

Verbos epistêmicos na fala espontânea: um estudo *corpus-based* do português europeu/*Epistemic verbs in spontaneous speech: a corpus-based study of European Portuguese*

Luciana Beatriz Ávila*
Giulliana Mendes Cária**

RESUMO

A modalidade é uma categoria semântica aplicada a uma produção linguística que qualifica e relativiza um material locutivo em termos de grau de certeza, possibilidade, necessidade, capacidade e volição (AVILA, 2014). Fundamentando-se nos pressupostos da Teoria da Língua em Ato (CRESTI, 2000) e utilizando a Linguística de Corpus como ferramenta de pesquisa, o objetivo do trabalho foi empreender uma descrição do comportamento dos verbos de caráter epistêmico, ou verbos de atitude proposicional (URMSON, 1969; HALL, 1958; HOOPER, 1975; HÜBLER, 1983; VENIER, 1991) ou predicados de estados mentais (NUYTS, 2001), em um *corpus* de fala espontânea do português europeu, o C-ORAL-ROM português (BACELAR DO NASCIMENTO et al., 2005). Foi utilizada uma amostra da parte informal desse *corpus*, composta de 20 textos de três tipologias interacionais (7 monólogos, 7 diálogos e 6 conversações, privados e públicos). Foi realizada uma busca manual de todos os índices que expressam modalidade, e sua posterior anotação nos tipos epistêmicos, deônticos e dinâmicos e seus respectivos subvalores. Na amostra, foram encontrados 936 itens modais lexicais, dos quais 237 ocorrências correspondem aos verbos epistêmicos. A partir de análises qualitativas e quantitativas de dados empíricos, observou-se que o verbo epistêmico mais frequente é o “achar”, responsável por 79,74% de um total de 237 ocorrências. Em termos sintáticos, um número de padrões é usado e o mais comum é o verbo como introdutor de uma oração encaixada. Semanticamente, estes verbos sinalizam para os diferentes graus comprometimento do conceptualizador em relação ao material locutivo enunciado. Pragmaticamente, podem sinalizar, em posição parentética, a atenuação de uma asserção anterior.

PALAVRAS-CHAVE: Modalidade; Verbos epistêmicos; Fala espontânea; Português europeu.

ABSTRACT

Modality is a semantic category applied to a linguistic production, which qualifies and relativizes a locutive material, in terms of degree of certainty, possibility, probability, capacity and volition (Author, 2014). Based on the Language into Act Theory (CRESTI, 2000) and using Corpus Linguistics as methodological orientation, this paper aims at describing the behavior of epistemic verbs, or verbs of propositional attitude (URMSON, 1969; HALL, 1958; HOOPER, 1975; HÜBLER, 1983; VENIER, 1991) or mental state predicates (NUYTS, 2001), in a spontaneous speech corpus of European Portuguese, the C-ORAL-ROM (BACELAR DO NASCIMENTO; BITTENCOURT GONÇALVES; VELOSO; ANTUNES; BARRETO; AMARO, 2005). The sample contains 20 texts of three different interactional typologies (7 monologues, 7 dialogues and 6 conversations, private and public). We proceed to a manual search of these modal indexes and its annotation of epistemic, deontic and dynamic modal values and their respective sub-values. We found 936 lexical modal markers, and 237 of all occurrences were epistemic verbs. From quantitative and qualitative analysis, we observe that the most frequent epistemic verb is “achar” (“to think”), responsible for 79,74% of all occurrences. In syntactic terms, different patterns are used, and the most common is the verb as introducer of an embedded clause. Semantically, these verbs signal to the different degrees of

* Doutora em Estudos Linguísticos pela Universidade Federal de Minas Gerais. Professora Adjunto II na Universidade Federal do Sul da Bahia – UFSB, luciana.avila@ufsb.edu.br.

** Graduada em Letras com habilitação em Português/Francês pela Universidade Federal de Viçosa - UFV. Bolsista no Programa Institucional de Bolsa de Iniciação à Docência – PIBID/CAPES, giullianamcaria@gmail.com.

commitment of the conceptualizer in relation to the uttered locutive material. Pragmatically, they may signal, in parenthetical position, the attenuation of a previous assertion.

KEYWORDS: Modality; Epistemic Verbs; Spontaneous Speech; European Portuguese.

1 Introdução

A modalidade é um fenômeno semântico que como afirmam Mello e colaboradoras (2009, p. 105), “até onde se sabe, foi mencionada pioneiramente na teoria lógica aristotélica, que apresenta seis valores: verdadeiro, falso, possível, impossível, necessário e contingente”. Esta é uma noção semântica definida, normalmente, a partir da ideia comum de que é comprometida com a verdade da proposição e com a opinião do falante em relação ao conteúdo proposicional de seu enunciado.

A maioria dos estudos sobre a modalidade têm-se centrado na análise de textos escritos (PALMER, 1986, BYBEE et al.; 1994; PAPAFRAGOU, 2000; NUYTS, 2001, para o inglês; PIETRANDREA, 2005, para o italiano; CORNILLIE, 2009, para o espanhol; HENDRICKX et al., 2012, para o português europeu). No que diz respeito à oralidade, especificamente para os dados de fala espontânea¹, destacamos os trabalhos de Tucci (2007, 2008, 2011), Cresti (2002), para o italiano; e os de Mello et al. (2009); Mello et al. (2010); Mello et al. (2011); Mello e Raso (2011), Mello e Caetano (2012); Avila e Mello (2013), Avila (2014), para o português brasileiro. Tanto os trabalhos para o italiano, quanto para o brasileiro se apoiam na Teoria da Língua em Ato (CRESTI, 2000), embasada empiricamente, que propõe que a unidade de análise da fala são os enunciados, divididos em unidades tonais, com correlatos prosódicos, e possuem uma contraparte ilocucionária. Na formulação de Moneglia e Cresti (2006, p. 91), o enunciado é “a entidade

¹ Fala espontânea é definida por Cresti e Scarano (1998, p. 5) “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.

linguística mínima que pode ser pragmaticamente interpretada; i.e. a entidade linguística que é ‘concluída’ e ‘autônoma’ desde um ponto de vista pragmático”.²

Este trabalho, portanto, alinhado aos estudos anteriores, visa a contribuir para o estudo da modalidade em dados orais do português europeu, especificamente a descrição do comportamento dos verbos de carácter epistêmico, em termos de frequência absoluta e relativa, escolhas lexicais, ocorrência em tipologia interacional e padrões de uso sintático e pragmático. Para o desenvolvimento da nossa proposta, utilizamos uma amostra de um corpus de fala espontânea do português europeu, o C-ORAL-ROM português (BACELAR DO NASCIMENTO et al., 2005).

Este artigo está organizado em mais quatro seções. Na primeira, apresentamos o conceito de modalidade e a definição dos valores modais; na segunda, expomos os procedimentos metodológicos, a apresentação do corpus de estudo e a organização dos dados; em seguida, passamos à discussão dos resultados e analisamos os dados e, por fim, concluímos com algumas considerações acerca do estudo em tela.

2 A definição de modalidade

A modalidade é um fenómeno mui caro aos estudos semânticos e, sob diferentes perspectivas, sejam da filosofia, da lógica modal ou da linguística, foi definida de forma diversa por pesquisadores de variadas persuasões.

Na tradição linguística, esta noção pode ser tomada a partir de uma abordagem mais formal, sob a perspectiva de uma semântica de mundos possíveis (KRATZER, 1991), como uma distinção entre *realis* e *irrealis* (GIVÓN, 1995) ou contraste entre factual e não factual (MITHUM, 1995), e como o nível semântico mais alto da organização da proposição e uma forma de o falante expressar a sua “subjatividade” (BALLY, 1932, 1942; LYONS, 1977; PALMER, 1990).

² Para mais informações sobre a Teoria da Língua em Ato, ver Cresti (2000).

Desde este último ponto de vista, assumimos neste trabalho que a modalidade é uma modificação, aplicada a uma produção linguística de um conceptualizador³, que qualifica e relativiza um material locutivo enunciado, em termos de grau de certeza, possibilidade, necessidade, capacidade e volição (AVILA, 2014). Em outras palavras, as expressões linguísticas que exprimem esses valores semânticos (verbos modais ou de nuance modal, verbos epistêmicos, expressões adjetivais, advérbios, expressões perifrásticas) permitem ao falante/conceptualizador qualificar o que enuncia como “possível”, “provável”, “necessário”, “de acordo com as suas crenças, opiniões e certezas”, “de acordo com seus desejos, vontades e capacidade”. Como exemplos, apresentamos ocorrências de um corpus oral do português brasileiro, o C-ORAL-BRASIL (RASO; MELLO, 2012):

- (a) *RUT: <mas eles> também deve ter condições / uai //
- (b) *JAE: mas é lógico que ea vai pôr ocês / uai //
- (c) *CAR: e / e eu achei que ia me dar trabalho //
- (d) *LUC: <isso pode ser / potencialmente> divertido //

No exemplo em (a), temos um item modal “deve” que indica uma probabilidade e modifica o material linguístico “ter condições”. Já em (b) temos uma expressão adjetival que exprime certeza e incide sobre “ea vai pôr ocês”. O terceiro exemplo em (c), é um verbo epistêmico ou verbo de crença “achei” que introduz uma crença do falante sobre “ia me dar trabalho”. Os marcadores modais em (d) “pode” e “potencialmente” carregam o significado de possibilidade, e qualificam “isso ser divertido”.

2.1 A definição dos valores modais

Assim como a definição de modalidade está longe de ser um consenso, a tipologia de valores modais também varia, ao longo do tempo, e entre pesquisadores. A tradição filosófica

³ O conceptualizador, em termos langackerianos (LANGACKER, 2008), corresponde, em primeiro lugar, ao falante; em segundo, ao ouvinte/endereçado e, derivativamente, a uma terceira pessoa cuja perspectiva é levada em conta.

estabelece o estudo das modalidades alética, epistêmica e deôntica, relacionadas, respectivamente, com as noções de verdade/falsidade, conhecimento e conduta. Já na perspectiva linguística, o valor epistêmico, que indica o grau de comprometimento com a verdade de uma proposição, é estável em todas as abordagens teóricas, em oposição aos não-epistêmicos, que variam, por exemplo, em deôntico ou raiz, que indica obrigação e permissão (SWEETSER, 1990); ou dinâmico, que aponta para a capacidade do indivíduo; bulético ou bulomaico, que indica a vontade ou intenção do falante; circunstancial, que expressa o que é possível ou necessário de acordo com determinadas circunstâncias (von FINTEL, 2006); e teleológica, que 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). Mais ainda, alguns autores seguem uma tipologia específica como Palmer (1986), que divide as modalidades em proposicional e de evento, ou Bybee et al. (1995) que propõem as modalidades orientada-para-o-agente; orientada-para-o-falante e epistêmica.

Em pesquisa mais ampla sobre a modalidade no português europeu (AVILA et al., 2015), consideramos, para fins de anotação dos dados, os valores epistêmico, deôntico e dinâmico e seus respectivos subvalores (possibilidade, probabilidade, conhecimento e crença; obrigação e permissão; capacidade e volição). No escopo deste trabalho, no entanto, levaremos em conta, em função do objeto estudado, apenas o tipo epistêmico, definido por Bybee e colaboradoras (1994, p. 179) como “a extensão até a qual o falante está comprometido com a verdade da proposição”. Nas palavras de Nuyts (2001, p. 21),

[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 que, no caso default, é o mundo real (ou antes, a interpretação do avaliador desse mundo [...])

O valor epistêmico, como mencionado anteriormente, pode ser subdividido em quatro subvalores: crença, conhecimento, possibilidade e probabilidade, os dois primeiros focos dessa pesquisa (AVILA, 2015). Abaixo, apresentamos quatro ocorrências, extraídas do mesmo corpus de fala espontânea do português brasileiro:

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

O exemplo em (e) representa a certeza que um estado-de-coisas vai ocorrer (“saber alguma coisa”), ao contrário de (f) e (g) que apontam para a certeza de que não vão ocorrer. Já em (h), temos um exemplo da nuance de possibilidade que Nuyts (2001) trata como um estado-de-coisas possível de ocorrer ou não, dadas determinadas circunstâncias, no caso a hipótese de “uma coordenação pedagógica melhor”.

Apresentaremos, a seguir, os procedimentos metodológicos e nosso *corpus* de pesquisa.

3 Procedimentos metodológicos

3.1 Abordagem da linguística de *corpus*

A Linguística de Corpus (doravante LC) ocupa-se da coleta e da exploração de corpora, isto é, de conjuntos de dados linguísticos coletados com critérios específicos, visando à realização de pesquisas sobre uma língua ou variedade linguística (SARDINHA, 2004). Dessa forma, um corpus não deve ser definido apenas como “um conjunto de material linguístico”. É necessário que este conjunto seja formado por dados autênticos, seja representativo de um determinado grupo linguístico ou uma determinada língua natural, não provenientes de programação ou produção artificial, com uma arquitetura específica, visando a um objetivo específico, e deve estar armazenado computacionalmente.

Para o autor, a Linguística de Corpus é uma ferramenta com a qual é possível realizar o processamento automático de textos, informatização de grandes bases de dados e a montagem de sistemas inteligentes de reconhecimento de voz, bem como o gerenciamento de informação.

De acordo com Shepherd (2009, p. 151), o objeto de estudo da LC não é um fenômeno mental, e sim um fenômeno social, algo passível de ser observado e acessível através de evidências que advêm de corpus digitalizado. McEnery e Wilson (1996) confirmam o status de metodologia da LC, ao afirmarem que abordagens que partem da análise de um corpus podem ser aplicadas a praticamente qualquer área de investigação linguística.

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

O C-ORAL-ROM (CRESTI; MONEGLIA, 2005) é um corpus multilíngue de fala espontânea, representativo das quatro principais línguas românicas europeias – italiano, português, francês e espanhol. Este conjunto de corpora busca suprir a carência de dados de fala oral espontânea no panorama das línguas românicas e fornecer uma base de dados adequada ao estudo da diamesia oral. Um aspecto importante na constituição desse projeto é a comparabilidade entre seus corpora, a qual garante a possibilidade da realização de estudos interlinguísticos de língua falada espontânea.

Os corpora são ditos comparáveis porque se encaixam na seguinte matriz esquemática: são divididos em uma parte formal e outra informal. A parte formal compreende arquivos de texto e áudio em “contexto natural” (discurso político; debate político; sermões; aulas; explicações profissionais; conferências; reuniões de negócio; interações na área de direito); “media” (noticiário; esportes; entrevistas; ciência; meteorologia (previsão do tempo); imprensa científica; reportagens; *talk shows*); e “conversas telefônicas”. A parte informal é dividida em contextos familiares/privados (eventos de fala no ambiente familiar ou em um contexto social privado) e públicos (eventos de fala em um contexto social público) que, por sua vez, se subdividem em três tipologias interacionais, igualmente distribuídas: monólogos (eventos de fala com apenas um participante desempenhando uma tarefa comunicativa); diálogos (eventos de fala com dois participantes); conversações (eventos de fala com mais de dois participantes). A estratégia da amostra do domínio da fala espontânea adotada no C-ORAL-ROM é baseada na representação de diferentes tipos de contexto de uso e não está balanceada no que se refere às

características dos falantes (origem geográfica, idade, sexo, escolaridade e profissão), descritas nos metadados de cada sessão.

No que diz respeito ao C-ORAL-ROM português (BACELAR DO NASCIMENTO; BETTENCOURT GONÇALVES; VELOSO; ANTUNES; BARRETO; AMARO, 2005), este apresenta um total de 114 textos que totalizam aproximadamente 30 horas de gravação e representam a diatopia da área metropolitana de Lisboa. Nem todo o material presente no corpus do PE foi gravado especialmente para o C-ORAL-ROM, tendo sido reaproveitada uma quantidade significativa de textos. No total, o corpus do PE contém gravações, com transcrições ortográficas, que cobrem um período de 30 anos (de 1970 a 2002).

3.2.1 Delimitação do corpus e organização dos dados

Como subcorpus, foi utilizada uma amostra da parte informal do C-ORAL-ROM português, contendo 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 (5 privadas e 1 pública).

A partir da organização dos dados em uma tabela com o nome de cada arquivo, a tipologia interacional, a ocorrência no corpus, o valor modal (epistêmico, deôntico e dinâmico) e os respectivos subvalores (epistêmico: conhecimento, crença, probabilidade e possibilidade; deôntico: necessidade, obrigação e permissão; dinâmico: capacidade e volição), foi realizada a classificação dos índices lexicais modais (verbos modais e de nuance modal; verbos epistêmicos, advérbios e expressões adverbiais, adjetivos e expressões adjetivais, e expressões emergentes) no minicorpus C-ORAL-ROM português. Na amostra, foram encontrados um total de 936 índices lexicais modais, sendo que, desses 585 verbos, 348 modais e de nuance modal e 237 epistêmicos.

Para este trabalho, como apontado anteriormente, recortamos a análise de enunciados contendo verbos de crença e conhecimento, como “achar”, “pensar”, “saber”, entre outros, realizada quantitativa e qualitativamente em termos de frequência absoluta e relativa, bem como de sua distribuição por tipologia interacional.

4 Resultados e discussão dos dados: os verbos epistêmicos

Os verbos epistêmicos, também chamados de verbos de crença ou, como descritos por Nuyts (2001), predicadores de estados mentais, funcionam 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” (VENIER, 1991, p. 68 *apud* TUCCI, 2007, p. 172)⁴. Além disso, sintaticamente, podem se configurar de diversas maneiras, permitindo o encaixamento de cláusulas e diferentes complementos, a própria omissão de complemento e, em termos de estrutura informacional, a parentetização.

Na nossa amostra, foram encontrados 237 *tokens* de verbos epistêmicos, distribuídos em 11 *types*⁵, conforme a Tabela 1 abaixo:

| Verbos epistêmicos | # ocorrências | % |
|--------------------|---------------|--------|
| Achar | 189 | 79,74% |
| Achar-se | 1 | 0,43% |
| Acreditar | 2 | 0,84% |
| Calcular | 1 | 0,43% |
| Considerar | 1 | 0,43% |
| Crer | 3 | 1,26% |
| Imaginar | 3 | 1,26% |
| Julgar | 2 | 0,84% |
| Pensar | 32 | 13,50% |
| Saber | 1 | 0,43% |

⁴ Tradução das autoras 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).

⁵ O termo “*token*” se refere ao número total de palavras em um corpus, não importa quantas vezes esteja repetido. O termo “*type*” se refere ao número de palavras distintas em um corpus.

| | | |
|-------|---|-------|
| Supor | 2 | 0,84% |
|-------|---|-------|

Tab. 1 – Frequência dos verbos epistêmicos na amostra do PE

Podemos observar a discrepância entre as ocorrências do verbo “achar” e as demais; este único verbo representa quase 80% da amostra de verbos epistêmicos, em suas diferentes formas ou *tokens* (“acho”, a mais frequente, “achei”, “achava” “achámos”, “achas”, “acha”, “achou”, “acham”, “achar” e “achando”). Em segundo lugar, aparece o verbo “pensar”, com 32 ocorrências, o que corresponde a 13,5% do total. Vejamos alguns exemplos:

- (i) eu **acho** que nunca tinha visto ninguém com um sono tão pesado como ele // (pfamd109)⁶
- (j) eu **achei** / &eh / muito interessante // (pfamd116)
- (k) *LUI: / têm / para / eu depois dizer / não / olhe / **acho** que esta é melhor / ou aquela é melhor / ou não sei quê // \$ (pfammn19)
- (l) mas/as nossas sobremesas são/ assim mais/ substanciais/ **acho** eu hhh // \$ (pfammn08)
- (m)*RIT: / &eh / **penso** que os portugueses / que lá estão / não terão os mesmos / hábitos / que os brasileiros / não é // \$ (pfamd104)
- (n) *PAU: [<] <eu / de início **pensava**> que era a gorda que ia ganhar // \$ (pfamcv04)

Em (i), (j), (k) e (l), temos quatro das ocorrências com o verbo “achar”, com diferentes conformações sintáticas e, conseqüentemente, diferentes funções semânticas. Nos exemplos (i) e (k), os verbos introduzem orações encaixadas, no entanto, na primeira ocorrência, a forma “acho” expressa uma crença do falante, enquanto na segunda, a forma é um introdutor de opinião. Em (j), a forma “achei” vem seguida de uma expressão adjetival, sem o complementizador “que”, e

⁶ Os nomes dos arquivos são construídos da seguinte forma: a primeira letra representa a língua do corpus, no nosso caso o português europeu (“p”); 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.

também indica uma avaliação do falante. Já na ocorrência em (l), temos o que Schneider (2007) denomina “cláusula parentética reduzida”, que se caracteriza, sintaticamente, por não ser uma cláusula subordinada (i.e., não há um elemento de ligação explícito), por ter distribuição livre no enunciado (neste caso, posição final), pelo verbo não apresentar um dos argumentos exigidos pela sua valência (no nosso exemplo, o argumento vem representado semanticamente na unidade anterior). Pragmaticamente, esta construção em posição parentética funciona como atenuador da asserção anterior.

Por último, em (m) e (n), temos duas ocorrências com o verbo “pensar” que introduzem uma oração encaixada e, semanticamente, introduzem a opinião ou julgamento do falante.

Os verbos “crer” e “imaginar”, por sua vez, apareceram três vezes; “acreditar”, “julgar” e “supor”, duas; e “achar-se”, “calcular”, “considerar” e “saber”, apenas uma vez cada.

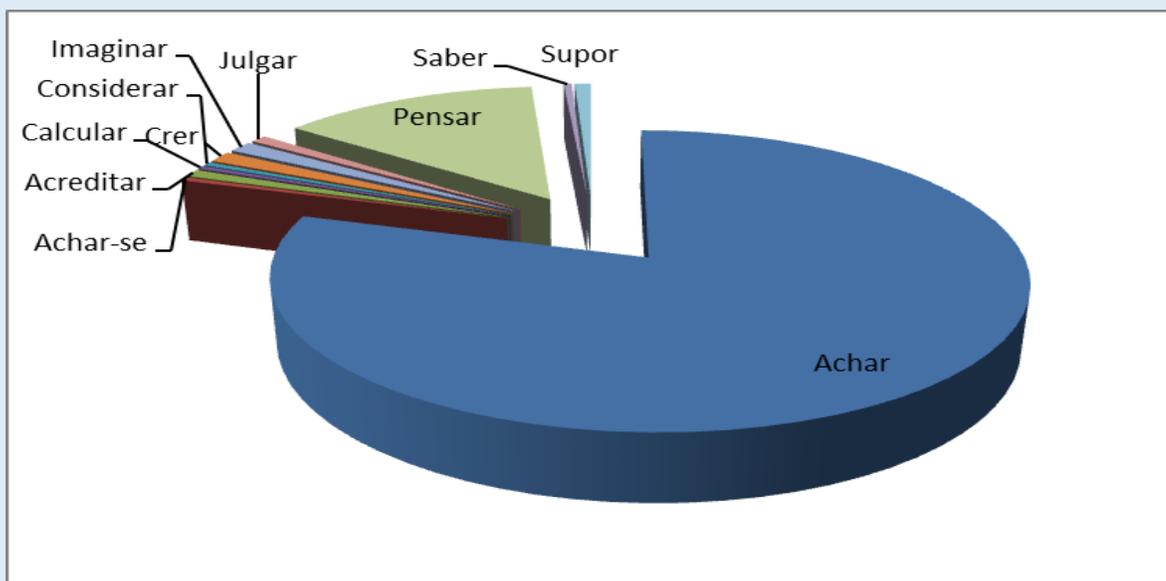


Fig. 1 – Frequência dos verbos epistêmicos na amostra do PE

Vejamos as ocorrências:

- (o) **creio** que / se chama mesmo "o abismo" // \$ (pfammn02)
- (p) *JOS: porque eu **imagino** que / cada / cliente / é um &c / é um caso especial / não é? (ppubl08)
- (q) e depois era o crente / que **acreditava** / que o homem / vai conseguir ser ainda / &ah / retrato de deus na terra / e ser / também capaz de ele / pôr o seu selo / e de ele poder conduzir esse / esse / esse animal / &ah / bravo que é a técnica e a &tecno / e as tecnologias // (pfammn02)
- (r) há muita gente que **julga** que o gajo é um / coiso // (pfamcv09)
- (s) *LUI: / &se / **suponho** que deva ser sul-americano // (pfammn19)
- (t) e achava que aquilo era / teatralizar / e era ser / &ah / actor de teatro / estar a conversar / sozinho / **achando-se** importantíssimo / e achando que o que diz / &eh / merece ficar para a posteridade // (pfammn02)
- (u) *JOS: e / eu **calculo** que / tratar / &eh / do cabelo duma pessoa / por um lado deve implicar no / no aspecto estético / etcétera / mas por outro lado / a natureza do próprio cabelo / não é / é preciso / ter <essas noções / não é>? (ppubl08)

- (v) *JOS: e / ao &se / ao estar no Brasil / &eh / acha que / de facto / podemos **considerar** a cultura brasileira / como uma extensão da cultura portuguesa? (pfamdl04)
- (w) mas / &ah / penso que / &eh / cada vez mais / &ah / &ah / se tem que investir / &ah / num cidadão / que **sabe** / à partida / que vai conviver / com muitos outros / em situações de desigualdade / &ah / económicas / &a / é óbvio / mas de outras desigualdades // (pfammn02)

Em todos os exemplos apresentados, os diferentes verbos expressam uma crença, uma avaliação, um julgamento ou uma opinião de um conceptualizador em relação ao material locutivo enunciado, e introduzem orações encaixadas.

Destacamos, neste conjunto, as ocorrências em (q) e (r), nas quais são empregadas a terceira pessoa do discurso (“o crente” e “muita gente”) como quem opera a modificação sobre o enunciado. Em (q), temos que o falante enuncia a crença de uma terceira pessoa “o crente”, e, em (r), o falante enuncia um julgamento generalizado, indicado pelo SN genérico “muita gente”. Parece que, há, nos dois casos, diferentes níveis de conceptualização: um primeiro nível relacionado ao falante que pode ser considerado uma “fonte aninhada”⁷ (WIEBE et al., 2005), uma vez que é enunciada uma avaliação filtrada pelos seus olhos, e, um segundo nível relacionado ao conceptualizador em terceira pessoa e à sua (suposta) crença/avaliação, portanto, uma interpretação epistêmica.

Observamos, nesses casos, o limite entre evidencialidade e modalidade epistêmica. Segundo afirma Pietrandrea (2005, p. 33), “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” ou como coloca de Haan (2005, p. 379), “a evidencialidade *assevera* a evidência, enquanto a modalidade epistêmica *avalia* a evidência”.

⁷ Segundo Wiebe e colaboradoras (2005, p. 9), a fonte de um evento de fala é o falante ou o escritor. No entanto, o falante ou o escritor pode enunciar ou escrever sobre a opinião ou emoção de uma outra pessoa. Neste caso, ocorre um aninhamento de fontes. O falante/escritor é considerado, então, uma fonte aninhada (no original, “*nested source*”).

4.1 Frequência dos verbos e tipologia interacional

As ocorrências dos verbos epistêmicos foram classificadas quantitativamente de acordo com a tipologia interacional (monólogo, diálogo ou conversação, privados ou públicos).

Analisando os números para tipologia interacional, os resultados são: 60 exemplares para monólogos, 7 em monólogos públicos e 53 em privados; 99 exemplares para diálogos, 23 para públicos versus 76 para diálogos privados; e 78 exemplares para conversações, 12 em públicas versus 66 em conversações privadas. É possível visualizar os dados na Tabela 2:

| | Privado | Público | TOTAL |
|---------------------|----------------|----------------|--------------|
| Monólogos | 53 | 7 | 60 |
| Diálogos | 76 | 23 | 99 |
| Conversações | 66 | 12 | 78 |
| TOTAL | 195 (82,3%) | 42 (17,7%) | 237 |

Tab. 2 – Distribuição de *tokens* de verbos epistêmicos por tipologia interacional

Observamos que há uma alta ocorrência de verbos epistêmicos em situações dialógicas privadas e públicas, 99 para diálogos e 78 para conversações, o que corresponde a 74,7% da amostra. Dadas as condições de coleta do corpus, temos que os participantes das interações são frequentemente elicitados pelos entrevistadores a emitirem sua opinião sobre um determinado tema. Dessa forma, o uso dos verbos de atitude proposicional é justificado pela sua própria definição, qual seja, a de sinalizar o comprometimento do falante em relação ao que se enuncia. Já a diferença de *tokens* encontrados nos textos públicos ou privados se dá pelo fato de que dentre os 20 textos selecionados, apenas quatro são referentes a interações públicas.

6 Considerações Finais

A partir de análises qualitativas e quantitativas, observou-se que o verbo epistêmico mais frequente é o “achar”, em suas diferentes formas, com 189 ocorrências, o que corresponde a

79,74% de um total de 237 de epistêmicos. Em seguida, temos o verbo “pensar”, com 32 exemplares, o que corresponde a 13,5% do total da amostra.

Apesar de não ter sido um parâmetro de análise, podemos atestar pelos exemplos que os enunciados que contêm estes verbos se configuram em diferentes padrões sintáticos, como introdutores de orações encaixadas, sem complementizador explícito, e com omissão de argumentos. Ainda há a ocorrência de construções cunhadas por Schneider (2007) como “cláusulas parentéticas reduzidas” (X [acho eu]) que, pragmaticamente, funcionam como mitigadoras da asserção anterior.

Em termos de distribuição por tipologia interacional, há uma maior ocorrência de epistêmicos em interações dialógicas (74,7% da amostra). Isto se dá pela própria constituição do corpus, em que os interactantes são elicitados a fornecerem sua opinião ou avaliação sobre um determinado tema.

Por fim, como já apontado por Avila (2009) para dados do português brasileiro, estes índices modais podem provocar a discussão sobre os limites entre a modalidade epistêmica e a evidencialidade. Além disso, este trabalho pode fomentar estudos comparativos sobre modalidade e evidencialidade com a variante brasileira do português e com outras línguas românicas.

Referências

AVILA, L. B. Pensar a modalidade: marcação epistêmica e evidencial no português brasileiro falado. In: X Encontro de Linguística de Corpus: aspectos metodológicos dos estudos de corpora, 2012, Belo Horizonte. Anais do X Encontro de Linguística de Corpus: aspectos metodológicos dos estudos de corpora. Belo Horizonte: Faculdade de Letras da UFMG, 2012. v. 1. p. 37-56.

ÁVILA, L. Modalidade em perspectiva: estudo baseado em corpus de fala espontânea do português brasileiro. 253f. Tese (Doutorado – Estudos Linguísticos). Belo Horizonte, Universidade Federal de Minas Gerais, 2014.

AVILA, L. B. MASS - Modal Annotation in Spontaneous Speech: Semantic annotation scheme for modality in a spontaneous speech Brazilian Portuguese corpus. Veredas (UFJF. Impresso), v. 19, p. 1-11, 2015.

ÁVILA, L. B.; MELLO, H. Challenges in modality annotation in a Brazilian Portuguese spontaneous speech corpus. In: *Proceedings of WAMM-IWCS2013*, Potsdam, Germany, 2013.

AVILA, L. B.; CÁRIA, G. M.; LUCHI, S. S. Modalidade na fala espontânea: um estudo corpus-based do português europeu. Projeto de pesquisa, Universidade Federal de Viçosa, 2015. (Manuscrito).

BACELAR DO NASCIMENTO, M. F., J. BETTENCOURT GONÇALVES, R. VELOSO, S. ANTUNES, F. BARRETO & R. AMARO. The Portuguese Corpus. In: CRESTI, E.; MONEGLIA, M. (Eds.). *C-ORAL-ROM: Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam / Philadelphia: John Benjamins, 2005, p. 163-207 (with DVD).

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.

BYBEE, J.; PERKINS, R.; PAGLIUCA, W. *The evolution of grammar*. Chicago: The University of Chicago Press, 1994.

CRESTI, E. *Corpus di italiano parlato*. Firenze: Accademia della Crusca, 2000. 2 v.

CRESTI E. Illocuzione e modalità. In: BECCARIA, P.; MARELLO, C. (Ed.). *La parola al testo*. Scritti per Bice Mortara-Garavelli. Torino: Ed. dell'Orso, 2002. p. 133-145

CRESTI, E.; MONEGLIA, M. *C-ORAL ROM: Integrated reference corpora for spoken Romance languages*. Amsterdam/Philadelphia: John Benjamins, 2005.

CRESTI, E.; SCARANO, A. Sur la notion de parlé spontané. In : BILGER, M. (a cura di). *Corpus. Méthodologie et applications linguistiques*, Actes du colloque Questions de méthode dans la linguistique sur corpus, Champion, Paris, 2000, p. 340-349.

von FINTEL, K. Modality and language. In: BORCHET, D. M. (ed.). *Enciclopedia of Philosophy*. Detroit: MacMillan, 2006.

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.

GIVÓN, T. Modal prototypes of truth and action. In: GIVÓN, T. *Functionalism and grammar*. Amsterdam / Philadelphia: John Benjamins, 1995, p. 111-174.

de HAAN, F. Typological approaches to modality. In: FRAWLEY, F. *The expression of modality*. Berlin/New York: Mouton de Gruyter, 2005.

HALL, R. Assuming: on the set of positing words. *Philosophical Review*, 67, p. 52-75, 1958.

HOOPER, J. B. On assertive predicates. In: KIMBALL, J. P. (Ed.). *Syntax and semantics*, 4. New York / London: Academic Press, 1975, p. 91-124.

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, 2012.

HÜBLER, A. *Understatements and hedges in English*. Amsterdam / Philadelphia: John Benjamins, 1983.

- KRATZER, A. Modality. In: VON STECHOW, A.; DIETER, W. *Semantics: an International Handbook of Contemporary Research*. Berlin: de Gruyter. 1991, p. 639–650.
- LANGACKER, R. W. *Cognitive grammar*. Cambridge: Cambridge University Press, 2008.
- LYONS, J. *Semantics*. vol. 1. Cambridge: Cambridge University Press, 1977.
- LYONS, J. *Semantics*. vol. 2. Cambridge: Cambridge University Press, 1977.
- McENERY, T., WILSON, A. *Corpus linguistics*. Edinburgh: Edinburgh University Press, 1996.
- MELLO, H.; MELO, E.; CARVALHO, J.; CÔRTEZ, P. Prolegômenos sobre modalidade. *Domínios da Linguagem: Revista Eletrônica de Linguística*. Ano 3, nº 1, p. 104-134, 2009.
- MELLO, H.; CARVALHO, J.; CÔRTEZ, P. Modalização na fala espontânea do português brasileiro: um primeiro mapeamento de índices morfolexicais. *Revista de Estudos Linguísticos*, Belo Horizonte, v. 18, n.2, p. 105-133, jul/dez. 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: FALE/UFMG, 2012. (Manuscrito).
- 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.
- MONEGLIA, M.; CRESTI, E. C-ORAL-ROM – Prosodic Boundaries for Spontaneous Speech Analysis In: KAWAGUCHI Y.; ZAIMA, S.; TAKAGAKI, T. (Eds.). *Spoken Language Corpus and Linguistic Informatics*. Amsterdam: John Benjamins, 2006, p. 89-113.
- NUYTS, J. *Epistemic modality, language and conceptualization: a cognitive-pragmatic perspective*. Amsterdam/Philadelphia: John Benjamins, 2001.
- PALMER, F. R. *Mood and modality*. Cambridge: Cambridge University Press, 1986.
- PALMER, F. R. *Mood and modality*. 2nd ed. Cambridge: Cambridge University Press, 1990.
- PAPAFRAGOU, A. *Modality: issues in the semantics-pragmatics interface*. Amsterdam/New York: Elsevier Science, 2000.
- PIETRANDREA, P. *Epistemic modality: functional properties and the Italian system*. Amsterdam / Philadelphia: John Benjamins, 2005.

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.

SARDINHA, T. B. *Linguística de corpus*. Barueri, SP: Manole, 2004.

SCHNEIDER, S. *Reduced parenthetical clauses as mitigators*. A corpus study of spoken French, Italian, and Spanish. Amsterdam / Philadelphia: John Benjamins, 2007.

SHEPHERD, T. M. G. O Estatuto da Linguística de Corpus: Metodologia ou Área da Linguística? *Matraga*, Rio de Janeiro, v.16, n.24, jan./jun. 2009. p. 150-172.

SWEETSER, E. *From etymology to pragmatics: metaphorical and cultural aspects of semantic structure*. Cambridge: Cambridge University Press, 1990.

TUCCI, I. L'espressione lessicale della modalità nel parlato spontaneo: Analisi del corpus CORAL- 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. 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.

URMSON, J. O. Parenthetical verbs. *Mind* 61: 480-496, 1969.

VENIER, F. *La modalizzazione assertiva. Avverbi modali e verbi parentetici*. Milano: Franco Angeli, 1991.

WIEBE, J., WILSON, T., CARDIE, C. Annotating expressions of opinions and emotions in language. Kluwer Academic Publishers, 2005, p. 1-54.

Data de recebimento: 15/09/2017

Data de aceite: 23/10/2017