

O número de índices lexicais é de 781 e corresponde a 65,7% de todos os índices, enquanto o de índices gramaticais é de 416 e corresponde a 34,6% dos modais. O número de itens gramaticais modais é praticamente idêntico ao número de verbos modalizadores.

A tabela indica que, independentemente do valor, são os verbos modalizadores a estratégia preferencial para a expressão da modalidade, num percentual correspondente a 33,5% do total de índices, com destaque para o tipo deôntico, em que esta estratégia predomina: 46,9% do uso de verbos modais correspondem a este valor; 31,9%, ao valor epistêmico; e 21,2%, ao dinâmico.

A modalidade epistêmica é a mais utilizada, como já visto, e está associada a todas as estratégias modalizadoras. Mello e colaboradores (2013), em análise quantitativa da distribuição dos marcadores de modalidade na fala espontânea do português brasileiro, construíram um gráfico em que quanto mais vetores estivessem associados a um determinado nó, mais representativo seria este nó em relação ao parâmetro analisado, no caso a relação entre a tipologia modal e os lemas modais. Desta forma, demonstraram a tendência para o uso, no português brasileiro, do tipo epistêmico e apontaram que este valor tem uma taxa de associação a diferentes índices modais muito mais elevada do que os tipos deôntico e dinâmico.

Os valores deôntico e dinâmico são realizados predominantemente pelos verbos modalizadores, com a ocorrência de uma expressão modal com valor deôntico e uma, para o valor dinâmico. Construções do tipo “vamos+Vinf” (também chamada de construção *let* ou construções hortativas), com 17 ocorrências, representam o futuro como volição/intenção, subvalores associados ao valor dinâmico.

4.4.2 Relação entre o valor modal e as unidades informacionais

Nesta seção, apresento os resultados para a frequência do valor modal em relação às unidades informacionais. Já deixamos claro que o escopo da modalidade não é o enunciado, mas a unidade informacional. Em um mesmo enunciado pode ocorrer mais de um elemento veiculador de modalidade, e a modalidade só se estende ao enunciado inteiro, se ele for do tipo simples, quer dizer, só composto pela unidade de Comentário.

A Tabela 4.7 traz os números para os valores modais nas unidades informacionais de Comentário, Tópico, Parentético e Introdutor Locutivo:

Unidades informacionais	epi	deo	dyn
COM	746	185	96
TOP	19	1	2
PAR	15	1	1
INT	16	3	3

Tabela 4.7 – Distribuição dos valores modais nas unidades informacionais

As unidades informacionais têm comportamento diverso e não-uniforme no que diz respeito à realização do valor modal. Para o Comentário, não há restrição quanto ao valor modal, a unidade assume significados epistêmicos, deônticos e dinâmicos. O Tópico, ao contrário, realiza principalmente o tipo epistêmico (86,3%) e, na amostra, assumiu o valor dinâmico em duas ocorrências e em uma ocorrência do tipo deôntico, correspondente ao item modal contido na unidade de Apêndice de Tópico que a integra. A unidade de Parentético tem comportamento semelhante: realiza, em sua maioria, significados epistêmicos (88,2%), e assume em apenas uma ocorrência o valor deôntico e o dinâmico. Quanto ao Introdutor Locutivo, parece apresentar, da mesma forma que o Comentário, a possibilidade de maior realização de todos os significados modais (73,9% para os epistêmicos e 13% para os deônticos e dinâmicos). Abaixo, o gráfico representa a distribuição absoluta das unidades informacionais relativo a cada valor modal:

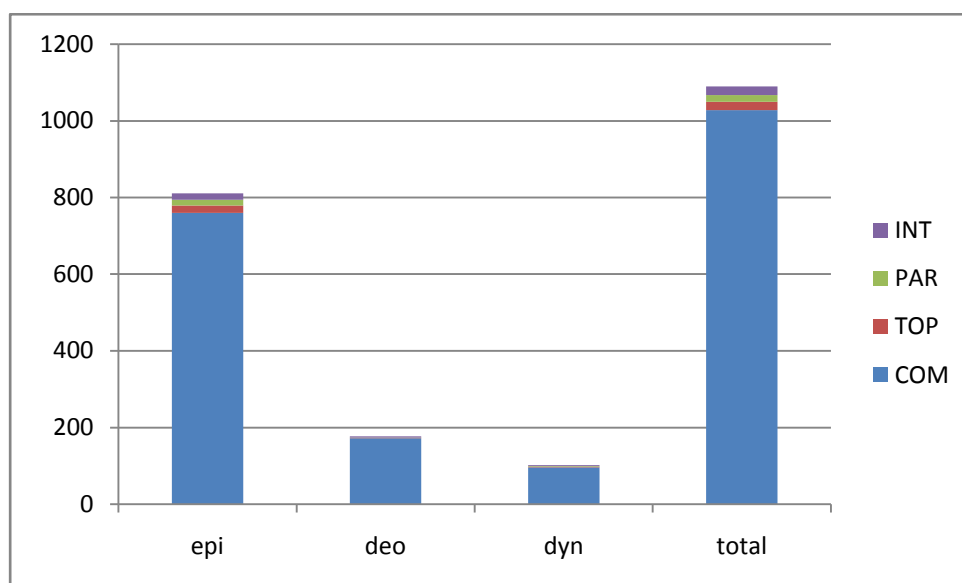


Figura 4.8 – Distribuição absoluta de unidades informacionais por valor modal

Abaixo, na Figura 4.9, a distribuição relativa das unidades no que diz respeito a cada valor modal, com base no número total de ocorrências:

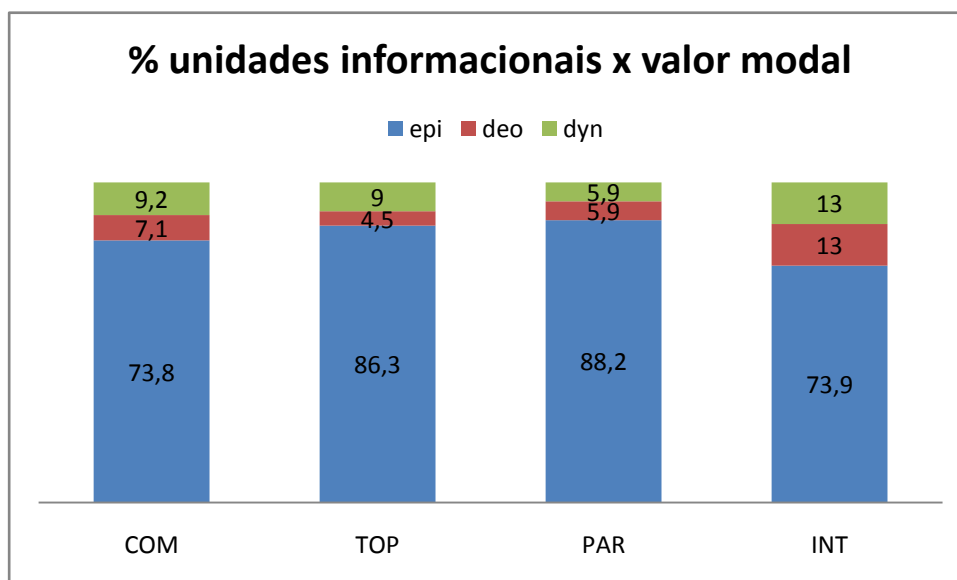


Figura 4.9 – Frequência relativa de unidades informacionais por valor modal

Segundo Tucci (2007, p. 270), em análise dos dados da modalidade para o italiano, o fato de as unidades de Comentário e Introdutor Locutivo poderem realizar todos os valores modais sem qualquer restrição não é surpreendente. O Comentário representa a unidade suficiente e necessária para a interpretabilidade de um enunciado e, da mesma forma, o Introdutor Locutivo, que introduz um discurso reportado, um elenco, uma comparação, uma exemplificação (no português brasileiro, introduz uma opinião ou consideração), assinalando a modalidade e suspendendo a operação ilocutiva.

Sobre o Tópico e o Parentético, Tucci (2007, p. 271-273) faz considerações sobre as duas unidades no que diz respeito à realização majoritariamente das modalidades epistêmica e alética e à restrição do uso do tipo deontico. De acordo com a autora, esta restrição está intimamente associada à sua função informativa.

Nas próximas seções, passo a analisar os dados para alguns dos índices lexicais marcadores de modalidade.

4.5 Análise de índices marcadores de modalidade:

Os significados modais no português brasileiro, como já apontado, podem ser expressos por vários meios lexicais, morfológicos e sintáticos. Esta seção apresenta a análise de quatro categorias lexicais – verbos modalizadores, verbos epistêmicos, advérbios modais – e uma categoria gramatical – as construções condicionais.

4.5.1 Os verbos modalizadores:

A grande maioria dos estudos sobre a categoria da modalidade se dedica à análise dos verbos modais auxiliares ou semi-auxiliares, como os de Papafragou (1998, 2000), Cornillie (2005, 2007), van der Auwera and Plungian (1998), Bybee et al. (1994), Almeida (2010) para nomear apenas uns poucos. Para a língua portuguesa, destaco os trabalhos de Oliveira (2003), Gonçalves (2004), Galvão (2000), Gonçalves e Galvão (2001), Salomão (2008).

4.5.1.1 Frequência dos verbos modalizadores:

Como apontado na seção 4.4.1, os verbos modalizadores são a estratégia preferencial utilizada para a expressão da modalidade. Relembrando, ilustro com um excerto da Tabela 4.6:

Estratégias modalizadoras	epi	deo	dyn
verbos modalizadores	128	188	85

Foram identificados na amostra, portanto, 401 verbos modalizadores, isto é, todos os verbos modais auxiliares, semimodais e verbos plenos que expressam modalidade. Abaixo, elenco todos os lemas encontrados e suas respectivas ocorrências:

verbos modalizadores	# ocorrências	%
adiantar	3	0,8
aguentar	2	0,5
conseguir	11	2,7
dar	30	7,5
dever	31	7,7
esperar	3	0,8
gostar	1	0,3
parecer	10	2,4
poder	132	32,9
precisar	17	4,2
querer	65	16,2
ter que	93	23,2
valer	3	0,8

Tabela 4.8 – Distribuição dos verbos modalizadores

Desta lista, observa-se a grande ocorrência de verbos que figuram em listas de várias línguas, como “poder”, “dever”, “ter que”, “precisar”, “querer” e “parecer”. No entanto, há formas outras que normalmente não são consideradas modais canônicos como os verbos “dar”, “adiantar”, “valer”, e “aguentar”, que parecem ser usos especializados do português brasileiro.⁸⁵ O gráfico abaixo mostra a distribuição dos itens:

⁸⁵ Restrinjo-me à variante brasileira do português, uma vez que não tenho dados suficientes para fazer afirmações sobre as outras variantes, seja a europeia ou as africanas.

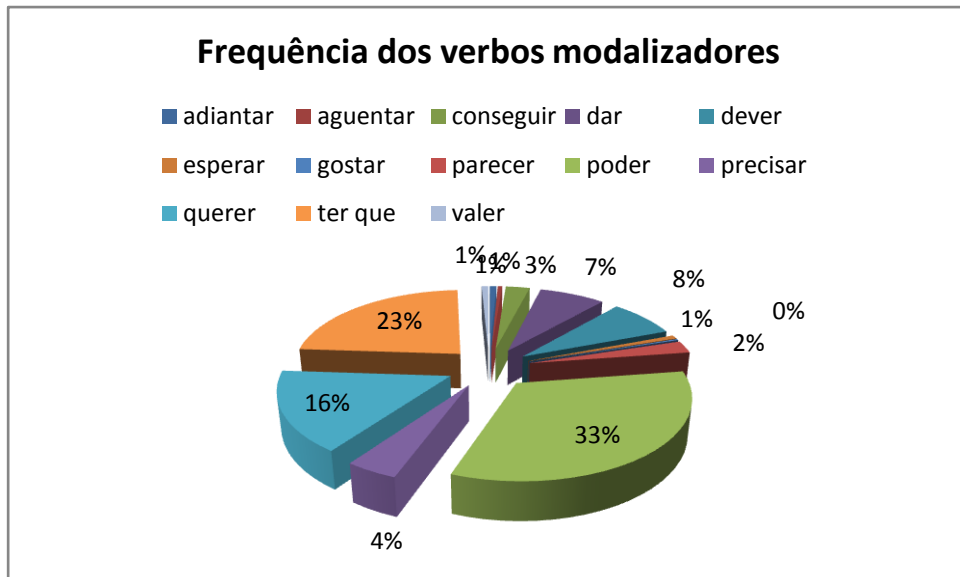


Figura 4.10 – Frequência dos verbos modalizadores

A seguir, apresento alguns exemplos de verbos que, por um viés metafórico, permitem a leitura como marcador modal:

(4.1) Adiantar

(a) [217]aí não **adianta** /=CMM= nũ pode desmanchar //=-CMM=\$ (*bpubdl01*)

(4.2) Conseguir

(b) *BRU: [383] <cê **conseguiu** ser> pior //=-COM=\$ (*bfamcv04*)

(c) *KAT: [32] **conseguiu** marcar /=CMM= não /=CMM= né //=-CMM=\$

(4.3) Dar (para)

(d) *REN: [266] **dá** pa fechar o jogo //=-COM=\$

(e) *LUZ: [210] <Nossa> /=EXP= esse negócio de terra nũ **dá** não //=-COM=\$

(f) [168] dois filho /=TOP_r= **dá** pa rir e chorar //=-COM_r=\$

(4.4) Esperar

(g) então assim /=INT= **espero** que /=SCA= isso nũ seja /=SCA= coisa pros times que jogam com a gente deixar de jogar com a gente //=-COM= \$ (*bfamcv01*)

(h) *REN: [518] **espero** que eu não tinha [/1]=SCA= tenha perdido //=-COM=\$ (*bfamdl01*)

(4.5) **Parecer**

(i) [307] ele <tem setenta-e-seis> metros quadrado me **parece** // =COM=\$ (bpubdl01)

(4.6) **Valer**

(j) *FLA: [444] não **vale** mais chorar por ele // =COM=\$

Destaco, na amostra, os números para as ocorrências dos principais verbos que dizem respeito às modalidades epistêmica, deôntica e dinâmica — “poder”, “dever”, “ter que” “querer” e “conseguir”:

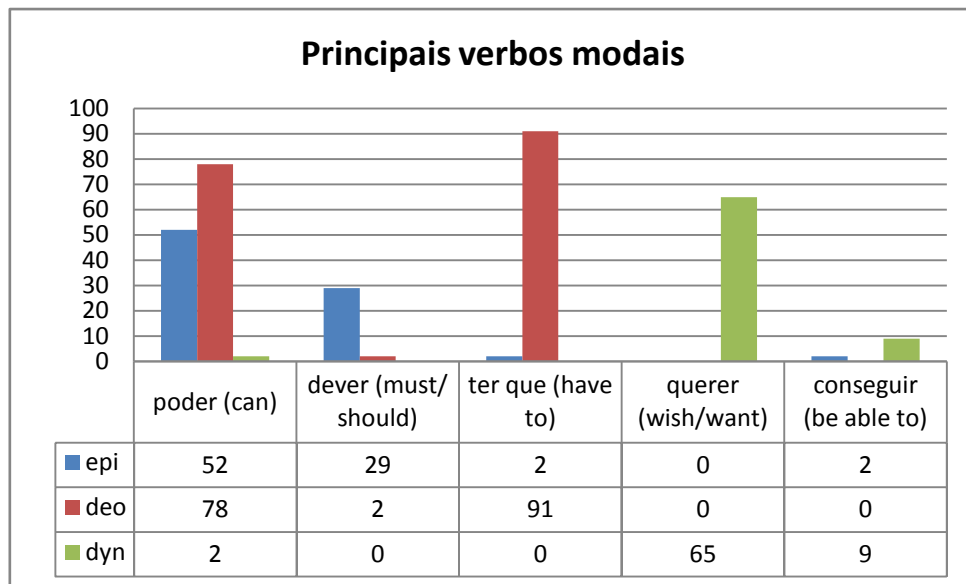


Figura 4.11 – Ocorrência dos principais verbos modais / tipo de modalidade

Observamos que o verbo “poder” é mais usado que o verbo “dever” como estratégia de modalidade, com 132 ocorrências comparadas a 31 ocorrências. Necessário salientar que o uso de “ter que” (93 ocorrências, 2 em sentido epistêmico e 91 em sentido deôntico), no PB, como veiculador de modalidade deôntica, indica que o sentido de obrigação está quase totalmente vinculado a ele. O verbo “querer” assume valor dinâmico de volição e o “conseguir” indica habilidade/capacidade dinâmica em 9 casos e possibilidade epistêmica em 2.

Na Tabela 4.9, os números de ocorrências de cada verbo e a frequência relativa aos verbos modalizadores e aos valores modais totais:

Verbos modais	epi	deo	dyn
poder (<i>can</i>)	52	78	2
dever (<i>must/ should</i>)	29	2	0
ter que (<i>have to</i>)	2	94	0
querer (<i>wish/want</i>)	0	0	65
conseguir (<i>be able to</i>)	2	0	9
total	85	174	76
% verbos modalizadores	20,5	42,1	18,4
% valor modal	9,4	91,5	73,7

Tabela 4.9 – Distribuição e frequência dos principais verbos modalizadores

Observa-se que a partir do cálculo da porcentagem para cada um dos valores modais, tomando em consideração todos os enunciados estes verbos em destaque assumem o valor deôntico (91,5%) e dinâmico (73,7%), contra 9,4% de epistêmicos. Em relação aos verbos modalizadores, os deônticos continuam a predominar (42,1%), seguido em números próximos dos epistêmicos (20,5%) e dos dinâmicos (18,4%). Conforme o gráfico em 4.12:

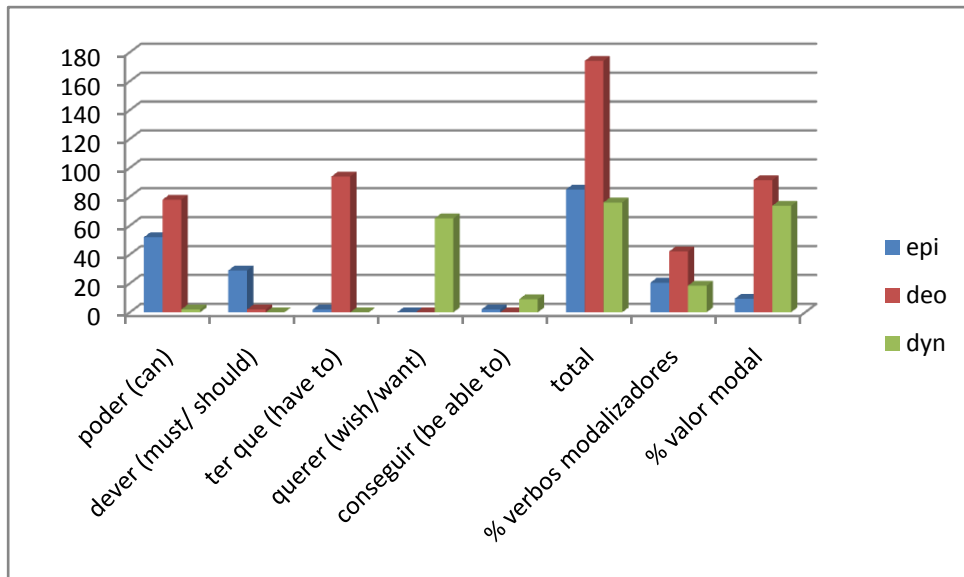


Figura 4.12 – Frequência relativa dos valores modais dos principais verbos modalizadores

A baixa ocorrência de “dever” com valor deôntico, relacionado com o sentido de uma obrigação fraca, e a sua maior ocorrência com o valor epistêmico, indica que este verbo, conforme Mello et al. (2010, p. 119):

[d]e típico modalizador deôntico (por conter a noção de obrigação, associada à dívida), o verbo passou a frequente modalizador epistêmico, veiculando menos a noção de obrigação do que a de possibilidade ou a de probabilidade. A noção de obrigação aparece fracamente na semântica do verbo, porque a avaliação de probabilidade (epistêmico) deriva de uma avaliação baseada na necessidade (deôntica, neste caso) das relações entre as coisas. É notável a ocorrência dessa nova acepção (epistêmica), mais frequente que a acepção de raiz (deôntica).

O espaço deixado pelo verbo “dever” em usos de obrigação deôntica tem sido ocupado pelo semimodal “ter que” como confirmam os números.

Em termos de distribuição dos verbos nas unidades informacionais, como estratégia mais produtiva, podem ocupar qualquer uma das UIs modalizáveis, com destaque, mais uma vez, para o grande número de ocorrências na unidade de Comentário (315), o que corresponde a 78,6% de todas as ocorrências. Se ao Comentário, acrescentamos as ocorrências para APC, CMM e COB, este número sobe para 390, 97,3% do total de ocorrências da amostra. É importante notar que todos os significados epistêmicos são realizados nesta unidade. Nas unidades de Tópico, Apêndice de Tópico, Parentético e Introdutor Locutivo, temos a realização de valores deônticos (tanto permissão quanto obrigação) e dinâmicos (volição).

4.5.2 Os verbos epistêmicos:

Os verbos de caráter epistêmico ou verbos de crença funcionam, segundo Venier (1991, p. 68 *apud* TUCCI, 2007, p. 172), como “sinais, para manifestar no ouvinte o grau de confiabilidade conferido pelo falante à proposição e para manter uma função sinalizadora também quando o enunciado que a contém vem reportado”⁸⁶. Além disso, desenvolvem sentidos atitudinais e interacionais, podem estar configurados sintaticamente de variadas maneiras, permitindo diferentes complementos, a própria omissão de complemento e, informacionalmente, a parentetização.

Vejamos ocorrências do Corpus do Português (DAVIES; FERREIRA, 2006) com o verbo de crença “considerar”, a fim de apontar nuances de com conceptualização no uso deste tipo de verbo:⁸⁷

⁸⁶ Tradução minha para “segnali, di manifestare all’ascoltatore il *grado di attendibilità assegnato dal parlante* alla proposizione e di mantenerne una funzione segnaletica anche quando l’enunciato che li contiene viene riportato” (VENIER, 1991, p. 68 *apud* TUCCI, 2007, p. 172).

⁸⁷ Os exemplos foram coletados do Corpus do Português, esse é um corpus de referência da língua portuguesa, projeto desenvolvido pelos professores Mark Davies (BYU) e Michael J. Ferreira (Georgetown University), com mais 45.000.000 de palavras, e um total de 57.000 textos do séc. XIV ao séc. XX. Esse corpus permite o cruzamento de dados e distribuição de palavras, frases e construções por **registro** (oral, ficção, jornalístico e

(4.7) os direitos com facilidade ou tem ciúme? João Ubaldo - Realmente não gostei de O Sorriso do Lagarto, mas gostei da adaptação de Sargento Getúlio para o cinema. Não tenho ciúme algum. Trata-se mesmo de outra obra, a obra do cineasta. iBEsp_242### **15 de outubro de 1997 Nahas se considera “um bode expiatório”** Estado - O sr. esperava esta sentença da Justiça? Naji Nahas - Não esperava de jeito nenhum. É mais uma violência, uma discriminação odiosa, entre as muitas que tenho sido vítima desde 89. Estado - Se o sr. é a vítima, quem (CDP:19Or:Br:Intrv:ISP)

(4.8) Nahas - Confio na Justiça e sei que esta sentença está sujeita a revisão. Terminando meu trabalho, eu volto porque quero me apresentar para a apelação, aproveitando o momento para esclarecer todo esse assunto e mostrar definitivamente quem são os culpados. **Sou a vítima. Estado - O sr. se considera um bode expiatório? Nahas - Sou. E o mais sofrido do Brasil.** iBEsp_244### 18 de outubro de 1997 Manoel de Barros faz do absurdo sensatez Estado - Como surgiu seu amor pelas coisas sem importância? Manoel de Barros - Quando eu era jovem, fiz uma longa viagem pela Bolívia. (CDP:19Or:Br:Intrv:ISP)

Em (4.7), confirmado por (4.8), tem-se um caso de discurso reportado e mostra-se a separação entre “quem enuncia” e, nos termos tradicionais explicitados anteriormente, “o falante que se compromete com a verdade da proposição enunciada”. Em última consequência, há uma separação, a meu ver, entre enunciado e proposição enunciada. Observemos outro exemplo:

(4.9) lugar (exceto Belo Horizonte) por mais de 2 anos. Sebastião: Nem no Rio de Janeiro? Prof. Eduardo: No Rio, eu peguei uma vez, mas minha mulher me gozou tanto. **Comecei a puxar o “s”, igual ao pessoal de Juiz de Fora, que se considera carioca..** Dizem aqui em Belo Horizonte que o pessoal de lá dá o endereço tipo Avenida Brasil, 9 milhões, 582 mil etc. (risos) Foi na época em que servi o Exército no Rio em Magalhães Bastos, subúrbio. Sebastião: Não entendi. Com 17 anos o senhor

A partir de (4.9), temos duas possíveis interpretações:

(4.9a) Os juizforanos efetivamente dizem que são cariocas (e aí se constituiria, como (xi), discurso reportado)

(4.9b) Os juizforanos se comportaram de uma determinada forma ou fizeram alguma consideração sobre serem cariocas que levam o falante a crer que o pessoal de Juiz de Fora “se considera carioca” (inclusive enfatizado pelo enunciado seguinte em que temos a presença do evidencial “*Dizem que*”).

Dessa forma, o que se pode depreender dos exemplos acima, quando empregada a terceira pessoa do discurso, é que um dos usos dessas construções é uma avaliação do falante sobre a avaliação (ou perspectiva ou, ainda, ponto de vista) de uma pessoa outra (no papel sintático de sujeito da cláusula principal).

Segundo Nuyts (2012), esta seria uma diferença ente subjetividade e intersubjetividade. Uma avaliação é tomada como subjetiva quando é responsabilidade exclusiva do avaliador, enquanto a intersubjetiva é compartilhada entre o avaliador e um grupo de pessoas. O autor sugere que são principalmente os verbos de predicados mentais que devem ser identificados com a subjetividade (NUYTS, 2001, p. 122-128) e pleiteia que a subjetividade seja tratada como uma categoria diferenciada da modalidade epistêmica (NUYTS, 2012), como, por exemplo, em (4.10) e (4.11):

(4.10) [94] **achei** aquele lugar incrível //COM=\$ (bfamcv01)

(4.11) [74] que eu **acho** que deu muito pau /=COM= nessa taça //APC=\$ (bfamcv01)

Na interpretação de Nuyts, exemplos como o apresentado em (4.10) parecem não carregar um sentido epistêmico, considerando que o verbo “achar”, neste caso, é um marcador puramente de subjetividade, já que indica uma avaliação estética ou moral e não uma. Em (4.11), ao contrário, o verbo estaria empregado como um marcador epistêmico, com a introdução de uma opinião baseada em evidência ou na estimativa das possibilidades de alguma coisa ser o caso.

Sob outra perspectiva, Almeida e Ferrari (2012), em artigo sobre a diferença entre os complementos de construções epistêmicas no inglês, do tipo [*X thinks that Y*] e [*X thinks Y*], levantam a questão de qual seria a diferença pragmática entre estas duas construções, baseado no Princípio da Não-Sinonímia (GOLDBERG, 1995). As autoras, a partir da análise de 382 construções epistêmicas de complementação, coletadas nas versões impressa e eletrônica da revista *Speak Up*, afirmam que estas construções são “operadores de subjetividade, que

apresentam o objeto de conceptualização (a cláusula complemento), tanto direta como indiretamente, desde a perspectiva do falante.”⁸⁸ (ALMEIDA; FERRARI, 2012, p. 123-124). Ainda, sustentam que as construções sinalizam intersubjetividade, uma vez que se referem implicitamente à perspectiva do falante em relação a outras perspectivas apresentadas em discurso anterior: as não-completivas indicam conjunção cognitiva e as completivas, disjunção cognitiva entre a perspectiva do falante e outras perspectivas disponíveis no fluxo discursivo. (ALMEIDA; FERRARI, 2012, p. 124). Esta posição, que leva em conta uma abordagem discursiva, contraria outras propostas discutidas na literatura sobre o tema, como a já mencionada de Nuyts (2012) e a de Verhagen (2005), justamente porque pleiteia que a intersubjetividade vai envolver uma cena mais ampla que inclui a relação entre falante e ouvinte, a cláusula complemento e as perspectivas anteriores disponíveis no discurso.

De fato, os exemplos da amostra apontam para uma diferença na qualidade da avaliação em construções completivas e não-completivas. No entanto, não temos dados suficientes para uma generalização que corrobore uma ou outra posição apresentada. Para fins da pesquisa, abarco crença e opinião sob o guarda-chuva de crença, tomando uma opinião como o resultado de uma determinada crença.

4.5.2.1 Os números para os verbos epistêmicos

Os *types* e *tokens* correlatos para os verbos epistêmicos foram classificados quantitativamente de acordo com a tipologia interacional (público versus privado; monólogos versus diálogos versus conversações) e a tipologia textual (narrativo, relato, expositivo, argumentativo, descritivo).

Além disso, foi utilizado um número de variáveis na tabulação dos dados, que levam em conta: o padrão de estrutura informacional (qual unidade informacional contém o marcador modal), composicionalidade (enunciado simples, dois índices no mesmo enunciado, dois ou mais índices em enunciados diferentes, referência contextual), tipo de modalidade e seus subvalores, o padrão sintático, a projeção pragmática dos índices, e o conceptualizador.

Os tipos encontrados foram: ‘achar’, ‘acreditar’, ‘crer’, ‘imaginar’, ‘pensar’ e ‘saber’ como exemplificado abaixo:

⁸⁸ No original: “[...] subjectivity operators which present the object of conceptualization (complement clause), either direct or indirectly, from the speaker’s perspective.” (ALMEIDA; FERRARI, 2012, p. 123-124).

(4.12) **Achar:**

(k) *GIL: [2] <ô /=CNT= mas> /=DCT= voltando à questão /=COB= falando em [2]=EMP= e também falando em povo mascarado /=COB= esse povo do Galáticos é muito palha /=COB= eu **acho** que es nũ deviam mais participar /=COM= e <tal> //UNC=\$ (bfamcv01)

(l) [44] aí /=PHA= passou um pouquim /=COB= o filho /=i-COB= **achando** que tava errado aquele negócio /=PAR= voltou lá outra vez //COM=\$ (bfammn03)

(m) *KAT: [43] então ela **acha** que é a meia que tá melhorando //COM=\$ (bfamd104)

(4.13) **Acreditar:**

(n) [46] tipo /=INT= eu [1]=EMP= eu **acredito** /=i-COM= tipo /=PAR= cem /=SCA= por cento /=SCA= nisso //COM=\$ (bfamcv01)

(o) [87] não /=INP_r= nũ **acredito** nisso não //COM_r=\$ (bfamd101)

(p) [49] e eu **acredito** que depois que eu terminar o EDUCONLE /=COB= eu acho que aí eu vou tar mais madura ainda /=COB= acho que mais preparada //COM=\$ (bpubmn01)

(4.14) **Crer:**

(q) *ENC: [208] eu **creio** que sim //COM (bfamd105)

(4.15) **Imaginar:**

(r) [171] não /=PHA= trinta reais /=TOP= aí eu &j [2]=SCA= eu [1]=EMP= eu fico **imaginando** que e' fica pensando assim /=INT= Nossa Sio' /=EXP_r= às vezes lá em casa tá precisando de fazer uma compra e tudo /=COM_r= né //PHA=\$ (bpubmn01)

(4.16) **Pensar:**

(s) *RUT: [208] cê **pensa** que ele /=SCA= participa da [1]=SCA= &d [1]=EMP= desses presente //COM=\$ (bfamcv02)

(4.17) **Saber:**

(t) *TER: [69] **sei** o tanto não //COM=\$ (bfamcv02)

(u) *BAL: [45] cê **sabe** que aquelas caxinhas ali /=TOP= ela [1]=EMP= eu descobri ontem /=COB=\$ (bfamd102)

(v) *ANE: [19] rua Joaquim Nabuco /=TOP= sô **sabe** me informar //COM=\$ (bfamdl03)

Como mencionado anteriormente, o número total de *tokens* de verbos epistêmicos encontrado no *subcorpus* foi de 210, distribuídos em 6 *types*, como a seguir: 105 para ‘achar’ (50%); 3 para ‘acreditar’; 1 para ‘crer’; 1 para ‘imaginar’; 2 para ‘pensar’ e 98 para ‘saber’ (47%). A Figura 4.13 mostra esta distribuição:

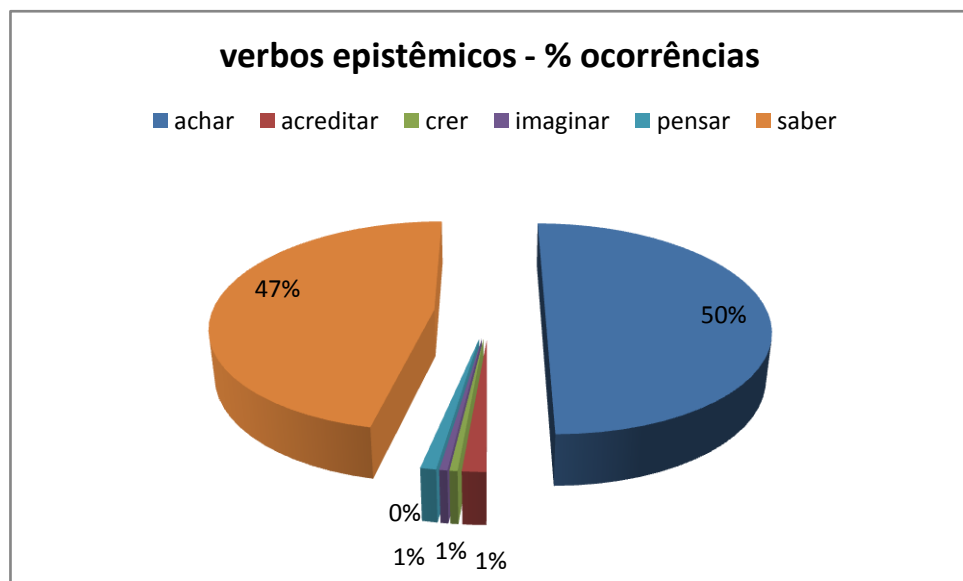


Figura 4.13 – Distribuição e frequência dos verbos epistêmicos na amostra

Analisando os números para tipologia interacional, os resultados são: 60 exemplares para monólogos, 21 em monólogos públicos e 39 em privados; 99 exemplares para diálogos, 16 para públicos versus 83 para diálogos privados; e 51 exemplares para conversações, 14 em públicas versus 37 em conversações privadas. É possível visualizar os dados na Tabela 4.10:

	Privado	Público	TOTAL
Monólogos	39	21	60
Diálogos	83	16	99
Conversações	37	14	51
TOTAL	159 (75,7%)	32 (24,3%)	210

Tabela 4.10: Distribuição de *tokens* de verbos epistêmicos por tipologia interacional

Esta tabela aponta para uma alta taxa de *tokens* em interações dialógicas, principalmente em diálogos privados. Isso se deve a características específicas desses tipos de texto e à relação dos participantes na situação comunicativa. Se levarmos em conta a tipologia textual, nota-se que 31,9% destes verbos são usados em textos argumentativos ou expositivos, o que coincide com a sua própria definição: sinalizar o comprometimento do conceptualizador em relação ao que se enuncia. As outras ocorrências distribuídas em textos narrativos, descritivos e relatos mostram os momentos em que os participantes expressam suas opiniões ou crenças ou requerem a opinião de seu interlocutor. Devo destacar que as ocorrências em monólogos públicos correspondem a apenas um texto: a participante, uma professora da escola básica relata sua atividade profissional e dá sua opinião sobre o processo de ensino-aprendizagem. É também necessário destacar que a diferença entre os números de textos privados e públicos é minimizada em termos de frequência relativa.

4.5.2.2 Padrões sintáticos, semântica e questões pragmáticas:

Vários padrões sintáticos são utilizados, principalmente:

- (i) os verbos epistêmicos introduzem orações encaixadas. Em nossa amostra, 57,1% de todas as ocorrências seguem este padrão.
- (ii) eles podem ocupar diferentes posições no enunciado, nos casos em que não introduzem uma encaixada.
- (iii) [SN]_{SUBJ} V [SN]_{OBJ} [SAdv] [SAdj]_{ATR}. Os casos em que não há a realização do objeto correspondem a 22,9% do total de ocorrências, principalmente com os verbos “achar” e “saber”.

Alguns exemplos:

(4.18) SN_{SUJ} V comp S

*LUI: [236] eu **acho** que a gente deve chamar os <times> legais // =COM=\$
(bfamcv01)

(4.19) SN_{SUJ} V SAdv S

*SIL: [154] eu **acho** assim /=INT= se a pessoa nã tem condições de fazer /=TOP= ele paga pra fazer // =COM=\$ (bfamdl04)

(4.20) SN_{SUJ} V SN_{OBJ} SAdv SAdj_{ATR}

[77] então eu **achava** aquilo muito interessante // =COM=\$ (bfammm06)

(4.21) SN_{SUJ} SAdv_{NEG}V SAdj_{ATR} SAdv_{NEG}

[283] eu nã **achei** ruim não /=COM= Jael // =ALL=\$ (bfamcv02)

Semanticamente, estes verbos expressam o grau de comprometimento do conceptualizador em relação ao material locutivo enunciado. Estas unidades lexicais estão ligadas a diferentes frames descritos para o inglês: *Opinion, Certainty, Awareness, Assessing and/or Cogitation*.⁸⁹

Em termos da distribuição dos verbos por unidades informacionais, podemos ver na Tabela 4.11 como o *subcorpus* está caracterizado:

Unidades informacionais	# ocorrências	% epistêmicos	% amostra
COM	182	86,6	17,7
INT	3	1,4	13,63
PAR	11	5,3	64,7
TOP	14	6,7	63,6

Tabela 4.11 – Distribuição dos verbos epistêmicos em unidades informativas

A primeira coisa a se observar é qual a unidade informacional que contém o índice modal: apenas o Comentário (incluindo as unidades Comentário Múltiplo e Comentário Ligado), o Tópico (inclusive a Lista de Tópico), o Parentético (inclusive a Lista de Parentético) e o Introdutor Locutivo podem ser modalizados. No caso dos verbos epistêmicos, a unidade de Comentário é modalizada em 86,6% de todas as ocorrências (182, para COM, 11 para CMM e 20 para COB), seguida pelo Tópico (6,7 % das ocorrências, sendo, em números absolutos, 12 para o TOP e 2 para TPL), o Parentético (5,3% das ocorrências, 8 para PAR e 3 para PRL). Em número bastante reduzido, o Introdutor Locutivo, com apenas 3 ocorrências. Em relação ao número total de ocorrências das unidades informacionais, o que se destaca são os números para o Tópico e o Parentético. As duas unidades têm como estratégia preferencial de modalização os verbos epistêmicos com 63,6% e 64,7% de todas as ocorrências do *minicorpus*. Quanto ao Comentário e o Introdutor Locutivo, como visto anteriormente, a estratégia principal são os verbos modalizadores, sendo os epistêmicos responsáveis por

⁸⁹ A descrição para o *frame* de Opinion, por exemplo, é a seguinte: “Cognizer holds a particular Opinion, which may be portrayed as being about a particular Topic (John THINKS that it looks better back)”. Disponível em: <https://framenet.icsi.berkeley.edu/fndrupal/>

menos de 20% das ocorrências. O gráfico na Figura 4.14 ilustra a distribuição dos epistêmicos nas unidades informacionais e a frequência desta estratégia no conjunto das unidades informacionais.

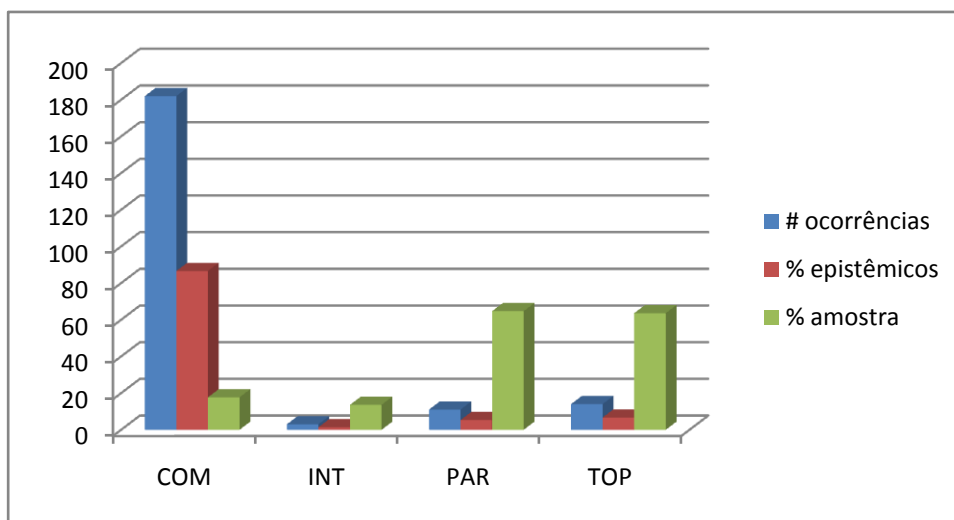


Figura 4.14 – Distribuição e frequência dos epistêmicos em unidades informacionais

Abaixo alguns exemplos:

(4.22) COM

*DFL: [56] aquilo que o /=SCA= professor **achava** mais importante //
=COM=\$ (bfammn02)

(4.23) TOP

*ANE: [324] **eu acho que quando eu vim** /=TOP= tinha esse /=CMM= tinha
o outro //CMM=\$ (bfamd105)

(4.24) PAR

[41] só que é de microondas /=COM= **eu acho** //PAR=\$ (bfamd101)

(4.25) INT

*SIL: [154] **eu acho assim** /=INT= se a pessoa nã tem condições de fazer
/=TOP= ele paga pra fazer //COM=\$ (bfamd104)

Mello e Raso (2013) discutem a necessidade de ampliação do enunciado e das unidades tonais como construtos analíticos básicos no tratamento da fala espontânea via *frames*. Como, segundo a Teoria da Língua em Ato (CRESTI, 2000), ao enunciado

corresponde uma ilocução, mesmo que esses verbos epistêmicos, objeto de nossa investigação, estejam, de alguma forma, representados em *frames* ligados por herança metafórica e no uso de atividade mental, ao falarmos, estamos cumprindo diferentes ações. Dessa forma, “a análise da linguagem pautada por *frames*, necessariamente deve levar em conta pelo menos dois níveis adicionais àqueles construcional e semântico, quais sejam, os níveis informacional e ilocucionário (MELLO; RASO, 2013, p. 106).

Transponho um princípio, talhado, até agora, para a análise da escrita, para a análise da fala, o Princípio da Não-Sinonímia, para discutir brevemente essa questão. Segundo esse Princípio (GOLDBERG, 1995, 2006):

Se duas construções são sintaticamente distintas, elas devem ser semântica ou pragmaticamente distintas.

Corolário A: Se duas construções são sintaticamente distintas e semanticamente sinônimas, então elas não devem ser pragmaticamente sinônimas.

Corolário B: Se duas construções são sintaticamente distintas e pragmaticamente sinônimas, então elas não devem ser semanticamente sinônimas.

Os verbos, próximos semanticamente, se organizam sintaticamente de várias maneiras, como vimos. Dessa forma, pelo PNS, o seu comportamento pragmático deve ser diferente. E de fato o é. Vejamos as possíveis funções e os seus respectivos exemplos com a mesma unidade lexical ‘achar’:

(i) Estrutura informacional: em posição parentética funciona como atenuador da asserção anterior:

Contexto: casal conversa sobre a distribuição de vagas em uma universidade pública.

*LAU: [51] não / tem um de ensino de arte +

*LUZ: [52] são duas vagas / **eu acho** // [53] não// [54] de ensino de artes // (bfamd103)

(ii) Marcadores de concordância / discordância: indica um (possível) padrão lexical. Corresponde a um fenômeno de gradiência de índice modal para um marcador discursivo:

Contexto: amigas no supermercado.

*REN: [413] <pode> // [414] tanto faz // [415] pode //

*FLA: [416] ou cê acha muito //

*REN: [417] uhn // [418] acho que não // [419] tá // [420] papel <higiênico / eu nunca> + (bfamd101)

Dessa forma, damos conta de que uma descrição de *frames* em termos exclusivamente sintáticos e semânticos pode ser, de fato, enriquecida com desdobramentos pragmáticos de diversas ordens.

4.5.3 Advérbios modais:

Os advérbios, assim como os verbos, receberam atenção especial como índice marcador de modalidade, principalmente na língua inglesa (GREENBAUM, 1969; PERKINS, 1983; SWAN, 1988; BIBER; FINEGAN, 1988; HOYE, 1997; SIMON-VANDERBERGEN; AIJMER, 2007; SIMON-VANDERBERGEN, 2008). No português brasileiro, destacam-se os trabalhos de Castilho (1993, 2010) e, para a fala espontânea, os de Mello, Couto e Ávila (2010) e Mello e Caetano (2012).

A classe de advérbios é uma classe flexível em termos sintáticos, uma vez que pode aparecer em diferentes posições nos enunciados. Na fala, como aponta Mello et al. (2010), esta complexidade de escopo se amplifica, dado que devem ser levados em consideração os traços prosódicos. Já foi salientado que a análise da prosódia é importante para a desambiguação de itens modais e, no caso específico dos advérbios modais, se estes são, de fato, itens modalizadores ou se cumprem outra função no enunciado, como a de intensificador ou marcador discursivo.

4.5.3.1 Frequência de advérbios e locuções adverbiais modais:

Nesta seção forneço os resultados para os advérbios modais, em relação à sua frequência na amostra. Os *types* e *tokens* também foram classificados quantitativamente de acordo com a tipologia interacional (público vs privado; monólogos vs diálogos vs conversações). Foram encontrados 16 *types* em 94 ocorrências na amostra, o que corresponde,

no universo dos índices modais, a 7,9% das ocorrências.⁹⁰ A Tabela 4.12 e 4.13 trazem os números para o número de ocorrências deste índice no *minicorpus* e sua distribuição em termos de tipologia interacional, respectivamente:

advérbios	# ocorrências	%
com certeza	7	7,5
mesmo	45	47,9
realmente	3	3,2
exatamente	4	4,4
talvez	8	8,5
sem dúvida	2	2,2
às vezes	5	5,3
na verdade	7	7,5
sinceramente	2	2,2
claro	5	5,3
potencialmente	1	1
óbvio	1	1
sem chance	1	1
logicamente	1	1
justamente	1	1
certamente	1	1
Total	94	100

Tabela 4.12 – Número de ocorrências dos advérbios modais

⁹⁰ Mello e Caetano (2012) investigaram a distribuição e frequência dos advérbios modais em toda a parte informal do C-ORAL-BRASIL. Através da busca pelo parser PALAVRAS de todos os advérbios, foram identificados 28.000 advérbios no *corpus*, correspondendo a 13,45% das palavras na amostra. Dentre os advérbios e locuções adverbiais, 763 foram destacados como modais, que representa 0,37% do total de palavras da parte informal do *corpus*. No que diz respeito à relação entre advérbios/advérbios modais, aproximadamente 2,72% do total de advérbios assume função modal. O estudo revela, dessa forma, que a frequência de advérbios modais é expressiva, mas as autoras sublinham que, ao olhar com lentes de aumento para a amostra, um único item, o advérbio “mesmo”, contribui com uma frequência de 55,7% de todas as ocorrências. O que coincide com os números para o *minicorpus*.

	Privado	Público	TOTAL
Monólogos	14	5	19
Diálogos	33	12	45
Conversações	25	5	30
TOTAL	72	22	94

Tabela 4.13 – Distribuição dos advérbios x tipologia interacional

Em termos absolutos, o número de ocorrências em contexto privado/familiar é bastante maior do que no contexto público. No entanto, como já observado para as outras estratégias, é possível um balanceamento na frequência dos itens, que apontam para um uso proporcional nos dois contextos interacionais. No que diz respeito à tipologia interacional, também como esperado, os advérbios ocorrem mais em trocas dialógicas (45, para diálogos e 30 para conversações) do que em monólogos (19). Estes números acompanham os encontrados por Mello e Caetano (2012), na análise de todos os advérbios modais da parte informal do C-ORAL-BRASIL.

4.5.3.2 Distribuição em unidades informacionais:

Em termos de distribuição pelas unidades informacionais, são estes os números:

Unidades informacionais	# ocorrências	% advérbios
COM	87	92,6
INT	2	2,1
PAR	3	3,2
TOP	2	2,1

Tabela 4.14 – Distribuição dos advérbios modais nas unidades informacionais

No que tange aos valores modais, todas as ocorrências da amostra são do tipo epistêmico, com o subvalor de possibilidade, como “às vezes” e “talvez”, ou de crença, indicando graus de certeza, em sua maioria, o que corrobora a posição de Perkins (1983, p. 89) para os advérbios modais em inglês. O autor afirma que estes marcadores são fundamentalmente de natureza epistêmica.

Vejam os exemplos para cada um dos 16 itens aqui encontrados e seu comportamento em relação a outros elementos co-textuais.

(4.26) **Mesmo:**

(w) *REN: <tem que ser o> Knorr mesmo //

*FLA: é // &va [/1] vai esse / né / Rena // (bfamdl01)

O item “mesmo” é o advérbio modal mais frequente na amostra, com quase a metade das ocorrências. Este item tem um comportamento particular, porque, muitas das vezes, pode funcionar como intensificador, como um operador que regula o fluxo da interação, ou como pronome adjetivo. Para o desempate, utilizamos a regra de substituição, observamos o contexto e, ainda, fazemos a oitiva do arquivo de som. Vejamos os exemplos, em que o advérbio não possui função modal:

(x) *TON: tinha que matar o cinco / sô //

*REN: não / aí e' lá //

*CAR: é o quatro mesmo / Jacaré //

*TON: <penei> //

*CAR: <quatro mesmo / sô> // purra ele // se ele nũ morrer / nũ tem problema não // <aí o'> //

(y) *CES: vão lá // é / e ele [/2] e é ele mesmo que tá ainda / viu // é o trezentos-e-três / viu Anete //

*ANE: é ele mesmo //

*CES: é ele mesmo // então // uhn // (bfamdl05)

(4.27) **Na verdade e com certeza:**

(z) *EUG: cê já escolheu / moça //

*JAN: hum hum // eu acho que eu vou + na verdade eu queria levar as duas / né // mas eu vou levar essa aqui // (bpubdl02)

(aa) *FLA: colorido é mais caro //

*REN: com certeza // deve ser // (bfamdl01)

(4.28) **Talvez e às vezes:** indicadores de possibilidade

(bb) *MAI: no norte de Mina / tinha esse [/2] antigamente / tinha esse tipo de cobra tudo / né // **talvez** agora já acabou / porque já desmataram muito / né // (bfammn01)

(cc) *ANE: [297] é / olha se tem gente aí //
 *CES: [298] xxx [/1] pode passar aqui //
 *ENC: [299] oi //
 *ANE: [300] <**às vezes**> [/2] às vezes fica <gente> //
 *CES: [301] <pode entrar> // [302] <po' / po' vim> / Anete // (bfamdl05)

“Às vezes” é uma locução que tanto pode exercer uma função temporal, quanto pode exercer uma função modal.⁹¹ A desambiguação também é feita pela análise do contexto, do co-texto e da prosódia. Por exemplo:

(dd) *PAU: cê tem que aprender // a olhar planta // porque / a hora que cê pegar um projeto maior aí pa fazer / que o dono nã tiver perto / aí cê tem que tocar // **às vezes** o engenheiro / também nã tem muito tempo / tá mexendo com outra coisa / só vem no final de semana // (bpubdl01)



No exemplo, o adverbial “às vezes” pode indicar uma situação possível, portanto, um uso modal, ou pode indicar a frequência com que o engenheiro visita a obra, um uso temporal, no caso.

(4.29) **Certamente e potencialmente:**

(ee) *MAI: a cobra tava&en[/2] continuou enrolada nele // **certamente** eatava querendo fazer o seguinte / eu / eu matei esse / eu vou matar o resto tudo / da [/1] &he / dentro da casa // nã sei / né / a imaginação hhh dum [/1] dum animal / o que que pode ser / né // (bfammn01)

(ff) *HEL: <é / tipo> / cês ficam<falando também> //
 *LUC: <isso **pode** ser / **potencialmente**> divertido /
 *BRU: <nã> //
 *LUC: <mas enfim / yyyhh> //
 *BRU: <nã acho> // (bfamcv04)

⁹¹ Não é do escopo deste trabalho investigar a gramaticalização dos itens modais. Parece-me, no entanto, que o caminho de adverbial de tempo para um adverbial modal é uma hipótese consistente. Como exemplo similar, temos o caso da locução “de repente” (“De repente ele chegou”, uso temporal; “De repente a gente toma um copo, assim que cê acabar a tese.”, uso modal, indicando possibilidade).

(4.30) **Sinceramente:**

- (gg) *TER: <oh / adorei> //
 *RUT: o' // eu ãu achei ruim ão / Jael // <sinceramente> //
 *JAE: <ñu gostei> de jeito nenhum // (bfamcv02)

(4.31) **Logicamente:**

- (hh) *JOR: e lá eu fiquei um período / desenvolvendo o mesmo tipo de trabalho / logicamente com um salário melhor / hhh e por amizade eu fui cair / em uma multinacional / que eu dei uma virada no produto // (bfammn06)

(4.32) **Exatamente e justamente:**

- (ii) *CEL: <fazer uma jogada> melhor //
 *CAR: é // que aí cê rola <o seis / fica atrás do quatro> //
 *CEL: <é> // <isso> // <exatamente> // (bfamcv03)

- (jj) *ANE: <que número> que é //
 *CES: setecentos-e-quatro //
 *ANE: aqui // é exatamente // po' parar o carro hhh //
 *CES: é // setecentos-e-quatro // (bfamd105)

(4.33) **Claro e óbvio:**

- (kk) *LUI: [151] ão // [152] e [/1] e [/1] e [/1] e principalmente / convocando a galera / falar assim / o' / galera / negócio seguinte / tá pensando em organizar um torneio / e a gente quer / que vocês participem / da organização / <mandando representante> +
 *LEO: [153] <e acima de tudo / eu acho que a gente tem> que chamar os times que / tipo / o [/1] realmente os times que / merecem a [/1] a nossa +
 *EVN: [154] é / a <gente tem que> <restringir também / isso> //
 *GIL: [155] <ño / claro> // (bfamcv01)

- (ll) *ROG: [222] é // [223] o projeto é bom por isso / uai // [224] a gente faz jeito que a pessoa pede / uai // [225] aquele carinha <do &Alexa> +
 *PAU: [226] <mas tem> que saber ler projeto também / né //
 *ROG: [227] aquele <carinha> do Alexandre lá ele me deu o desenho dele //
 *PAU: [228] <claro> // (bpubdl01)

- (mm) *BRU: tá falando do meu pé / né //
 *CEL: óbvio // (bfamcv04)

Os índices modais “claro” e “óbvio”⁹², apesar de pertencerem, formalmente, à categoria dos adjetivos, foram incluídos na análise das construções adverbiais modais, uma vez que seu comportamento se assemelha dos advérbios modais “claramente” e “obviamente”. São considerados advérbios somente quando ocorrem sozinhos, sem a presença do verbo ser e/ou da conjunção. Quando este é o caso, são analisados como adjetivos em posição predicativa.

Estes itens tendem a cumprir um papel de confirmação, quer dizer, um falante normalmente os utiliza para confirmar o que foi enunciado pelo falante anterior.

4.5.4 As construções condicionais: *se p, então q*

O caráter epistêmico das construções condicionais na variante brasileira do português é apontado em um número de trabalhos relevantes para o tema, entre eles os de Hirata-Vale (2001, 2008); Ferrari (2007, 2008); Bezerra e Meireles (2009). Hirata-Vale afirma que “as condicionais relacionam-se a mundos concebíveis, sejam eles reais, eventuais e contrafactuais” e se tornam mais abstratas à medida que passam por um processo de subjetivização, por isso estão situadas no eixo do conhecimento e da avaliação subjetiva.

Nas condicionais de forma canônica “*se p, então q*”, em que *p* é a prótase e *q*, a apódose, “o evento *p* é uma condição suficiente (e, em alguns casos, necessária) para a ocorrência do evento *q*.” (SWEETESER, 1990). As condicionais são, pois, projeções hipotéticas de manifestações causais diretas.

A sua hipoteticidade está ligada ao grau de probabilidade de realização das situações referidas na prótase (cf. COMRIE, 1986).

4.5.4.1 Os números para as construções condicionais de forma canônica:

As construções condicionais, dentro do quadro teórico da Teoria da Língua em Ato, constituem-se como um desafio, uma vez que, como apontado anteriormente, o escopo da modalidade é a unidade informacional. No caso deste tipo de construção, como afirmam Côrtes e Mello (2013, p. 2), ela “pode ultrapassar os limites de um enunciado ou estar completamente acomodada nele, dividida, ou não, em mais de uma de suas unidades tonais”.

⁹² Poderiam estar incluídos os adjetivos “lógico” e exato” que têm comportamento similar ao dos advérbios “logicamente” e “exatamente”.

A frequência das construções condicionais no *minicorpus* foi considerada de acordo com a tipologia interacional; a estrutura informacional; a distribuição de prótase/apódose e a relação entre a estrutura sintática e informacional.

Como já dito, ao todo foram analisados 5.484 enunciados, sendo que 1.046 deles estão modalizados. Foram encontradas 109 ocorrências de construções condicionais entre as de forma canônica. O percentual de condicionais na amostra é de 9,09% entre todos os índices modais.

Passando à análise para a tipologia interacional, temos os seguintes resultados: 24 exemplares para monólogos, 8 em monólogos públicos e 16 em privados; 40 exemplares para diálogos, 9 para públicos versus 25 para diálogos privados; e 50 exemplares para conversações, 12 em públicas versus 38 em conversações privadas. É possível visualizar os dados na Tabela 4.15:

Tipologia	Contexto	Frequência
Monólogo	Privado	16
	Público	8
Diálogo	Privado	25
	Público	10
Conversação	Privado	38
	Público	12

Tabela 4.15 - Distribuição de *tokens* de condicionais por tipologia interacional

No que diz respeito à tipologia interacional, ainda não temos dados suficientes para verificar quais as variáveis em jogo e como as condicionais se comportam em cada contexto.

Com relação à organização sintática, encontramos três tipos de padrões: prótase/apódose; apódose/prótase e prótase somente. Abaixo as frequências para cada um dos tipos:

Organização sintática	Frequência
Prótase-apódose	76
Apódose-prótase	10
Prótase	23

Tabela 4.16 – Frequência dos padrões sintáticos de condicionais

Os números para cada tipo de organização não surpreendem exatamente. Em frequência absoluta, o padrão prótase/apódose é o mais recorrente, já o apódose/prótase apresenta ocorrência bem reduzida. Sobre a prótase sozinha, parece-me um número interessante, pois não segue o padrão estabelecido na escrita e corresponde a 21,1% do total de ocorrências, e se dividem, na amostra, em três tipos: (a) prótases cuja apódose não são realizadas, com efeito suspensivo, mas possível de ser compreendida entre os interlocutores (em termos de ilocução, parecem cumprir um ato do tipo expressivo, do subtipo expressão de obviedade); (b) prótases que são perguntas parciais (em termos de ato, são do tipo diretivo, do subtipo pedido de explicação); (c) prótases cuja apódose foi realizada anteriormente, diluída no contexto ou realizada por outro falante.

Antes de apresentar os exemplos das condicionais, mostro os números referentes à distribuição dos índices em termos de estrutura informacional, na Tabela 4.17, e a relação entre padrões sintáticos e organização informacional, na Tabela 4.18:

Estrutura informacional	Frequência
Enunciados diferentes	10
Mesma unidade de COM	30
TOP/COM	51
CMM/CMM	11
Outros	7

Tabela 4.17 – Frequência das condicionais quanto à estrutura informacional

Estrutura informacional	TOP/COM	Mesmo COM	CMM/CMM	Enunciados diferentes	Outros	Total
Estrutura sintática						
Pró - após	49	7	9	8	3	76
Após - pró	2	2	2	2	2	10
Pró	---	21	---	---	2	23
Total						109

Tabela 4.18 – Relação do padrão sintático de condicionais e estrutura informacional

Seguem exemplos das possíveis combinações sintáticas das condicionais e de sua distribuição em termos de estrutura informacional:

(a) **Prótase / apódose:**

(a.1) prótase /=TOP= apódose //COM=

(4.34) *BRU: [268] <e se for uma palavra composta_{PRO} /=TOP= cê faz assim_{APO}> //COM=\$ (bfamcv04)

(a.2) prótase e apódose na mesma unidade de COM: prótase apódose //COM=

(4.35) *PAU: [72] esse tipo de muro /=TOP= se ficar baixo demais_{PRO} ele fica feio_{APO} //COM=\$ (bpubdl01)

(a.3) prótase // apódose // em enunciados diferentes

(4.36) *SIL: [233] se ocê não tem relógio dentro do quarto_{PRO} //COM=\$
[234] cê vai comprar um relógio_{APO} /=CMM= pôr lá pa despertar //CMM=\$ (bfamdl04)

(a.4) prótase /=CMM= apódose //CMM=

(4.37) *CAR: [75] se ele nũ morrer_{PRO} /=CMM= nũ tem problema não_{APO} //CMM=\$ (bfamcv03)

(b) **Apódose / prótase**

(b.1) apódose e prótase em enunciados diferentes: apódose // prótase //

(4.38) *ANE: [120] como é que a gente vai adivinhar_{APO} //COM=\$
[121] se for isso_{PRO} /=COM= né //PHA=\$ (bfamdl05)

(b.2) apódose prótase //COM=

(4.39) *RUT: [191] <lá em casa> /=TOP= <aviso_{APO} se cê for lá_{PRO}> //COM=\$ (bfamcv02)

(b.3) apódose /=TOP= prótase //COM=

(4.40) *BAL: [58] <dá> <problema_{APO} /=TOP= se você ligar o aparelho de> &duz [1]=SCA= cento-e-dez na tomada de duzentos-e-vinte_{PRO} //COM=\$ (bfamdl02)

(c) **Prótase**

(c.1) expressão de obviedade:

(4.41) *ANE: [38] eh /=PHA= se cê nũ tiver um carrinho que [/1]=SCA= que sobe aqui_{PRO} //COM=\$ (bfamd105)

(c.2) pergunta parcial, pedido de explicação:

(4.42) [318] <e se precisar de um esclarecimento_{PRO}> /=COM= <assim /=PAR= no> meio do negócio hhh //APC=\$ (bfamcv04)

(c.3) conjugação entre turnos:

(4.43) *TER: [22] mas /=INP= gente velha /=TOP= **já prometeu o** [/1]=SCA= **os presente** =TOP= <já /=SCA= **pode**> **garantir que ganhou_{APO}** //COM=\$*RUT: [23] <ah /=CMM= é> //CMM=\$ [24] se nũ morrer antes <deles_{PRO} /=COM=\$ (bfamcv02)

Observa-se que a estrutura mais frequente é o perfil prótase no Tópico e apódose no Comentário, o que coincide com a função informacional das unidades. Em termos discursivos, a prótase de uma condicional atua como um angulador, que estabelece as condições de validação do discurso subsequente. Conforme Ferrari (2008, p. 121), desde uma perspectiva cognitiva:

no domínio epistêmico, as condicionais expressam a ideia de que o conhecimento do evento ou do estado de coisas expresso na prótase seria uma condição suficiente para o estabelecimento da conclusão na apódose.

Os exemplos também nos mostram que é possível a combinação de outros perfis com as estruturas prótase/apódose, apódose/prótase, distribuídas em unidades informacionais diferentes, como Comentários Múltiplos, Comentários Ligados, Apêndices de Tópico e, também, em enunciados diferentes: neste caso, a estrutura se completa em atos de fala diferentes, cada um veiculando sua própria ilocução.

Podem ser encontradas, ainda, ocorrências em que a prótase tenha como âncora o contexto interacional, sem a realização necessária da apódose, como no caso em (4.41), o que denota um comportamento peculiar destas construções na fala espontânea

Os dados apontam para o fato de que a prótase carrega o não-factual da construção condicional (cf. COMRIE, 1986) e, nos casos em que ocorre a ordem inversa, “o foco da

construção está no conteúdo veiculado pela apódose, isto é, a parte mais saliente informativamente é a ação presente na apódose” (CORTES, MELLO, 2013, p. 4).

Por último, gostaria de retomar que as condicionais, como índices que expressam de modalidade, desafiam a premissa posta pela Teoria da Língua em Ato de que o marcador modal incide dentro da unidade informacional. Como vimos, este tipo de construção ultrapassa as fronteiras da unidade informacional e, inclusive, em alguns casos, do enunciado. Uma das hipóteses que levanto é que a estrutura condicional poderia ser tratada como uma construção padronizada, nos termos de Cresti (no prelo, p. 17), estas são “construções realizadas através de UTs, cada uma desenvolve uma função informacional diferente”⁹³. A autora ainda completa que incluem também construções que são realizadas em diferentes enunciados, com cada uma cumprindo sua própria ilocução.

Neste capítulo, apresentei a descrição da modalidade no *minicorpus* da parte informal do C-ORAL-BRASIL. Apresentei e discuti os resultados em termos de frequência por tipologia interacional, por distribuição dos índices nas unidades informacionais, a distribuição e frequência dos valores modais epistêmicos, deônticos e dinâmicos, e a relação de cada valor com as unidades informacionais. Em seguida, empreendi a análise para os verbos modalizadores, os verbos epistêmicos, os advérbios e as construções condicionais de forma canônica.

Passo agora à Parte III deste trabalho que traz a proposta de esquema de anotação para a modalidade na fala espontânea do português brasileiro que, como afirmam Ávila e Mello (2013, p. 1), vai fornecer

um ponto de partida confiável para pesquisadores interessados em desenvolver metodologias associadas a PLN que resulta na extração no discurso oral de marcadores de confiabilidade, certeza e factualidade, ou proceder à análise de sentimentos, modelagem da modalidade e objetivos afins.⁹⁴

⁹³ Tradução minha para: “[...] constructions performed across TUs, with each developing a different information function [...]” (CRESTI, no prelo, p. 17).

⁹⁴ No original: “[...] reliable starting point for researchers that might be interested in developing methodologies associated to NLP that ensue the extraction of oral discourse reliability, certainty and factuality markers, or carrying sentiment analysis, modeling modality and similar objectives.” (ÁVILA; MELLO, 2013, p.1).

PARTE III – ANOTAÇÃO SEMÂNTICA DA MODALIDADE

Capítulo 5

O PROJETO MASS (MODAL ANNOTATION IN SPONTANEOUS SPEECH): anotação semântica da modalidade no C-ORAL-BRASIL⁹⁵

O presente capítulo apresenta o projeto MASS (Modal Annotation in Spontaneous Speech), um esquema de anotação semântica da modalidade desenvolvido para dados de fala espontânea do Português Brasileiro, descreve e explica em detalhes os passos da anotação, bem como discute os casos problemáticos e as decisões tomadas para solucioná-los.

A anotação linguística de corpora, nas palavras de Leech (1993, p. 275), pode ser definida como “a prática de adicionar informação *interpretativa* (especialmente linguística) a um corpus eletrônico de dados linguísticos falados e/ou escritos, por algum tipo de código anexado a, ou intercalados com, a *representação* eletrônica do próprio material linguístico”⁹⁶. E também pode se referir ao produto final deste processo.

A Linguística Computacional fornece meios para o desenvolvimento de *softwares* para o processamento de linguagem natural escrita e falada. Os corpora anotados não são um fim em si mesmo: a resolução de tarefas específicas, bem definidas e limitadas, como a etiquetagem morfosintática (cf. o parser PALAVRAS, BICK, 2000), é pré-requisito para a construção de sistemas complexos, como os de tradução automática e de sumarização.

Especificamente, o interesse por se distinguir uma informação que é real e certa de uma informação especulativa e modal resulta do fato de esta ser tarefa necessária para aplicações em PLN, como a extração de informação (KARTUNNEN; ZAENNEN, 2005), modelagem de incerteza em textos clínicos (MOWERY et al., 2012), resposta a perguntas (SAURÍ et al., 2006); classificação de *hedges* (MEDLOCK; BRISCOE, 2007; MORANTE; DAELEMANS, 2009); e análise de sentimentos (WIEBE et al., 2005).

Anotar a modalidade, com a finalidade de permitir o seu reconhecimento automático, inclui identificar os índices modais, classificá-los em uma determinada tipologia (por exemplo, em significados epistêmicos e não-epistêmicos), definir a sua fonte e o seu escopo semântico. Vários têm sido os projetos desenvolvidos para a anotação de expressões modais,

⁹⁵ Parte da pesquisa sobre a anotação semântica foi desenvolvida em meu Estágio Doutoral no Centro de Linguística da Universidade de Lisboa, no período de janeiro a agosto de 2013, sob a supervisão da Dra. Amália Mendes, e financiada pela Capes, Proc. nº BEX 9537-12-0.

⁹⁶ No original: “[...] the practice of adding *interpretative* (especially linguistic) information to an existing corpus of spoken and/or written language, by some kind of coding attached to, or interspersed with, the electronic *representation* of the language material itself” (LEECH, 1993, p. 275).

em sua grande maioria para a língua inglesa e focados nos auxiliares modais, com destaque para a relação entre modalidade e negação (MORANTE; SPORLEDER, 2012; BAKER et al., 2012), a anotação de sentido de verbos modais (RUPPENHOFER; REHBEIN, 2012), o desenvolvimento de um léxico para a modalidade e a construção de etiquetadores automáticos (BAKER et al., 2010). Há ainda esforços de anotação sendo empreendidos em outras línguas, como os trabalhos para o chinês (CUI; CHI, 2013) e para o português europeu (HENDRICKX et al., 2012; MENDES et al., 2013).

A anotação da modalidade para dados do português brasileiro é terreno a ser explorado tanto para *corpora* escritos, quanto para *corpora* de fala. Segundo Nurmi (2007, p.1), a anotação linguística contribui para a recuperação de diferentes elementos linguísticos, no entanto, a natureza multifacetada da modalidade “é ainda um obstáculo para a pesquisa assistida por computador”⁹⁷. Na mesma clave, Baker et al. (2010, p. 1403) afirmam que o desafio para a criação de um esquema de anotação de modalidade reside em “lidar com o complexo escopo das modalidades entre cada uma delas e com a negação e, ao mesmo tempo, criar um procedimento operacional simplificado que possa ser seguido por um especialista da linguagem sem treinamento especial”⁹⁸. Como já discutido na parte I desta tese, está longe de haver um consenso sobre como definir e caracterizar a modalidade: ela pode ser tomada como expressão da subjetividade, como uma distinção entre *realis* e *irrealis*, ou ainda como uma quantificação sobre mundos possíveis, restringidas por uma relação de acessibilidade. Dessa forma, o entendimento do que é essa categoria semântica é fundamental, bem como ter claro quais são os elementos que podem veicular a modalidade.

Na seção seguinte, contextualizo o nosso projeto dentro dos estudos anteriores de anotação semântica das expressões modais.

5.1 Trabalhos anteriores em anotação semântica

Os trabalhos desenvolvidos em anotação semântica centram-se nos fenômenos da factualidade e da modalidade, uma vez que, como foi dito, para um bom número de aplicações em PLN é necessário distinguir uma informação factual de uma não-factual e, também, as certezas das incertezas.

⁹⁷ Tradução minha para: “[...] is still a hurdle in computer assisted-research”.

⁹⁸ No original: “[t]he challenge of creating a modality annotation scheme was to deal with the complex scopig of modalities with each other and with negation, while, at the same time creating a simplified operational procedure that could be followed by language experts without special training”.

Do ponto de vista computacional, a definição de modalidade envolve um número de conceitos diferentes relacionados a ela, a depender da tarefa que se deve cumprir e de fenômenos específicos que são levantados como, por exemplo, subjetividade, factualidade, certeza/incerteza, *hedging*, como já explicitado.

O trabalho de Saurí, Verhagen e Pustejovsky (2006) objetiva identificar o escopo da modalidade e propor uma sua solução para sua identificação automática. Os autores utilizam a linguagem TimeML (PUSTEJOVSKY et al., 2005) para sua anotação, que codifica com diferentes etiquetas (*tags*), nos níveis lexical e sintático, vários tipos de modalidade. Os eventos são identificados no TimeML como as expressões que participam de uma narrativa em um dado documento que podem ser ordenados temporalmente. No nível sintático, os seguintes valores são levados em conta: *factive* (para eventos implicados ou pressupostos); *counterfactive* (para um evento que pressupõe a não-veracidade de seu argumento); *evidential* (introduzido por eventos reportados ou perceptuais); *negative evidential* (introduzido por eventos reportados ou perceptuais que expressam polaridade negativa); *modal* (para eventos que introduzem uma referência a mundo possível); e *conditional* (para construções condicionais).

Tendo em vista a tarefa de reconhecer implicaturas textuais, Saurí e Pustejovski (2009) apresentam uma ferramenta que fornece eventos com os seus valores de factualidade. Os autores identificam os valores de factualidade, baseados na análise de Horn (1989) para a modalidade epistêmica, em que o valor factual é apresentado pelo par $\langle mod, pol \rangle$, contendo um valor modal (certo, provável, possível e desconhecido) e um valor de polaridade (positivo ou negativo). Há ainda a possibilidade do participante estar completamente descomprometida com a factualidade de um determinado evento ($\langle UN, UN \rangle$).

No entanto, os autores destacam que o valor assinalado para os eventos estão diretamente relacionados com os participantes (fontes da modalidade) em jogo, quer dizer, há um ato de comprometimento em relação à factualidade de um evento, desempenhado por um determinado participante. O conjunto de valores factuais que diferentes participantes assinalam para um evento é denominado perfil de factualidade (*factuality profile*).

Saurí (2008) e Saurí e Pustejovski (2009, 2012) adicionaram ao TimeBank corpus (PUSTEJOVSKY et al., 2005) uma nova camada de informação semântica, o FactBank. Este corpus de eventos, constituído por 208 documentos com 9.488 eventos, é anotado para a factualidade de eventos (ou factividade), definida por Saurí e Pustejovsky (2012, p. 263) como “[...] o nível de informação que expressa a natureza factual de eventualidades mencionadas em um texto. Isto é, expressar se elas correspondem a um fato (...), a uma

possibilidade (...), ou a uma situação que não ocorre no mundo (...) [...]”.⁹⁹ Segundo os autores, a factualidade é resultado da interação entre polaridade e certeza e se relacionam com outras categorias como a modalidade epistêmica, a evidencialidade, a postura epistêmica e *hedging*.

A anotação de modalidade para a caracterização de eventos pode também ser utilizada em processos analíticos automáticos. Baker et al. (2010) desenvolveram um esquema de anotação de modalidade, um léxico da modalidade e dois etiquetadores automatizados construídos a partir do léxico e do esquema de anotação. O esquema é aplicado a exemplos do inglês, com mapeamentos possíveis para o Urdu. Os autores consideram a modalidade como um componente extra-proposicional do significado e argumentam que pode ser tomada de forma mais ampla para incluir vários tipos de atitude (no sentido de “posição”). Assim, segundo eles, a modalidade é definida como “uma atitude por parte do falante em relação a uma ação [...] ou um estado”¹⁰⁰ (BAKER et al., 2010, p. 1) e pode indicar factividade (relacionada ao fato de um evento, um estado ou uma proposição acontecer ou não acontecer); evidencialidade (relacionada à fonte da informação, cf. seção 2.4 do segundo capítulo deste trabalho); ou sentimento (relacionado aos sentimentos negativos ou positivos do falante em relação ao evento, estado ou proposição).

O esquema de anotação reconhece três elementos: o *trigger* (a palavra ou sequência de palavras que expressam modalidade), o *target* (é a unidade de anotação - o evento, estado ou relação no escopo do *trigger*) e o *holder* (o experienciador ou o cognoscente da modalidade). O anotador seleciona apenas o *target* e a modalidade relacionada a ele; nenhuma anotação é feita no *holder* ou no *trigger*. Esta unidade de anotação está contida em uma oração, e o verbo principal da oração é o único a ser marcado.

Para fins de etiquetagem, são consideradas oito modalidades relacionadas estritamente à factividade, que podem, no entanto, se sobrepor às categorias de evidencialidade e sentimento: *Requirement*, *Permissive*, *Success*, *Effort*, *Intention*, *Ability*, *Want*, *Belief*.

Da mesma forma, Ruppenhofer e Rehbein (2012) propõem um esquema de anotação para os verbos modais em inglês aplicados a documentos do MPQA Opinion Corpus¹⁰¹ (WIEBE et al., 2005). Em seu esquema, também levam em conta três elementos de

⁹⁹ No original: “[...] level of information expressing the factual nature of eventualities mentioned in text. That is, expressing whether they correspond to a fact in the world (...), a possibility (...), or a situation that does not hold (...) [...]” (SAURÍ e PUSTEJOVSKY, 2012, p. 263).

¹⁰⁰ No original: “an attitude on the part of the speaker toward an action [...] or state.” (BAKER et al., 2010, p. 1).

¹⁰¹ Disponível em: <http://mpqa.cs.pitt.edu/>. Último acesso em: 28 out. 2013. O corpus contém notícias e outros documentos anotados manualmente para opiniões e outros estados privados, como crenças, especulações, sentimentos. Sua mais nova versão inclui a anotação de atitudes e *targets*. (WILSON, 2008).

significado modal: a expressão modal, a *source* e o *target*. Para a tarefa de anotação, utilizam a ferramenta SALTO¹⁰² (BURCHARDT et al., 2006), e identificam seis categorias (epistêmico, deôntico, dinâmico, optativo, concessivo e condicional) para a anotação de cinco verbos modais (*can/could, may/might, must, ought, shall/should*).

Como o trabalho está restrito aos verbos modais do inglês, é utilizado um número menor de categorias comparado a Baker et al. (2010). Importante destacar que, na descrição das instruções de anotação, os autores distinguem, para os verbos ‘*must*’, ‘*should*’ e ‘*ought*’, dois sentidos – o epistêmico e o deôntico –; nestes casos, para os epistêmicos, não distinguem entre as inferências subjetivas (*‘The light is on. He must be home’*) e as objetivas (*John is 35 and Peter is only a year or two older than John so he must be under 40 still’*) (cf. HUDDLESTON; PULLUM, 2002). No primeiro exemplo, não há uma inferência lógica que leve à conclusão de que John está em casa e, no segundo, chega-se à conclusão sobre a idade de Peter por meio de operações matemáticas. Para o valor deôntico, não fazem qualquer subcategorização no que diz respeito à força imposta sobre os atores, se externa (*‘Dogs must be leashed here. A city ordinance requires it’*) ou interna (*I really must call him. He will be worried’*), e também não fazem distinção entre uma obrigação reportada (*‘Mom says you must go home now. It’s past 10 p.m.’*) e uma obrigação imposta pelo ato de fala (*You must go home now. I want you gone.’*).

Ainda, para o verbo ‘*may*’ identificam três sentidos: epistêmico, deôntico (relacionado à permissão) e optativo (relacionado a um desejo). Para o verbo ‘*can*’, três são os valores: dinâmico (relacionado à habilidade e ao potencial de envolvimento em eventos ou comportamento), deôntico (relacionado à permissão) e epistêmico (relacionado à possibilidade).

Já Matsuyoshi et al. (2010), a partir de estudos sobre modalidade e trabalhos em PLN, apresentam um esquema para a anotação da modalidade estendida de eventos modais para um corpus do japonês, constituído de 50.018 eventos, recolhidos em diferentes recursos, como blogs, documentos da web, o corpus de Murakami et al. (2009) e posts de sites de perguntas e respostas. Um evento é definido como “consistindo de um predicado central e seus argumentos (complementos e adjuntos) em uma sentença”¹⁰³ (MATSUYOSHI et al., 2010, p. 1458).

¹⁰² Esta ferramenta de anotação foi originalmente concebida para a anotação manual de papéis semânticos na moldura teórica da semântica de frames no contexto do projeto Salsa (ERK et al., 2003).

¹⁰³ No original: “(...) consisting of a core predicate and its arguments (complements and adjuncts) in the sentence.” (MATSUYOSHI et al., 2010, p. 1458).

Para a construção do esquema, com vistas para aplicação em PLN, são colocadas quatro condições necessárias: 1) a informação da modalidade deve estar reunida em um único elemento, especificamente o predicado central; 2) o sistema deve ser independente de língua; 3) a polaridade deve comportar duas classes: a polaridade da realidade e a polaridade do ponto de vista da avaliação da fonte, para capturar explicitamente a factualidade do evento; 4) as etiquetas em cada componente não podem ser muito refinadas, porque, segundo os autores, “as classificações de modalidade restrita em Linguística, por exemplo, (Palmer, 2001) e sistemas de lógica modal (Portner, 2009) são muito sofisticadas, e é muito difícil executar analisadores de modalidade estendida nelas baseados no atual estágio de tecnologia em PNL.”¹⁰⁴ (MATSUYOSHI et al., 2010, p. 1458). O esquema de modalidade proposto é constituído de sete componentes: *Source*, *Time*, *Conditional*, *Primary modality type* (*assertion*, *volition*, *wish*, *imperative*, *permission*, *interrogative*), *Actuality*, *Evaluation* e *Focus*.

Este sistema encontrou alguns desafios como a representação de advérbios de frequência, de construções de dificuldade e de potencial. No entanto, apresentou um acordo entre dois anotadores razoavelmente aceitável (em média um $\kappa=0.71$ para os componentes).

Comparado aos projetos de Saurí et al. (2006), Saurí e Pustejovski (2009, 2012) e ao de Baker (2010), o sistema proposto por Matsuyoshi et al. (2010), apesar de também anotar eventos, é mais rico em termos de elementos a serem anotados. Isso se deve ao fato de que os sistemas construídos tanto por Saurí e seus colegas, quanto por Baker e colegas atendem a um propósito específico e, dessa forma, a escolha dos componentes a serem marcados fica condicionada a este objetivo final.

Em uma perspectiva mais ampla do que seja modalidade, Wiebe e suas colaboradoras (2005) exploram um esquema de anotação de opiniões e emoções, baseado em estudo de um corpus de artigos da imprensa internacional, o MPQA Opinion Corpus (WIEBE; WILSON; CARDIE, 2005), composto de 10.657 sentenças em 535 documentos em língua inglesa. Segundo as autoras (2005, p. 1), a motivação para a identificação e extração de opiniões, reconhecimento de emoções e análise de sentimentos nasce do “desejo de fornecer ferramentas para analistas da informação nos domínios governamental, comercial e político, que querem rastrear automaticamente atitudes e sentimentos em notícias e em fóruns on-line”.

¹⁰⁴ Nas palavras dos autores: “[...] classifications of restricted modality in Linguistics, e.g., (Palmer, 2001) and systems of modal logic (Portner, 2009) are too sophisticated, and it is very difficult to implement analyzers of extended modality based on them with the current level of technology in NLP.” (MATSUYOSHI, 2010, p. 1458).

As pesquisadoras propõem um esquema de anotação refinado para a etiquetagem de componentes e propriedades de opiniões, sentimentos, emoções, estados privados, especulações, que abriga sob o rótulo de estados privados (*'private states'*).

Os estados privados são definidos como “estados internos que não podem ser diretamente observados pelos outros”¹⁰⁵ (WIEBE et al., 2005, p. 2). Para cada expressão de estado privado é definido um frame de estado privado, que se constitui pelos seguintes elementos: a *source*¹⁰⁶ (cujo estado privado está sendo expresso); o *target* (sobre o quê é o estado privado); as propriedades que envolvem: *intensity*, *significance* e *type of attitude*.

Há dois tipos de frames de estado privado: (a) elementos subjetivos expressivos (*'expressive subjective elements'*): representam elementos subjetivos expressivos; (b) subjetivos diretos: representam menções explícitas a estados privados (*'explicit mentions of private states'*) e eventos de fala que expressam estados privados (*'speech events expressing private states'*).

O esquema proposto por Wiebe, Wilson e Cardie é bastante detalhado, assim como o de Matsuyoshi et al. (2010) e aplicado a um grande corpus. Como se pode observar, a noção de modalidade nos esquemas apresentados acima não segue a noção corrente na literatura sobre o tema (cf. PALMER, 1986; ou PORTNER, 2009). Os trabalhos que apresento em seguida, ao contrário, se apóiam na ideia de modalidade restrita e na definição, se assim posso dizer, mais tradicional da categoria, qual seja “a atitude do falante”, além de considerar, geralmente, em sua tipologia, a oposição entre significados epistêmicos e não-epistêmicos e não entre informação factual e não-factual (o que, evidentemente, não exclui a análise destas características). Estes próximos projetos não foram desenvolvidos com um objetivo específico e definido de aplicação em PLN, no entanto, reconhecem que são pontos de partida para cumprir as tarefas de extração de informação, mineração de opinião, resposta a perguntas, por exemplo.

Nirenburg e McShane (2008) anotaram um corpus com informação sobre a modalidade, no âmbito do projeto OntoSem¹⁰⁷ (NIRENBURG; RASKIN, 2004) e desenvolveram um analisador, alimentado com textos não-tratados, e realizaram várias tarefas

¹⁰⁵ Tradução para: “[...] internal states that cannot be directly observed by others.” (WIEBE et al., 2005, p. 2).

¹⁰⁶ Uma propriedade importante das *sources* é que elas podem estar encaixadas, isto é, eventos privados e eventos de fala podem estar frequentemente encaixados um no outro.

¹⁰⁷ Este projeto foi desenvolvido para o tratamento computacional da representação do significado de um texto. Nas palavras dos autores: “[o]ntological semantics is a theory of meaning in natural language and an approach to natural language processing (NLP) which uses a constructed world model, or ontology, as the central resource for extracting and representing meaning of natural language texts, reasoning about knowledge derived from texts as well as generating natural language texts based on representations of their meaning.” (NIRENBURG; RASKIN, 2004, p. 10).

de análise linguística, a modalidade incluída. Para a codificação da modalidade, quatro propriedades são consideradas: o tipo de modalidade (MODALITY TYPE), o valor escalar (SCALAR VALUE), o escopo (SCOPE), e a-quem-é-atribuída (ATTRIBUTED-TO).

Os tipos de modalidade centrais incluem: (i) factividade-epistêmica (*‘epistemic-factivity’*); (ii) crença (*‘belief’*); (iii) obrigação (*‘obligative’*); (iv) permissão (*‘permissive’*); (v) potencial (*‘potential’*); (vi) avaliativo (*‘evaluative’*); (vii) intencional (*‘intentional’*); (viii) epitéutico - grau de sucesso (*‘epiteuctic – degree of success’*); (ix) esforço (*‘effort’*); (x) volição (*‘volitive’*).

A cada um destes significados modais estão distribuídos valores que variam em uma escala de zero a um, considerados qualquer valor decimal válido neste intervalo. Por exemplo, para a factividade-epistêmica, o valor ‘0’ corresponde a “não aconteceu” (*‘didn’t happen’*) e o valor ‘1’ a “definitivamente aconteceu” (*‘definitely happen’*) ou, para o epitéutico - grau de sucesso, o ‘zero’ equivale ao “fracasso” (*‘fail’*) e o “um”, ao “completamente bem-sucedido” ou “sucesso absoluto” (*‘succeed fully’*).

Assim como nos outros projetos, o escopo (‘SCOPE’) é o predicado afetado pela modalidade, e a propriedade a-quem-é-atribuída (‘ATTRIBUTED-TO’) aponta para a quem a modalidade é atribuída, sendo o falante o valor padrão. É importante notar que as propriedades da modalidade estão explicitamente presentes no texto.

O sistema desenvolvido para o português europeu (HENDRICKX et al., 2012a; 2012b) baseia-se no esquema de anotação proposta por Nirenburg e McShane (2008). Enquanto no OntoSem a informação sobre a modalidade é um módulo semântico nas entradas lexicais que expressam modalidade, o modelo para o PE centra-se na anotação de eventos, e não de entidades. São anotadas as expressões modais, que incluem verbos, advérbios, nomes, adjetivos, sintagmas preposicionais e orações, entretanto a tarefa é restrita à anotação de sentenças.

Entre os componentes anotados estão: o *trigger*, o elemento que expressa o valor modal¹⁰⁸; o *target*, a expressão no escopo do *trigger* (corresponde ao atributo ‘SCOPE’); a *source of the modality*, um agente ou experienciador (correspondente à propriedade ‘ATTRIBUTED-TO’); e a *source of the event mention*, que pode ser o falante ou o escritor¹⁰⁹.

¹⁰⁸ Para o *trigger* são assinalados dois atributos: os valores modais e a polaridade (positiva ou negativa).

¹⁰⁹ A decisão de anotar duas *sources* se deve à necessidade de se distinguir entre ‘aquele que produz a sentença’ e ‘aquele que expressa a modalidade’. Em muitas ocorrências estes dois elementos são coincidentes, mas não necessariamente este é o caso, como em “*Os portugueses* necessitam, em média, de 180 contos por mês para a manutenção de uma família de quatro pessoas”. Neste exemplo, “*Os portugueses*” é a entidade com a necessidade interna disparada pelo verbo “necessitar”. O produtor do evento não está explícito aqui, e assume-se que é o produtor da frase.

Para os significados modais, são considerados sete valores e alguns subvalores correspondentes, a saber:

Valores	Subvalores
Epistemic	knowledge belief doubt possibility interrogative
Deontic	obligation permission
Participant-internal	necessity: personal needs capacity: personal capacity
Evaluation	evaluation of the proposition
Volition	hopes and wishes
Effort	attempt of the participant to make sth. happen
Success	results of the commitment of the participant

Tabela 5.1 – Valores e subvalores para o esquema do português europeu

Se comparado aos tipos de modalidade selecionados no OntoSem, é possível observar grande semelhança entre os dois esquemas. O tipo “*intentional*” foi incluído no valor “*effort*” e o subvalor “*doubt*” do PE, incluído nos epistêmicos, é parte do tipo “*belief*” no OntoSem.

A tarefa de anotação foi empreendida primeiramente por um anotador (em papel) e, em um segundo momento, foi revisada por um segundo anotador, utilizando o software MMAX2 (MÜLLER; STRUBE, 2006)¹¹⁰, em uma amostra de aproximadamente 2000 sentenças da parte escrita do Corpus de Referência do Português Contemporâneo – CRPC (GÉNÉREUX et al., 2012)¹¹¹.

Os anotadores utilizam a estratégia min-max (cf. FARKAS et al., 2010). Para o *trigger*, é anotada a menor unidade possível, já para o *target*, o máximo de unidades é

¹¹⁰ Disponível para download em <http://mmax2.net>. Esta ferramenta de anotação será detalhada na seção 5.3.1.1 deste capítulo.

¹¹¹ Este é um corpus de aproximadamente 312 milhões de palavras, entre textos e escritos e registros orais, constituído por uma ampla gama de textos de diferentes gêneros textuais em todas as variedades da língua portuguesa. Disponível em: <http://www.clul.ul.pt/pt/recursos/183-reference-corpus-of-contemporary-portuguese-crpc>. Último acesso em: 13 nov. 2013.

considerado e este elemento também pode ser anotado em descontinuidade. No que diz respeito às *sources*, são marcados sintagmas nominais inteiros ou verbos. No esquema proposto, ainda há um campo adicional para comentários (*Comment*) para que quaisquer dificuldades sejam registradas, principalmente casos de ambiguidade.

Em trabalho mais recente (MENDES et al., 2013), foi integrada a este esquema de anotação a interação entre o foco e a modalidade, especialmente o foco envolvendo partículas exclusivas como “só”. Os autores extraíram os contextos e aplicaram o esquema a 100 sentenças de documentos escritos do CRPC.

Dois projetos desenvolvidos na Georgetown University, os trabalhos de Rubinstein e colegas (2013), para a anotação da modalidade no inglês, e o de Cui e Chi (2013), para a anotação de dados do chinês, seguem a mesma orientação teórica (cf. PORTNER, 2009) e utilizam um quadro bastante similar de elementos a serem anotados.

Rubinstein et al. (2013) propõem um esquema de anotação de sentidos modais, construído a partir de trabalhos anteriores e acrescentado de novos traços, para a anotação do MPQA Corpus. No esquema, os anotadores, em primeiro lugar, codificavam cada modal em três categorias mais amplas: Epistêmica ou Circunstancial (*Epistemic* ou *Circumstantial*); Habilidade ou Circunstancial (*Ability* ou *Circumstantial*); e Prioridade (*Priority*). Em seguida, sete tipos modais mais refinados foram individualizados: Epistêmica (*Epistemic*); Circunstancial (*Circumstantial*); Habilidade (*Ability*); Deontica (*Deontic*); Boulética (*Bouletic*); Teleológica (*Teleological*); e Bulética/Teleológica (*Bouletic/Teleological*).

Cui e Chi (2013) descrevem em seu trabalho a tentativa de refinar a anotação de alguns aspectos dos modais para o Penn Chinese Treebank e apontam seus primeiros resultados para a primeira fase de anotação. Para a tarefa de anotação, realizada por dois anotadores, também foi utilizado o software MMAX2 (MÜLLER; STRUBE, 2006), com um esquema de dez traços e uma lista de inicial com onze índices modais que, à medida que o projeto avança, pode crescer. Como contribuição, os autores argumentam que o projeto demonstrou que “é possível usar um esquema e um conjunto de instruções para a anotação translinguística.”¹¹² (CUI; CHI, 2013, p. 8).

¹¹² No original: “(...) it is possible to use one scheme and set of guidelines for cross-linguistic annotation.” (CUI e CHI, 2013, p. 8).

5.2 Panorama geral (ou quem identifica e/ou anota o quê):

Em resumo, os componentes identificados e/ou anotados nos projetos apresentados na seção anterior:

Autores	Trigger	Target	Source	Factuality	Certainty	Transition of certainty	Focus	Time	Conditional
Saurí et al., 2006, 2007, 2009, 2012	x	x	x	x	x				
Baker et al., 2010	x	x	x	x					
Matsuyoshi et al., 2010	x	x	x		x	x	x	x	x
Wiebe et al., 2005	x	x	x	x	x				
Niremburg; McShane, 2008	x	x	x						
Hendrixx et al., 2012 Mendes et al., 2013	x	x	x				x		
Rubinstein et al. 2013	x	x	x						
Cui; Chi, 2013	x	x	x						

Tabela 5.2 – Elementos identificados e/ou anotados em diferentes projetos de anotação semântica

Entre os diferentes projetos, apesar de a terminologia usada para a categorização dos significados modais diferirem, é possível encontrar correspondências entre os valores modais correntes na literatura linguística e aqueles valores denominados não-padrão. Abaixo, resumidamente, elenco na Tabela 5.2 todos os valores assinalados nos esquemas de anotação apresentados. Na coluna da esquerda, listo os valores modais utilizados na tradição linguística (cf. PALMER, 2001) e na coluna da esquerda os valores correspondentes encontrados nos esquemas aqui apresentados e seus respectivos autores.

Valores modais	
tradição linguística	esquemas de anotação
Epistemic	factuality (SAURÍ et al., 2006; BAKER et al., 2010) actuality (MATSUYOSHI et al., 2010) evidentiality (SAURÍ et al., 2006) assertion (MATSUYOSHI et al., 2010) transition of certainty (MATSUYOSHI et al., 2010) degrees of possibility (SAURÍ et al., 2006) potential (NIREMBURG; McSHANE, 2008) belief (SAURÍ et al., 2006; BAKER et al., 2010 ; NIREMBURG; McSHANE, 2008) requirement (BAKER et al., 2010) interrogative (MATSUYOSHI et al., 2010) evaluation (HENDRICKX et al., 2012b; NIREMBURG; McSHANE, 2008)
Deontic modality	command (SAURÍ et al., 2006) requirement (BAKER et al., 2010) imperative (MATSUYOSHI et al., 2010) obligative (NIREMBURG; McSHANE, 2008) permissive (BAKER et al., 2010; NIREMBURG; McSHANE, 2008) permission (MATSUYOSHI et al., 2010)
Dynamic modality Participant-internal modality	ability (BAKER et al., 2010) success (BAKER et al., 2010) epiteuctic (NIREMBURG; McSHANE, 2008) requirement (BAKER et al., 2010) volition (MATSUYOSHI et al., 2010 ; McSHANE & NIREMBURG, 2008 ; HENDRICKX et al., 2012 ; NIREMBURG; McSHANE, 2008) expectation (SAURÍ et al., 2006) want (BAKER et al., 2010) wish (MATSUYOSHI et al., 2010) intention (BAKER et al., 2010 ; NIREMBURG; McSHANE, 2008) attempting (SAURÍ et al., 2006) effort (BAKER et al., 2010; HENDRICKX et al., 2012 ; NIREMBURG; McSHANE, 2008)

Tabela 5.3 – Correspondência entre valores na tradição linguística e nos esquemas de anotação

5.3 Proposta para anotação da modalidade no C-ORAL-BRASIL: o projeto MASS (Modal Annotation in Spontaneous Speech)

Este projeto de esquema de anotação de modalidade em dados orais do português brasileiro segue diretamente o esquema proposto para o português europeu (HENDRICKX et al., 2012a; 2012b; MENDES et al., 2013) e, igualmente, inspira-se em outros esquemas de anotação previamente explorados para a língua inglesa (BAKER et al., 2010; SAURÍ et al., 2006; RUBINSTEIN et al., 2013), o japonês (MATSUYOSHI et al., 2010) e o chinês (CUI; CHI, 2013).

Este esquema se baseia nos pressupostos da Teoria da Língua em Ato e tem como unidade de referência de análise, portanto, o enunciado e as unidades informacionais (cf. CRESTI, 2000, v. capítulo 3, seção 3.1). Desta forma, difere-se de outros projetos, uma vez que não centra a sua análise na diamesia escrita e, portanto, não fornece uma anotação no domínio da sentença.

Nas subseções seguintes, apresento a metodologia empregada para a anotação, no que diz respeito à preparação dos textos, o corpus utilizado e a ferramenta para anotação — o software livre MMAX2 (MÜLLER; STRUBE, 2006); sigo com a escolha dos valores modais e os elementos a serem anotados; a marcação da polaridade aplicada ao elemento *Trigger*; exemplos de anotação; os resultados e, finalmente, algumas considerações sobre o processo de anotação.

5.3.1 Metodologia:

5.3.1.1 O software de anotação MMAX2:

O MMAX2 (MÜLLER; STRUBE, 2006) é um software livre para anotação linguística de corpora em múltiplos níveis. Esta é uma ferramenta flexível para criar, navegar por e visualizar as anotações linguísticas. MMAX oferece uma interface visual para anotar sentenças pela marcação sequências textuais e criação de links entre os elementos marcados. O MMAX está escrito em Java (por razões de independência de plataforma) e as anotações são armazenadas em XML.

Abaixo, na Figura 5.1, um exemplo da tela principal do programa que se refere a um dos arquivos anotados (*bfamcvo4*) deste projeto de anotação de modalidade no corpus C-ORAL-BRASIL:



Figura 5.1 – Tela principal do software MMAX2

Na tela, as linhas azuis são as sequências que contêm os elementos anotados (*trigger*, *target*, *source of the modality*, *source of the event*)¹¹³. Em amarelo, temos marcado um dos componentes anotados (no caso, o *trigger*); em cinza os outros elementos da sequência modal; e, por último, em verde, os *links* entre os elementos que compõem um *set* modal.

Segundo Müller e Strube (2006), são cinco os passos cruciais de um projeto de anotação, o que chamam de “ciclo de vida da anotação”, a saber:

- (i) preparação do corpus;
- (ii) definição e formalização da tarefa de anotação;
- (iii) manual de anotação;
- (iv) checagem da viabilidade da anotação;
- (v) utilização da anotação completa.

Apresento, nas próximas seções, alguns passos do ciclo de vida do projeto MASS (*Modal Annotation in Spontaneous Speech*).

¹¹³ Estes elementos serão descritos em detalhes nas seções subsequentes deste capítulo.

5.3.1.2 A preparação dos textos

Para fins de pré-processamento e utilização do software MMAX2, foi necessária a limpeza dos arquivos em extensão “.txt”. Os elementos a serem eliminados foram detectados através da própria ferramenta de busca do editor de texto e os textos foram limpos manualmente. O corpus C-ORAL-BRASIL é transcrito segundo o sistema CHAT (MAcWHINNEY, 2000), que fornece um formato padronizado para produzir transcrições computadorizadas de interações conversacionais face a face, implementado para a anotação prosódica (MONEGLIA; CRESTI, 1997), e é segmentado em enunciados e unidades tonais.

Algumas das convenções da segmentação prosódica e também notações da transcrição tiveram que ser removidas, tais como interrupções, *retractings*, sobreposições, fragmentos fonológicos, números de enunciados e os asteriscos que precedem os nomes de cada participante. Abaixo segue um resumo dos elementos eliminados:

- (a) números dos enunciados [xxx]: os arquivos são transcritos por enunciados e a cada enunciado corresponde um número.

Ex.: *LUA: **[72]** tá parecendo aqueas história daquele livro que cê tava contando
//=COM=\$

- (b) asterisco (*): o asterisco indica a “linha principal”, isto é, a linha que codifica um único enunciado produzido por um determinado falante. Este sinal antecede o código em três letras maiúsculas que indicam o participante produtor daquele enunciado.

Ex.: *CES: [104] uhn /=PHA= tô achando que não //COM=\$

- (c) *ampersand* ou e-comercial (&): este símbolo é usado para indicar que o material seguinte é apenas um fragmento fonológico ou um pedaço de palavra. Esta forma de notação é usada quando o falante gagueja ou para de falar antes de completar uma palavra reconhecível (falsos começos).

Ex.: *ANE: [224] **&**he /=TMT= essa localização aqui eu nũ acho ruim //COM=\$

(d) anguladores (< >): sinalizam que o material locutivo de um falante em um determinado turno contido entre os anguladores, está sendo produzido ao mesmo tempo que o material angulado do seu interlocutor no turno seguinte.

Ex.: *ANE: [298] ≤às vezes≥ [2]=EMP= às vezes fica <gente> //COM=\$
*CES: [299] ≤pode entrar≥ //COM=\$

(e) sinal de adição (+): indica uma interrupção, seja por tomada de turno pelo interlocutor ou por decisão do próprio falante.

Ex.: [99] então /=INP= eu /=TOP= eu sei /=COB= Mara /=ALL= que /=DCT= ela me tirou /=i-COB= assim ± =PAR=\$

(f) [/n]: indica *retracting*. O numeral “n” ao lado da barra indica o número de palavras envolvidas no *retracting* e canceladas pelo falante. É usado quando o falante começa a dizer alguma coisa, para, repete o sintagma básico, muda a sintaxe, mas mantém a mesma ideia; pode envolver também a completa reformulação da mensagem sem quaisquer correções específicas; ou ainda indicar que o falante começou a dizer alguma coisa, parou e repetiu o material anterior sem qualquer mudança.

Ex.: [147] essa história &de [1]=SCA= da menina /=TOP= eu tenho certeza que e' nã te conta não //COM=\$

Apresento abaixo, em (a), um exemplo de enunciado antes da limpeza, com alguns dos elementos eliminados e, a seguir, em (b), o resultado da limpeza deste mesmo enunciado¹¹⁴:

(a) *JOR: [17] e aí /=TOP= eu consegui /=i-COB= &he /=TMP= com a experiência que eu tinha dentro da multinacional /=PAR= concorrer à vaga e &f [1]=SCA= isso me facilitou e eu passei pra área comercial da empresa pra vender /=COB_s= disjuntores /=CMB= transformadores /=CMB= motores de /=INT= corrente contínua /=CMB= corrente alternada /=CMB= isoladores /=CMB= e /=TXC= relés de proteção secundária /=CMB= e assim foi iniciando a minha vida comercial //COM

(b) JOR: e aí / eu consegui / he / com a experiência que eu tinha dentro da multinacional /=PAR= concorrer à vaga e f isso me facilitou e eu passei pra área comercial da empresa pra vender / disjuntores / transformadores / motores de / corrente contínua / corrente alternada / isoladores / e / relés de proteção secundária / e assim foi iniciando a minha vida comercial //

¹¹⁴ As unidades informacionais já não constavam na versão dos arquivos de texto utilizada para o pré-processamento.

5.3.1.3 Definição e formalização da tarefa de anotar:

(a) Os *markables* (marcadores):

Um *markable* é o componente básico de anotação no MMAX2. É uma porção de texto que é marcada, normalmente uma sequência de texto ou partes descontínuas de um texto.

Estes marcadores podem se expandir em diferentes sentenças, podem se sobrepor, podem corresponder exatamente a mesma porção de texto, e podem estar anotados em descontinuidade, isto é, duas partes do texto podem estar divididas e serem anotadas como um único marcador.

Em nosso projeto, defino quatro tipos de *markables*, os quais descreverei em detalhes mais adiante, que são: o *trigger*, o *target*, a *source of the event mention* e a *source of the modality*. Estes componentes, em conjunto, formam um “*set modal*”. Os componentes *trigger* e *target* possuem traços que os distinguem. O primeiro carrega os traços *modal_value*, *polarity*, *information unit* e *comment*; o segundo, os traços *information unit* (i.e., a unidade informacional em que o componente está contido) e *polarity*.

Para a criação de um “*set modal*” é necessário, em primeiro lugar, criar cada um dos *markables* e, em seguida, estabelecer os *links* entre eles. Para a criação dos *markables* (Figura 5.2), **arraste** com o mouse sobre o elemento que se deseja marcar, **clique com o botão direito** sobre o texto selecionado, **selecione** “*Create Markable at level 'modal'*”, por último, na caixa de anotação, **faça a escolha** dos atributos apropriados para cada *markable*.

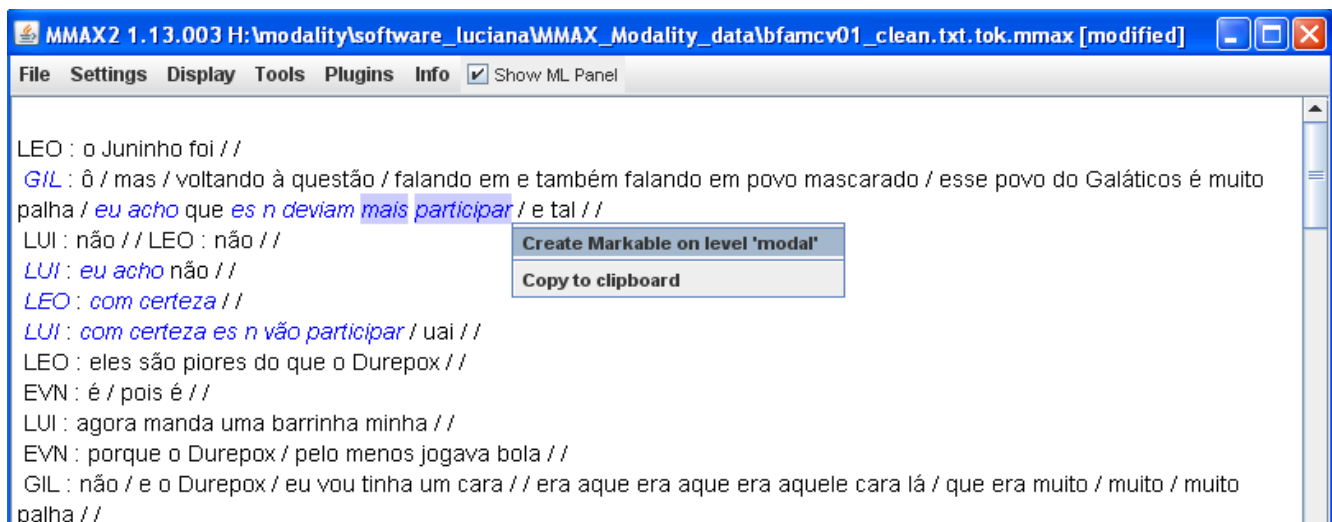


Figura 5.2 – Criação de *markable*

Os passos para estabelecer os *links* entre os *markables* são os seguintes (Figuras 5.3 e 5.4): **clique** sobre o componente marcado como o *trigger*, que ganha **destaque em amarelo**; em seguida, **clique com o botão direito** sobre um dos componentes, por exemplo, a *source of the modality*, e **selecione** “Mark as link”. **Repita** o mesmo procedimento para os outros componentes. Uma linha verde vai surgir entre os *markables* e o *set* modal está constituído.

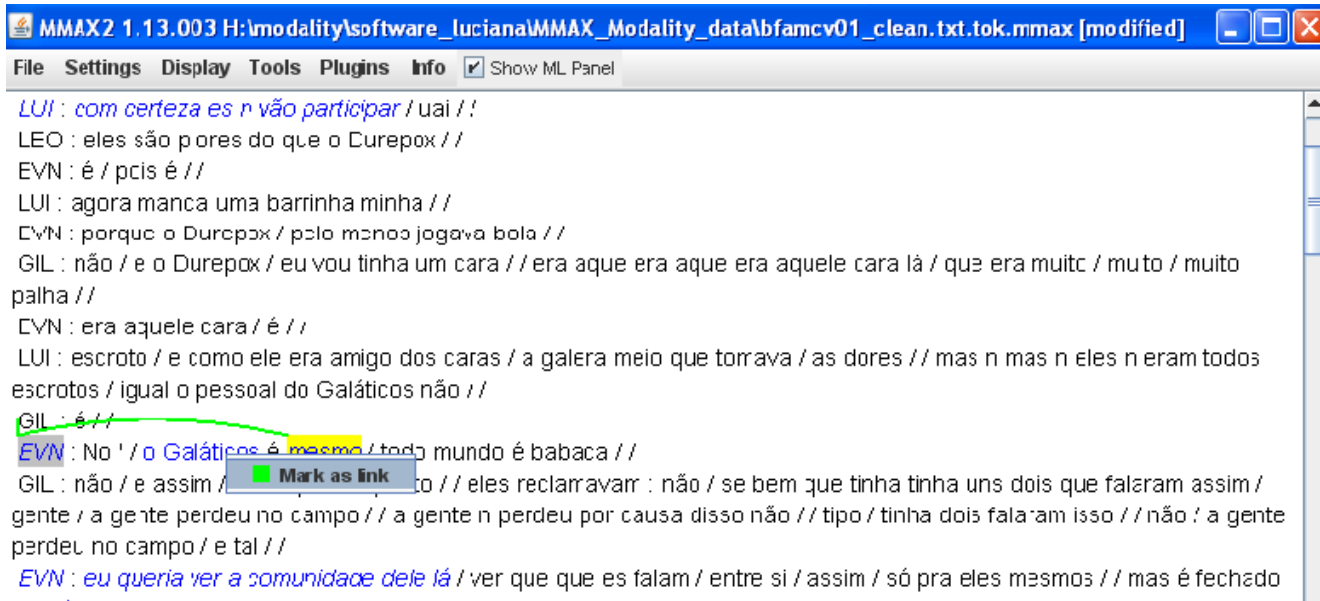


Figura 5.3 – Estabelecer *links* entre *markables*

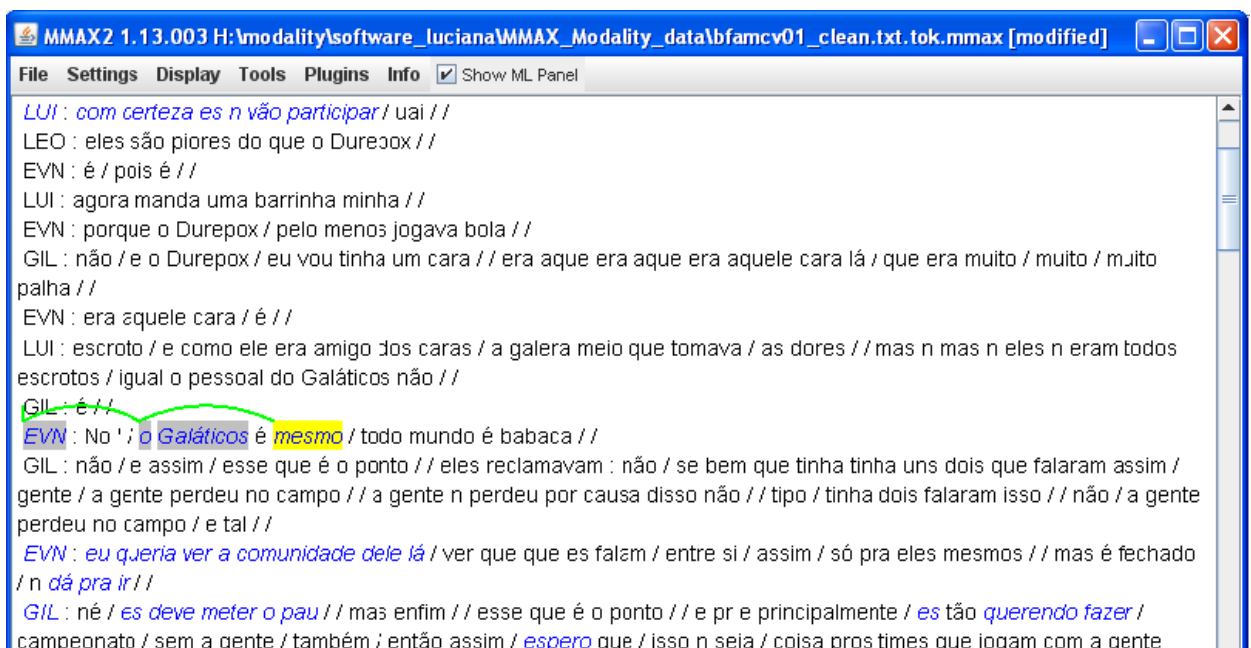


Figura 5.4 – *Links* entre *markables* estabelecido

No campo “*modal_class*” da caixa de anotação (Figura 5.5), aparecerá “*set_x*”, em que “*x*” corresponde ao número do *set* criado.

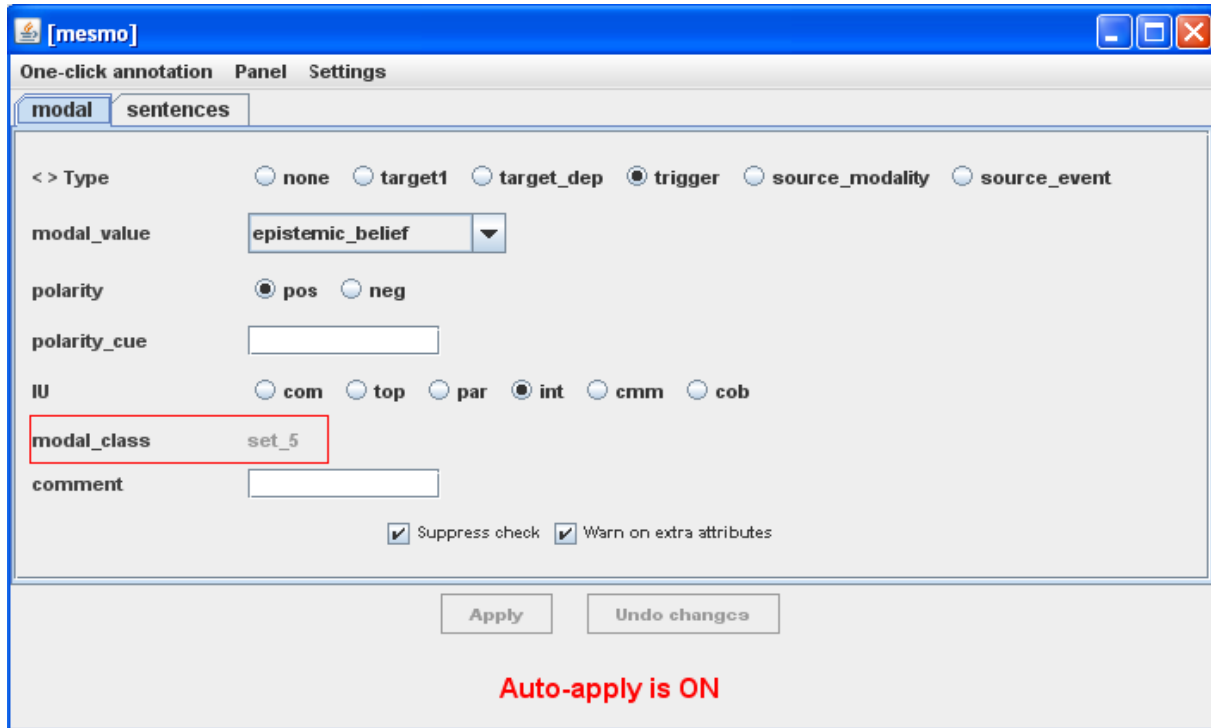


Figura 5.5 – *Modal-class*: número do *set* modal

No caso de haver uma mesma sequência de texto com dois ou mais papéis diferentes, isto é, uma sequência de texto correspondente a dois ou mais *markables*, basta proceder da forma indicada acima. Para selecionar os *links* apropriados, **clique** em cada uma das porções de texto disponíveis para criar o *set* modal.

(b) Regras básicas de anotação:

(b.1) Criar: como explicitado acima, para um novo *markable* ser criado (Figura 5.2), é necessário selecionar uma porção de texto, arrastando o mouse com o botão esquerdo e clicando com o botão direito na seleção. Outros elementos não devem estar selecionados, do contrário este material será adicionado ao *markable* em amarelo.

(b.2) Remover: um *link* pode ser removido (Figura 5.6), primeiro selecionando o componente com o botão esquerdo do mouse, depois, com o botão direito sobre a seleção, escolher “*Unmark as link*”.

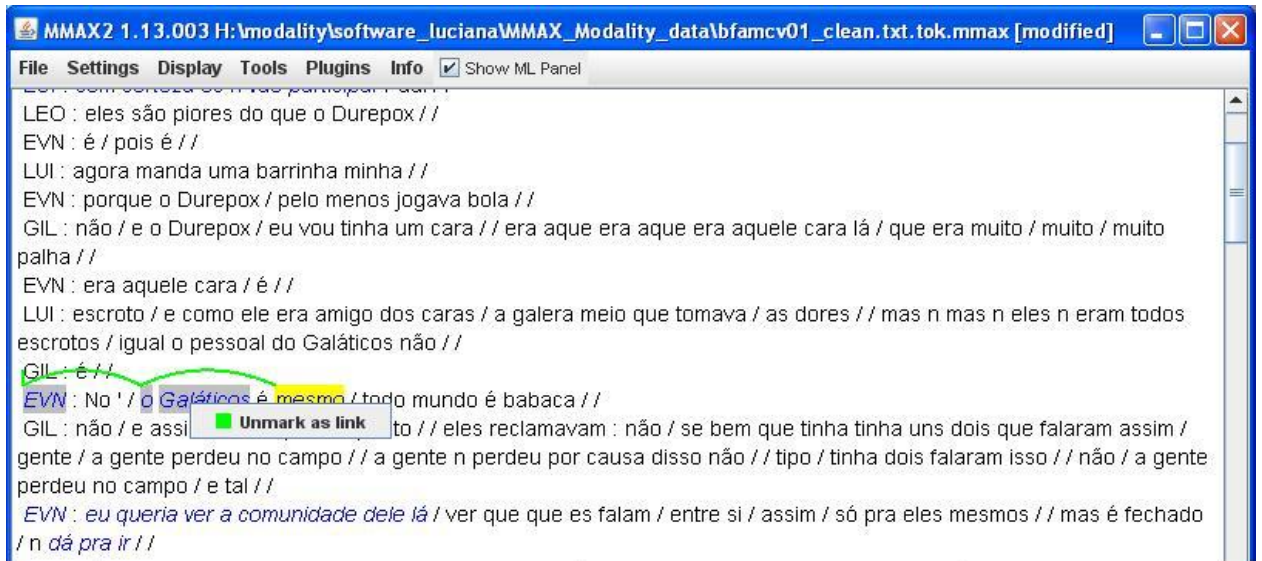


Figura 5.6 - Remover o link

(b.3) Apagar: para apagar completamente um *markable* (Figura 5.7), basta clicar em cima do elemento a ser apagado. O elemento não pode estar em amarelo.

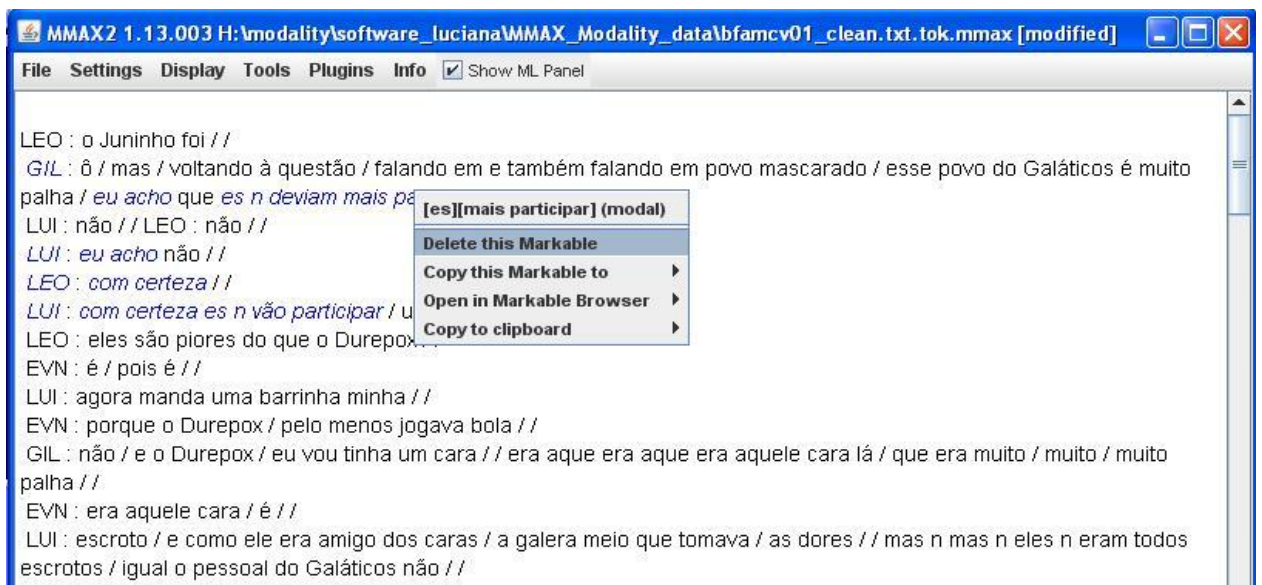


Figura 5.7 - Apagar um markable

(b.4) *Markable* descontínuo: a partir de um *markable* existente, é possível adicionar a ele uma porção de texto não-contínua. É necessário clicar sobre o *markable* criado, que fica em amarelo, então selecionar com o botão esquerdo a segunda porção de texto a ser adicionada (Figura 5.8), clicar com o botão direito sobre esta seleção e escolher “Add to this markable”. Para o caso de toda a porção de texto adicionada ser removida ou apenas parte dela (Figura

5.9), basta selecionar o elemento que deseja desvincular do *markable*, clicar com o botão direito em cima dele e escolher “*Remove from this markable*”.

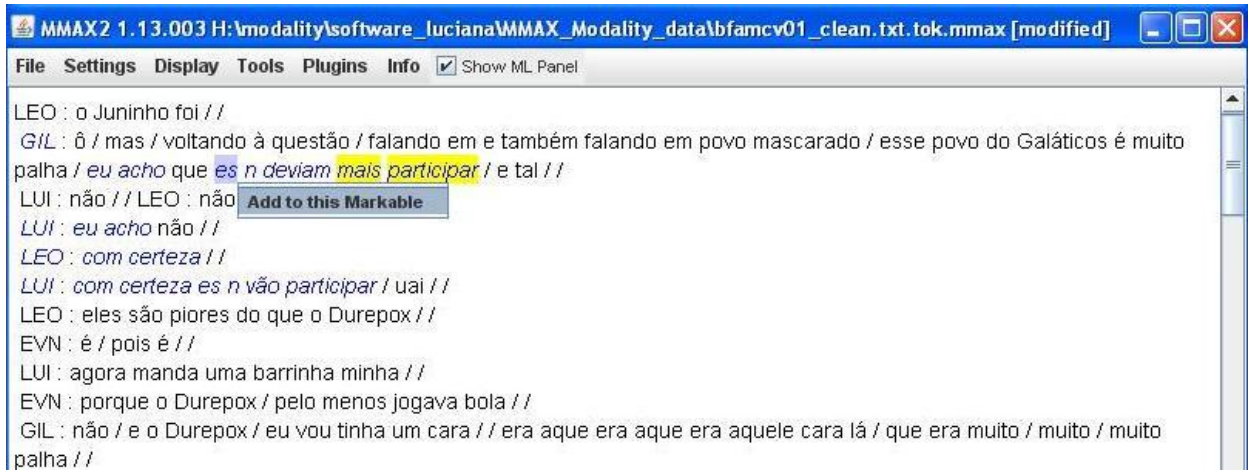


Figura 5.8 - Adicionar uma sequência ao *markable*

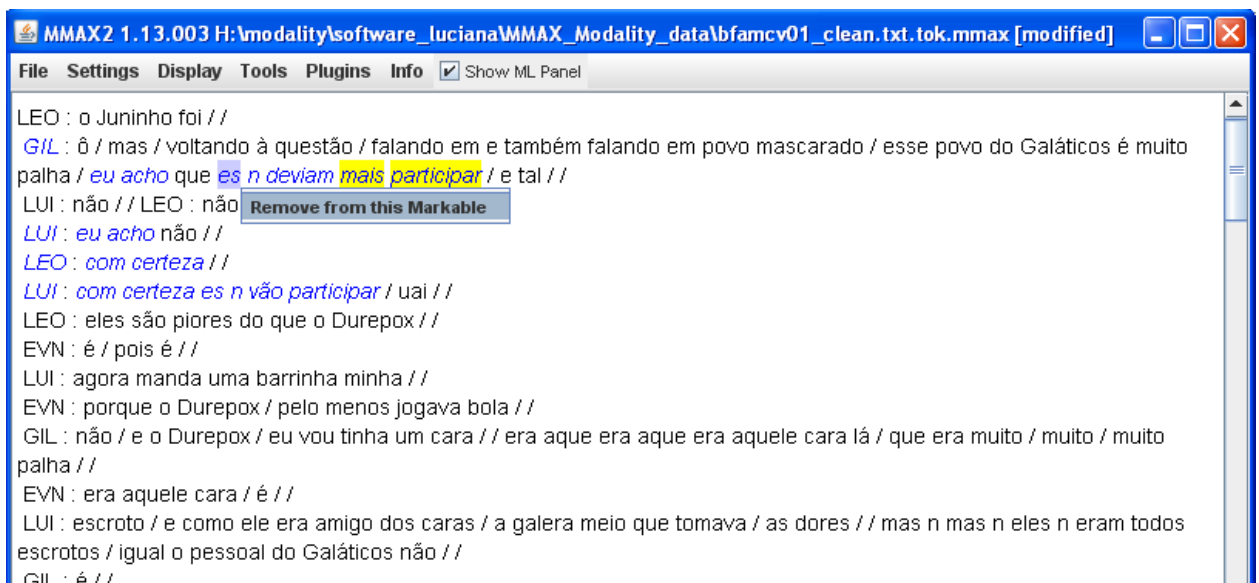


Figura 5.9 – Remover sequência de um *markable*

(c) Interface visual do MMAX2 e configurações básicas iniciais:

O MMAX2 possui uma apresentação visual bastante amigável e algumas configurações básicas devem ser aplicadas para se iniciar a tarefa de anotação. Ao carregar o software (“mmax.bat”), vão surgir quatro caixas na tela: (a) uma diz respeito ao que está sendo rodado no sistema; (b) a caixa de anotação dos valores e atributos dos *markables*; (c) a tela principal com o texto a ser anotado; e (d) o painel de controle do nível do *markable*.

Também surgirá uma janela em “*pop-up*” perguntando se o anotador deseja validar as anotações.

Abaixo apresento instantâneos das quatro principais telas a serem configuradas como e quando necessário pelos anotadores.

(a.1) *Snapshot* da tela de anotação dos valores e atributos:



Figura 5.10- Caixa de anotação de valores e atributos

Nesta tela, em “*Settings*”, selecione “*Auto-apply on*” para evitar a ação de “*Apply*” a cada modificação. Quando os campos “*comment*” ou “*polarity_cue*” forem modificados, o “*Apply*” deve ser feito manualmente.

(a.2) *Snapshot* da tela principal de anotação:

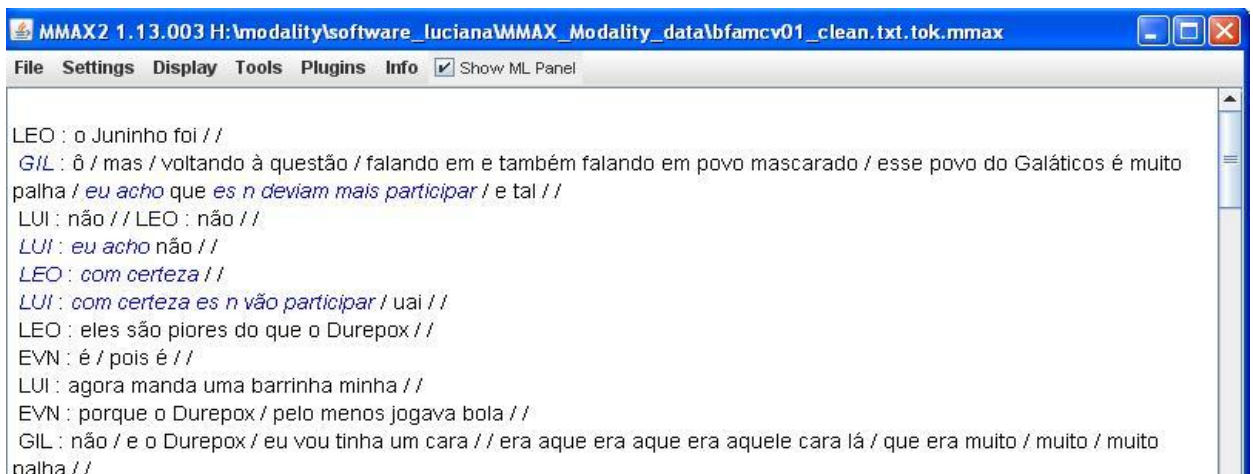


Figura 5.11 - Tela principal de anotação

No caso de os anotadores considerarem o espaçamento entre as linhas de anotação pequeno, é possível utilizar a opção “Font” > “Line spacing” para a configuração desejada.

Na parte esquerda superior desta tela, no menu “File”, há a opção “Auto-save”, onde o intervalo para salvar as modificações realizadas pode ser controlado. O ideal é configurá-lo para 5 ou 10 minutos.

(a.3) *Snapshot* do painel de controle do nível do *markable*:

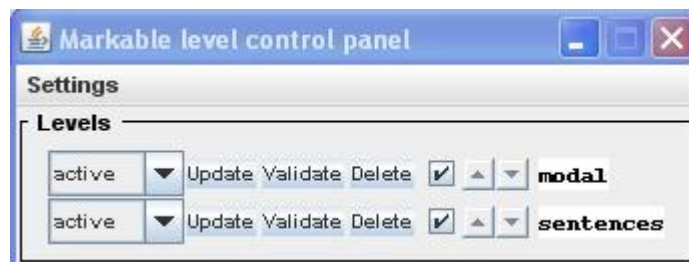


Figura 5.12 - Painel de controle

Nesta caixa, temos dois níveis: “modal” e “sentence”. Este último nível deve estar configurado como “visible” e não “active”, uma vez que nenhuma anotação no nível da sentença será realizada.

(a.4) *Snapshot* da tela de validação da anotação:

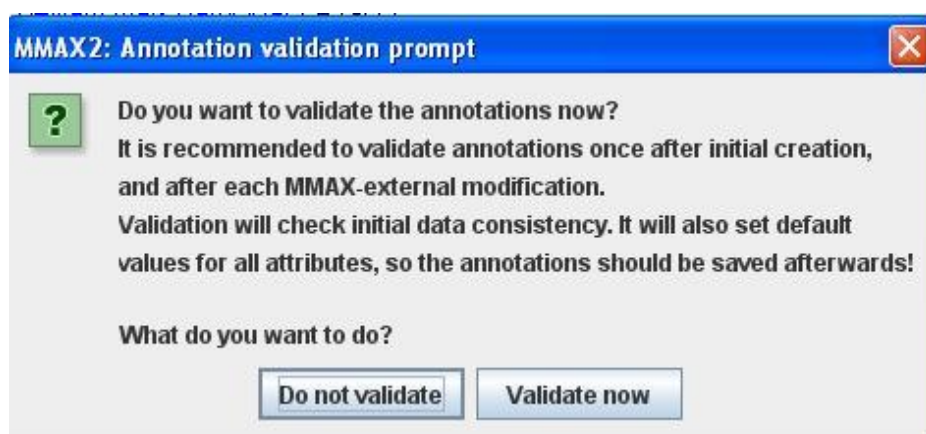


Figura 5.13 - Tela de validação da anotação

O anotador deve clicar em “Do not validate” nesta fase inicial.

5.3.1.4 O esquema de anotação em versão XML:

No MMA2, qualquer tipo de anotação pode ser representada pela associação de elementos individuais com atributos e pela conexão dos vários *markables* por meio de relações (MÜLLER; STRUBE, 2006, p. 202). Abaixo apresento os atributos e as relações definidos no esquema de anotação de modalidade para a fala espontânea em sua versão XML.

No primeiro nível está definido o atributo *Type* e os seus respectivos valores (*none*, *target1*, *target_dependent*, *trigger*, *source of the event*, *source of the modality*) que apareceram como “*nominal_buttons*” (botões nominais). A cada valor corresponde um conjunto de propriedades e estas dependências são expressas pelo atributo “*next*”. Isso significa que, ao ser selecionado um determinado valor, o seu conjunto de propriedades estará disponível. Para o *target1*, temos as propriedades “polaridade” e “unidade informacional”; para o *trigger*, as propriedades “valor modal”, “polaridade”, “índice de polaridade” e “unidade informacional”.

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<annotationscheme>
<attribute id="level_type" name="Type" type="nominal_button" text="Choose type of expression!">
<value id="value_type_none" name="none"/>
<value id="value_type_target1" name="target1" next="level_polarity_target1,level_info_unit_target1"/>
<value id="value_type_target_dependent" name="target_dep" next="level_polarity_target_dependent,
level_info_unit_target_dependent"/>
<value id="value_type_trigger" name="trigger" next="level_modal_value,level_polarity, level_polarity_cue,
level_info_unit_trigger"/>
<value id="value_type_source_modality" name="source_modality"/>
<value id="value_type_event" name="source_event"/>
</attribute>
```

No segundo nível, está definido o atributo “valor modal” e seus respectivos valores, em uma lista nominal:

```
<attribute id="level_modal_value" name="modal_value" type="nominal_list" text="">
<value id="value_modal_value_none" name="none"/>
<value id="value_modal_value_epistemic_knowledge" name="epistemic_knowledge"/>
<value id="value_modal_value_epistemic_belief" name="epistemic_belief"/>
<value id="value_modal_value_epistemic_possibility" name="epistemic_possibility"/>
<value id="value_modal_value_epistemic_probability" name="epistemic_probability"/>
```

```

<value id="value_modal_value_epistemic_necessity" name="epistemic_necessity"/>
<value id="value_modal_value_epistemic_verification" name="epistemic_verification"/>
<value id="value_modal_value_deontic_necessity" name="deontic_necessity"/>
<value id="value_modal_value_deontic_permission" name="deontic_permission"/>
<value id="value_modal_value_deontic_obligation" name="deontic_obligation"/>
<value id="value_modal_value_deontic_prohibition" name="deontic_prohibition"/>
<value id="value_modal_deontic_restriction" name="deontic_necessity"/>
<value id="value_modal_value_dynamic_volition" name="dynamic_volition"/>
<value id="value_modal_value_dynamic_ability" name="dynamic_ability"/>
</attribute>

```

No terceiro, quarto quinto e sexto níveis, estão definidos os atributos “polaridade do *trigger*”, “polaridade do *target1*”, “polaridade do *target_dependent*”, com os valores “positivo” e “negativo” (em botões nominais), e o “índice de polaridade”, o qual é um atributo “*freetext*”:

```

<attribute id="level_polarity" name="polarity" type="nominal_button" text="Choose the polarity of the
trigger.">
  <value id="value_polarity_pos" name="pos"/>
  <value id="value_polarity_neg" name="neg"/>
</attribute>

```

```

<attribute id="level_polarity_target1" name="polarity_target1" type="nominal_button" text="Choose the
polarity of the target1.">
  <value id="value_polarity_pos" name="pos"/>
  <value id="value_polarity_neg" name="neg"/>
</attribute>

```

```

<attribute id="level_polarity_target_dependent" name="polarity_target_dep" type="nominal_button"
text="Choose the polarity of the target1.">
  <value id="value_polarity_pos" name="pos"/>
  <value id="value_polarity_neg" name="neg"/>
</attribute>

```

```

<attribute id="level_polarity_cue" type="freetext" name="polarity_cue">
  <value id="value_polarity_cue" name="polarity_cue"/>
</attribute>

```


Os sétimo, oitavo e nono níveis definem os atributos para a “unidade informacional do *trigger*”, a “unidade informacional do *target1*” e a “unidade informacional do *target_dependent*”, em botões nominais:

```
<attribute id="level_info_unit_trigger" type="nominal_button" name="IU">
  <value id="value_IU_comment" name="COM"/>
  <value id="value_IU_topic" name="TOP"/>
  <value id="value_IU_parenthetical" name="PAR"/>
  <value id="value_IU_locutive_introducer" name="INT"/>
  <value id="value_IU_multiple_comment" name="CMM"/>
  <value id="value_IU_bound_comment" name="COB"/>
</attribute>
```

```
<attribute id="level_info_unit_target1" type="nominal_button" name="IU">
  <value id="value_IU_comment" name="COM"/>
  <value id="value_IU_topic" name="TOP"/>
  <value id="value_IU_parenthetical" name="PAR"/>
  <value id="value_IU_locutive_introducer" name="INT"/>
  <value id="value_IU_multiple_comment" name="CMM"/>
  <value id="value_IU_bound_comment" name="COB"/>
  <value id="value_IU_appendix_comment" name="APC"/>
</attribute>
```

```
<attribute id="level_info_unit_target_dependent" type="nominal_button" name="IU">
  <value id="value_IU_comment" name="COM"/>
  <value id="value_IU_topic" name="TOP"/>
  <value id="value_IU_parenthetical" name="PAR"/>
  <value id="value_IU_locutive_introducer" name="INT"/>
  <value id="value_IU_multiple_comment" name="CMM"/>
  <value id="value_IU_bound_comment" name="COB"/>
  <value id="value_IU_appendix_comment" name="APC"/>
</attribute>
```

O décimo nível descreve o atributo “*modal_class*”, que define o *set* modal, e suas propriedades:


```
<attribute id="level_modalclass" name="modal_class" color="green" width="2" type="markable_set"
add_to_markableset_text="Mark as link" style="lcurve" remove_from_markableset_text="Unmark as link"
adopt_into_markableset_text="Move this into current modal set" merge_into_markableset_text="Merge both
modal sets into one">
<value id="value_modalclass" name="modal_class"/>
</attribute>
```

O último nível define o atributo “*comment*”, do tipo “*freetext*”:

```
<attribute id="level_comment" type="freetext" name="comment">
<value id="value_comment" name="comment"/>
</attribute>

</annotationscheme>
```

5.3.2 Escolha dos valores modais:

Este trabalho, como explicitado anteriormente na parte teórica, filia-se à tradição linguística sobre modalidade, que apresenta, a depender da orientação teórico-metodológica, diferentes tipologias para o estudo dos modais. O repertório de significados modais está longe de ser estável entre as línguas, no entanto, observamos que há consenso entre todas as abordagens em relação ao significado epistêmico e o que as difere são as nuances nos significados não-epistêmicos (cf. seção 2.3, do capítulo 2).

Retomo brevemente aqui os valores modais que serão utilizados para fins de anotação, a saber, os valores epistêmico, deôntico e dinâmico¹¹⁵. Para o estabelecimento dos subvalores de cada tipo de modalidade, partiu-se da definição de modalidade discutida neste trabalho e dos valores apresentados em diferentes projetos de anotação, como descrito na seção 5.2 deste capítulo, além de os próprios dados analisados revelarem, pelo contexto, sutilezas de interpretação que aqui foram consideradas.

¹¹⁵ Os termos que se referem ao esquema de anotação da modalidade, a saber, os valores e subvalores modais, o índice modal disparador, a expressão no escopo do item modal e a fonte (do evento e da modalidade), serão mantidos em língua inglesa, uma vez que no domínio da anotação semântica é consenso o uso do inglês nos sistemas de anotação. Desta forma, quando necessário, será apresentada em nota de pé de página uma legenda com a tradução dos rótulos para o português e das suas definições.

(a) Epistêmicos: o significado epistêmico está relacionado com o grau de certeza de um conceptualizador sobre o material locutório enunciado e também se refere às noções de possibilidade, probabilidade e necessidade. Como exemplificado em (5.1), (5.2), (5.3) e (5.4):

- (5.1) *LAU: [195] também nã **tem certeza** de nada //COM=\$ (bfamdl03)
- (5.2) *BAL: [107] <só que é aquela coisa /=INT= eu nã &po> [/3]=EMP= eu nã **posso** esperar crescer dentro disso //COM=\$ (bfamdl02)
- (5.3) *GIL: [36] né / es **deve** meter o pau //\$ (bfamcv01)
- (5.4) *REN: [223] oito //COM=\$ [224] mas olha o preço //COM=\$ [225] quatorze e oitenta-e-nove //COM=\$ [226] esse daqui é quatro e noventa-e-oito //COM=\$
 *FLA: [227] é //COM=\$ [228] oh //COM=\$ [229] não /=CMM= mas aí tem [/2]=SCA= aqui tem o [/1]=SCA= <o dobro> //CMM=\$
 *REN: [230] o <dobro> //COM=\$ [231] pois é /=CMM= mas **teria que** ser uns dez reais /=CMM= né //PHA=\$ (bfamdl01)

Em (5.1), o falante “LAU” descreve uma situação que é apresentada como uma crença ou opinião com nenhum (ou baixo) grau de certeza. O exemplo seguinte (5.2), dada determinada evidência, “BAL” apresenta a impossibilidade (“nã posso”) de que “esperar crescer dentro disso” seja o caso. O verbo “dever” em (5.3) carrega o significado de que, a partir de um conhecimento prévio sobre “es”, chega-se à conclusão de que é provável que um grupo (“es”) critique uma determinada situação, em outras palavras, o exemplo expressa um julgamento epistêmico da probabilidade de uma determinada situação ocorrer. No excerto em (5.4), os participantes comparam o preço de dois produtos. A participante “REN”, chega, então, à conclusão de que há a necessidade (disparada pelo semimodal “ter que”) do preço do produto ser “uns dez reais”, para que fosse coerente a quantidade do produto com o seu preço.

Para este tipo, foram identificados seis subvalores: conhecimento, crença, possibilidade, probabilidade, necessidade e verificação. Segundo Palmer (2001, p. 24), existem três tipos de julgamento epistêmico possíveis: o especulativo (que expressa incerteza); o dedutivo (que expressa uma inferência a partir de uma evidência); e o assumptivo (que indica uma inferência no senso comum). Assim, os subvalores que serão descritos abaixo, foram definidos a partir destes três tipos de julgamento.

(a.1) epistemic_knowledge: o conceptualizador (o falante ou uma outra entidade) expressa o seu conhecimento ou compreensão sobre algo.

(5.5) *ANE: [152] <a> gente nũ <**sabia** que era> essa [/1]=SCA= essa <rua> //COM=\$ (bfamd105)

(a.2) epitemic_belief: o conceptualizador expressa a sua crença ou opinião sobre algo.

(5.6) *JAE: [84] mas **é lógico** que ea vai pôr ocês /=COM= uai //PHA=\$ (bfamcv02)

(a.3) epistemic_possibility: O conceptualizador apresenta o que enuncia como uma possibilidade.

(5.7) *TON: [173] dá pra &ma [/3]=EMP= **dá** pra jogar ela aqui /=CMM= ela vem na frente da quatro o' //CMM=\$ (bfamcv03)

(a.5) epistemic_probability: O conceptualizador apresenta o que enuncia como uma probabilidade, baseado em alguma evidência.

(5.8) *CAR: [95] **deve** ser <um> [/1]=SCA= alguns milhares de reais a conta /=COM= né //PHA=\$ (bpubcv02)

(a.6) epistemic_necessity: O conceptualizador apresenta o que enuncia como uma necessidade, baseado em conhecimento anterior.

(5.9) *DFL: [121] meu avô falava com papai /=INT= ele é doido /=CMM_r= só **pode** ser doido /=CMM_r= né /=PHA= &po +=EMP=\$ (bfammn02)

(a.7) epistemic_verification: O conceptualizador expressa incerteza em relação a um estado-de-coisas, evento ou atividade em foco.

(5.10) *ANE: [390] **olha** aí se nũ tem ninguém /=COM= César //ALL=\$ (bfamd105)

(b) Deônticos: o valor deôntico normalmente está associado às convenções morais e sociais e, portanto, normalmente é usado para indicar obrigação, permissão ou proibição. Os exemplos (5.11) e (5.12) ilustram, respectivamente, uma obrigação e uma permissão:

(5.11) *RUT: [257] cê **tem que** ir chique //COM=\$ (bfamcv02)

(5.12) *BAL: [127] <cê **pode**> deixar o cabo bater no chão //COM=\$ (bfamd102)

Na ocorrência em (5.11), o falante impõe a obrigação ao sujeito “cê” de “ir chique” a um determinado evento. Já em (5.12), excerto de um diálogo em que um participante dá instruções para uma outra pessoa de como utilizar um equipamento, é o falante “BAL” que garante a permissão, para que o sujeito da sentença “cê” esteja autorizado a realizar uma determinada ação, no caso “deixar o cabo bater no chão”.

Como subvalores, os deônticos abrangem: obrigação, permissão, proibição e necessidade. Apesar de a literatura sobre os deônticos normalmente classificarem tais modais como uma necessidade que gera uma obrigação (imposta pelo falante ou por uma autoridade, mas também uma mera obrigação a ser cumprida por um conceptualizador) ou uma permissão, a partir dos dados em contexto, foram identificados significados que sugeriam proibição (uma permissão mais fortemente negada) e necessidade (pessoal ou de um grupo) que não gera necessariamente uma obrigação ou permissão. Estes subvalores estão descritos abaixo:

(b.1) deontic_obligation: O conceptualizador obriga a alguém, se vê obrigado ou obriga a si mesmo a realizar uma atividade por uma determinada razão.

(5.13) *CAR: [220] <aí> **tem que** olhar os vizinhos //COM=\$ (bpubcv02)

(b.2) deontic_permission: O conceptualizador permite alguém ou se permite realizar uma atividade, ou permite que algo aconteça.

(5.14) *BAL: [134] <cê **pode**> deixar o cabo bater no chão //COM=\$ (bfamdl02)

(b.3) deontic_prohibition: O conceptualizador proíbe alguém ou a si mesmo de fazer algo, ou proíbe que algo aconteça.

(5.15) *CAR: [250] hhh <palavrão> **nũ po'** falar não //COM=\$ (bfamcv03)

(b.4) deontic_necessity: O conceptualizador expressa suas necessidades ou a necessidade de uma outra pessoa ou grupo.

(5.16) *JOR: [50] nós **precisamo** criar esse hábito //COM=\$ (bfammn06)

(c) Dinâmicos: o tipo dinâmico, apesar de ser menos central nos trabalhos sobre modalidade (HUDDLESTON; PULLUM, 2002; KIEFER, 1994; SALKIE, 2009), se relaciona à capacidade, habilidade e volição/intenção de um conceptualizador, como mostram as ocorrências em (5.17), (5.18) e (5.19):

(5.17) *BRU: [383] <cê **conseguiu** ser> pior //COM=\$ (bfamcv04)

(5.18) *JOR: [51] é a mesma coisa que você comprar hoje um celular com tanta tecnologia /=CMB= e você nã **consegue** usá-lo porque você nã aprende a usar /=SCA= aqueles manuais tão extenso pra poder /=SCA= ter /=SCA= o =EMP= o &us [/1]=SCA= uso /=i-COB= né /=PHA= da própria tecnologia //COM=\$ (bfammn06)

(5.19) *GIL: [78] cês **querem** olhar com a gente /=COB= cês **querem** sugerir um lugar /=COB= <qualquer> coisa desse tipo //COM=\$

No exemplo (5.17), o verbo “conseguir” se refere à capacidade do experienciador de realizar uma atividade, no caso, a capacidade de “ser pior” dentre os participantes de um grupo, expressa pelo falante “BRU”. Em (5.18), o mesmo verbo indica que o falante expressa a falta de habilidade (generalizada) de um conjunto de pessoas (marcado pelo uso impessoal do pronome “você”) de cumprir uma determinada tarefa (“usar o celular”). Por último, em (5.19), “GIL” reporta a necessidade de se consultar um grupo para que expressem sua vontade. A perspectiva levada em conta aqui é a do endereçado “cês”.

O tipo dinâmico expressa dois subvalores: habilidade/capacidade e volição/intenção. Palmer (2001, p. 76) define duas categorias para a modalidade dinâmica: habilitativo e volitivo. De acordo com Palmer (2001, p. 10), a categoria da habilidade “tem que ser interpretada mais amplamente do que em termos dos poderes físico e mental dos sujeitos, para incluir circunstâncias que imediatamente os afetam (mas, não, claro, a permissão deôntica)”¹¹⁶. Abaixo, a descrição de cada um dos subvalores:

(c.1) dynamic_ability: O conceptualizador expressa a sua própria habilidade/capacidade ou a habilidade/capacidade de uma outra pessoa para realizar ou alcançar algo.

(5.20) *CEL: [85] e' nã <**consegue** fazer isso nunca tá meio de hhh> [/3]=SCA= <tá de ladinho> //COM=\$ (bfamcv03)

¹¹⁶ No original: “has to be interpreted rather more widely than in terms of the subject’s physical and mental powers, to include circumstances that immediately affect them (but not, of course, deontic permission.” (PALMER, 2001, p. 10).

(5.21) *CAR: [150] porque ele ama /=COB= ele chama ela de /=SCA= fuminho /=COB= né /=COB= meu pretinho /=COB= papai nã **güenta** carregar mais /=COB= mas /=DCT= nã **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=\$ (bfammn05)

(c.2) dynamic_volition: O conceptualizador expressa as suas vontades, desejos, esperanças e intenções.

(5.22) *REN: [550] **espero** que eu não tinha [/1]=SCA= tenha perdido //COM=\$ (bfamd101)

Na tabela abaixo, apresento, em resumo, os valores modais, os subvalores a eles associados e a definição de cada um:

Valores	Subvalores	Definição
Epistêmico	conhecimento	O conceptualizador (o falante ou uma outra entidade) expressa o grau de conhecimento ou compreensão sobre algo.
	crença	O conceptualizador expressa a sua crença ou sua opinião sobre algo.
	possibilidade	O conceptualizador apresenta o que enuncia como uma possibilidade.
	probabilidade	O conceptualizador apresenta o que enuncia como uma probabilidade.
	necessidade	O conceptualizador apresenta o que enuncia como uma necessidade.
	verificação	O conceptualizador expressa incerteza em relação a um estado-de-coisas, evento ou atividade em foco.
Deôntico	obrigação	O conceptualizador se vê obrigado ou obriga a si mesmo a realizar uma atividade por uma determinada razão.
	permissão	O conceptualizador permite alguém ou a si mesmo a fazer algo, ou permite que algo aconteça.
	proibição	O conceptualizador proíbe alguém ou a si mesmo a fazer algo, ou proíbe que algo aconteça.
	necessidade	O conceptualizador expressa suas necessidades ou a necessidade de uma outra pessoa ou grupo.
Dinâmico	habilidade	O conceptualizador expressa a sua própria habilidade/capacidade ou a habilidade/capacidade de uma outra pessoa para realizar ou alcançar algo.
	volição	O conceptualizador expressa as suas vontades, necessidades, desejos, esperanças e intenções.

Tabela 5.4 – Valores e subvalores modais e suas definições

5.3.3 *Triggers, Sources e Targets*: elementos a serem anotados

5.3.3.1 *Triggers*:

Os *triggers* são as “palavras ou sequência de palavras que expressam modalidade” (BAKER et al., 2010). Considero como *triggers* os verbos auxiliares e semi-auxiliares modais, os verbos epistêmicos, os advérbios modais, as expressões adjetivas, as expressões lexicais que carregam modalidade. Para este esquema que proponho, serão anotados apenas os índices lexicais, deixando de lado, por ora, os gramaticais.

Como descrito no capítulo 3 deste trabalho, os índices modais mais frequentes são os verbos auxiliares e semi-auxiliares (“poder”, “dever”, “ter que”, “parecer”, “precisar”, “dar”, “conseguir”, “aguentar”, “querer”, “valer”, “adiantar”), bem como os verbos de crença ou epistêmicos (“achar”, “acreditar”, “crer”, “imaginar”, “julgar”, “pensar”), verbos de conhecimento e compreensão (“saber”, “ver”, “perceber”). Na nossa amostra, ainda encontramos os seguintes *types* para os advérbios e as locuções adverbiais: “certamente”, “com certeza”, “exatamente”, “justamente”, “logicamente”, “mesmo”, “na verdade”, “óbvio”, “potencialmente”, “realmente”, “sem chance”, “sem dúvida”, “sinceramente”, “talvez”, “às vezes”, e para os adjetivos e locuções adjetivas: “capaz”, “é capaz”, “claro”, “é claro”, “é lógico”, “é verdade”, “é verdadeira”, “foi verdade”, “lógico”, “mais certo é que tá”, “verdade”. Ainda, as expressões com valor modal, “era pra”, “pode saber”, “será”, “será que”, “tem certeza”, “tem chance”, “tem condição”, “tem condições”, “tem jeito”, “tenho certeza”, “tinha condições”, “tinha jeito”.

Para fins de anotação, levo em conta aqui apenas os índices lexicais. Não serão anotados os marcadores morfológicos e gramaticais, nem os itens que expressam evidencialidade (cf. definição no capítulo 2 desta tese). Desta forma, não estão codificados os marcadores de futuro perifrástico, futuro do presente, futuro do pretérito (com significado condicional), o subjuntivo e as construções condicionais¹¹⁷.

A partir de agora, apresento as regras para anotação do *trigger*.

¹¹⁷ Avalio que, para a anotação das construções condicionais, do tipo factual e contrafactual, por exemplo, um sistema de anotação deveria levar em conta não apenas a sua epistemicidade, mas também um conjunto de valores distintivos da factualidade dos eventos, nos moldes do FactBank (cf. SAURÍ; PUSTEJOVSKY, 2009), que definem a factualidade de um evento como “o nível de informação que expressa o comprometimento de fontes relevantes em relação à natureza factual de eventos mencionados no discurso” (SAURÍ; PUSTEJOVSKY, 2009, p. 231). O termo evento é tratado em termos amplos, para designar tanto processos como estados, bem como outros objetos abstratos como proposições, fatos e possibilidades. Adicionar uma segunda camada de anotação de factualidade para a interpretação de eventos é um dos objetivos para trabalhos futuros.

A. Se o *trigger* é:

A.1 um *verbo auxiliar ou semi-auxiliar modal*, anota-se o verbo modalizador como *trigger*.

(5.23) *FLA: [104] <hhh não /=INP= e a gente **tem que**> pesar a bolsa de novo
//=COM=\$ (bpubcv01)

trigger: tem que

(5.24) *CAR: [38] e /=DCT= falei com Deus também que eu nã **queria** buscar
//=COM_r (bfammn05)

trigger: queria

Se o verbo modalizador está em uma construção com os verbos auxiliares *ser*, *estar* ou *ter/haver*, anota-se apenas o modal.

(5.25) [99] já sofri o suficiente /=COB_r= agora eu tô **querendo** relaxar //COM_r=\$
trigger: querendo

(5.26) [271] eu tô **achando** que vai <chover> //COM=\$

trigger: achando

A.2 um *advérbio*, anota-se o advérbio ou locução adverbial como *trigger*.

(5.27) *LEO: [247] <**talvez** o Racing> //COM=\$ (bfamcv01)

trigger: talvez

(5.28) [44] que **às vezes** a gente sente uma dor numa hora //COM=\$ (bfamd102)

trigger: às vezes

A.3 um *adjetivo* ou uma *construção adjetival*, anota-se o adjetivo ou toda a expressão adjetival, inclusive o auxiliar “ser”, porque é parte do predicado nominal. Apenas os adjetivos com valor modal são considerados neste esquema de anotação, uma vez que não consideramos como modais os “avaliativos”¹¹⁸.

¹¹⁸ O esquema de anotação para o português europeu considera o valor “*Evaluation*”, definido como “[t]he speaker expresses his or someone else’s evaluation of facts and propositions” (HENDRICKX et al., 2012a).

(5.29) *PAU: [146] **capaz** /=COM=\$ (bpubdl01)

trigger: capaz

(5.30) *JAE: [84] mas **é lógico** que ea vai pôr ocês /=COM= uai /=PHA=\$ (bfamcv02)

trigger: é lógico

A.4. uma *expressão modal*, anota-se toda a expressão, inclusive, se for o caso, o verbo suporte .

(5.31) [64] esquentá /=SCA= comida /=CMB= <yyyy /=CMB= **tem a possibilidade** de um> monte de produto /=COM= o' /=CNT=\$ (bfamdl01)

trigger: tem a possibilidade

(5.32) *LUZ: [1] porque /=DCT= eu só soube que eu ã eu **tive certeza absoluta** que eu ã era daqui quando eu saí /=COM=\$ (bfamdl03)

trigger: tive certeza absoluta

B. O *trigger* não inclui as partículas negativas¹¹⁹. A negação é um fenômeno que interage diretamente com a modalidade.

(5.33) [16] ã **sei** se eu jogo aí na três /=COB= ã posso /=COM=\$ (bfamcv03)

trigger: sei

C. As preposições e conectivos não são incluídos no *trigger*. A única exceção é o verbo semimodal *ter que*.

(5.34) [108] e eu **sei** que ea devia /=TOP= porque /=SCA= &he /=TMT= foi [/1]=EMP= foi &q [/1]=SCA= nas véspera d' eu vim embora /=COM=\$

trigger: sei

(5.35) *BAL: [38] tá vendo /=CNT= a setinha **tem que** tá no cento-e-dez /=COM (bfamdl02)

trigger: tem que

¹¹⁹ No PB existem três possibilidades de realização da negação: negação simples preposta, negação simples posposta e dupla negação.

(5.36) [80] / e /=DCT= **teve que** amputar as duas perna aqui /=COB= hoje anda numa cadeira de roda /=COB= então a gente levava ela pra avó //COM=\$ (bfammn05)

trigger: teve que

D. Se no *trigger* vier intercalada qualquer partícula, não se considera esta palavra como parte do *trigger*.

(5.37) [81] porque eu nunca confundo letras com <informática> /=COB= nã **tem** nem **como** //COM=\$ (bfamdl02)

trigger: tem @ como¹²⁰

E. Quando uma unidade informacional ou um enunciado contém mais de um *trigger*, anota-se cada um dos *triggers* separadamente.

(5.38) [169] vão> vão /=CNT= <eu **acho** que **tem que** ser esses> //COM=\$ (bfamcv01)

trigger: acho

target: tem que ser esses

trigger: tem que

target: ser esses

(5.39) [220] e nã **pode** falar e nã **pode** <apontar o bagulho> //COM=\$ (bfamcv04)

trigger: poder

target: falar

trigger: pode

target: apontar o bagulho

F. Quando um *trigger* está em unidade de escansão (SCA), anota-se a unidade da qual ela herda as propriedades.¹²¹

¹²⁰ Utilizo o sinal “@” para marcar sequências não-contínuas. Como proceder à anotação de sequências descontínuas com o software MMAX2, ver a seção 5.3.1.3 deste capítulo, item (b.4).

¹²¹ A unidade de escansão constitui partes tonais diferentes de uma mesma unidade informacional. Ela introduz a unidade informacional que escande (escansão à esquerda), de maneira que a parte final da UI escandida especifica a função informacional do todo.

(5.40) [33] então assim /=INT= **espero** que /=SCA= isso nũ seja /=SCA= coisa pros times que jogam com a gente deixar de jogar com a gente //=**COM**= \$ (bfamcv01)

trigger: espero

IU: COM

(5.41) *EVN: [134] acho <que a gente> **tem** /=SCA= **que** olhar direito //=**COM**=\$ (bfamcv02)

trigger: tem que

IU: COM

G. Se um *trigger* estiver expresso em uma unidade de Apêndice de Comentário (APC) ou no Apêndice de Tópico (APT), seja como uma repetição de um índice expresso no Comentário ou Tópico ou como informação atrasada, quer dizer, quando se adiciona informação para facilitar a compreensão do enunciado pelo endereçado, anota-se a unidade de Comentário ou Tópico do qual é dependente.

(5.42) [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**=\$ (bfammn01)

trigger: deve

IU: COM (APC como “eco”)

(5.43) [157] distancia /=COM= que es **queria** <colocar> //=**APC**=\$ (bpubcv02)

trigger: queria

IU: COM (APC como “informação atrasada”)

Como o *trigger* é o componente central na expressão modal, atribui-se as especificações do tipo de modalidade a ele. Especifica-se para cada *trigger* as seguintes características:

- (a) valor modal (‘*modal value*’);
- (b) polaridade (‘*polarity*’);
- (c) índice de modalidade (‘*polarity_cue*’);
- (d) unidade informacional em que está contido (‘*IU*’).

A **polaridade** (*'polarity'*) é o componente usado para marcar se há uma negação agindo sobre o valor modal. Os valores atribuídos a este traço são “positivo” e “negativo”. O **índice de polaridade** (*'polarity_cue'*) é um campo “*freetext*”, isto é, um campo em que se pode digitar qualquer informação desejada, neste caso, identificar a palavra ou palavras que expressam a polaridade que afeta o *trigger*.

Uma **unidade informacional** pode coincidir com um enunciado ou apenas ser uma parte dele. Cada unidade cumpre uma função textual ou dialógica. As unidades informacionais textuais que podem conter um *trigger* são o Comentário (COM), Comentário Múltiplo (CMM), Comentário Ligado (COB), o Tópico (TOP), o Parentético (PAR) e o Introdutor Locutivo (INT)¹²². Como os textos do minicorpus alimentados no MMAX2 não contêm a anotação da estrutura informacional, para a anotação deste traço é necessária a utilização da plataforma DB-IPIC (Database of Information Pattern of Italian C-ORAL-ROM)¹²³, uma plataforma de busca disponível gratuitamente, composta de:

- (i) um corpus da seção informal do C-ORAL-ROM italiano, com 124.735 palavras e 74 sessões gravadas;
- (ii) um minicorpus da seção informal do C-ORAL-ROM italiano, com 32.589 palavras e 20 sessões, para comparação interlinguística com o minicorpus brasileiro;
- (iii) um minicorpus da seção informal do C-ORAL-BRASIL, com 31.464 palavras e 20 sessões.

Os arquivos estão divididos por enunciados e, para cada enunciado, além da transcrição, estão disponíveis o arquivo de som, a anotação da estrutura informacional e a anotação de PoS.

¹²² Excepcionalmente, a unidade de Apêndice de Comentário (APC) pode conter um índice modal. Esta unidade integra o texto do Comentário e conclui o enunciado. É dependente da unidade informacional de Comentário. Para a anotação das ocorrências do *trigger* nesta unidade, ver regra “G” desta seção.

¹²³ Disponível em: <http://lablita.dit.unifi.it/app/dbipic/index.php>. Último acesso: 01 dez. 2013.

5.3.3.2 Sources:

Uma ‘*source*’ é definida, em sentido amplo, como “um agente ou uma organização que toma uma atitude em relação a um evento em uma sentença”¹²⁴ (MATSUYOSHI et al., 2010, p. 1459). Vários estudos levam em consideração este componente em suas análises, tais como Bethard et al. (2004); Breck & Cardie (2004); Choi et al. (2005); Wiebe et al. (2005); Rubin et al. (2005); Prasad et al. (2007); Saurí e Pustejovsky (2007); Inui et al. (2008); Baker et al. (2010). Wiebe et al. (2005), em seu trabalho sobre a anotação de opiniões, emoções e outros estados privados, consideram a *source* do evento de fala, isto é, o falante ou escritor; a *source* do estado privado, isto é, a *source* cujo estado privado está sendo expresso, e apontam o aninhamento de *sources*, uma propriedade que reflete o fato de que eventos e estados privados podem estar encaixados um no outro.

De acordo com Matsuyoshi et al. (2010, p. 1459), a *source* é um importante traço da modalidade estendida, porque “[esta] informação ajuda o leitor a julgar a credibilidade dos conteúdos expressos em uma determinada sentença”¹²⁵.

O esquema desenvolvido para o português europeu (HENDRICX et al., 2012a, 2012b) anota a *source of the event mention* (o falante ou o produtor do evento) e a *source of the modality* (a quem a modalidade é atribuída). Segundo os autores, a decisão de anotar duas *sources* se deve à necessidade de se distinguir entre ‘aquele que produz a sentença’ e ‘aquele que expressa a modalidade’. Em muitas ocorrências estes dois elementos são coincidentes, mas não necessariamente este é o caso, como em “*Os portugueses* necessitam, em média, de 180 contos por mês para a manutenção de uma família de quatro pessoas. Neste exemplo, “Os portugueses” é a entidade com a necessidade interna disparada pelo verbo “necessitar”. O produtor do evento não está explícito aqui, e assume-se que é o produtor da sentença.

Inspirada neste último trabalho e ciente dos diferentes níveis de conceitualização na expressão da modalidade, utilizo as mesmas categorias para o esquema proposto aqui.

5.3.3.2.1 Source of the event mention:

A *source of the event mention* é o produtor do evento.

¹²⁴ No original: “an agent or an organization that takes an attitude toward an event mention in a sentence (MATSUYOSHI et al., 2010, p. 1459).

¹²⁵ Tradução para: “[this] information helps a reader judge credibility of contents conveyed from a given sentence” (MATSUYOSHI et al., 2010, p. 1459).

- A. O produtor é, normalmente, o falante que enuncia um determinado conteúdo locutório.
 B. A pessoa que produz o evento modal pode ser identificada, no texto, como a palavra, em caixa alta, no início de cada enunciado.

(5.44) ***CAR**: [98] <eu> consigo /=COM= né //PHA=\$ (bfamcv03)

trigger: consigo
modal value: dynamic_ability
polarity: pos
IU: COM
source of the event mention: CAR

(5.45) ***JAN**: [291] **na verdade** eu queria levar as duas /=COM= né //PHA=\$
 (bpubdl02)

trigger: na verdade
modal value: epistemic_belief
polarity: pos
IU: COM
source of the event mention: JAN

(5.46) ***EMM**: [79] **esperamos** <que esse> novo programa <que> vai vim /=TOP= ele
 /=TOP=\$ (bpucv01)

trigger: esperamos
modal value: dynamic_volition
polarity: pos
IU: TOP
source of the event mention: EMM

5.3.3.2.2 *Source of the modality*:

A *source of the modality* é o agente, experienciador ou cognoscente que veicula a modalidade.

As *sources* correspondentes a cada um dos valores modais são assim descritas:

Valores	Subvalores	Source
Epistêmico	conhecimento	O conceptualizador (o falante ou uma outra entidade) que expressa o grau de conhecimento ou compreensão sobre algo.
	crença	O conceptualizador que expressa a sua crença ou sua opinião sobre algo.
	possibilidade	O conceptualizador que apresenta o que enuncia como uma possibilidade.
	probabilidade	O conceptualizador que apresenta o que enuncia como uma probabilidade, baseado em alguma evidência
	necessidade	O conceptualizador que apresenta o material enunciado como uma necessidade, baseado em conhecimento anterior.
	verificação	O conceptualizador que expressa incerteza em relação a um estado-de-coisas, evento ou atividade em foco.
Deôntico	obrigação	O conceptualizador que obriga alguém, se vê obrigado ou obriga a si mesmo a realizar uma atividade por uma determinada razão.
	permissão	O conceptualizador que permite alguém ou a si mesmo a fazer algo, ou permite que algo aconteça.
	proibição	O conceptualizador que proíbe alguém ou a si mesmo a fazer algo, ou proíbe que algo aconteça.
	necessidade	O conceptualizador que expressa suas necessidades ou a necessidade de uma outra pessoa ou grupo.
Dinâmico	habilidade	O conceptualizador que expressa a sua própria habilidade/capacidade ou a habilidade/capacidade de uma outra pessoa para realizar ou alcançar algo.
	volição	O conceptualizador que expressa as suas vontades, necessidades, desejos, esperanças e intenções.

Tabela 5.5 – Sources correspondentes a cada valor e subvalor modais

As regras para a *source of the modality* são as seguintes:

A. A *source of the modality* e a *source of the event mention* são, normalmente, coincidentes:

(5.47) ***LUI**: [53] **eu** quero fazer o próximo campeonato no Arnaldinum /=CMM= e foda-se <pro seu Joaquim> //CMM=\$ (bfamcv01)

trigger: quero

modal value: dynamic_volition

polarity: pos

IU: CMM

source of the event mention: LUI

source of the modality: eu

- (5.48) ***GIL**: [74] que **eu** acho que deu muito pau /=COM= nessa taça //APC=\$
 (bfamcv01)
trigger: acho
modal value: epistemic_belief
polarity: pos
IU: COM
source of the event mention: GIL
source of the modality: eu
- (5.49) ***GIL**: [94] **a gente** podia fazer a taça aqui /=COB= todo mundo vai adorar
 /=COB= e tal //COM=\$ (bfamcv01)
trigger: podia
modal value: epistemic_possibility
polarity: pos
IU: COB
source of the event mention: GIL
source of the modality: A gente

B. Casos difíceis:

1. Se a *source of the modality* não está explícita, anota-se:

- (a) o falante, se coincidir com ele ou com o grupo em que está inserido:

- (5.50) ***TIQ**: [227] deve ser da /=SCA= <da irmã> Geni /=COM= <né> //PHA=\$
 (bpubcv02)
trigger: deve
modal value: epistemic_probability
polarity: pos
IU: COM
source of the event mention: TIQ
source of the modality: TIQ
- (5.51) ***REN**: [136] sabão em pó ã precisa não /=COM= né //PHA=\$
trigger: precisa
modal value: deontic_necessity
polarity: pos
IU: COM
source of the event mention: REN
source of the modality: REN

2. Quando a *source of the modality* é um pronome, ele é anotado:

(5.52) *CES: [307] ele <tem setenta-e-seis> metros quadrado **me** parece //COM=\$

trigger: parece

modal value: epistemic_belief

polarity: pos

IU: COM

source of the event mention: CES

source of the modality: me

3. Quando a *source* de uma *deontic_permission* é externa ao participante, ela não é anotada:

(5.53) *BRU: [147] esse aqui não /=CMM= porque esse aqui é quando for desenhar //CMM=\$ [148] aí /=DCT= o [/1]=EMP= no jogo do desenho /=TOP= por exemplo /=INT= um /=TOP= cê ã pode tirar o [/1]=SCA= o [/1]=EMP= <o> [/1]=EMP= o lápis do papel /=CMM= <o outro tem que desenhar com a mão> esquerda //CMM=\$

*HEL: [149] <não /=CMM= nós ã> vão fazer o desenho //CMM=\$

*LUC: [150] <eu posso desenhar ao invés de fazer mímica> //COM=\$

(bfamcv04)

trigger: posso

modal value: deontic_permission

polarity: pos

IU: COM

source of the event mention: LUC

source of the modality: -

(5.54) *BRU: [88] <hhh aí /=DCT= passa um tiquim /=CMM= fala de novo /=CMM= né> //PHA=\$

*HEL: [89] <ham ham> //COM=\$ [90] a pessoa &fa +=COM=\$

*LUC: [91] <é> //COM=\$

*HEL: [92] <ah> //COM=\$

*LUC: [93] <é> //COM=\$

*HEL: [94] pessoa faz que não //COM=\$

*LUC: [95] é //COM=\$

*HEL: [96] pois é //COM=\$ [97] o que que <ode> <que que ã pode> //COM=\$

*LUC: [98] <xxx> //UNC=\$

*BRU: [99] <mas aí /=TOP= a> gente escolhe /=SCA= qual //COM=\$ [100] o laranja /=CMM= ou o amarelo //CMM=\$

*HEL: [101] <tá> \$

(bfamcv04)

trigger: pode
modal value: deontic_permission
polarity: pos
IU: COM
source of the event mention: HEL
source of the modality: -

4. Se é uma pergunta retórica, anota-se o cognoscente. Se não está explícito, anota-se o verbo:

(5.55) *REG: [134] e ocê acha que eu nũ te conheço // =COM_r=\$ (bfammn04)

trigger: acha
modal value: epistemic_belief
polarity: pos
IU: COM
source of the event mention: REG
source of the modality: ocê

(5.56) *GIL: [54] sabe que que eu penso /=COM= velho // =ALL=\$ (bfamcv01)

trigger: sabe
modal value: epistemic_knowledge
polarity: pos
IU: COM
source of the event mention: GIL
source of the modality: sabe

5.3.3.3 Targets:

O *target* em textos falados é a expressão afetada pelo índice modal expresso pelo *trigger* dentro de uma unidade informacional. Portanto, diferente de outros esquemas de anotação que anotam o evento no escopo do item modal, não é necessário um predicado completo, como é usualmente tomado em textos escritos (uma cláusula subordinada ou um evento com todos os seus complementos e adjuntos). Na fala, como argumenta Cresti (no prelo), “um grande número de *chunks* falados, de fato, não podem ser definidos como cláusulas, mas como fragmentos, interjeições, advérbios, sintagmas, no entanto, funcionam perfeitamente do ponto de vista comunicativo”¹²⁶.

Assim, o *target* é anotado maximamente, admitindo-se descontinuidade, dentro do domínio da unidade informacional que contém o índice modal.

¹²⁶ No original: “a large number of spoken chunks, indeed, cannot be defined as clauses, but are rather fragments, interjections, adverbs, phrases, while nevertheless functioning properly from a communicative point of view.” (CRESTI, no prelo).

A. Se o *target* é:

A.1. um *sintagma*, anota-se todo o sintagma.

(5.57) *REN: [136] **sabão em pó** nũ precisa não /=COM= né //=-PHA=\$ (bfamd101)

trigger: precisa
modal value: deontic_necessity
polarity: neg
IU: COM
target: **sabão em pó**
polarity_tgt: pos
IU_tgt: COM
source of the event mention: REN
source of the modality: REN

(5.58) *REN: [463] precisando **xxx** //=-COM=\$ (bfamd101)

trigger: precisando
modal value: deontic_necessity
polarity: pos
IU: COM
target: **xxx**
polarity_tgt: pos
IU_tgt: COM
source of the event mention: REN
source of the modality: REN

(5.59) *DFL: [86] não /=INP_r= nũ acredito **nisso** não //=-COM_r=\$ (bfammn02)

trigger: acredito
modal value: epistemic_belief
polarity: neg
IU: COM
target: **nisso**
polarity_tgt: pos
IU_tgt: COM
source of the event mention: DFL
source of the modality: DFL

A.2. uma *cláusula*, anota-se a cláusula, excluído o complementizador que a introduz.

(5.60) *GIL: [2] <ô /=CNT= mas> /=DCT= voltando à questão /=COB= falando em [2]=EMP= e também falando em povo mascarado /=COB= esse povo do Galáticos é muito palha /=COB= eu acho que **es nũ deviam mais participar** /=COM= e <tal> //=-UNC=\$ (bfamcv01)

trigger: acho

modal value: epistemic_belief
polarity: pos
IU: COM
target: **es nũ deviam mais participar**
polarity_tgt: neg
IU_tgt: COM
source of the event mention: GIL
source of the modality: GIL

(5.61) *LUI: [7] com certeza **es nũ vão participar** /=COM= uai //PHA=\$

trigger: com certeza
modal value: epistemic_belief
polarity: pos
IU: COM
target: **es nũ vão participar**
polarity_tgt: neg
IU_tgt: COM
source of the event mention: LUI
source of the modality: LUI

(5.62) *JAE: [84] mas é lógico que **ea vai pôr ocês** /=COM= uai //PHA=\$
(bfamcv02)

trigger: é lógico
modal value: epistemic_belief
polarity: pos
IU: COM
target: ea vai pôr ocês
polarity_tgt: pos
IU_tgt: COM
source of the event mention: JAE
source of the modality: JAE

B. Quando o *target* tem polaridade negativa, incluir a partícula de negação no *target*:

(5.63) *CAR: [200] <acho que **nũ> deu muito certo pra ele não** /=COM= Toninho
//ALL=\$ (bfamcv03)

trigger: acho
modal value: epistemic_belief
polarity: pos
IU: COM
target: nũ deu muito certo pra ele
polarity_tgt: neg
IU_tgt: COM
source of the event mention: CAR
source of the modality: CAR

B.1. No caso de dupla negação e negação posposta simples, não incluir a segunda partícula de negação no *target*:

- (5.64) *LUZ: [210] <Nossa> /=EXP= esse negócio de terra nũ dá não //COM=\$
(bfamd103)
- trigger*: dá
modal value: epistemic_possibility
polarity: neg
IU: COM
target: **esse negócio de terra**
polarity_tgt: pos
IU_tgt: COM
source of the event mention: LUZ
source of the modality: LUZ

C. Quando há um caso de *retracting*¹²⁷ no *target*, anotar apenas a última parte em que está completa a enunciação:

- (5.65) *GIL: [170] <eu &a [/2]=EMP= eu acho que é> esse [/2]=SCA= é esse aqui o' //COM=\$ (bfamcv01)
- trigger*: acho
modal value: epistemic_belief
polarity: pos
IU: COM
target: **é esse aqui o'**
polarity_tgt: neg
IU_tgt: COM
source of the event mention: GIL
source of the modality: GIL

D. Se o *target* está em uma unidade de escansão (SCA), anotá-lo como uma única unidade informacional.

- (5.66) *RUT: [208] cê pensa que ele /=SCA= participa da [/1]=SCA= &d
[/1]=EMP= desses presente //COM=\$ (bfamcv02)
- trigger*: pensa
modal value: epistemic_belief
polarity: pos
IU: COM
target: **ele@participa@desses presente**
polarity_tgt: pos

¹²⁷ Como para a utilização do MMAX2, foi necessária a limpeza das barras invertidas e colchetes, sinais indicativos de *retractings*, o material linguístico que fica como vestígio é reconhecido pela repetição de *chunks* e/ou mudança de planejamento.

IU_tgt: COM
source of the event mention: RUT
source of the modality: CE

E. Se o *trigger* está em unidade de Parentético (PAR) e a expressão em seu escopo está em uma unidade informacional diferente, não se anota o *target*:

(5.67) *LUZ: [52] são duas vagas /=COM= eu acho //=PAR=\$ (bfamdl03)
trigger: acho
modal value: epistemic_belief
polarity: pos
IU: COM
target: -
polarity_tgt: -
IU_tgt: -
source of the event mention: LUZ
source of the modality: LUZ

(5.68) *FLA: [296] <oito /=CMM= né /=CMM= na verdade> //=PAR=\$ (bfamdl01)
trigger: na verdade
modal value: epistemic_belief
polarity: pos
IU: COM
target: -
polarity_tgt: -
IU_tgt: -
source of the event mention: FLA
source of the modality: FLA

F. *Target-dependent*:

O atributo *target-dependent* foi criado para os casos em que o *target* não está explícito em um determinado enunciado, mas é recuperável na cadeia referencial do texto.

Segundo Cresti (no prelo, p. 12), “[c]ada *chunk* linguístico, concebido para desempenhar uma determinada função textual (UT) dentro de um padrão informacional (PI), corresponde a uma *cena* (Barwise & Perry 1981; Fauconnier 1984) de um ponto de vista semântico. Como já dito, de um ponto de vista sintático uma UT pode até mesmo

corresponder a uma coleção de fragmentos, mas a fim de permitir o desenvolvimento de uma função textual, as expressões participantes devem estar reunidas na mesma cena”.¹²⁸

(5.69) *LUZ: [6] passei a vida toda num lugar errado //COM=\$ [7] que que é isso //COM=\$ [8] passei a vida toda fora d’ água hhh //COM=\$ [9] que loucura //COM=\$ [10] aí que ocê sabe /=COM= né //PHA=\$ [11] porque **quando cê chega num lugar que cê se sente em casa** /=TOP= cê sabe imediatamente //COM=\$ [12] é um +=EMP=\$

*LAU: [13] ham ham //COM=\$

*LUZ: [14] &he /=TMT= o corpo **sabe** /=CMM= tudo **sabe** //CMM=\$ [15] o seu humor /=CMB= a sua +=EMP=\$ [16] tudo //COM=\$

(bfamd103)

trigger: sabe

modal value: epistemic_knowledge

polarity: pos

IU: COM

target_dep: **quando cê chega num lugar que cê se sente em casa**

polarity_tgt: pos

IU_tgt: TOP

source of the event mention: LUZ

source of the modality: LUZ

(5.70) *GIL: [2] <ô /=CNT= mas> /=DCT= voltando à questão /=TOP= falando em [2]=EMP= e também falando em povo mascarado /=TOP= esse povo do Galáticos é muito palha /=COB= eu acho que **es nũ deviam mais participar** /=COM= e <tal> //UNC=\$

[...]

*LUI: [5] <eu **acho** não> //COM=\$

*LEO: [6] <**com certeza**> //COM=\$

(bfamcv01)

trigger: acho

modal value: epistemic_belief

polarity: neg

IU: COM

target_dep: **es nũ deviam mais participar**

polarity_tgt: neg

IU_tgt: COM

source of the event mention: LUI

source of the modality: LUI

¹²⁸ No original: “Each linguistic chunk, conceived to perform a certain textual function (TU) within an information pattern (IP), corresponds to a *scene* (Barwise & Perry 1981; Fauconnier 1984) from a semantic point of view. As we have already said, from a syntactic point of view a TU can even correspond to a collection of fragments, but in order to allow the development of a textual function, the participating expressions must be gathered within the same scene.” (CRESTI, no prelo, p. 12).

trigger: com certeza
modal value: epistemic_belief
polarity: pos
IU: COM
target_dep: **es nũ deviam mais participar**
polarity_tgt: neg
IU_tgt: COM
source of the event mention: LEO
source of the modality: LEO

Apesar de, em termos da anotação em seu conjunto, esta ser uma informação redundante, a decisão se justifica para fins de recuperação das informações referenciais mais facilmente. A linguagem natural possui muitos recursos e a anotação linguística tem como função permitir a melhor compreensão da linguagem natural pela máquina. A máquina não dispõe dos recursos do humano para a interação acional da linguagem, dessa forma, são necessárias categorias que a ajudem a recuperar a referencialidade da linguagem natural. Destaco que esta não é uma tentativa para a anotação de anáforas ou correferências, para o qual um projeto específico deve ser empreendido.

G. Casos difíceis:

G.1. No caso de usos formulaicos, em que a expressão do *target* envolve aspectos não-verbais (como gestos, expressões faciais etc.), o *target* é inespecífico e, portanto, não é anotado:

(5.71) *CEL: [138] <então tá /=CMM= amarelo> /=CMM=\$
 *HEL: [139] é /=COM=\$ [140] tá /=COM=\$
 *BRU: [141] amarelo /=COM=\$
 *LUC: [142] beleza <xxx> /=UNC=\$
 *BRU: [143] <tá> /=COM=\$
 *LUC: [144] <**posso**> /=COM=\$
 *BRU: [145] <e> +=EMP=\$
 *LUC: [146] e tem um dadinho diferente ali /=COM=\$
 (bfamcv01)
trigger: posso
modal value: deontic_permission
polarity: pos
IU: COM
target: **inespecífico**
polarity_tgt: -
IU_tgt: -
source of the event mention: LUC
source of the modality: **inespecífica**

Nesta ocorrência em (5.71), os participantes da interação estão discutindo, esclarecendo e conversando sobre as regras de um jogo. Após as explicações iniciais, um dos participantes ‘LUC’ pede a permissão a uma pessoa para realizar alguma ação. Não podemos precisar a expressão no escopo do *trigger* modal, nem o conceptualizador da permissão, que poderia ser qualquer um dos envolvidos.

G.2. Se o *trigger* está em uma unidade informacional e o *target* afetado por ele está em uma unidade subsequente, considera-se o todo como uma “construção padronizada”, ou seja, “construções realizadas através de UTs [unidades textuais], com cada qual desenvolvendo uma função informacional diferente” (CRESTI, no prelo, p. 18)¹²⁹, e anota-se a expressão afetada contida na unidade informacional subsequente e seu valor:

(5.72) *LUC: [1] pois é então /=INT= cê **sabe** que /=INT= lá na Letras /=TOP= **se eu conto essa história** /=SCA= **que eu sou parente do Drummond** /=COM= né //=-PHA=\$ (bfammn02)

trigger: sabe

modal value: epistemic_knowledge

polarity: pos

IU: COM

target: se eu conto essa história que eu sou parente do Drummond

polarity_tgt: pos

IU_tgt: COM

source of the event mention: LUC

source of the modality: cê

(5.73) *FLA: [239] só que a gente não **sabe** /=COB_s= **se elas têm agaivê** /=CMB¹³⁰= **se elas <têm hepatite>** /=CMB= se elas +=EMP=\$

trigger: sabe

modal value: epistemic_knowledge

polarity: pos

IU: COB

target: **se elas têm agaivê@se elas têm hepatite**

polarity_tgt: pos

IU_tgt: COB

source of the event mention: FLA

source of the modality: a gente

¹²⁹ Tradução para: “constructions performed across TUs, with each developing a different information function” (CRESTI, no prelo, p. 18)

¹³⁰ O CMB é uma categoria a ser revisada e está sendo considerada como COB.

Para o *target*, atribuem-se as seguintes características:

- (a) polaridade (*'polarity'*);
- (b) índice de modalidade (*'polarity_cue'*);
- (c) unidade informacional em que está contido (*"IU"*).

5.3.3.4 Identificação de *targets*, de acordo com o valor modal:

(a) epistemic_knowledge

(5.74) *OSV: [67] <a de telefone> ã veio ainda /=CMM= que eu ã sei **quando que vai** /=CMM=\$ (*bpubcv02*)

trigger: sei
modal value: epistemic_knowledge
polarity: neg
IU: CMM
target: quando que vai
polarity_tgt: pos
IU_tgt: CMM
source of the event mention: OSV
source of the modality: eu

(b) epistemic_belief

(5.75) *REN: [321] **o Neve é caro** <mesmo> /=COM=\$ (*bfamdl01*)

trigger: mesmo
modal value: epistemic_belief
polarity: pos
IU: COM
target: o Neve é caro
polarity_tgt: pos
IU_tgt: COM
source of the event mention: REN
source of the modality: REN

(c) epistemic_possibility

(5.76) *CEL: [229] cê pode **ter feito <LIBRAS>** /=COM=\$ (*bfamcv04*)

trigger: pode
modal value: epistemic_possibility
polarity: pos
IU: COM
target: ter feito LIBRAS
polarity_tgt: pos
IU_tgt: COM
source of the event mention: CEL
source of the modality: cê

(d) epistemic_probability

- (5.77) *MAI: [13] o diâmetro dea deve dar uns [1]=SCA= uns quarenta a cinquenta centímetro de [1]=SCA= de &s [2]=EMP= de grossura /=COM= o diâmetro dela /=APC=\$ (bfammn01)

trigger: sabe

modal value: epistemic_probability

polarity: pos

IU: COM

target: o diâmetro dea@dar@ uns quarenta a cinquenta centímetro@ de grossura

polarity_tgt: pos

IU_tgt: COM

source of the event mention: MAI

source of the modality: MAI

(e) epistemic_necessity

- (5.78) *REN: [231] pois é /=CMM= mas teria que ser uns dez reais /=CMM= né /=PHA=\$ (bfamd101)

trigger: teria que

modal value: epistemic_necessity

polarity: pos

IU: CMM

target: ser uns dez reais

polarity_tgt: pos

IU_tgt: CMM

source of the event mention: REN

source of the modality: REN

(f) epistemic_verification

- (5.79) *ANE: [393] olha aí se nũ tem ninguém /=CMM= César /=CMM=\$ (bfamd105)

trigger: olha

modal value: epistemic_verification

polarity: pos

IU: CMM

target: nũ tem ninguém

polarity_tgt: neg

IU_tgt: CMM

source of the event mention: ANE

source of the modality: ANE

(g) deontic_obligation

(5.80) *JAN: [239] depois eu tem que **comprar uma** //COM=\$ (bpubdl02)

trigger: tem que
modal value: deontic_obligation
polarity: pos
IU: COM
target: comprar uma
polarity_tgt: pos
IU_tgt: COM
source of the event mention: JAN
source of the modality: eu

(h) deontic_permission

(5.81) *PAU: [247] ele não pode **chegar no final da semana e** /=SCA=
desmanchar tudo o que foi feito durante a semana /=COM= né //PHA=\$
 (bpubdl01)

trigger: pode
modal value: deontic_permission
polarity: neg
IU: COM
target: ele@chegar no final da semana e@ desmanchar tudo o que foi feito durante a semana
polarity_tgt: pos
IU_tgt: COM
source of the event mention: PAU
source of the modality: PAU

(i) deontic_prohibition

(5.82) *CES: [82] nã pode **subir** /=CMM= é contramaõ //CMM=\$ (bfamdl05)

trigger: pode
modal value: deontic_prohibition
polarity: pos
IU: CMM
target: subir
polarity_tgt: pos
IU_tgt: CMM
source of the event mention: CES
source of the modality: CES

(j) deontic_necessity

(5.83) *JOR: [50] nós precisamos **criar esse hábito** //COM=\$ (bfammn06)

trigger: precisamos
modal value: deontic_necessity
polarity: pos
IU: COM
target: criar esse hábito
polarity_tgt: pos
IU_tgt: COM
source of the event mention: JOR
source of the modality: nós

(h) dynamic_ability

(5.84) *CAR: [150] porque ele ama /=COB= ele chama ela de /=SCA= fuminho
 /=COB= né /=COB= meu pretinho /=COB= papai ã güenta **carregar mais**
 /=COB= mas /=DCT= ã 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=\$ (bfammn05)

trigger: güenta
modal value: dynamic_ability
polarity: neg
IU: COB
target: carregar mais
polarity_tgt: pos
IU_tgt: COB
source of the event mention: CAR
source of the modality: papai

(i) dynamic_volition

(5.85) *DFL: [63] mas ele quis que **todos os filhos estudassem** //COM=\$
 (bfammn02)

trigger: quis
modal value: dynamic_volition
polarity: pos
IU: COM
target: todos os filhos estudassem
polarity_tgt: pos
IU_tgt: COM
source of the event mention: DFL
source of the modality: ele

5.3.4 Polaridade:

O traço da polaridade, como explicitado nas seções sobre os componentes *trigger* e o *target*, é marcado para o *trigger* e o *target*. Possui dois valores: “positivo” e “negativo” e o valor *default* é o positivo.

A polaridade global do enunciado não é computada. Se um *trigger* e um *target* ambos possuam polaridade negativa, a polaridade dos componentes será marcada separadamente.

Nos casos de valores deônticos de permissão e proibição marcados por uma partícula negativa, o que os diferencia é a força da permissão ([-forte] ou [+forte], respectivamente). Para a *deontic_permission*, marca-se a polaridade como “negativa”, já para a *deontic_prohibition*, marca-se a polaridade como “positiva”.

(5.86) *PAU: [247] ele não pode chegar no final da semana e /=SCA= desmanchar tudo o que foi feito durante a semana /=COM= né //=-PHA=\$ (bpubdl01)

trigger: pode

modal value: deontic_permission

polarity: neg

IU: COM

target: ele@chegar no final da semana e@ desmanchar tudo o que foi feito durante a semana

polarity_tgt: pos

IU_tgt: COM

source of the event mention: PAU

source of the modality: PAU

(5.87) *CAR: [250] hhh <palavrão> nũ po' falar não //=-COM=\$ (bfamcv03)

trigger: po'

modal value: deontic_prohibition

polarity: pos

IU: COM

target: palavrão @ falar

polarity_tgt: pos

IU_tgt: COM

source of the event mention: CAR

source of the modality: CAR

A decisão de se individualizar o valor de proibição deve-se à tentativa de se cobrir exemplos do tipo “Não é proibido X” ou “Você não está proibido de X”. Mais que uma permissão, estes tipos de ocorrência se configurariam como uma proibição deôntica de polaridade negativa.

5.4 Resultados:

Em nossa amostra, encontramos 1.088 marcadores modais (lexicais e gramaticais, excluídas as condicionais e as construções de subjuntivo) e, destes, foram anotados 781 *triggers*. A Tabela 5.6 mostra a distribuição dos valores modais e respectivos subvalores no minicorpus:

Valores	Subvalores	Freq	%
Epistêmicos		506	64,7
	conhecimento	108	21,3
	crença	223	44
	possibilidade	122	24,1
	probabilidade	24	4,7
	necessidade	10	1,9
	verificação	19	3,7
Deônticos		189	
	obrigação	96	50,8
	permissão	70	37
	proibição	6	3,2
	necessidade	17	9
Dinâmicos		86	
	habilidade	17	19,8
	volição	69	80,2

Tabela 5.6 – Frequência dos valores modais na amostra

A modalidade epistêmica é o valor mais frequente (64,7%) e o valor “crença”, que compreende os verbos epistêmicos e advérbios modais, é o mais numeroso, representando 44% de todos os índices epistêmicos. Dentre os itens deônticos, o seu maior uso é o de “obrigação”, metade de todos os itens, e a proibição deôntica tem o número menor de ocorrências, possivelmente por se tratar de uma nuance mais forte da permissão deôntica (37% dos deônticos). Os casos de volição correspondem a 80,2% da modalidade do tipo dinâmico e, em número menor, os casos de habilidade/capacidade, o que pode confirmar o caráter menos central desta categoria, como já mencionado anteriormente.

Os *triggers* são, em sua maioria, verbos modais, 81,5% de todos os índices (65%, auxiliares e semi-auxiliares; 31%, modais epistêmicos), seguido pelos advérbios (11,8%). Em número bastante menor, os adjetivos (2,8%) e expressões modais (3,9%).

A *source of the event mention* é normalmente o falante. A *source of modality* coincide com o falante em 88,3% das ocorrências, mas encontramos diferentes níveis de

conceptualização envolvendo a perspectiva do endereçado ou outra entidade, além da perspectiva do falante. No entanto, há que se observar que estas perspectivas de conceptualizadores diferentes da do enunciador são normalmente filtradas pelos olhos do falante, principalmente em relação ao uso dos verbos de crença, confirmando o argumento de Wiebe e suas colaboradoras (2005, p. 9) de que o aninhamento é uma propriedade importante das *sources*.

Quanto aos *targets*, em 79,1% de todas as ocorrências eles são realizados dentro da mesma unidade informacional. Em onze ocorrências ele não foi marcado, sendo que em uma delas é inespecífico (o caso em que poderia ser recuperado apenas se toda a cena estivesse registrada em imagem). Por fim, cinco são os casos em que estão em construções padronizadas (três com unidades de INT contendo o *trigger*, e duas em unidade de COB, com os *targets* realizados na unidade de Comentário).

Foram encontradas 108 ocorrências com polaridade negativa do *trigger*, e 27, para o *target*, disparada, em sua maioria, pela partícula de negação “não” (ou em sua forma “nũ”).

5.4.1 Acordo entre anotadores:

A anotação dos textos do minicorpus nos termos descritos neste capítulo foi realizada por apenas um anotador até o momento. Entretanto, para fins de validação do esquema e checagem de sua viabilidade, em uma etapa posterior a este trabalho, serão realizados testes com pelo menos mais dois anotadores, estudantes de pós-graduação.

No primeiro teste, um dos anotadores receberá um treinamento mínimo, com diretrizes gerais da anotação, para a anotação de 3 textos (um texto de cada tipologia interacional). No segundo experimento, um anotador diferente do primeiro receberá instruções detalhadas e procederá à anotação dos mesmos textos, acompanhando o manual de anotação. Ambas as anotações serão revisadas pelo anotador original. O acordo entre anotadores será computado utilizando a Estatística Kappa (COHEN, 1960), para todos os componentes da anotação.

Neste capítulo, apresentei as regras para a anotação de um *subcorpus* de 20 textos de fala espontânea do português brasileiro. Este esquema foi inspirado por outros já elaborados para a língua inglesa e se aproxima do projeto proposto para o português europeu. No entanto, diferencia-se deste último pelas opções teóricas adotadas aqui em termos de valores modais e, fundamentalmente, em termos de conceber os *targets* não como uma proposição completa,

mas *chunks* linguísticos, dado que a modalidade se realiza dentro do escopo da unidade informacional.

Na primeira seção, trago os trabalhos anteriores em anotação semântica relevantes para o projeto aqui apresentado e, em seguida apresento um quadro resumo de quem anota o quê e quais os valores são utilizados por cada um deles. Na seção 5.3, introduzo a proposta, descrevendo a metodologia utilizada (o software de anotação, a preparação dos textos e regras básicas de anotação), para então explicitar a escolha dos valores e subvalores modais e os elementos a serem anotados (*trigger*, *target*, *sources*) e o detalhamento das regras para cada um. Por fim, apresento e discuto os resultados, além de mostrar os experimentos a serem realizados para computar o acordo entre os anotadores.

No capítulo seguinte, caminho para a conclusão deste trabalho, resumindo os principais pontos discutidos, os avanços para o tema da modalidade e o horizonte futuro desta pesquisa.

Capítulo 6

CONSIDERAÇÕES FINAIS

A título de conclusão deste trabalho, gostaria de evidenciar alguns pontos centrais para a discussão do conceito de modalidade, a partir de uma análise baseada em *corpus*. Este estudo seguiu um fio narrativo, por assim dizer, que compreende três eixos que estão intrinsecamente relacionados: o primeiro, a discussão do conceito de modalidade; o segundo, a descrição e a análise dos itens modais; e o terceiro, a elaboração de um esquema de anotação semântica da modalidade.

No capítulo teórico, busquei apresentar e, de certa maneira confrontar, as diferentes abordagens para o estudo da categoria da modalidade – a semântica dos mundos possíveis; a distinção entre *realis* e *irrealis*, e o caráter subjetivo do fenômeno. Esta noção é apresentada na literatura linguística de variadas maneiras. Claro está, no entanto, que, seja em uma interpretação mais ampla (e vaga) como a de Palmer (2001), que afirma que a modalidade se preocupa com o estatuto da proposição que descreve o evento, seja em uma mais estrita como a de van der Auwera e Plungian (1998), que a definem a partir dos domínios da possibilidade e da necessidade (e excluem outros domínios), não há um consenso sobre os limites da categoria e que tipo de modificação ela abarca.

Coloquei-me, portanto, as seguintes questões: de que qualidade é a modificação definida como modalidade? Que critérios para delimitar esta modificação poderiam ser propostos? Na tentativa de alcançar uma resposta, o primeiro passo foi apresentar o seu avesso: “o que não é modalidade”, ou melhor explicando, apresentei as categorias com as quais a modalidade conversa e dela se diferenciam, como a negação, o modo, a ilocução e a atitude.

Alinhando-me à proposta de Cresti (2000) e Tucci (2007), para quem a modalidade é inerente a qualquer enunciado e pertence ao nível semântico, não ao pragmático, e seguindo a tradição ballyniana que leva em consideração a dimensão do enunciado e a noção de subjetividade, propus a modalidade como uma **modificação avaliativa operada por um sujeito conceptualizador (o falante, o endereçado ou uma terceira pessoa em cena), que relativiza o material locutório enunciado, de acordo com o grau de certeza/comprometimento e baseada em noções de possibilidade, necessidade, capacidade e volição.**

A delimitação da categoria pela subjetividade e pela inclusão dos domínios semânticos da possibilidade, necessidade, capacidade e volição, levou-nos a uma opção teórico-metodológica de estudar o fenômeno a partir de uma tipologia já bem definida na literatura e que está no centro de muitas análises: os significados epistêmicos, deônticos e dinâmicos. Esta escolha exclui a modalidade alética, dado que, segundo Palmer (1986), não há distinção entre o que é logicamente verdadeiro e aquilo que o falante considera como verdadeiro. Completando este raciocínio, pergunto: se a linguagem é, *per se*, argumentativa, isto é, se sempre há uma perspectiva de um conceptualizador sobre uma determinada cena, de que maneira seria possível alcançar uma objetividade tal que levaria a uma “verdade”?

A seleção dos itens modais, portanto, foi baseada nos critérios acima mencionados, com lemas que figuram em listas de modais em várias línguas e outros que emergem, por um viés metafórico, como uso específico da língua portuguesa. Como dito, faço esta afirmação desde os dados pesquisados na variante brasileira, já que não possuo informação suficiente para generalizar para as outras variantes de Portugal e África.

Definida a concepção de modalidade, a tipologia a ser utilizada, e os pressupostos da Teoria da Língua em Ato (CRESTI, 2000), procedi à descrição da modalidade em uma amostra de 20 textos da parte informal do C-ORAL-BRASIL, *corpus* da fala espontânea do português brasileiro (especificamente da região de Belo Horizonte, MG). Este *minicorpus*, compilado a partir de critérios de “máxima qualidade”, tenta preservar a estrutura da parte informal e representa uma ampla variação diastrática e diafásica.

Esta descrição lançou luz a pontos importantes, como:

- a categoria semântica da modalidade se comporta de forma distinta da escrita na fala espontânea, uma vez que o seu escopo é a unidade informacional, que não necessariamente abriga uma proposição, mas antes *chunks* linguísticos que não obedecem aos critérios sintáticos propostos para a análise da escrita;
- os índices modais são amplamente utilizados em situações dialógicas, sejam públicas ou privadas;
- apenas algumas unidades informacionais podem ser modalizadas: o Comentário incluídos os Comentários Múltiplos e Comentários Ligados), o Parentético (e a Lista de Parentéticos), o Tópico (e a Lista de Tópicos), e o Introdutor Locutivo;
- A unidade de Comentário é a mais modalizada e não há restrição de realização de itens nesta UI;

- o Introdutor Locutivo, no português brasileiro, também pode marcar uma introdução de uma avaliação epistêmica;
- as unidades de Tópico e Parentético abrigam como estratégia preferencial de modalização os verbos epistêmicos com 63,6% e 64,7%, respectivamente, de todas as ocorrências do *minicorpus*.
- a modalidade epistêmica é a mais frequente entre os tipos de significado, com uma taxa de associação a diferentes índices modais muito mais elevada do que a dos tipos deontico e dinâmico (cf. MELLO et al., 2013);
- os verbos modalizadores são a estratégia preferencial para marcação de modalidade. Sua distribuição está concentrada na unidade informacional de Comentário, com uma porcentagem de 97,3% de todas as ocorrências, incluídos o Comentário Múltiplo, o Comentário Ligado e o Apêndice de Comentário;
- o verbo “dever” com valor deontico, relacionado com o sentido de uma obrigação fraca, tem baixíssima ocorrência. O espaço em usos de obrigação deontica tem sido ocupado pelo semimodal “ter que”.
- os verbos epistêmicos, que sinalizam a opinião ou grau de comprometimento de um falante em relação ao que está enunciado, e se organizam sintaticamente de várias maneiras, possuem funções pragmáticas distintas: (a) em termos de estrutura informacional, em posição parentética funciona como atenuador da asserção anterior e (b) podem ser marcadores de concordância/discordância, o que indica um (possível) padrão lexical, correspondente a um fenômeno de gradiência de índice modal para um marcador discursivo.
- os itens “claro” e “óbvio”, apesar de pertencerem, formalmente, à categoria dos adjetivos, foram incluídos na análise das construções adverbiais modais, uma vez que seu comportamento se assemelha dos advérbios modais “claramente” e “obviamente”. São considerados advérbios somente quando ocorrem sozinhos, sem a presença do verbo ser e/ou da conjunção.
- as construções condicionais, dentro do quadro teórico da Teoria da Língua em Ato, constituem-se como um desafio, uma vez que podem ultrapassar os limites de um enunciado ou estar completamente acomodada nele, dividida, ou não, em mais de uma de suas unidades tonais.

Por fim, nasceu o MASS – Modal Annotation in Spontaneous Speech, projeto de anotação da modalidade na fala espontânea do português brasileiro. Os desafios para cumprir esta tarefa foram muitos, fundamentalmente relacionados ao fato de que todos os esquemas de anotação desenvolvidos até agora são aplicados a textos escritos, normalmente a *corpora* de textos jornalísticos ou textos biomédicos.

Como já visto, anotar a modalidade, com a finalidade de permitir o seu reconhecimento automático, inclui identificar os índices modais, classificá-los em uma determinada tipologia (por exemplo, em significados epistêmicos e não-epistêmicos), definir a sua fonte e o seu escopo semântico. Nosso esquema se inspirou no esquema proposto para o português europeu, mas dele se diferencia, na medida em que algumas decisões cruciais para a anotação foram baseadas na moldura teórico-metodológica da TLA, o que significa tomar o enunciado como unidade analítica de referência e a unidade informacional o campo de aplicação da modalidade.

Foram anotados 781 itens modais lexicais. Como valores modais, escolhemos o tipo epistêmico, com seis subvalores: conhecimento, crença, possibilidade, probabilidade, necessidade e verificação; o tipo deontico que abriga quatro subvalores: obrigação, permissão, proibição e necessidade e, por fim, o tipo dinâmico, com dois subvalores: habilidade e volição. Como elementos a serem anotados, o *trigger*, a palavra ou sequência de palavras que carrega(m) a modalidade; o *target*, a expressão que está no escopo do *trigger* dentro da unidade informacional; a *source of the modality*, o conceptualizador; e a *source of the event*, o falante ou produtor.

A anotação foi realizada, por um único anotador até o momento, com a ferramenta (MÜLLER; STRUBE, 2006), um *software* livre para a anotação em múltiplos níveis. valores e subvalores modais e os elementos a serem anotados (*trigger*, *target*, *sources*) e o detalhamento das regras para cada um.

O estudo da modalidade, como campo profícuo, pode inspirar futuras possibilidades e projetos. Tendo em vista que o C-ORAL-BRASIL é um *corpus* construído com a mesma arquitetura dos *corpora* do C-ORAL-ROM, é possível propor um estudo comparativo entre a expressão da modalidade no português brasileiro e nas outras quatro línguas românicas que compõem o C-ORAL-ROM. Mais imediatamente, faz sentido a comparação com o trabalho de Ida Tucci (2007) que descreveu a modalidade para a fala espontânea do italiano, em uma análise *corpus-based*.

Uma segunda tarefa premente é a aplicação dos experimentos para a computação do acordo entre anotadores, o que vai indicar a viabilidade do esquema proposto. A porcentagem

de itens anotados alcançada na amostra leva-nos a crer na possibilidade também de se buscar algoritmos para o treinamento de máquinas para alguns dos índices mais representativos, como os verbos modalizadores.

Outras pesquisas sobre a modalidade e a sua relação com outras categorias, como modo, tempo e aspecto podem ser exploradas. Assim como pode ser considerada a possibilidade de inclusão de construções avaliativas do tipo “é osso //” e ser estudada a relação entre a modalidade e as ilocuições.

No que concerne aos trabalhos em anotação semântica, como mencionado, é terreno a ser semeado e cultivado. Trabalhos que envolvam a modalidade e outras categorias, como o foco, e trabalhos comparativos, por exemplo, entre as diamesias oral e escrita, e entre os diferentes projetos propostos para outras línguas.

Entendo que o trabalho realizado contribui para uma melhor compreensão da noção semântica de modalidade e sua realização na fala espontânea do português brasileiro. Além disso, o esquema de anotação da modalidade é um passo importante e relevante para os estudos em Processamento em Linguagem Natural, e poderá contribuir com ferramentas para anotação automática da categoria no que diz respeito a textos falados e, por que não, sua ampliação (e adaptação) para textos escritos.

REFERÊNCIAS BIBLIOGRÁFICAS

ALMEIDA, S. A. F. Subjetividade e intersubjetividade: as construções completivas epistêmicas em inglês. 2010. 209f. Tese (Doutorado em Linguística) - Faculdade de Letras da Universidade Federal do Rio de Janeiro, Rio de Janeiro.

ALMEIDA, S.; FERRARI, L. Subjectivity, Intersubjectivity and Epistemic Complementation Constructions. Selected Papers from UK-CLA Meetings, <http://uk-cla.org.uk/proceedings>, Vol 1: 110-127, 2012.

ARENDT, H. [1958]. *A condição humana*. Rio de Janeiro: Editora Forense Universitária, 2010.

ARISTOTE. *Catégories. Sur l'interprétation. Organon I-II*. Paris : Éditions Flammarion, 2007.

AUSTIN, J. *How to do things with words*. Oxford: Clarendon, 1962.

van der AUWERA, J., PLUNGIAN, V. Modality's semantic map. *Linguistic Typology* 2, 79-124, 1998.

ÁVILA, L.B.; MELLO, H. Challenges in modality annotation in a Brazilian Portuguese spontaneous speech corpus. In: *Proceedings of WAMM-IWCS2013*, Potsdam, Germany, 2013.

BAKER, K., BLOODGOOD, M., DORR, B. J., FILARDO, N. W., LEVIN, L., PIATKO, C. A modality lexicon and its use in automatic tagging, in Proceedings of the Seventh Language Resources and Evaluation Conference (LREC'10), 2010.

BAKER, K.; DORR, B.; BLOODGOOD, M.; CALLISON-BURCH, C.; FILARDO, N. W.; PIATKO, C.; LEVIN, L.; MILLER, S. Use of modality and negation in semantically-informed syntactic MT. *Computational Linguistics*. 38(2): 411-438, 2012.

BALLY, Ch. *Linguistique générale et linguistique française*. Berna: Francke Verlag, 1932.

BALLY, Ch. Syntaxe de la modalité explicite. *Cahiers Ferdinand de Saussure* 2 :3-13, 1942.

BARWISE, J.; PERRY, J. Situation and attitudes. *Journal of Philosophy*, 78 (11):668-69, 1981.

BETHARD, S. YU, H.; THORNTON, A.; HATZIVASSILOGLU, V.; JURAFSKY, D. Automatic extraction of opinion propositions and their holders. In: *2004 AAAI Spring Symposium on Exploring Attitude and Affect in Text*, 2004.

BEZERRA, W. S.; MEIRELES, F. A. R. Um estudo sobre construções condicionais no Português do Brasil. In: SALOMÃO, M. M. M.; MIRANDA, N. S. *Construções do Português do Brasil: da gramática ao discurso*. Belo Horizonte: Editora da UFMG, 2009.

BRECK, E.; CARDIE, C. Playing the Telephone Game: Determining the Hierarchical Structure of Perspective and Speech Expressions. In: *The 20th International Conference on Computational Linguistics (COLING 2004)*, 2004.

- BICK, E. The Parsing System "PALAVRAS": Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework. Aarhus: Aarhus University Press, 2000.
- BIBER, D.; FINEGAN, E. Adverbial stance types in English. *Discourse Processes* 11: 1–34, 1988.
- BOYE, K. *Epistemic meaning*. Berlin/New York: Mouton de Gruyter, 2012.
- BROWN; P.; LEVINSON, S. *Politeness: some universals in language use*. Cambridge: Cambridge University Press, 1987.
- BURCHARDT, A.; ERK, K.; FRANK, A.; KOWALSKI, A.; PADO, S.; PINKAL, M. SALTO - A Versatile Multi-Level Annotation Tool. In: Proceedings of LREC'06, Genoa, Italy, 2006.
- BYBEE, J. *Morphology: a study of the relation between meaning and form*. Amsterdam / Philadelphia: John Benjamins, 1985.
- BYBEE, J.; PERKINS, R.; PAGLIUCA, W. *The evolution of grammar*. Chicago: The University of Chicago Press, 1994.
- BYBEE, J.; FLEISCHMAN, S. *Modality and grammar in discourse*. Amsterdam / Philadelphia: John Benjamins, 1995.
- CASTILHO, A. T. de; MORAES DE CASTILHO, C. M. Advérbios modalizadores. In: ILARI, R. (org.). *Gramática do português Falado*, vol. II. Campinas: Editora da UNICAMP 1993, p. 213-260.
- CASTILHO, A. T. de. O sintagma adverbial. In: *Nova Gramática do português brasileiro*. São Paulo: Contexto, 2010, p. 541-558.
- CHAFE, W. Evidentiality in English conversation and academic writing. In: CHAFE, W.; NICHOLS, J (eds.). *Evidentiality: The Linguistic Coding of Epistemology*. New York: Ablex, 1986, p. 261-272.
- CHAFE, W. 1995. The realis-irrealis distinction in Caddo, the Northern Iroquoian Languages, and English. In: BYBEE, J.; FLEISCHMAN, S. (eds.). *Modality in Grammar and Discourse*. Typological Studies in Language 32. Amsterdam: John Benjamins, 1995, p. 349-365.
- CHILDES - Child Language Data Exchange System. Disponível em: <http://childes.psy.cmu.edu/>. Último acesso: 29 nov. 2012.
- CHOI, Y.; CARDIE, C.; RILOFF, E.; PATWARDHAN, S. Identifying Sources of Opinions with Conditional Random Fields and Extraction Patterns. In: *Proceedings of HLT-EMNLP 2005*, 2005.
- CHUNG, S.; TIMBERLAKE, A. Tense, aspect, and mood. In: SHOPEN, T. (ed.). *Language Typology and Syntactic Description, III: Grammatical Categories and the Lexicon*. Cambridge: Cambridge University Press, 1985, p. 202-258.
- COATES, J. *The semantics of the modal auxiliaries*. London; Routledge, 1983.

COHEN, J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 1960, 20, p. 37-46.

COMRIE, B. Conditionals: A typology. In: TRAUGOTT, E. (ed.). *On conditionals*. Cambridge: Cambridge University Press, 1986, p.77-99.

CORNILLIE, B. *Evidentiality and epistemic modality in Spanish (semi-)auxiliaries. A cognitive-functional approach*. Berlin/New York: Mouton de Gruyter, 2007.

CORNILLIE, B. Evidentiality and epistemic modality: on the close relationship of two different categories. *Functions of Language*. Amsterdam/Philadelphia: John Benjamins, 16 (1): 44-32, 2009.

CORNILLIE, B.; PIETRANDREA, P. Modality at work. Cognitive, interactional and textual functions of modal markers. *Journal of Pragmatics* 44(15) 2109-2115, 2012.

CORTES, P. O.; MELLO, H. R. Modalidade no português brasileiro: as estruturas condicionais. In: *Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology*, 2013, p. 163-167.

CRESTI, E. *Corpus di italiano parlato*. Firenze: Accademia della Crusca, 2000.

CRESTI, E. Illocuzione e modalità. In: BECCARIA, P.; MARELLO, C. (eds.). *La parola al testo*. Scritti per Bice Mortara-Garavelli. Torino: Ed. dell'Orso, 2002, p. 133-145.

CRESTI, E.; SCARANO, A. Sur la notion de parlé spontané. CRESTI, E.; SCARANO, A. Sur la notion de parlé spontané. Disponível em: <http://lablita.dit.unifi.it/preprint/preprint-98coll03.pdf>. Último acesso em: 15 nov. 2012.

CRESTI, E.; MONEGLIA, M. *C-ORAL ROM: Integrated reference corpora for spoken Romance languages*. Amsterdam/Philadelphia: John Benjamins, 2005.

CRESTI, E. Syntactic properties of spontaneous speech in The Language into Act Theory: data on Italian complements and relative clauses. In: RASO, T.; MELLO, H. (eds.). *Spoken Corpora and Linguistic Studies*. Amsterdam/Philadelphia: John Benjamins, no prelo.

CROFT, W.; CRUSE, D. A. *Cognitive linguistics*. Cambridge: Cambridge University Press, 2004.

CUI, Y.; CHI, T. Annotating Modal Expressions in the Chinese Treebank. In: *Proceedings of WAMM-IWCS2013*, Potsdam, Germany, 2013.

DAVIES, M.; FERREIRA, M.. (2006-). *Corpus do Português* (45 million words, 1300s-1900s). Disponível em: <http://www.corpusdoportugues.org>. Acesso em: 12 mai. 2011.

DENDALE, P.; TASMOWSKY, L. Introduction: evidentiality and related notions. *Journal of Pragmatics*, 33, 339-348, 2001.

DOLZ, J.; SCHNEUWLY, B. *Gêneros orais e escritos na escola*. Campinas: Mercado de Letras, 2004.

FALLER, M. Semantics and Pragmatics of Evidentials in Cuzco Quechua. 2002. Tese (Doctor of Philosophy) – Department of Linguistics, Stanford University, Stanford, California.

FARKAS, R.; VINCZE V.; MÓRA, G.; CSIRIK, J.; SZARVAS, G. The CoNLL-2010 shared task: learning to detect hedges and their scope in natural language text. In: *Proceedings of the Fourteenth Conference on Computational Natural Language Learning - Shared Task*, 2010, p. 1-12.

FAUCONNIER, G. *Espaces mentaux*. Paris: Editions de Minuit, 1984.

FERRARI, L. V. Condicionais reportadas e flexibilidade de ponto de vista. *Gragoatá (UFF)*, v. 23, p. 95-109, 2007.

FERRARI, L. V. Reportar condicionais: uma questão de ponto de vista. *Revista de Estudos da Linguagem*, v. 16, p. x, 2008.

von FINTEL, K. Modality and language. In: BORCHET, D. M. (ed.). *Encyclopedia of Philosophy*. Detroit: MacMillan, 2006. Versão mais recente online, disponível em: <http://mit.edu/fintel/www/modality.pdf>. Último acesso: 13 set. 2012.

von FINTEL, K.; IATRIDOU, S. Anatomy of a Modal. In: GAJEWSKI, J.; HACQUARD, V.; NICKEL, B.; YALCIN, S. (eds). *New Work on Modality*. No. 52, in: *MIT Working Papers in Linguistics*. Department of Linguistics and Philosophy, MIT, 2005.

FITNEVA, S. Epistemic marking and reliability judgments: evidence from Bulgarian. *Journal of Pragmatics*, 33, 401-420, 2001.

FRAWLEY, W. *Linguistic Semantics*. Hillsdale, NJ / Hove / London: Lawrence Erlbaum, 1992.

FRAWLEY, W. (ed.). *The expression of modality*. Berlin: Mouton de Gruyter, 2005.

GALVÃO, V. C. C. . O ACHAR no português do Brasil: um caso de gramaticalização. *Revista Sínteses*, Campinas, v. 6, p. 143-156, 2000.

GALVÃO, V. C. C. ; GONÇALVES, S. C. L. . Modalidade e Gramaticalização: os casos do ACHAR e do PARECER. *Estudos Lingüísticos (Revista do GEL - CD ROM)*, Artigo 43, Marília, v. XXX, 2001.

GÉNÉREUX, M.; HENDRICKX, I.; MENDES, A. Introducing the Reference Corpus of Contemporary Portuguese On-Line. In: *Proceedings of the Eighth International Conference on Language Resources and Evaluation - LREC 2012*, Istanbul, May 21-27, 2012, p. 2237-2244.

GIVÓN, T. Modal prototypes of truth and action. In: GIVÓN, T. *Functionalism and grammar*. Amsterdam / Philadelphia: John Benjamins, 1995, p. 111-174.

GOLDBERG, A. *Constructions: a construction grammar approach to argument structure*. Chicago: University of Chicago Press, 1995.

GOLDBERG, A. *Constructions at work*. Chicago: University of Chicago Press, 2006.

GONÇALVES, S. C. L. Gramaticalização de construções com o verbo “parecer” no português brasileiro: de verbo pleno a satélite atitudinal.. Veredas (UFJF), Juiz de Fora, v. 8, n.1/2, p. 195-214, 2004.

GONZALVEZ-GARCIA, F. Modulating Grammar through modality: a discourse approach. Disponível em: <http://institucional.us.es/revistas/elia/1/9-gonzalvez.pdf>. Acesso em: 20 jul. 2010.

GREENBAUM, S. *Studies in English adverbial usage*. London: Longman, 1969.

GREEN, M. Speech Acts. In: ZALTA, E. N. (ed.) *The Stanford Encyclopedia of Philosophy* (Spring 2009 Edition). Disponível em: <http://plato.stanford.edu/entries/speech-acts/>. Último acesso 27 abr. 2014.

GRIES, S. Th. Corpus-based methods and cognitive semantics: The many senses of *to run*. In: GRIES, S. Th.; STEFANOWITSCH, A. *Corpora in Cognitive Linguistics: corpus-based approaches to syntax and lexis*. Berlin/New York: Mouton de Gruyter, 2006.

HAAK, S. *Filosofia das lógicas*. São Paulo: Editora UNESP, 2002.

HACQUARD, V. Aspects of Modality. Ph.D. Thesis, MIT, 2006. Disponível em: http://ling.umd.edu/~hacquard/papers/hacquard_thesis.pdf. Último acesso em: 13 jan. 2012.

HALLIDAY, M.A.K. Functional diversity in language, as seen from a consideration of modality and mood in English. *Foundations of Language*, 6.3. 322-361, 1970.

de HAAN, F. Evidentiality and epistemic modality: setting boundaries. *Southwest Journal of Linguistics*, 1999.

de HAAN, F. Typological approaches to modality. In: FRAWLEY, F. *The expression of modality*. Berlin/New York: Mouton de Gruyter, 2005.

HARA, Y. Non-propositional modal meaning. University of Delaware, 2006. (Manuscrito). Disponível em: http://cpct13.cityu.edu.hk/~yurihara/download/darou_hara.pdf. Último acesso em: 10 out. 2012.

HENDRICKX, I.; MENDES, A.; MENCARELLI, S.; SALGUEIRO, A. et al. 2012a. Modality Annotation Manual, version 1.0. Centro de Linguística da Universidade de Lisboa, (Manuscript).

HENDRICKX, I.; MENDES, A.; MENCARELLI, S. 2012b. Modality in Text: a Proposal for Corpus Annotation. In: Proceedings of the Eighth Conference on the International Language Resources and Evaluation (LREC'12), Istanbul, Turkey, p.1805-1812.

HENGEVELD, K. Illocution, mood and modality in a functional grammar of Spanish. *Journal of Semantics*, v. 6, 1988, p. 227-269.

HINTIKKA, J. *Knowledge and Belief: An Introduction to the Logic of the Two Notions*, Ithaca, NY: Cornell University Press, 1962.

HIRATA-VALE, F. B. M. Articulação de orações no português escrito do Brasil: as orações condicionais. *Estudos Linguísticos* (São Paulo), Belo Horizonte - MG, v. 5, n. 9, 2001.

- HIRATA-VALE, F. B. M. A condicionalidade em construções paratáticas: um uso argumentativo. *Estudos Linguísticos* (São Paulo), v. 37, p. 204-213, 2008.
- HORN, L. *A Natural History of Negation*. Chicago: University of Chicago Press, 1989.
- HOYE, L. *Adverbs and modality in English*. London / New York: Longman, 1997.
- HUDDLESTON, R. ; PULLUM, G. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press, 2002.
- JESPERSEN, O. [1924]. *The philosophy of grammar*. Chicago: Chicago University Press, 1992.
- INUI, K.; ABE, S.; MORITA, H.; EGUCHI, M. SUMIDA, A.; SAO, C.; HARA, K.; MURAKAMI, K.; MATSUYOSHI, S. Experience mining: Building a large-scale database of personal experiences and opinions from web documents. In: 2008 *IEEE/WIC/ACM International Conference on Web Intelligence*, 2008, p. 314-321.
- KÄRKKÄINEN, E. *Epistemic stance in English conversation. A description of its interactional functions, with a focus on I think*. Amsterdam - Philadelphia: Amsterdam, 2003.
- KARTTUNEN, L.; ZAENEN, A. Veridicity. In: KATZ, G.; PUSTEJOVSKY, J.; SCHILDER, F. (eds.). *Dagstuhl seminar proceedings. Annotating, extracting and reasoning about time and events*. Dagstuhl, Germany, 2005.
- KIEFER, F. Modality. In: ASHER, R. E. (ed.). *The Encyclopedia of Language and Linguistics*. Oxford: Pergamon Press, 1994, p. 2515-2520.
- KIEFER, F. Modality and pragmatics. *Folia Linguistica* 31 (3-4), 241-253, 1997.
- KIEFER, F. Modality. In: BRISARD, F.; ÖSTMAN, J-O.; VERSCHUEREN, J. *Grammar, meaning and pragmatics*. Amsterdam/Philadelphia: John Benjamins, 2009, p. 179-207.
- KRATZER, A. The notional category of modality. In: EIKMEYER, H. J.; RIESER, H. (ed.). *Words, worlds, and contexts - new approaches in Word Semantics*. Berlin: de Gruyter, 1981, p. 38-74.
- KRATZER, A. Modality. In: *Semantics: an International Handbook of Contemporary Research*. VON STECHOW, A.; DIETER, W. Berlin: de Gruyter. 1991, p. 639-650.
- KRATZER, A. *Modals and Conditionals: New and Revised Perspectives*. Oxford: Oxford University Press, 2012.
- KRIPKE, S. Semantical Considerations on Modal Logic, *Acta Philosophica Fennica*, 16: 83-94, 1963.
- LANGACKER, R. W. *Foundations of Cognitive Grammar* (Volume I). Stanford: Stanford University Press, 1987.
- LANGACKER, R. W. *Foundations of Cognitive Grammar* (Volume II). Stanford: Stanford University Press, 1991.

- LANGACKER, R. W. Viewing and Experiential Reporting in Cognitive Grammar. In: SILVA, A. S. (org.). *Linguagem e Cognição: a perspectiva da Linguística Cognitiva*. Braga: Editora Associação Portuguesa de Linguística, 2001, p. 19-49.
- LANGACKER, R. W. Cognitive Grammar. GEERAERTS, D.; CUYCKENS, H. (eds.). *The Oxford Handbook of Cognitive Linguistics*. New York: Oxford University Press, 2007, p. 421-462.
- LANGACKER, R. W. *Cognitive grammar*. Cambridge: Cambridge University Press, 2008.
- LARREYA, P. Towards a typology of modality in language. In: SALKIE, R.; BUSUTIL, P.; van der AUWERA, J. *Modality in English; theory and description*. Berlin/New York: Mouton de Gruyter, 2009, p. 9-29.
- LAWLER, J. Negation and negative polarity. Cambridge Encyclopedia of the Language Sciences, 2007. Disponível em: <http://www.umich.edu/~jlawler/CELS-Negation.pdf>. Último acesso em: 16 nov. 2012.
- LEECH, G. Corpus Annotation Schemes. In: *Literacy and Linguistic Computing* 8(4), 275-281, 1993.
- LEE, H. S. Cognitive constraints on expressing newly perceived information, with reference to epistemic modal suffixes in Korean. *Cognitive Linguistics* 4: 135–6, 1993.
- LOCAL, J. Phonetic and interactional features of attitude in everyday conversation. Research proposal. Disponível em: <http://www-users.york.ac.uk/~lang4/emotion-proposal.html>. Último acesso em: 22 out. 2012.
- LUNGUINHO, M. V. Sobre a concordância modal em português. *Cadernos de Linguagem e Sociedade*, v. 11, n. 2, 2010.
- LYONS, J. *Semantics*. vol. 1. Cambridge: Cambridge University Press, 1977.
- LYONS, J. *Semantics*. vol. 2. Cambridge: Cambridge University Press, 1977.
- MACHADO, N. J.; CUNHA, M. O. *Lógica e linguagem cotidiana: verdade, coerência, comunicação, argumentação*. Belo Horizonte: Autêntica, 2005.
- MACWHINNEY, B. *The CHILDES Project: Tools for Analyzing Talk*. 3rd Edition. Mahwah, NJ: Lawrence Erlbaum Associates, 2000.
- MARTIN, P. WinPitch Corpus: A text to Speech Alignment Tool for Multimodal Corpora. In: *Proceedings of the Fourth Conference on the International Language Resources and Evaluation (LREC'04)*, Lisbon, Portugal, 2004, p. 537-540.
- MATLOCK, T. Metaphor and the grammaticalization of evidentials. *Berkeley Linguistics Society*, 15: 215–25, 1989.
- MATSUYOSHI, S., EGUCHI, M., SAO, C., MURAKAMI, K., INUI, K., MATSUMOTO, Y. Annotating Event Mentions in Text with Modality, Focus, and Source Information. In: *Proceedings of the Seventh Conference on the International Language Resources and Evaluation (LREC'10)*, Malta, 2010.

MEDLOCK, B.; BRISCOE, T. Weakly supervised learning for hedge classification in scientific literature. In: Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, Prague, Czech Republic, Association for Computational Linguistics, p. 992–999, 2007.

MEJÍAS-BIKANDI, E. Space accessibility and moods in Spanish. In: SWEETSER, E.; FAUCONNIER, G. (eds.). *Spaces, worlds and grammar*. Chicago: University of Chicago Press, 1996, p. 157-178.

MELLO, H. R.; MELO, E. L.; CARVALHO, J. M.; CORTES, P. O. Prolegômenos sobre modalidade. *Domínios de Linguagem*, v. 5, p. 104-134, 2009.

MELLO, H. R.; CARVALHO, J. M.; CORTES, P. O. Modalização na fala espontânea do português brasileiro: um primeiro mapeamento de índices morfolexicais. *Revista de Estudos da Linguagem*, v. 18, p. 105-133, 2010.

MELLO, H. R.; RAMOS, A.; AVILA, L. Probing modal adverbs in Brazilian and European Portuguese: sociocultural variability in a pluricentric language. In: SILVA, A. S.; TORRES, A.; GONÇALVES, M. (org.). *Pluricentric languages: Linguistic Variation and Socicognitive Dimensions*. 1ed. Braga, Portugal: Publicações da Universidade Católica Portuguesa, 2011, v. 1, p. 473-486.

MELLO, H. R.; RASO, T. Illocution, modality, attitude: different names for different categories. In: MELLO, H. R.; PANUNZI, A.; RASO, T (eds.). *Pragmatics and prosody: illocution, modality, attitude, information patterning and speech annotation*. Firenze: Firenze University Press, 2011, p. 1-18.

MELLO, H.; CAETANO, R. Mapeamento de construções adverbiais modais na fala espontânea. Belo Horizonte: FALÉ/UFGM, 2012. (Manuscrito).

MELLO, H. R. ; RASO, T. Frames e Fala Espontânea. *Cadernos de Estudos Linguísticos (UNICAMP)*, v. 55, p. 99-108, 2013.

MELLO, H. R.; PAIXÃO, C. A.; COELHO, F. C.; SOUZA, R. R. . Distribution of modality markers in Brazilian Portuguese spontaneous speech. In: Quantitative Investigations in Theoretical Linguistics 5, 2013, Leuven. Proceedings QITL 5. Leuven: KU Leuven, 2013. p. 64-67.

MENDES, A.; HENDRIKX, I.; SALGUEIRO, A.; ÁVILA, L. Annotating the Interaction between Focus and Modality: the case of exclusive particles. Proceedings of the 7th Linguistic Annotation Workshop & Interoperability with Discourse, Sofia, Bulgaria, Association for Computational Linguistics, p. 228–237, 2013.

MELAC, E. Modality and evidentiality in English and Tibetan. Apresentação na Modality, Corpus, Discourse Conference, Lund University, Lund, Suécia, 2012.

MITHUN, M. The relativity of irrealty. In: BYBEE, J.; FLEISCHMAN, S. (eds.). *Modality in Grammar and Discourse*. Typological Studies in Language 32. Amsterdam: John Benjamins, 1995, p. 367-388.

MITHUN, M. *The languages of native North America*. Cambridge: Cambridge University Press, 1999.

MITTMAN, M. M. O C-ORAL-BRASIL e o estudo da fala informal: um novo olhar sobre o Tópico no Português Brasileiro. 2012. (Tese de Doutorado – Estudos Linguísticos). Faculdade de Letras da Universidade Federal de Minas Gerais, Belo Horizonte, MG.

MONEGLIA, M. Spoken corpora and pragmatics. *Revista Brasileira de Linguística Aplicada* [online]. 2011, vol.11, n.2, p. 479-519.

MONEGLIA, M e CRESTI, E. L'intonazione e i criteri di trascrizione del parlato adulto e infantile. In: BORTOLINI, U., PIZZUTO, E. Il Progetto CHILDES Italia. Pisa: Del Cerro, 1997, p. 57-90.

MORAES, J.A.; RILLIARD, A.; MOTA, B.; SHOCHI, T. Multimodal perception and production of attitudinal meaning in Brazilian Portuguese. In: *Proceedings Speech Prosody 2010*. Chicago, 2010.

MORANTE, R.; DAELEMANS, W. Learning the scope of hedge cues in biomedical texts. In: *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing*, Boulder, Colorado, USA, 2009, p. 28-36.

MORANTE, R.; DAELEMANS, W. Annotating modality and negation for a machine reading evaluation. In: P. Forner, J. Karlgren, & C. Womser-Hacker (eds.), CLEF 2012 Conference and Labs of the Evaluation Forum - Question Answering For Machine Reading Evaluation (QA4MRE), Rome, Italy, 2012.

MORANTE, R.; SPORLEDER, C. Modality and Negation: An Introduction to the Special Issue, *Computational Linguistics*, 38:2, 2012.

MORTELMANS, T. Modality in Cognitive Linguistics. In: GEERAETS, D.; CUYCKENS, H. *The Oxford Handbook of Cognitive Linguistics*. New York: Oxford University Press, 2007, p. 869-889.

MOWERY, D. L.; VELUPILLAI, S.; CHAPMAN, W. W. Medical diagnosis lost in translation – Analysis of uncertainty and negation expressions in English and Swedish clinical texts. In: *Proceedings of the 2012 Workshop on Biomedical Natural Language Processing (BioNLP 2012)*, Montreal, Canada, Association for Computational Linguistics, p. 56-64, 2012.

MOZZICONACCI, S.J.L. Emotion and attitude conveyed in speech by means of prosody. For the *2nd Workshop on Attitude, Personality and Emotions in User-Adapted Interaction*. Sonthofen, Germany, 2001.

MURAKAMI, K.; MASUDA, S.; MATSUYOSHI, S.; NICHOLS, E.; INUI, K.; MATSUMOTO, Y. Annotating semantic relations combining facts and opinions. In: *Proceedings of the Third Linguistic Annotation Workshop*, 2009, p. 150-53.

MÜLLER, C.; STRUBE, M. Multi-level annotation of linguistic data with MMAX2. In *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods*, p. 197–214. Peter Lang, 2006.

NARROG, H. On defining modality again. *Language Sciences* 27, p. 165-192, 2005.

NIRENBURG, S.; RASKIN, V. *Ontological Semantics*. MIT Press, 2004.

NIRENBURG, S.; McSHANE, M. Annotating modality. Technical report, University of Maryland, Baltimore County, March 19, 2008.

NITTA, Y. Ninshiki no modariti to sono shu^hen [Epistemic modality and its periphery]. In: MORIYAMA, T., NITTA, Y., KUDO^, H. (eds.), *Modariti [Modality]* (Nihongo no bunpo^ 3). Iwanami, Tokyo, 2000, p. 81-159.

NURMI, A. Employing and elaborating annotation for the study of modality. Disponível em: www.helsinki.fi/varieng/journals/volumes/01/nurmi/. Último acesso: 27 mar. 2012.

NUYTS, J. Epistemic modal qualifications: On their linguistic and conceptual structure. *Antwerp Papers in Linguistics (APiL)*, n. 81, 1994. Antuérpia: University of Antwerp. Disponível em: <http://webhost.ua.ac.be/apil/list.html>. Último acesso em: 16 nov. 2012. (out of print).

NUYTS, J. *Epistemic modality, language and conceptualization: a cognitive-pragmatic perspective*. Amsterdam/Philadelphia: John Benjamins, 2001.

NUYTS, J. Grounding and the system of epistemic expressions in Dutch: A cognitive-functional view. In: BRISARD, F. (ed.). *Grounding*. The epistemic footing of deixis and reference. Berlin: Mouton de Gruyter, 2002, p. 433-466.

NUYTS, J. The modal confusion: On terminology and the concepts behind it. In: KLINGE, A.; MÜLLER, H.H. (eds.). *Modality: Studies in form and function*, London: Equinox, 2005, p. 5-38.

NUYTS, J. Notions of (inter)subjectivity. *English Text Construction* 5, 53-76, 2012.

OLIVEIRA, F. Modalidade e modo. In: MATEUS, M. H. M.; BRITO, A. M.; DUARTE, I.; FARIA, I. H. *Gramática da Língua Portuguesa*. 6. ed. Lisboa: Caminho, 2003.

PALMER, F. R. *Mood and modality*. Cambridge: Cambridge University Press, 1986.

PALMER, F. R. *Modality and the English modals*. London: Longman, 1990.

PALMER, F. R. *Mood and modality*. 2nd edition. Cambridge: Cambridge University Press, 2001.

PAPAFRAGOU, A. Inference and word meaning: the case of modal auxiliaries. *Lingua* 105: 1-47, 1998.

PAPAFRAGOU, A. *Modality: issues in the semantics-pragmatics interface*. Amsterdam/New York: Elsevier Science, 2000.

PERKINS, M. R. *Modal Expressions in English*. London: Pinter, 1983.

PIETRANDREA, P. *Epistemic modality: functional properties and the Italian system*. Amsterdam / Philadelphia: John Benjamins, 2005

PLUNGIAN, V. The place of evidentiality within the universal grammatical space. *Journal of Pragmatics*, 33, 349-357, 2001.

- PORTNER, P. Imperative and modals. *Natural Language Semantics*, 4: 351-383, 2007.
- PORTNER, P. *Modality*. Oxford: Oxford University Press, 2009.
- PRASAD, R.; DINESH, N.; LEE, A.; JOSHI, A.; WEBBER, B. Attribution and its Annotation in the Penn Discourse TreeBank. *Traitement Automatique des Langues, Special Issue on Computational Approaches to Document and Discourse*, 47(2), 2007.
- PUSTEJOVSKY, J.; INGRIA, B.; SAURI, R.; CASTANO, J.; LITTMAN, J.; GAIZAUSKAS, R.; SETZER, A.; KATZ, G.; MANI, I. The specification language TimeML. *The Language of Time: A reader*, 2005, p. 545–557.
- RASO, T.; MELLO, H. R. Parâmetros de compilação de um corpus oral: o caso do C-ORAL-BRASIL. *Veredas*, v. 13, n. 2, p. 20-35, 2009.
- RASO, T.; MELLO, H. *C-ORAL-BRASIL I: Corpus de referência do português brasileiro falado informal e DVD multimedia*. Belo Horizonte: Editora UFMG, 2012. v. 1.
- ROCHA, B. N. R. de M. 2011. 264f. Características prosódicas do tópico em PE e o uso do pronome lembrete. (Dissertação de Mestrado – Estudos Linguísticos). Belo Horizonte, Faculdade de Letras da Universidade Federal de Minas Gerais, Belo Horizonte, MG.
- RUBIN, V.; LIDDY, E.; KANDO, N. Certainty Identification in Texts: Categorization Model and Manual Tagging Result. New York Springer-Verlag. 2005, p. 61-74.
- RUBINSTEIN, A. *Roots of modality*. 2012. Tese (Doctor of Philosophy) – Department of Linguistics, University of Massachusetts, Amherst.
- RUBINSTEIN, A.; HARNER, H.; KRAWCZYK, E.; SIMONSON, D.; KATZ, G.; PORTNER, P. Toward fine-grained annotation of modality in text. [with] In: *Proceedings of WAMM-IWCS2013*, Potsdam, Germany, 2013.
- RUPPENHOFER, J., & REHBEIN, I. Yes we can!?! Annotating English modal verbs. In: *Proceedings of the Eighth conference on the International Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, 2012.
- SALKIE, R. *Degrees of modality*. In: SALKIE, R.; BUSUTIL, P.; van der AUWERA, J. *Modality in English; theory and description*. Berlin/New York: Mouton de Gruyter, 2009, p. 79-102.
- SALOMÃO, M. M. M. Polysemy, aspects and modality in Brazilian Portuguese — the case for a cognitive explanation of grammar. (Tese de Doutorado – Linguística). 1990. Berkeley, University of California System, 1990.
- SALOMÃO, M. M. M. A construção modal com dar no Português do Brasil: metáfora, uso e gramática. *Revista de Estudos da Linguagem*, v. 16, p. 3, Belo Horizonte, 2008.
- SAURÍ, R., VERHAGEN, M., PUSTEJOVSKY, J. Annotating and recognizing event modality in text, in *Proceedings of the 19th International FLAIRS Conference*, FLAIRS 2006.
- SAURÍ, R.; PUSTEJOVSKY, J. Determining modality and factuality for textual entailment. In: *First IEEE International Conference on Semantic Computing*. Irvine, California, 2007.

SAURÍ, R. A Factuality Profiler for Eventualities in Text. Ph.D. Thesis. Brandeis University, 2008.

SAURÍ, R.; PUSTEJOVSKY, J. FactBank: A Corpus Annotated with Event Factuality. *Language Resources and Evaluation*, 2009. Disponível em: <http://link.springer.com/article/10.1007/s10579-009-9089-9/fulltext.html>. Último acesso: 28 jan. 2013.

SAURÍ, R.; PUSTEJOVSKY, J. Assessing the factuality degree of events in text. *Computational Linguistics*, 38(2), 2012, p. 261-299.

SEARLE, J. *Speech acts*. Cambridge: Cambridge University Press, 1969.

SEARLE, J. [1979]. *Expressão e significado*. São Paulo: Martins Fontes, 2002.

SIMON-VANDENBERGEN, A.; AIJMER, K. *The semantic field of modal certainty: a corpus-based study of English adverbs*. Berlin / New York: Mouton de Gruyter, 2007.

SIMON-VANDENBERGEN, A.-M. Almost certainly and most definitely: Degree modifiers and epistemic stance. *Journal of Pragmatics*, 40 (2008), p. 1521–1542.

SWAN, T. *Sentence Adverbials in English: A Synchronic and Diachronic Investigation*. Oslo: Novus, 1988.

SWEETSER, E. *From etymology to pragmatics: metaphorical and cultural aspects of semantic structure*. Cambridge: Cambridge University Press, 1990.

SZARVAS, G., VINCZE, V., FARKAS, R., CSIRIK, J. The BioScope corpus: annotation for negation, uncertainty, and their scope in biomedical texts, in *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing*, Association for Computational Linguistics, 2008, p. 38-45.

TALMY, L. Force Dynamics. Paper presented at conference on Language and Mental Imagery. May 1981, University of California, Berkeley.

TALMY, L. Force Dynamics in language and cognition. *Cognitive Science*, 12, 1, 49-100, 1988.

TALMY, L. *Toward a Cognitive Semantics*. Cambridge, MA: MIT Press, 2000.

TUCCI, I. L'espressione lessicale della modalità nel parlato spontaneo: Analisi del corpus C-ORAL-ROM. 2007. (Tese de Doutorado – Linguística Italiana). Firenze, Università degli Studi di Firenze, 2007, 335p.

TUCCI, I. *The informational structure and the scope of lexical modality in spoken Italian*, 2008. (Manuscrito).

TUCCI, I. La modalità nel parlato spontaneo e il suo dominio de pertinenza. Una ricerca corpus-based (C-ORAL-ROM italiano). *Actes du XXVe CILPR*. (Innsbruck 3-8 September 2007). Disponível em: <http://lablita.dit.unifi.it/preprint/preprint.2008-02-06.3658867998>. Último acesso: 13 jan. 2013.

TUCCI, I. Illocution and modality in spoken Italian: performing a speech act through words and judging their semantic content – a corpus-based analysis. In; MELLO, H.; PANUNZI, A.; RASO, T. (eds.). *Pragmatics and prosody. Illocution, Modality, Attitude, Information Patterning and Speech Annotation*. Firenze: Firenze University Press, 2011, p. 83-100.

VALE, H. P. *A unidade informacional de Parentético no português do Brasil: uma análise baseada em corpus*. 2010. Dissertação (Mestrado em Estudos Linguísticos) – Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, MG.

VENIER, F. *La modalizzazione assertiva. Avverbi modali verbi parentetici*. Milano: Franco Angeli, 1991.

VERHAGEN, A. *Constructions of Intersubjectivity. Discourse, Syntax, and Cognition*. Oxford: Oxford University Press, 2005.

von WRIGHT, G. H. *An essay in modal logic*. Amsterdam: North Holland Publishing Company, 1951.

VORKUKOVA, E. *Epistemic Modalities in Spoken Standard Tibetan*. 2011. 377f. Tese (Doctorat en Sciences du Langage) – Université Paris 8, Paris.

WIEBE, J.; BRUCE, R.; BELL, M.; MARTIN, M.; WILSON, Th. A corpus study of evaluative and speculative language. *Proceedings of the Second ACL SIGdial Workshop on Discourse and Dialogue*, 2001.

WIEBE, J.; WILSON, Th.; BRUCE, R.; BELL, M.; MARTIN, M. Learning subjective language. *Computational Linguistics*, 30(3):277-308, 2004.

WIEBE, J., WILSON, T., CARDIE, C. *Annotating expressions of opinions and emotions in language*. Kluwer Academic Publishers, 2005, pp. 1-54.

WILSON, Th. *Fine-grained subjectivity and sentiment analysis: recognizing the intensity, polarity, and attitudes of private states*. 2008. Ph.D. thesis, University of Pittsburgh, Pittsburgh, PA, USA, 214p.

ANEXOS

MASS 1.0
MANUAL DE ANOTAÇÃO

LUCIANA ÁVILA

POS LIN – UNIVERSIDADE FEDERAL DE MINAS GERAIS
CENTRO DE LINGUÍSTICA DA UNIVERSIDADE DE LISBOA

02 DE JANEIRO DE 2013

SUMÁRIO:

1 INTRODUÇÃO	1
2 NOÇÕES IMPORTANTES	2
2.1 MODALIDADE	2
2.2 VALORES MODAIS	2
O PROJETO MASS (MODAL ANNOTATION IN SPONTANEOUS SPEECH)	5
3.1 O SOFTWARE MMAX2	6
3.2 <i>TRIGGERS</i> , <i>SOURCES</i> E <i>TARGETS</i> : ELEMENTOS A SEREM ANOTADOS	6
3.2.1 <i>TRIGGERS</i>	6
3.2.2 <i>SOURCES</i>	11
3.2.2.1 <i>SOURCE OF THE EVENT</i>	11
3.2.2.2 <i>SOURCE OF THE MODALITY</i>	12
3.3 <i>TARGET</i>	15
3.4 POLARIDADE	27
3.5 REFERÊNCIAS	28

1 INTRODUÇÃO

Este manual apresenta o projeto MASS (Modal Annotation in Spontaneous Speech), um esquema de anotação semântica da modalidade desenvolvido para dados de fala espontânea do Português Brasileiro, descreve e explica em detalhes os passos da anotação, bem como discute os casos problemáticos e as decisões tomadas para solucioná-los.

2 NOÇÕES IMPORTANTES

2.1 MODALIDADE:

A definição da modalidade está longe de ser consensual. Neste projeto, defino este fenômeno como o julgamento ou a avaliação de um sujeito conceptualizador, que relativiza o que está sendo enunciado, em termos de grau de certeza e das noções de possibilidade, probabilidade, necessidade, obrigação e permissão. Ela está ancorada em uma situação comunicativa e vai cumprir diferentes funções pragmático-discursivas no curso da interação. (AVILA & MELLO, 2013; AVILA, 2014).

2.2 VALORES MODAIS:

(a) Epistêmicos: o significado epistêmico está relacionado com o grau de certeza de um conceptualizador sobre o material locutório enunciado e também se refere às noções de possibilidade, probabilidade e necessidade. Como exemplificado em (1), (2), (3) e (4):

- (1) *LAU: [195] também nũ **tem certeza** de nada //COM=\$ (bfamdl03)
- (2) *BAL: [107] <só que é aquela coisa /=INT= eu nũ &po> [/3]=EMP= eu nũ **posso** esperar crescer dentro disso //COM=\$ (bfamdl02)
- (3) *GIL: [36] né / es **deve** meter o pau //\$ (bfamcv01)
- (4) *REN: [230] o <dobro> //COM=\$ [231] pois é /=CMM= mas **teria que** ser uns dez reais /=CMM= né //PHA=\$ (bfamdl01)

(a.1) epistemic_knowledge: o conceptualizador (o falante ou uma outra entidade) expressa o seu conhecimento ou compreensão sobre algo.

- (5) *ANE: [152] <a> gente nũ <**sabia** que era> essa [/1]=SCA= essa <rua> //COM=\$ (bfamdl05)

(a.2) epitemic_belief: o conceptualizador expressa a sua crença ou opinião sobre algo.

- (6) *JAE: [84] mas **é lógico** que ea vai pôr ocês /=COM= uai //PHA=\$ (bfamcv02)

(a.3) epistemic_possibility: O conceptualizador apresenta o que enuncia como uma possibilidade.

(7) *TON: [173] dá pra &ma [/3]=EMP= dá pra jogar ela aqui /=CMM= ela vem na frente da quatro o' //CMM=\$ (bfamcv03)

(a.5) epistemic_probability: O conceptualizador apresenta o que enuncia como uma probabilidade, baseado em alguma evidência.

(8) *CAR: [95] deve ser <um> [/1]=SCA= alguns milhares de reais a conta /=COM= né //PHA=\$ (bpubcv02)

(a.6) epistemic_necessity: O conceptualizador apresenta o que enuncia como uma necessidade, baseado em conhecimento anterior.

(9) *DFL: [121] meu avô falava com papai /=INT= ele é doido /=CMM_r= só **pode** ser doido /=CMM_r= né /=PHA= &po +=EMP=\$ (bfammn02)

(a.7) epistemic_verification: O conceptualizador expressa incerteza em relação a um estado-de-coisas, evento ou atividade em foco.

(10) *ANE: [390] **olha** aí se nũ tem ninguém /=COM= César //ALL=\$ (bfamdl05)

(b) Deônticos: o valor deôntico normalmente está associado às convenções morais e sociais e, portanto, normalmente é usado para indicar obrigação, permissão ou proibição. Os exemplos (11) e (12) ilustram, respectivamente, uma obrigação e uma permissão:

(11) *RUT: [257] cê **tem que** ir chique //COM=\$ (bfamcv02)

(12) *BAL: [127] <cê pode> deixar o cabo bater no chão //COM=\$ (bfamdl02)

(b.1) deontic_obligation: O conceptualizador obriga a alguém, se vê obrigado ou obriga a si mesmo a realizar uma atividade por uma determinada razão.

(13) *CAR: [220] <aí> **tem que** olhar os vizinhos //COM=\$ (bpubcv02)

(b.2) deontic_permission: O conceptualizador permite alguém ou se permite realizar uma atividade, ou permite que algo aconteça.

(14) *BAL: [134] <cê **pode**> deixar o cabo bater no chão //COM=\$ (bfamdl02)

(b.3) deontic_prohibition: O conceptualizador proíbe alguém ou a si mesmo de fazer algo, ou proíbe que algo aconteça.

(15) *CAR: [250] hhh <palavrão> **nũ po'** falar não //COM=\$ (bfamcv03)

(b.4) deontic_necessity: O conceptualizador expressa suas necessidades ou a necessidade de uma outra pessoa ou grupo.

(16) *JOR: [50] nós **precisamo** criar esse hábito //COM=\$ (bfammn06)

(c) Dinâmicos: o tipo dinâmico, apesar de ser menos central nos trabalhos sobre modalidade (HUDDLESTON; PULLUM, 2002; KIEFER, 1994; SALKIE, 2009), se relaciona à capacidade, habilidade e volição/intenção de um conceptualizador, como mostram as ocorrências em (17), (18) e (19):

(17) *BRU: [383] <cê **conseguiu** ser> pior //COM=\$ (bfamcv04)

(18) *JOR: [51] é a mesma coisa que você comprar hoje um celular com tanta tecnologia /=CMB= e você **nũ consegue** usá-lo porque você **nũ aprende** a usar /=SCA= aqueles manuais tão extenso pra poder /=SCA= ter /=SCA= o =EMP= o &us [/1]=SCA= uso /=i-COB= né /=PHA= da própria tecnologia //COM=\$ (bfammn06)

(19) *GIL: [78] cês **querem** olhar com a gente /=COB= cês **querem** sugerir um lugar /=COB= <qualquer> coisa desse tipo //COM=\$

(c.1) dynamic_ability: O conceptualizador expressa a sua própria habilidade/capacidade ou a habilidade/capacidade de uma outra pessoa para realizar ou alcançar algo.

(20) *CEL: [85] e' **nũ consegue** fazer isso nunca tá meio de hhh> [/3]=SCA= <tá de ladinho> //COM=\$ (bfamcv03)

(21) *CAR: [150] porque ele ama /=COB= ele chama ela de /=SCA= fuminho /=COB= né /=COB= meu pretinho /=COB= papai nã **güenta** carregar mais /=COB= mas /=DCT= nã **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=\$ (bfammn05)

(c.2) dynamic_volition: O conceptualizador expressa as suas vontades, desejos, esperanças e intenções.

(22) *REN: [550] **espero** que eu não tinha [/1]=SCA= tenha perdido //COM=\$ (bfamd101)

Na Tabela 1 abaixo, apresento, em resumo, os valores modais, os subvalores a eles associados e a definição de cada um:

Valores	Subvalores	Definição
Epistêmico	conhecimento	O conceptualizador (o falante ou uma outra entidade) expressa o grau de conhecimento ou compreensão sobre algo.
	crença	O conceptualizador expressa a sua crença ou sua opinião sobre algo.
	possibilidade	O conceptualizador apresenta o que enuncia como uma possibilidade.
	probabilidade	O conceptualizador apresenta o que enuncia como uma probabilidade.
	necessidade	O conceptualizador apresenta o que enuncia como uma necessidade.
	verificação	O conceptualizador expressa incerteza em relação a um estado-de-coisas, evento ou atividade em foco.
Deôntico	obrigação	O conceptualizador se vê obrigado ou obriga a si mesmo a realizar uma atividade por uma determinada razão.
	permissão	O conceptualizador permite alguém ou a si mesmo a fazer algo, ou permite que algo aconteça.
	proibição	O conceptualizador proíbe alguém ou a si mesmo a fazer algo, ou proíbe que algo aconteça.
	necessidade	O conceptualizador expressa suas necessidades ou a necessidade de uma outra pessoa ou grupo.
Dinâmico	habilidade	O conceptualizador expressa a sua própria habilidade/capacidade ou a habilidade/capacidade de uma outra pessoa para realizar ou alcançar algo.
	volição	O conceptualizador expressa as suas vontades, necessidades, desejos, esperanças e intenções.

Tabela 1 – Valores e subvalores modais e suas definições

3 O PROJETO MASS (MODAL ANNOTATION IN SPONTANEOUS SPEECH)

Este projeto de esquema de anotação de modalidade em dados orais do português brasileiro segue diretamente o esquema proposto para o português europeu (HENDRICKX et al., 2012a; 2012b; MENDES et al., 2013) e, igualmente, inspira-se em outros esquemas de anotação previamente explorados para a língua inglesa (BAKER et al., 2010; SAURÍ et al., 2006, 2009; RUBINSTEIN et al., 2013), o japonês (MATSUYOSHI et al., 2010) e o chinês (CUI; CHI, 2013).

Este esquema se baseia nos pressupostos da Teoria da Língua em Ato e tem como unidade de referência de análise, portanto, o enunciado e as unidades informacionais (cf. CRESTI, 2000). Desta forma, difere-se de outros projetos, uma vez que não centra a sua análise na diamesia escrita e, portanto, não fornece uma anotação no domínio da sentença.

Nas subseções seguintes, apresento o software livre MMAX2¹³¹ (MÜLLER; STRUBE, 2006); sigo com a escolha dos valores modais e os elementos a serem anotados (*trigger*, *source of the event*, *source of the modality*, *target*) e a marcação da polaridade aplicada ao elemento *trigger*.

3.1 O software de anotação MMAX2

O MMAX2 (MÜLLER; STRUBE, 2006) é um software livre para anotação linguística de corpora em múltiplos níveis. Esta é uma ferramenta flexível para criar, navegar por e visualizar as anotações linguísticas. MMAX oferece uma interface visual para anotar sentenças pela marcação sequências textuais e criação de links entre os elementos marcados. O MMAX está escrito em Java (por razões de independência de plataforma) e as anotações são armazenadas em XML.

3.2 TRIGGERS, SOURCES E TARGETS: ELEMENTOS A SEREM ANOTADOS

3.2.1 TRIGGERS:

Os *triggers* são as “palavras ou sequência de palavras que expressam modalidade” (BAKER et al., 2010). Considero como *triggers* os verbos auxiliares e semi-auxiliares

¹³¹ <http://mmax2.sourceforge.net/>

modais, os verbos epistêmicos ou de atitude proposicional, os advérbios modais, as expressões adjetivas, as expressões lexicais que carregam modalidade.

Os índices modais considerados são:

- (a) verbos auxiliares e semi-auxiliares: “poder”, “dever”, “ter que”, “parecer”, “precisar”, “dar”, “conseguir”, “aguentar”, “querer”, “valer”, “adiantar”;
- (b) verbos de crença: “achar”, “acreditar”, “crer”, “imaginar”, “julgar”, “pensar”;
- (c) verbos de conhecimento e compreensão: “saber”, “ver”, “perceber”.
- (d) advérbios e as locuções adverbiais: “certamente”, “com certeza”, “exatamente”, “justamente”, “logicamente”, “mesmo”, “na verdade”, “óbvio”, “potencialmente”, “realmente”, “sem chance”, “sem dúvida”, “sinceramente”, “talvez”, “às vezes” “claro”;
- (e) adjetivos e locuções adjetivas: “capaz”, “é capaz”, “é claro”, “é lógico”, “é verdade”, “é verdadeira”, “foi verdade”, “lógico”, “mais certo é que tá”, “verdade”;
- (f) as expressões com valor modal, “era pra”, “pode saber”, “será”, “será que”, “tem certeza”, “tem chance”, “tem condição”, “tem condições”, “tem jeito”, “tenho certeza”, “tinha condições”, “tinha jeito”.

A. Se o *trigger* é:

A.1 um *verbo auxiliar ou semi-auxiliar modal*, anota-se o verbo modalizador como *trigger*.

(23) *FLA: [104] <hhh não /=INP= e a gente **tem que**> pesar a bolsa de novo
//=COM=\$ (bpubcv01)

trigger: tem que

(24) *CAR: [38] e /=DCT= falei com Deus também que eu nã **queria** buscar
//=COM_r (bfammn05)

trigger: queria

Se o verbo modalizador está em uma construção com os verbos auxiliares *ser*, *estar* ou *ter/haver*, anota-se apenas o modal.

(25) [99] já sofri o suficiente /=COB_r= agora eu tô **querendo** relaxar //COM_r=\$

trigger: querendo

(26) [271] eu tô **achando** que vai <chover> //COM=\$

trigger: achando

A.2 um **advérbio**, anota-se o advérbio ou locução adverbial como *trigger*.

(27) *LEO: [247] <**talvez** o Racing> //COM=\$ (bfamcv01)

trigger: talvez

(28) [44] que **às vezes** a gente sente uma dor numa hora //COM=\$ (bfamdl02)

trigger: às vezes

A.3 um **adjetivo** ou uma **construção adjetival**, anota-se o adjetivo ou toda a expressão adjetival, inclusive o auxiliar “ser”, uma vez que é parte do predicado nominal. Apenas os adjetivos com valor modal são considerados neste esquema de anotação, uma vez que não consideramos como modais os “avaliativos”.

(29) *PAU: [146] **capaz** //COM=\$ (bpubldl01)

trigger: capaz

(30) *JAE: [84] mas é lógico que ea vai pôr ocês /=COM= uai //PHA=\$
(bfamcv02)

trigger: às vezes

A.4. uma **expressão modal**, anota-se toda a expressão.

(31) [64] esquenta /=SCA= comida /=CMB= <yyyy /=CMB= **tem a possibilidade** de
um> monte de produto /=COM= o’ //CNT=\$ (bfamdl01)

trigger: tem a possibilidade

(32) *LUZ: [1] porque /=DCT= eu só soube que eu nũ [/6]=EMP= eu tive certeza absoluta que eu nũ era daqui quando eu saí //COM=\$ (bfamdl03)

trigger: tive certeza absoluta

B. O *trigger* não inclui as partículas negativas¹³². A negação é um fenômeno que interage diretamente com a modalidade.

(33) [16] nũ **sei** se eu jogo aí na três /=COB= nũ posso //COM=\$ (bfamcv03)

trigger: sei

C. As preposições e conectivos não são incluídos no *trigger*. A única exceção é o verbo semimodal *ter que*.

(34) [108] e eu **sei** que ea devia /=TOP= porque /=SCA= &he /=TMT= foi [/1]=EMP= foi &q [/1]=SCA= nas véspera d' eu vim embora //COM=\$

trigger: sei

(35) *BAL: [38] tá vendo /=CNT= a setinha **tem que** tá no cento-e-dez //COM (bfamdl02)

trigger: tem que

(36) [80] / e /=DCT= **teve que** amputar as duas perna aqui /=COB= hoje anda numa cadeira de roda /=COB= então a gente levava ela pra avó //COM=\$ (bfammn05)

trigger: teve que

D. Se no *trigger* vier intercalada qualquer partícula, não se considera esta palavra como parte do *trigger*.

¹³² No PB existem três possibilidades de realização da negação: negação simples preposta, negação simples posposta e dupla negação.

(37) [81] porque eu nunca confundo letras com <informática> /=COB= nũ **tem** nem **como** //COM=\$ (bfamdl02)

trigger: tem @ como¹³³

E. Quando uma unidade informacional ou um enunciado contém mais de um *trigger*, anota-se cada um dos *triggers* separadamente.

(38) [169] vão> vão /=CNT= <eu **acho** que **tem que** ser esses> //COM=\$ (bfamcv01)

trigger: acho

target: tem que ser esses

trigger: tem que

target: ser esses

(39) [220] e nũ **pode** falar e nũ **pode** <apontar o bagulho> //COM=\$ (bfamcv04)

trigger: poder

target: falar

trigger: pode

target: apontar o bagulho

F. Quando um *trigger* está em unidade de escansão (SCA), anota-se a unidade da qual ela herda as propriedades.¹³⁴

(40) [33] então assim /=INT= **espero** que /=SCA= isso nũ seja /=SCA= coisa pros times que jogam com a gente deixar de jogar com a gente //COM= \$ (bfamcv01)

trigger: espero

IU: COM

¹³³ Utilizo o sinal “@” para marcar sequências não-contínuas.

¹³⁴ A unidade de escansão constitui partes tonais diferentes de uma mesma unidade informacional. Ela introduz a unidade informacional que escande (escansão à esquerda), de maneira que a parte final da UI escandida especifica a função informacional do todo.

(41) *EVN: [134] acho <que a gente> **tem** /=SCA= **que** olhar direito //COM=\$
(bfamcv02)

trigger: tem que

IU: COM

G. Se um *trigger* estiver expresso em uma unidade de Apêndice de Comentário (APC) ou no Apêndice de Tópico (APT), seja como uma repetição de um índice expresso no Comentário ou Tópico ou como informação atrasada, quer dizer, quando se adiciona informação para facilitar a compreensão do enunciado pelo endereçado, anota-se a unidade de Comentário ou Tópico do qual é dependente.

(42) [4] ele nã 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=\$ (bfammn01)

trigger: deve

IU: COM (APC como “eco”)

(43) [157] distancia /=COM= que es **queria** <colocar> //APC=\$ (bpubcv02)

trigger: queria

IU: COM (APC como “informação atrasada”)

Como o *trigger* é o componente central na expressão modal, atribui-se as especificações do tipo de modalidade a ele. Especifica-se para cada *trigger* as seguintes características:

- (a) valor modal (‘*modal value*’);
- (e) polaridade (‘*polarity*’);
- (f) índice de modalidade (‘*polarity_cue*’);
- (g) unidade informacional em que está contido (“IU”).

A **polaridade** (‘*polarity*’) é o componente usado para marcar se há uma negação escopando o valor modal. Os valores atribuídos a este traço são “positivo” e “negativo”. O

índice de polaridade (“*polarity_cue*”) é um campo “*freetext*”, em que está identificado a palavra ou palavras que expressam a polaridade que afeta o *trigger*.

Uma **unidade informacional** pode coincidir com um enunciado ou apenas ser uma parte dele. Cada unidade cumpre uma função textual ou dialógica. As unidades informacionais textuais que podem conter um *trigger* são o Comentário (COM), Comentário Múltiplo (CMM), Comentário Ligado (COB), o Tópico (TOP), o Parentético (PAR) e o Introdutor Locutivo (INT)¹³⁵. Como os textos do minicorpus alimentados no MMAX2 não contêm a anotação da estrutura informacional, para a anotação deste traço é necessária a utilização da plataforma DB-IPIC (Database of Information Pattern of Italian C-ORAL-ROM)¹³⁶, plataforma de busca disponível gratuitamente

3.2.2 SOURCES:

Uma ‘*source*’ é definida, em sentido amplo, como “um agente ou uma organização que toma uma atitude em relação a um evento em uma sentença”¹³⁷ (MATSUYOSHI et al., 2010, p. 1459).

De acordo com Matsuyoshi et al. (2010, p. 1459), a *source* é um importante traço da modalidade estendida, porque “[esta] informação ajuda o leitora julgar a credibilidade dos conteúdos expressos em uma determinada sentença”¹³⁸.

3.2.2.1 SOURCE OF THE EVENT MENTION:

A *source of the event mention* é o produtor do evento. O produtor é, normalmente, o falante que enuncia um determinado conteúdo locutório.

A. A pessoa que produz o evento modal pode ser identificada, no texto, como a palavra, em caixa alta, no início de cada enunciado.

(44) *CAR: [98] <eu> **consigo** /=COM= né // =PHA=\$ (bfamcv03)

¹³⁵ Excepcionalmente, a unidade de Apêndice de Comentário (APC) pode conter um índice modal. Esta unidade integra o texto do Comentário e conclui o enunciado. É dependente da unidade informacional de Comentário. Para a anotação das ocorrências do *trigger* nesta unidade, ver regra “G” desta seção.

¹³⁶ Disponível em: <http://lablita.dit.unifi.it/app/dbipic/index.php>. Último acesso: 01 dez. 2013.

¹³⁷ No original: “an agent or an organization that takes an attitude toward an event mention in a sentence (MATSUYOSHI et al., 2010, p. 1459).

¹³⁸ Tradução para: “[this] information helps a reader judge credibility of contents conveyed from a given sentence” (MATSUYOSHI et al., 2010, p. 1459).

trigger: consigo
modal value: dynamic_ability
polarity: pos
IU: COM
source of the event mention: CAR

- (45) *JAN: [291] **na verdade** eu queria levar as duas /=COM= né //PHA=\$
(bpubdl02)

trigger: na verdade
modal value: epistemic_belief
polarity: pos
IU: COM
source of the event mention: JAN

- (46) *EMM: [79] **esperamos** <que esse> novo programa <que> vai vim /=TOP=
ele /=TOP=\$ (bpucv01)

trigger: esperamos
modal value: dynamic_volition
polarity: pos
IU: TOP
source of the event mention: EMM

3.5.2.2 SOURCE OF THE MODALITY:

A *source of the modality* é o agente, experienciador ou cognoscente que veicula a modalidade. As *sources* correspondentes a cada um dos valores modais são assim descritas:

Valores	Subvalores	Source
Epistêmico	conhecimento	O conceptualizador (o falante ou uma outra entidade) que expressa o grau de conhecimento ou compreensão sobre algo.
	crença	O conceptualizador que expressa a sua crença ou sua opinião sobre algo.
	possibilidade	O conceptualizador que apresenta o que enuncia como uma possibilidade.
	probabilidade	O conceptualizador que apresenta o que enuncia como uma probabilidade, baseado em alguma evidência
	necessidade	O conceptualizador que apresenta o material enunciado como uma necessidade, baseado em conhecimento anterior.
	verificação	O conceptualizador que expressa incerteza em relação a um estado-de-coisas, evento ou atividade em foco.
Deôntico	obrigação	O conceptualizador que obriga alguém, se vê obrigado ou obriga a si mesmo a realizar uma atividade por uma determinada razão.
	permissão	O conceptualizador que permite alguém ou a si mesmo a fazer algo, ou permite que algo aconteça.
	proibição	O conceptualizador que proíbe alguém ou a si mesmo a fazer algo, ou proíbe que algo aconteça.
	necessidade	O conceptualizador que expressa suas necessidades ou a necessidade de uma outra pessoa ou grupo.
Dinâmico	habilidade	O conceptualizador que expressa a sua própria habilidade/capacidade ou a habilidade/capacidade de uma outra pessoa para realizar ou alcançar algo.
	volição	O conceptualizador que expressa as suas vontades, necessidades, desejos, esperanças e intenções.

Tabela 2 – Sources correspondentes a cada valor e subvalor modais

As regras para a *source of the modality* são as seguintes:

A. A *source of the modality* e a *source of the event mention* são, normalmente, coincidentes:

(47) ***LUI**: [53] **eu** quero fazer o próximo campeonato no Arnaldinum /=CMM= e foda-se <pro seu Joaquim> //CMM=\$ (bfamcv01)

trigger: quero

modal value: dynamic_volition

polarity: pos

IU: CMM

source of the event mention: LUI

source of the modality: eu

(48) ***GIL**: [74] que **eu** acho que deu muito pau /=COM= nessa taça //APC=\$ (bfamcv01)

trigger: acho
modal value: epistemic_belief
polarity: pos
IU: COM
source of the event mention: GIL
source of the modality: eu

- (49) ***GIL**: [94] **a gente** podia fazer a taça aqui /=COB= todo mundo vai adorar
/=COB= e tal //COM=\$ (bfamcv01)

trigger: podia
modal value: epistemic_possibility
polarity: pos
IU: COB
source of the event mention: GIL
source of the modality: A gente

B. Casos difíceis:

B.1. Se a *source of the modality* não está explícita, anota-se:

- (a) o falante, se coincidir com ele ou com o grupo em que está inserido:

- (50) ***TIQ**: [227] deve ser da /=SCA= <da irmã> Geni /=COM= <né> //PHA=\$
(bpubcv02)

trigger: deve
modal value: epistemic_probability
polarity: pos
IU: COM
source of the event mention: TIQ
source of the modality: TIQ

- (51) ***REN**: [136] sabão em pó ã precisa não /=COM= né //PHA=\$

trigger: precisa
modal value: deontic_necessity
polarity: pos

IU: COM

source of the event mention: REN

source of the modality: REN

2. Quando a *source of the modality* é um pronome, ele é anotado:

(52) *CES: [307] ele <tem setenta-e-seis> metros quadrado **me** parece //=**COM**=\$

trigger: parece

modal value: epistemic_belief

polarity: pos

IU: COM

source of the event mention: CES

source of the modality: me

3. Quando a *source* de uma *deontic_permission* é externa ao participante, ela não é anotada:

(53) *BRU: [147] esse aqui não /=CMM= porque esse aqui é quando for desenhar
//=**CMM**=\$ [148] aí /=DCT= o [/1]=EMP= no jogo do desenho /=TOP= por
exemplo /=INT= um /=TOP= cê ã pode tirar o [/1]=SCA= o [/1]=EMP= <o>
[/1]=EMP= o lápis do papel /=CMM= <o outro tem que desenhar com a mão>
esquerda //=**CMM**=\$

*HEL: [149] <não /=CMM= nós ã> vão fazer o desenho //=**CMM**=\$

*LUC: [150] <eu posso desenhar ao invés de fazer mímica> //=**COM**=\$

(*bfamcv04*)

trigger: posso

modal value: deontic_permission

polarity: pos

IU: COM

source of the event mention: LUC

source of the modality: -

(54) *BRU: [88] <hhh aí /=DCT= passa um tiquim /=CMM= fala de novo /=CMM= né> //PHA=\$

[...]

*HEL: [96] pois é //COM=\$ [97] o que que <pode> <que que nũ pode> //COM=\$

(bfamcv04)

trigger: pode

modal value: deontic_permission

polarity: pos

IU: COM

source of the event mention: HEL

source of the modality: -

4. Se é uma pergunta retórica, anota-se o cognoscente. Se não está explícito, anota-se o verbo:

(55) *REG: [134] e ocê acha que eu nũ te conheço //COM_r=\$ (bfammn04)

trigger: acha

modal value: epistemic_belief

polarity: pos

IU: COM

source of the event mention: REG

source of the modality: ocê

(56) *GIL: [64] sabe que que eu penso /=COM= velho //ALL=\$ (bfamcv01)

trigger: sabe

modal value: epistemic_knowledge

polarity: pos

IU: COM

source of the event mention: GIL

source of the modality: sabe

3.2.3 TARGETS:

O *target* em textos falados é a expressão afetada pelo índice modal expresso pelo *trigger* dentro de uma unidade informacional. Portanto, diferente de outros esquemas de anotação que anotam o evento no escopo do item modal, não é necessário um predicado completo, como é usualmente tomado em textos escritos (uma cláusula subordinada ou um evento com todos os seus complementos e adjuntos). Na fala, como argumenta Cresti (no prelo), “um grande número de *chunks* falados, de fato, não podem ser definidos como cláusulas, mas como fragmentos, interjeições, advérbios, sintagmas, no entanto, funcionam perfeitamente do ponto de vista comunicativo”¹³⁹.

Assim, o *target* é anotado maximamente, admitindo-se descontinuidade, dentro do domínio da unidade informacional que contém o índice modal.

A. Se o *target* é:

A.1. um *sintagma*, anota-se todo o sintagma.

(57) *REN: [136] **sabão em pó** nũ precisa não /=COM= né // =PHA=\$ (bfamdl01)

trigger: precisa

modal value: deontic_necessity

polarity: neg

IU: COM

target: **sabão em pó**

polarity_tgt: pos

IU_tgt: COM

source of the event mention: REN

source of the modality: REN

¹³⁹ No original: “a large number of spoken chunks, indeed, cannot be defined as clauses, but are rather fragments, interjections, adverbs, phrases, while nevertheless functioning properly from a communicative point of view.” (CRESTI, no prelo).

(58) *REN: [463] precisando xxx //COM=\$ (bfamd101)

trigger: precisando
modal value: deontic_necessity
polarity: pos
IU: COM
target: **xxx**
polarity_tgt: pos
IU_tgt: COM
source of the event mention: REN
source of the modality: REN

(59) *DFL: [86] não /INP_r= nũ acredito **nisso** não //COM_r=\$ (bfammn02)

trigger: acredito
modal value: epistemic_belief
polarity: neg
IU: COM
target: **nisso**
polarity_tgt: pos
IU_tgt: COM
source of the event mention: DFL
source of the modality: DFL

A.2. uma *cláusula*, anota-se a cláusula, excluído o complementizador que a introduz.

(60) *GIL: [2] <ô /CNT= mas> /DCT= voltando à questão /COB= falando em [2]=EMP= e também falando em povo mascarado /COB= esse povo do Galáticos é muito palha /COB= eu acho que **es nũ deviam mais participar** /COM= e <tal> //UNC=\$ (bfamcv01)

trigger: acho
modal value: epistemic_belief
polarity: pos
IU: COM

target: es nũ deviam mais participar

polarity_tgt: neg

IU_tgt: COM

source of the event mention: GIL

source of the modality: GIL

(61) *LUI: [7] com certeza **es nũ vão participar** /=COM= uai> //=-PHA=\$

trigger: com certeza

modal value: epistemic_belief

polarity: pos

IU: COM

target: es nũ vão participar

polarity_tgt: neg

IU_tgt: COM

source of the event mention: LUI

source of the modality: LUI

(62) *JAE: [84] mas é lógico que **ea vai pôr ocês** /=COM= uai //=-PHA=\$ (bfamcv02)

trigger: é lógico

modal value: epistemic_belief

polarity: pos

IU: COM

target: ea vai pôr ocês

polarity_tgt: pos

IU_tgt: COM

source of the event mention: JAE

source of the modality: JAE

B. Quando o *target* tem polaridade negativa, incluir a partícula de negação no *target*:

(63) *CAR: [200] <acho que nũ> deu muito certo pra ele não /=COM= Toninho
//=-ALL=\$ (bfamcv03)

trigger: acho
modal value: epistemic_belief
polarity: pos
IU: COM
target: ã deu muito certo pra ele
polarity_tgt: neg
IU_tgt: COM
source of the event mention: CAR
source of the modality: CAR

B.1. No caso de dupla negação, não incluir a segunda partícula de negação no *target*:

(64) *LUZ: [210] <Nossa> /=EXP= esse **negócio de terra** ã dá ão /=COM=\$
(bfamd103)

trigger: dá
modal value: epistemic_possibility
polarity: neg
IU: COM
target: **esse negócio de terra**
polarity_tgt: pos
IU_tgt: COM
source of the event mention: LUZ
source of the modality: LUZ

C. Quando há um caso de *retracting*¹⁴⁰ no *target*, anotar apenas a última parte em que está completa a enunciação:

(65) *GIL: [170] <eu &a [/2]=EMP= eu acho que é> esse [/2]=SCA= **é esse aqui o'**
/=COM=\$ (bfamcv01)

¹⁴⁰ Como para a utilização do MMAX2, foi necessária a limpeza dos *retractings*, os seus vestígios são reconhecidos pela repetição de *chunks* e/ou mudança de planejamento.

trigger: acho
modal value: epistemic_belief
polarity: pos
IU: COM
target: **é esse aqui o'**
polarity_tgt: neg
IU_tgt: COM
source of the event mention: GIL
source of the modality: GIL

D. Se o *target* está em uma unidade de escansão (SCA), anotá-lo como uma única unidade informacional.

(66) *RUT: [208] cê pensa que **ele** /=SCA= **participa** da [/1]=SCA= &d [/1]=EMP=
desses presente //COM=\$ (bfamcv02)

trigger: pensa
modal value: epistemic_belief
polarity: pos
IU: COM
target: **ele@participa@desses presente**
polarity_tgt: pos
IU_tgt: COM
source of the event mention: RUT
source of the modality: CE

E. Se o *trigger* está em unidade de Parentético (PAR) e a expressão em seu escopo está em uma unidade informacional diferente, não se anota o *target*:

(67) *LUZ: [52] são duas vagas /=COM= eu acho //PAR=\$ (bfamd103)

trigger: acho
modal value: epistemic_belief
polarity: pos

IU: COM
target: -
polarity_tgt: -
IU_tgt: -
source of the event mention: LUZ
source of the modality: LUZ

(68) *FLA: [296] <oito /=CMM= né /=CMM= na verdade> //PAR=\$ (bfamd101)

trigger: na verdade
modal value: epistemic_belief
polarity: pos
IU: COM
target: -
polarity_tgt: -
IU_tgt: -
source of the event mention: FLA
source of the modality: FLA

F. *Target-dependent*:

O atributo *target-dependent* foi criado para os casos em que o *target* não está explícito em um determinado enunciado, mas é recuperável na cadeia referencial do texto.

Segundo Cresti (no prelo, p. 12), “[c]ada *chunk* linguístico, concebido para desempenhar uma determinada função textual (UT) dentro de um padrão informacional (PI), corresponde a uma *cena* (Barwise & Perry 1981; Fauconnier 1984) de um ponto de vista semântico. Como já dito, de um ponto de vista sintático uma UT pode até mesmo corresponder a uma coleção de fragmentos, mas a fim de permitir o desenvolvimento de uma função textual, as expressões participantes devem estar reunidas na mesma cena”.¹⁴¹

¹⁴¹ No original: “Each linguistic chunk, conceived to perform a certain textual function (TU) within an information pattern (IP), corresponds to a *scene* (Barwise & Perry 1981; Fauconnier 1984) from a semantic point of view. As we have already said, from a syntactic point of view a TU can even correspond to a collection of fragments, but in order to allow the development of a textual function, the participating expressions must be gathered within the same scene.” (CRESTI, no prelo, p. 12).

(69) *LUZ: [6] passei a vida toda num lugar errado //COM=\$ [7] que que é isso //COM=\$ [8] passei a vida toda fora d' água hhh //COM=\$ [9] que loucura //COM=\$ [10] aí que ocê sabe /COM= né //PHA=\$ [11] porque **quando cê chega num lugar que cê se sente em casa** /TOP= cê sabe imediatamente //COM=\$ [12] é um +=EMP=\$

*LAU: [13] ham ham //COM=\$

*LUZ: [14] &he /TMT= o corpo **sabe** /CMM= tudo **sabe** //CMM=\$ [15] o seu humor /CMB= a sua +=EMP=\$ [16] tudo //COM=\$

(bfamd103)

trigger: sabe

modal value: epistemic_knowledge

polarity: pos

IU: COM

target: **quando cê chega num lugar que cê se sente em casa**

polarity_tgt: pos

IU_tgt: TOP

source of the event mention: LUZ

source of the modality: LUZ

(70) *GIL: [2] <ô /CNT= mas> /DCT= voltando à questão /TOP= falando em [2]=EMP= e também falando em povo mascarado /TOP= esse povo do Galáticos é muito palha /COB= eu acho que **es nũ deviam mais participar** /COM= e <tal> //UNC=\$

[...]

*LUI: [5] <eu **acho** não> //COM=\$

*LEO: [6] <**com certeza**> //COM=\$

(bfamcv01)

trigger: acho

modal value: epistemic_belief

polarity: neg

IU: COM

target: es nũ deviam mais participar

polarity_tgt: neg

IU_tgt: COM

source of the event mention: LUI

source of the modality: LUI

trigger: com certeza

modal value: epistemic_belief

polarity: pos

IU: COM

target: es nũ deviam mais participar

polarity_tgt: neg

IU_tgt: COM

source of the event mention: LEO

source of the modality: LEO

Apesar de, em termos da anotação em seu conjunto, esta ser uma informação redundante, a decisão se justifica para fins de recuperação das informações referenciais mais facilmente. Destaco que esta não é uma tentativa para a anotação de anáforas ou correferências, para o qual um projeto específico deve ser empreendido.

G. Casos difíceis:

G.1. No caso de usos formulaicos, em que a expressão do *target* envolve aspectos não-verbais (como gestos, expressões faciais etc.) que poderiam ser recuperados apenas se disponíveis arquivos de imagem, o *target* é inespecífico e, portanto, não é anotado:

- (71) *CEL: [138] <então tá /=CMM= amarelo> //CMM=\$
*HEL: [139] é //COM=\$ [140] tá //COM=\$
*BRU: [141] amarelo //COM=\$
*LUC: [142] beleza <xxx> //UNC=\$
*BRU: [143] <tá> //COM=\$
*LUC: [144] <posso> //COM=\$
*BRU: [145] <e> +=EMP=\$
*LUC: [146] e tem um dadinho diferente ali //COM=\$
(bfamcv01)

trigger: posso
modal value: deontic_permission
polarity: pos
IU: COM
target: **inespecífico**
polarity_tgt: -
IU_tgt: -
source of the event mention: LUC
source of the modality: **inespecífica**

Nesta ocorrência em (71), os participantes da interação estão discutindo, esclarecendo e conversando sobre as regras de um jogo. Após as explicações iniciais, um dos participantes ‘LUC’ pede a permissão a uma pessoa para realizar alguma ação. Não podemos precisar a expressão no escopo do *trigger* modal, nem o conceptualizador da permissão, que poderia ser qualquer um dos envolvidos.

G.2. Se o *trigger* está em uma unidade informacional e o *target* afetado por ele está em uma unidade subsequente, considera-se o todo como uma “construção padronizada”, ou seja, “construções realizadas através de UTs [unidades textuais], com cada qual desenvolvendo uma função informacional diferente” (CRESTI, no prelo, p. 18)¹⁴² e com modalidades também distintas, e anota-se a expressão afetada contida na unidade informacional subsequente e seu valor:

(72) *LUC: [1] pois é então /=INT= cê **sabe** que /=INT= lá na Letras /=TOP= **se eu conto essa história** /=SCA= **que eu sou parente do Drummond** /=COM= né
//=PHA=\$ (bfammn02)

trigger: sabe
modal value: epistemic_knowledge
polarity: pos
IU: COM
target: se eu conto essa história que eu sou parente do Drummond
polarity_tgt: pos
IU_tgt: COM

¹⁴² Tradução para: “constructions performed across TUs, with each developing a different information function” (CRESTI, no prelo, p. 18)

source of the event mention: LUC

source of the modality: cê

(73) *FLA: [239] só que a gente não **sabe** /=COB_s= **se elas têm agaivé** /=CMB¹⁴³=
se elas <têm hepatite> /=CMB= se elas +=EMP=\$

trigger: sabe

modal value: epistemic_knowledge

polarity: pos

IU: COB

target: **se elas têm agaivé@se elas têm hepatite**

polarity_tgt: pos

IU_tgt: COB

source of the event mention: FLA

source of the modality: a gente

Para o *target*, atribuem-se as seguintes características:

- (d) polaridade (*'polarity'*);
- (e) índice de modalidade (*'polarity_cue'*);
- (f) unidade informacional em que está contido (*"IU"*).

✓ Identificação de *targets*, de acordo com o valor modal:

(a) epistemic_knowledge

(74) *OSV: [67] <a de telefone> ãu veio ainda /=CMM= que eu ãu sei **quando que vai** //CMM=\$ (*bpubcv02*)

trigger: sei

modal value: epistemic_knowledge

polarity: neg

IU: CMM

target: quando que vai

polarity_tgt: pos

IU_tgt: CMM

¹⁴³ As unidades de CMB, não descritas na literatura, correspondem, em sua grande maioria, à unidade de COB e foram, portanto, substituídas por ela na anotação.

source of the event mention: OSV
source of the modality: eu

(b) epistemic_belief

(75) *REN: [321] **o Neve é caro** <mesmo> //COM=\$ (bfamdl01)

trigger: mesmo

modal value: epistemic_belief

polarity: pos

IU: COM

target: o Neve é caro

polarity_tgt: pos

IU_tgt: COM

source of the event mention: REN

source of the modality: REN

(c) epistemic_possibility

(76) *CEL: [229] **cê pode ter feito <LIBRAS>** //COM=\$ (bfamcv04)

trigger: pode

modal value: epistemic_possibility

polarity: pos

IU: COM

target: ter feito LIBRAS

polarity_tgt: pos

IU_tgt: COM

source of the event mention: CEL

source of the modality: cê

(d) epistemic_probability

(77) *MAI: [13] **o diâmetro dea** deve **dar** uns [1]=SCA= **uns quarenta a cinquenta centímetro** de [1]=SCA= de &s [2]=EMP= **de grossura** /=COM= o diâmetro dela //APC=\$ (bfammn01)

trigger: sabe

modal value: epistemic_probability

polarity: pos

IU: COM

target: o diâmetro dea@dar@ uns quarenta a cinquenta centímetro@ de grossura

polarity_tgt: pos

IU_tgt: COM

source of the event mention: MAI

source of the modality: MAI

(e) epistemic_necessity

(78) *REN: [231] pois é /=CMM= mas teria que **ser uns dez reais** /=CMM= né
//=-PHA=\$ (bfamd101)

trigger: teria que

modal value: epistemic_necessity

polarity: pos

IU: CMM

target: ser uns dez reais

polarity_tgt: pos

IU_tgt: CMM

source of the event mention: REN

source of the modality: REN

(f) epistemic_verification

(79) *ANE: [393] olha aí se **nũ tem ninguém** /=CMM= César //=-CMM=\$
(bfamd105)

trigger: olha

modal value: epistemic_verification

polarity: pos

IU: CMM

target: nũ tem ninguém

polarity_tgt: neg

IU_tgt: CMM

source of the event mention: ANE

source of the modality: ANE

(g) deontic_obligation

(80) *JAN: [239] depois eu tem que comprar uma //COM=\$ (bpubdl02)

trigger: tem que

modal value: deontic_obligation

polarity: pos

IU: COM

target: comprar uma

polarity_tgt: pos

IU_tgt: COM

source of the event mention: JAN

source of the modality: eu

(h) deontic_permission

(81) *PAU: [247] ele não pode chegar no final da semana e /=SCA= desmanchar tudo o que foi feito durante a semana /=COM= né //PHA=\$ (bpubdl01)

trigger: pode

modal value: deontic_permission

polarity: neg

IU: COM

target: ele@chegar no final da semana e@ desmanchar tudo o que foi feito durante a semana

polarity_tgt: pos

IU_tgt: COM

source of the event mention: PAU

source of the modality: PAU

(i) deontic_prohibition

(82) *CES: [82] nã pode **subir** /=CMM= é contramão //CMM=\$ (bfamdl05)

trigger: pode

modal value: deontic_prohibition

polarity: pos

IU: CMM

target: subir
polarity_tgt: pos
IU_tgt: CMM
source of the event mention: CES
source of the modality: CES

(j) deontic_necessity

(83) *JOR: [50] nós precisamos **criar esse hábito** //COM=\$ (bfamnn06)
trigger: precisamos
modal value: deontic_necessity
polarity: pos
IU: COM
target: criar esse hábito
polarity_tgt: pos
IU_tgt: COM
source of the event mention: JOR
source of the modality: nós

(h) dynamic_ability

(84) *CAR: [150] porque ele ama /=COB= ele chama ela de /=SCA= fuminho
/=COB= né /=COB= meu pretinho /=COB= papai ãü güenta **carregar mais**
/=COB= mas /=DCT= ãü 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=\$ (bfamnn05)
trigger: güenta
modal value: dynamic_ability
polarity: neg
IU: COB
target: carregar mais
polarity_tgt: pos
IU_tgt: COB
source of the event mention: CAR
source of the modality: papai

(i) dynamic_volition

(85) *DFL: [63] mas ele quis que **todos os filhos estudassem** //COM=\$
(bfammn02)

trigger: quis

modal value: dynamic_volition

polarity: pos

IU: COM

target: todos os filhos estudassem

polarity_tgt: pos

IU_tgt: COM

source of the event mention: DFL

source of the modality: ele

3.3 POLARIDADE:

O traço da polaridade, como explicitado em seções anteriores, é marcado para o *trigger* e o *target*. Possui dois valores: “positivo” e “negativo” e o valor *default* é o positivo.

A polaridade global do enunciado não é computada. Se um *trigger* e um *target* ambos possuam polaridade negativa, a polaridade dos componentes será marcada separadamente.

Nos casos de valores deônticos de permissão e proibição marcados por uma partícula negativa, o que os diferencia é a força da permissão ([-forte] ou [+forte], respectivamente). Para a *deontic_permission*, marca-se a polaridade como “negativa”, já para a *deontic_prohibition*, marca-se a polaridade como “positiva”.

(86) *PAU: [247] ele não pode chegar no final da semana e /=SCA= desmanchar tudo o que foi feito durante a semana /=COM= né //PHA=\$ (bpubdl01)

trigger: pode

modal value: deontic_permission

polarity: neg

IU: COM

target: ele@chegar no final da semana e@ desmanchar tudo o que foi feito durante a semana

polarity_tgt: pos

IU_tgt: COM
source of the event mention: PAU
source of the modality: PAU

(87) *CAR: [250] hhh <palavrão> nũ po' falar não //COM=\$ (bfamcv03)
trigger: po'
modal value: deontic_prohibition
polarity: pos
IU: COM
target: palavrão @ falar
polarity_tgt: pos
IU_tgt: COM
source of the event mention: CAR
source of the modality: CAR

A decisão de se individualizar o valor de proibição deve-se à tentativa de se cobrir exemplos do tipo “Não é proibido X” ou “Você não está proibido de X”. Mais que uma permissão, estes tipos de ocorrência se configurariam como uma proibição deontica de polaridade negativa.

3.7 REFERÊNCIAS:

- AVILA, L.; MELLO, H. Challenges in modality annotation in a Brazilian Portuguese spontaneous speech corpus. In: *Proceedings of WAMM-IWCS2013*, Potsdam, Germany, 2013.
- AVILA, L. Modalidade em perspectiva: estudo baseado em corpus de fala espontânea do Português Brasileiro. Tese (Doutorado em Estudos Lingüísticos). Belo Horizonte, Universidade Federal de Minas Gerais, 2014.
- BAKER, K., BLOODGOOD, M., DORR, B. J., FILARDO, N. W., LEVIN, L., PIATKO, C. A modality lexicon and its use in automatic tagging. In: *Proceedings of the Seventh Language Resources and Evaluation Conference (LREC'10)*, Valletta, Malta. ELRA, 2010, p. 1402-1407.
- BARWISE, K. J.; PERRY, J. *Situations and attitudes*. Cambridge: MIT Press, 1981.
- CRESTI, E. *Corpus di italiano parlato*. Firenze: Accademia della Crusca, 2000.
- CRESTI, E. Syntactic properties of spontaneous speech in The Language into Act Theory: data on Italian complements and relative clauses. In: Raso, T. & Mello, H. (eds.). *Spoken Corpora and Linguistic Studies*. Amsterdam/Philadelphia: John Benjamins, no prelo.

- CUI, Y.; CHI, T. Annotating Modal Expressions in the Chinese Treebank. In: *Proceedings of WAMM-IWCS2013*, Potsdam, Germany, 2013.
- FAUCONNIER, G. *Espaces mentaux*. Paris: Les éditions de Minuit, 1984.
- HENDRICKX, I.; MENDES, A.; MENCARELLI, S.; SALGUEIRO, A. Modality Annotation Manual, version 1.0. Centro de Linguística da Universidade de Lisboa, Lisboa, Portugal, 2012a.
- HENDRICKX, I.; MENDES, A.; MENCARELLI, S.. Modality in Text: a Proposal for Corpus Annotation. In: *Proceedings of the Eighth International Conference on Language Resources and Evaluation - LREC 2012*, Istanbul, May 21-27, 2012b.
- HUDDLESTON, R.; PULLUM, G. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press, 2002.
- KIEFER, F. Modality. In: ASHER, R. E. (ed.). *The Encyclopedia of language and linguistics*. Oxford: Pergamon Press, 1994, p. 2515–2520.
- MATSUYOSHI, S., EGUCHI, M., SAO, C., MURAKAMI, K., INUI, K., MATSUMOTO, Y., (2010), Annotating Event Mentions in Text with Modality, Focus, and Source Information, in Proceedings of the Seventh conference on the International Language Resources and Evaluation (LREC'10).
- A.; HENDRICKX, I.; SALGUEIRO, A.; ÁVILA, L. Annotating the interaction between focus and modality: the case of exclusive particles. In: *Proceedings of the 7th Linguistic Annotation Workshop & Interoperability with Discourse*, Sofia, Bulgaria, 2013, p. 228–237.
- MÜLLER, C.; STRUBE, M. Multi-level annotation of linguistic data with MMAX2. In: BRAUN, S.; KOHN, K.; MUKHERJEE, J. (eds.). *Corpus technology and language pedagogy*. New Resources, New Tools, New Methods. Frankfurt: Peter Lang, p. 197-214. (English Corpus Linguistics, Vol.3)
- RUBINSTEIN, A.; HARNER, H.; KRAWCZYK, E.; SIMONSON, D.; KATZ , G.; PORTNER, P. Toward fine-grained annotation of modality in text. In: *Proceedings of WAMM-IWCS2013*, Potsdam, Germany, 2013.
- SALKIE, R. Degrees of modality. In: SALKIE, R.; BUSUTTIL, P.; VAN DER AUWERA, J. (eds.). *Modality in English: theory and description*. Topics in English Linguistics, 58. Berlin: Mouton de Gruyter, 2009, p. 79-103.
- SAURÍ, R., VERHAGEN, M., PUSTEJOVSKY, J. Annotating and recognizing event modality in text, in Proceedings of the 19th International FLAIRS Conference, FLAIRS 2006.
- SAURÍ, R.; PUSTEJOVSKY, J. FactBank: a corpus annotated with event actuality. *Language Resources & Evaluation*, 43, 227–268, 2009.