

**A PRESENÇA DE UNIDADES FRASEOLÓGICAS
NO ATLAS LINGUÍSTICO DO CEARÁ (ALECE, 2010)**

Vicente Martins (UVA e UFC)
vicente.martins@uol.com.br

Este trabalho, sob a orientação da Professora Dra. Rosemeire Monteiro-Plantin (UFC), insere-se na linha de pesquisa Aquisição, Desenvolvimento e Processamento da Linguagem do Programa de Pós-Graduação em Linguística da Universidade Federal do Ceará (UFC). A partir de aportes da dialetologia, etnolinguística, sociolinguística, lexicologia e psicologia cognitiva, procedemos com um levantamento de unidades fraseológicas (locuções e compostos nominais), registradas pelo Atlas Linguístico do Ceará, doravante ALECE (2010), decorrentes do universo vocabular dos informantes escolarizados e dos informantes analfabetos.

As unidades fraseológicas analisadas, em nosso trabalho, são frases feitas, combinatórias sintagmáticas, locuções, compostos ou expressões cristalizadas na memória declarativa de longo prazo dos informantes, cujo sentido, geralmente, não é literal e se caracterizam por sua fixação, não-composicionalidade, idiomaticidade e metafóricidade. Para esta pesquisa, consideramos, assim, para maior ênfase na discussão, as seguintes abordagens para a descrição das unidades fraseológicas no ALECE (2010): (a) a *etnolinguística*, avaliando os procedimentos dialetológicos adotados pela pesquisa, através das técnicas de pré-questionário e questionário, para a coleta e a análise dos dados linguísticos e extralinguísticos; b) a *cognitiva*, as estratégias psicolinguísticas, utilizadas pelos documentadores, para a recuperação de itens lexicais armazenados, nas memórias declarativas de longo prazo (semântica e episódica) de seus entrevistados.

As definições fraseológicas dos dicionários de linguística mais antigos, como as de Camara Junior (1981, p. 142) e Dubois et alii (1993, p. 93 e p. 330), definem-nas como frases cristalizadas numa língua cuja combinação ou sintagma tem um caráter *estabilizado*. Para nós, contrastando, brevemente, com as acepções mais “estáveis” e reproduzidas em grande parte pelos dicionaristas, postulamos que, por apresentarem uma gama de variações léxicas e fonéticas, as unidades fraseológicas são uma *enunciação* do ponto de vista linguístico, o que equivale a dizer serem

eventos de fala, em que se ressalta o papel dos falantes num dado contexto comunicativo.

A questão da composição das unidades fraseológicas, especialmente as locuções e compostos nominais, é um dos temas centrais da linguística cognitiva. A linguista Nicole Delbecque (2008) afirma que, na estrutura interna dos compostos nominais, por exemplo, há uma estreita dependência do significado de base das diferentes classes de palavras que se combinam. Unidades fraseológicas do tipo “chuva forte”, chuva grande, “chuva grossa”, “chuvas torrenciais” e “chuva-tempestade”, evocadas, no ALECE (2010), por informantes analfabetos para o conceito de *temporal*, denotariam algo de caráter estável, por oposição, numa frase como “fazer finca-pé”, que nos sugere “manter-se firme em resolução, em opinião; fincar-se, obstar-se, teimar”, que evocaria coisa menos estável. (MARTINS, 2010)

Com base ainda nas postulações cognitivistas de Nicole Delbecque, diríamos que no grupo de unidades fraseológicas, entre os quais podemos encontrar muitos exemplos no ALECE (2010), como nos lemas “alco-celeste”, “alquiceleste”, “arca-da-aliança”, produzidos por informantes escolarizados, ou “acso-irpe” e “arco-ili”, relacionados ao lema *arco-íris*, sugere-nos, por força do princípio de pluriverbalidade, toda uma espécie de relações semânticas entre os dois componentes do frasema, indicando, também, o princípio de iconicidade observado no estabelecimento de “certa semelhança entre a forma do enunciado e aquilo que ele representa” (DELBECQUE, 2008, p. 26). Esse princípio de iconicidade se manifesta, ainda, do ponto de vista cognitivo, através de outros princípios de estruturação da linguagem, encontrados nas unidades fraseológicas, selecionadas em nossa pesquisa, como o *princípio da ordem linear*, onde observamos, a disposição linear dos componentes no interior do vocábulo, como no caso de “galão-d’água”, podendo ser assim segmentado em seus três componentes: galão + de + água.

Os procedimentos onomasiológicos e semasiológicos utilizados no ALECE (2010), com prevalência daqueles sobre estes, como nos indica Bessa (1982), levam-nos a considerar as unidades fraseológicas, presentes no ALECE (2010) como enunciados fraseológicos, diríamos, sociolinguística e psicolinguisticamente, em que verificamos, a partir das respostas dadas pelos informantes aos inquiridores, formas *memorizadas* resultantes de meio social e de experiências já vividas. E, como tais, não podem ser vistas apenas e, formalmente, como “frases feitas”, definitivamente acabadas, isto é, *frames* estocados em nossa mente, mas, subs-

tancialmente são expressões que “metaforizam” e, portanto, “corporificam” nossas experiências e vivências pessoais.

1. *Objetivos e metodologia do ALECE (2010)*

Do ponto de vista metodológico, o *Atlas Linguístico do Estado do Ceará* fundamentou-se em premissas e pressupostos de diversos modelos, entre os quais a teoria da variação ou sociolinguística quantitativa, a etnolinguística, a geolinguística e a dialetologia diatópica. Para este trabalho, interessou-nos, em particular, verificar, entre os aspectos supra citados, os de viés sociolinguístico indicados na pesquisa dialetológica, por evidenciar, de forma nítida, a relação (ou correlação) entre formas de dizer e os níveis de escolaridade dos falantes. O viés psicolinguístico decorre da recuperação das unidades fraseológicas serem recuperadas diante das perguntas de natureza onomasiológica do Questionário do ALECE (2010) como, por exemplo, o documentador faz a pergunta ”Se são dois dias antes de hoje, a gente chama” com o objetivo de obter a resposta ou decisão lexical “*anteontem*”.

O ALECE (2010) teve por objetivo apurar a realidade linguística do Ceará, de modo a possibilitar uma descrição dos traços linguísticos dos falantes cearenses, nomeadamente fonológicos, gramaticais e lexicais. Entre os objetivos específicos do ALECE, os linguistas assinalam os seguintes: (a) proporcionar, aos sistemas estadual e municipal de ensino, informações linguísticas para a adequação do ensino de língua materna na educação básica, a partir de suas peculiaridades regionais; (b) detecção dos traços linguísticos da cultura cearense; (c) possibilidade, com base nos dados do ALECE, de um exame crítico do material pedagógico adotado pela rede de ensino no Estado Ceará; (d) contribuição para o desenvolvimento e aperfeiçoamento dos métodos de dialetologia diatópica ou regional; e (e) formação e aperfeiçoamento de pesquisadores na área dialetológica.

Em substância, os pesquisadores do ALECE (2010) presumem que a visibilidade ou alcance social e cultural do ALECE (2010) é dependente do cumprimento de metas socioeducacionais: (a) detecção dos problemas sociolinguísticos e político-educacionais das comunidades; b) realização de uma análise (socio)linguística e educacional dos problemas diagnósticos; e c) apresentação de propostas linguístico-educacionais e metodológicas alternativas (BESSA, 2010, p. 67)

Acreditamos que, para a repercussão do ALECE nos sistemas de ensino, inicialmente, os governos, estadual e municipais deem um tratamento pedagógico aos dados pesquisa dialetológica, disseminando as informações linguísticas sobre o falar cearense ou falares cearenses, ali contidas, junto aos gestores educacionais e docentes, especialmente os pedagogos que, a partir de uma discussão sobre o que pode ser contemplado nas suas propostas pedagógicas, considerem ou reconsiderem, para o cumprimento dos objetivos educacionais, previstos na legislação educacional, as novas habilidades e competências curriculares requeridas da educação básica, com base nos resultados do Atlas Linguístico.

2. *Perfil social da amostra do ALECE (2010)*

No perfil social da amostra do ALECE (2010) foram descritas as variáveis escolaridade, idade, sexo e localização geográfica. Para nosso estudo, levamos em conta, apenas, para o levantamento de unidades fraseológicas, a variável escolaridade que nos permitiu pôr em evidência premissas sociolinguísticas relacionadas às relações entre língua e sociedade, com ênfase ao comportamento linguístico dos membros da comunidade cearense e de como as variações linguísticas são determinadas pelas relações sociais, culturais e econômicas existentes. A escolaridade para efeito cartográfico ou lexicográfico foi codificada em informante escolarizado (ou alfabetizado) e informante analfabeto. No caso do informante escolarizado, foi considerado o indivíduo que, tendo ou não concluído o 1º ciclo do ensino fundamental (1º ao 5º ano) ou outro nível de ensino, sabe ler e escrever. Já o analfabeto, foi assim considerado quando o informante não frequentou a escola ou o que, tendo-a frequentado, não sabe ler nem escrever. (BESSA, 2010, p. 70).

Outras variáveis, em um trabalho futuro, poderão ser consideradas tais como idade, sexo e localização geográfica. Em termos quantitativos, o perfil de informantes, por nível de escolaridade, em ALECE (2010), indica o seguinte: a) *quanto aos informantes escolarizados*: dos 133 informantes do sexo masculino, 67 eram escolarizados, sendo a maioria com o ensino fundamental incompleto; das 132 informantes do sexo feminino, 66 eram escolarizadas, sendo a maioria, também, com ensino fundamental incompleto e b) *quanto aos informantes analfabetos*: 66 entrevistados do sexo masculino; 66 do sexo feminino. Em substância, com relação aos participantes do inquérito dialetológico, os pesquisadores aplicaram o questionário a 265 falantes, sendo que 133 eram escolariza-

dos, o equivalente a 50,2% e a 132 analfabetos, o equivalente, em termos percentuais, a 49,8% dos informantes.

3. *Categorias linguísticas e onomasiológicas*

Para a definição das categorias e conceitos operatórios de nossa investigação de modo a favorecer uma melhor sistematização das unidades fraseológicas no ALECE (2010), tomamos como base os aportes dialetológicos de Maria do Socorro Silva de Aragão e Cleusa Palmeira Bezerra de Menezes (1984a) e os fraseológicos, de Glória Corpas Pastor (1997) e Mário García-Page Sánchez (2008), assim descritos:

Atlas: representação gráfica de cartas ou mapas de uma dada região que mostra a distribuição geográfica das variedades dialetais mais representativas dos falares ali existentes, e também das isoglossas que os separam. No âmbito da geografia linguística e dialetologia, o Atlas destina-se, primordialmente, a recensear e a localizar os falantes de uma região e a estudar suas variedades, variações e evoluções linguísticas, através de um questionário que trata da expressão dos falantes, de certos campos conceituais (natureza, partes do corpo etc.), e da interpretação de certos enunciados (perguntas). Como trabalho coletivo, o atlas destina-se, fundamentalmente, a um levantamento paradigmático-sintagmático do léxico de determinada língua, que compreende, ainda, a análise da língua falada por pessoas de nível de escolaridade (alfabetizado e analfabeto), de modo a se obter uma descrição, a mais completa possível, da língua falada de uma comunidade de falantes. (NEVEU, 2008; ARAGÃO e MENEZES, 1984a)

a) *Fraseologia*: no âmbito da lexicografia, trata-se de uma disciplina linguística que estuda unidades fraseológicas. Fraseologia, em sentido restrito, diz respeito ao estudo das combinações dos lexemas e a identificação de características específicas das construções próprias de determinado indivíduo, grupo ou língua. Como fazemos, aqui, um levantamento dessas unidades fraseológicas se constitui, também, fraseologia um conjunto de frases feitas, locuções figuradas, metáforas e comparações fixadas, modismos, existentes em uma língua, no ensino individual ou regional (CORPAS-PASTOR, 1997, p. 17)

b) *Colocações*: no âmbito da disciplina Fraseologia, as colocações são um dos tipos de unidades fraseológicas mais recorrentes no ALECE (2010). Referem-se a uma combinação pré-fabricada na norma,

isto é, no sentido coseriano de uso corrente numa língua relativamente estabilizada pelas instituições sociais, e não no sentido saussuriano de sistema, ou seja, de redes de relações pelas quais uma língua se organiza e se estrutura. As colocações apresentam restrições combinatórias no campo da linguagem. (CORPAS PASTOR, p. 1997, p. 53 e 87)

c) *Compostos*: assim como as locuções, os compostos, no ALECE (2010), apontam uma regularidade na estrutura sintagmática de sua formação lexical. Os compostos selecionados para nossa pesquisa apresentam características próprias de unidades fraseológicas: (1) *pluriverbalidade*, isto é, uma combinação de duas ou mais palavras; (2) *Fixação*, resultado de um processo histórico-diacrônico, evolutivo, da conversão de uma construção livre em uma construção fixa ou estável, sendo considerados, dentro de uma perspectiva psicolinguística, fórmulas memorizáveis, facilmente recuperáveis pelos falantes de uma língua. (3) *Idiomatidade*, caracterizada pela não composicionalidade, isto é, seu significado não resulta da soma do significado normal de seus componentes e (4) *institucionalização*, entendida como processo pela qual uma comunidade linguística adota uma expressão fixa, a sanciona como algo próprio e passa a ser um componente de seu acervo linguístico-cultural, de seu código idiomático. (SÁNCHEZ, 2008, p. 25-35).

Para o levantamento das unidades fraseológicas, recorreremos aos campos semânticos já estabelecidos pelo ALECE (2010), a saber: (a) *natureza* (ventania, neblina, temporal, redemoinho, orvalho, arco-íris); (b) *tempo* (anteontem, trasanteontem); c) *parentesco* (órfão, tio) e d) *partes do corpo* (tornozelo, estômago, nádegas). No âmbito dos campos conceituais, o ALECE recorreu, para a coleta de dados, a premissas onomasiológicas, as designações, as noções ou os conceitos para determinar as formas linguísticas a eles correspondentes. A seguir, vamos comentar alguns dos lexemas que nos ofereceram uma maior produtividade fraseológica no ALECE (2010).

Para o lexema *anteontem*, advérbio datado por Houaiss (2009) de 1575, traz acepção de “no dia que precedeu o de ontem” ou “antes de ontem”, tendo sua primeira forma histórica registrada como “ant’ontem”. Para capturar este lexema do léxico mental dos informantes, os pesquisadores fizeram a seguinte pergunta: “Se são dois dias antes de hoje, a gente chama”(=anteontem). Obtiveram os seguintes dados linguísticos, ortográfica e fraseologicamente assim expressos: a) *informantes escolarizados*: andeontes, ante de ontem, ante de ontem, anteontem, anteontem, antes de ontem, antes de ontem, antontem, ontem-ontem, ontem-ontem, ontem-ontem,

onteontem, ontonte e b) *informantes analfabetos*: ante de onte, anteonte, anteontem, antes de ontem, antonte, antonte de ontem, numero um, onte de onte, onte de ontem, ontem-onte, ontem-ontem, onte-onte, onteontem, ontontem, ontreonte, ternanteontem, tentanteontem, ternanteonte, ternanteonte e ternanteontem. (ALECE, 2010)

No campo conceitual relacionado ao *tempo*, registramos muitas unidades fraseológicas relacionadas ao substantivo *ventania*. Datado por Houaiss (2009) de 1553, uma das acepções para ventania é a de “vento forte, que sopra com ímpeto e continuamente”, também chamada “ventaneira” cuja etimologia pode ser assim segmentada: vento + -ana + -ia. Houaiss traz uma rica sinonímia para ventania a partir do lexema vento: guieira, harmatão, lés-nordeste, lés-sudeste, lés-sueste, lestada, leste. (HOUAISS, 2009). Os pesquisadores do Atlas ao fizeram pergunta “Vento muito forte”, obtiveram os seguintes dados linguísticos, especialmente unidades fraseológicas do tipo colocação, com nítida influência de conhecimentos prévios dos entrevistados: a) *informantes escolarizados*: geral forte, vento brabo, vento forte, vento geral e vento leste e b) *informantes analfabetos*: vento brabo, vento da serra grande, vento forte, vento grande e vento guaxo (ALECE, 2010).

Também, no campo conceitual relacionado ao *tempo*, *neblina* é um dos lexemas também com alto registro de unidades fraseológicas. Trata-se de uma palavra datada de 1660 com a acepção, segundo Houaiss (2009), de “névoa baixa e fechada”, “nevoeiro”, com registro regionalista no nordeste brasileiro e significado de “chuva miúda, chuvisco”. No Piauí, tem a acepção regional de “pancada de chuva forte e rápida” e “aguaceiro”. Sua etimologia indica que é uma palavra de origem espanhola “neblina” (1220-1250) decorrente da forma latina “nebula,ae”, que quer dizer, “névoa, nevoeiro”. Segundo Houaiss (2009), as formas leblina, lebrina, liblina, librina, lubrina, nebrina ocorrem por dissimilação consonântica e/ou assimilação vocálica. A respeito da dissimilação, vale recorrer, ainda Houaiss (2009), quando diz que este fenômeno fonético se refere ao “processo de mudança linguística em que um ou mais traços fonéticos de um fonema, que ele tinha em comum com outro fonema vizinho (contíguo ou não), são trocados, tornando os dois fonemas diferentes”. Essa tendência seria atribuída, segundo Houaiss(2009), à necessidade de mais contraste entre os elementos fônicos da língua, para maior clareza. A sinonímia de neblina é leblina, lebrina, liblina, librina, lubrina, nebrina, nebrinha. (HOUAISS, 2009)

Importante observar que há uma aproximação de neblina ao lexema *bruma* como podemos atestar a seguir: cacimba, cerração, corrubiana, corrupiana, neblina, nebrina, nébula, névoa, nevoeiro, ruço, senceno. Tal comportamento linguístico ainda é mais estreito quando nos deparamos com a sinonímia de *chuva*: aguaceiro, apaga-pó, carga-d'água, chuva-criadeira, chuvada, chuva-de-caju, chuva-de-carçoço, chuva-de-rama, chuva-de-santa-luzia, chuva-dos-cajueiros, chuva-dos-imbus, chuvão, chuvarada, chuveirão, chuveiro, chuvinha, chuvisco, chuvisqueiro, cinzeiro, corda-d'água, dilúvio, garoa, garua, lebreia, librina, manga-d'água, maria-das-pernas-compridas, molhe-molhe, molinha, molinheira, moinheiro, morrinha, neblina, neblinas, nimbo, ondeirada, orvalheira, orvalho, pancada, parajá, patameira, pé-d'água, peneira, troz-troz, xixixi, zerbada. Finalmente, neblina pode ser associada, semanticamente, a *chuvisco* a partir do seguinte registro de sinonímia deste: apaga-pó, arenga-de-mulher, borraceiro, borriço, bulceira, chivilho, chuvinha, chuvisqueiro, cruviana, curviana, garoa, garua, jereré, lebreia, librina, meruja, merujo, mijaceiro, molhe-molhe, molinha, molinheira, moinheiro, morrinha, neblina, orvalho, peneira, xereré, xererém, xixi, xixixi, zimbro (HOUAISS, 2009). Os pesquisadores do ALECE ao fazerem a pergunta “Chuva miúda e fina”, registram dados linguísticos indicadores da visão de mundo dos informantes, o que confirma a hipótese psicolinguística de influência da memória semântica na produção das unidades fraseológicas: a) *Informantes escolarizados*: chuva fina, chuvas finas, chuvas forte, chuvas pesada, chuvas torrenciais, chuva torrenciais, chuva torrencial, chuva velha, (es)tá só neblinando, tempo neblinado e b) *informantes analfabetos*: chuva finíssima, chuva manerinha, chuvinha fina, peneirazinha de chuva (ALECE, 2010).

O lexema *transanteontem*, segundo Houaiss (2009), é datada de 1789. Trata-se de um advérbio com a acepção de “o dia que precedeu ao de anteontem”. Sua etimologia indica a seguinte segmentação de seus componentes: tras- + anteontem (este, resultante de ante- + ontem). Sua forma histórica é *trasantehontem*. Uma variante para *transanteontem* é *trasantontem*, mais antiga, datada de 1721, com o registro da seguinte forma histórica: *trazantontem*. Para obter o lexema *transanteontem*, no ALECE, os pesquisadores fizeram a seguinte pergunta “Se são três dias antes de hoje, a gente chama”. Foram obtidos os seguintes dados fraseológicos: a) *informantes escolarizados*: anteontem, antes de anteontem, antes de antes de ontem, antes de ontem, antes de ontem-ontem antevéspera, fazem três dias, quarto dia, ternandeontem, ternanteonte, ternanteontem, ternantontem, ternontonte, ternontontem, tesnanteontem, tesnaonte-

ontem, transeontem, trasanteontem, transonteontem, treanteonte, tremnanteontem, trernantonte, tesanteonte, tresanteontem, tresantonte, tresenteontem, tresnanteontem, tresnantonte e b) *informantes analfabetos*: ante de onteontem, antes de anteonte, antes de anteontem, antes de ontonte, antonte, atrás de anteontem, onte de antonte, antes de antonte, ontes de ontem, ontonte, ternantes de ontem, ternantonte, ternantontem, ternonteonte, ternonteontem, ternontonte, tirontonte, tredantonte, tremnanteonte, tremnanteontem, tremnionte-onte, tresanteontem, tresantonte, três de antonte, três de ontonte, três dias atrás, tresnanteonte (ALECE, 2010).

O lexema *arco-íris* é dos mais frequentes lexemas nos atlas linguísticos brasileiros. Datado de 1712, o composto arco-íris é definido, segundo Houaiss (2009), como “ arco luminoso que se origina em fenômenos físicos e meteorológicos e é produzido quando a luz solar é refratada, dispersa e internamente refletida por gotículas de água provenientes da chuva e suspensas na atmosfera” Acrescenta o dicionário que arco-íris é “visível como um conjunto de bandas coloridas adjacentes na forma de arcos de circunferência (mais raramente como anéis) com as cores do espectro solar” Entre suas formas históricas, são registradas, em 1712, “iris”, e, em 1789, o composto, sem hífen, “arco Iris”. Os sinônimos de arco-íris são arco-celeste, arco-da-aliança, arco-da-chuva, arco-da-velha, arco-de-deus, olho-de-boi, sinônimos de arco-íris (HOUAISS, 2009) e sofrem variações léxicas e fonéticas no registro fraseológico do ALECE (2010). Para obter a variação diatópica de *arco-íris*, os pesquisadores fizeram, no Atlas, a seguinte pergunta: “Barras coloridas que aparecem no céu depois de uma chuva”. As respostas à pergunta dos pesquisadores se aproximam muito das restrições combinatórias do lexema arco-íris, assim distribuídas por informantes: a) *informantes escolarizados*: alço-celeste, alquiceleste, arca-da-aliança, arco-celeste, arco-da-velha, arca-da-aliança, arco-celeste, arco-da-velha, arco-ílis, arco-íris, arco-íris, arcoíro, arquiceleste, as barras, os torreame e b) *informantes analfabetos*: arco-celeste, acso-irpe, aico-celeste, alço-celeste, aliança-do-céu, arco-celeste, arco-celesto, arcoí, arcoil, arco-íli, arcoílo, arco-íri, arcoírio, arco-íris, arcoíro, arcori, arquiceleste, as nuve escura, asto-celeste, barra do vento, barra do dia, escama-de-peixe, (es) ta barreando o tempo, oio-de-boi, os aico, sinal da nuve, torre de capelo, torre de chuva, torre de chuva, uma torre, é. (ALECE, 2010).

Situado no campo conceitual da *natureza, redemoinho*, diz respeito a uma palavra, segundo Houaiss (2009), datada de 1105, com as seguintes acepções: a) “movimento de rotação em espiral”; b) “turbilhão de

água que se forma no mar ou no rio, devido a cruzamento de correntes contrárias de águas; voragem, sorvedouro” e c) “mudança brusca de direção do vento ao encontrar um obstáculo”, ainda “rajada, tufão, pé-de-vento”. Sua etimologia indica que se trata de palavra derivada regressivamente de remoinhar. Suas formas históricas são: em 1105, remoino; no século XIV, rremuno; sXIV rremuño e em 1632, remuinho. Seus sinônimos são: redomoinho, redemunho. A palavra redemoinho se relaciona com a palavra rebojo, que tem acepções regionais de: a) “movimento de rotação em espiral causado por queda de cachoeira”; b) “remoinho de água que se forma no mar ou no rio e leva coisas para o fundo”; c) “sorvedouro, turbilhão, voragem” e d) “remoinho de vento, quando muda repentinamente de direção ao encontrar obstáculo”. Os sinônimos de rebojo são, por sua vez: peganho, redemoinho, remoinho, rodamento, rodamoinho, rodomoinho, rolo, sorvedouro, torvelim, torvelinho, torvelino, turbilhão, voragem, vórtice; ver tb. sinonímia de vento. O lexema *redemoinho* foi onomasiologicamente alcançado através da seguinte pergunta: “Água que roda no rio, assim – gesto”. Os pesquisadores do Atlas, quanto a esta pergunta, obtiveram, além das lexias simples, as seguintes unidades fraseológicas: a) *informantes escolarizados*: onda da água, queda-d’água e redemunho d’água e b) *informantes analfabetos*: correnteza forte e galão-d’água.

O lexema *tornozelo* é, segundo Houaiss (2009), datado de 1543, relacionado com anatomia ou parte do corpo, definido como “segmento ou saliência óssea que, em cada um dos membros inferiores, se situa entre a perna e o pé”. Os pesquisadores, para evocação do lexema, utilizaram uma pista elicitativa representada por desenho ou indicação. Em termos de dados coletados, além das lexias simples, foram registradas as seguintes unidades fraseológicas: a) *informantes escolarizados*: junta-do-pé, maria-joaquina, ossim-da-risada, osso-da-maria-joaquina, osso-do-cacanã e osso-do-vintém e b) *informantes analfabetos*: junta-do-pé, maria-joaquia, maria-joaquina, ossim-do-vintém, osso da canela, osso-da-joaquina, osso-da-madrugada, osso-da-maria-joaquina, osso-da-risada, osso-do-pé e osso-de-vintém.

Estômago é um dos lexemas examinados no ALECE (2010). Datado do século XV, Houaiss (2009) informa que se trata de um “órgão oco do tubo digestivo, de estrutura musculomembranosa, situado abaixo do diafragma, entre o esôfago e o duodeno, onde os alimentos são depositados, pré-digeridos e esterilizados antes de serem enviados ao intestino, para ali serem absorvidos”. Sua etimologia grega “stómakhos” sugere

“orifício, abertura”, daí, 'boca, esôfago, o orifício do estômago, estômago'. Para recolher dados lexicais sobre estômago, os pesquisadores do ALECE não utilizaram pista elicitativa no Q-ALECE, obtendo, além das lexias simples, os seguintes resultados: a) *informante escolarizado*: boca do estômago e b) *informante analfabeto*: boca do estombo. Curioso comparar os dados dos informantes analfabetos com as seguintes formas históricas de estômago, repletas de metaplasmos de supressão: no século XIV, estamago; XV, estamaguo; XV stamago; XV stomago e XVI, estomago.

Os pesquisadores do ALECE (2010) fizeram a pergunta “Aquele que não tem pai nem mãe” para o conceito de *órfão*. Além da obtenção de lexias simples, os dados fraseológicos granjeados pelos falantes cearenses foram assim classificados: a) *Informantes escolarizados*: criança abandonada, desagarrado do mundo, filho adotivo, filho enjeitado, filho sem dono, filho sem pai sem mãe, filho único e num tem arrimo e b) *Informantes analfabetos*: fi bandonado, fi desvalido, filho adotivo, filho sem pai sem mãe, filho unco, filho único, fi sem pai, pessoa desvalida e sem parte.

Datado do século XIV, o lexema **orvalho**, em Houaiss (2009), tem as seguintes acepções aplicadas ao Questionário do ALECE (2010): a) “condensação do vapor da água da atmosfera que se deposita em gotículas sobre superfícies horizontais e resfriadas (terra, telhados, folhagens etc.), pela manhã e à noite” e b) “espécie de chuva fina, leve, miúda” e “chuvisco”. Entre seus sinônimos de maior interesse para uma comparação entre a forma lexicalizada do Houaiss (2009) e a forma registrada na variação diatópica do ALECE (2010) podemos citar: aljôfar, aljofre, brandura, cacimba, caruja, carujeira, carujeiro, irroração, lentura, mangra, molúria, orvalhada, orvalheira, relento, rociada, rocio, rol, rorejo. (HOUAISS, 2009).

O lexema *orvalho* foi obtido pelos pesquisadores do ALECE (2010) por meio da seguinte pergunta: “O que faz com que as plantas amanheçam molhadas sem ter chovido à noite”. Foram registradas, do ponto de vista fraseológico, as seguintes respostas dos falantes cearenses: a) *informantes escolarizados*: (es)tá orvalhado, gelo da noite, orvalho da noite, orvalho da noite fria e tempo frio e b) *informantes analfabetos*: frieza da noite, librina da noite, orvalho da noite, penera de chuva, sereno da noite, serenezim da noite e vaim da noite.

Datado do XIII, o lexema *temporal* traz, entre suas acepções do Houaiss (2009), o de “chuva forte com vento; tempestade, estado de violenta agitação atmosférica que, segundo sua intensidade, se caracteriza por ventos cuja velocidade pode atingir de 62 km a 102 km por hora, na escala de Beaufort”. O conceito de *temporal* aproxima-se ao de chuva e este avizinha-se à sinonímia de *quantidade* como acúmulo e enxurrada. Na pesquisa, o conceito *temporal* foi obtido pelos pesquisadores do ALECE (2010) através da seguinte pergunta: “Chuva muito grossa e demorada”. Além das *lexias* simples, as respostas dos falantes cearenses foram assim, do ponto de vista de dados fraseológicos, classificadas: a) *Informantes escolarizados*: caiu um toró de chuva, chuva agitada, chuva boa, chuva de peda, chuva de pedra, chuva extênsia, chuva forte, chuva grande, chuva grossa, chuva mais forte, chuva pesada, chuvas de muita água, enchente grande, (es)tá chovendo forte, né, sistema de diluve, *temporal* forte, *temporal* forte demais, tromba-d’água, trombas-d’água e b) *informantes analfabetos*: a chuva (es) tá pesada, chovendo grosso, chuva aturada, chuva demorada, chuva de peda, chuva de tempestade, chuva forte, chuva grande, chuva grossa, chuva rendosa, chuvas pesada, chuvas pesadas, chuvas torrenciais, chuva-tempestade, chuvona grande, (es)tá chovendo muito grosso, passou o dia chovendo, pau-d’água, rolo-d’água e tempestade de chuva e tromba d’água.

Datado do século XIV, o lexema *nádega* é definido por Houaiss (2010) como “cada uma das duas partes musculo adiposas da região traseira da pelve humana, formada por cada lado dos glúteos “Sua etimologia indica que a palavra vem do latim vulgar “natica” e do latim clássico “nates,ium” (singular natis, is, mais raro) com significado de “nádegas; rabadilha, uropígio”. Suas formas históricas são: no século XIV, *nadegas*; XIV *nadigas*, XV, *madeguas*, tendo ainda o sinônimo *nalga*.

Nádegas é, no âmbito do tabuísmo, uma das interdições linguísticas para ânus. Se tomarmos como referência Houaiss (2009), podemos constituir um corpus de tabus linguísticos (*lexia* simples) relacionados com o ânus: *alvado*, *anilha*, *apito*, *berba*, *boga*, *bogueiro*, *caqueiro*, *cesta*, *cu*, *diferencial*, *feofó*, *finfa*, *fiofó*, *fiota*, *fite*, *fito*, *foba*, *frasco*, *fueiro*, *furico*, *loto*, *oritimbó*, *panela*, *pêssego*, *pevide*, *quiosque*, *rosa*, *rosca*, *roscofe*, *sim-senhor*, *tutu*, *viegas*.

Os pesquisadores do ALECE (2010) ao fazerem pergunta “Quando não se toma injeção no músculo ou na veia, a parte do corpo mais indicada são as...”, cremos que induziu o falante a evocar, de forma *geltática*, a resposta *nádegas*, resposta-alvo dos pesquisadores, o que veio a

revelar, no nosso entendimento, um grave equívoco ou erro de procedimento dialetológico.

As respostas dadas pelos falantes cearenses registraram as seguintes unidades fraseológicas: a) *Informantes escolarizados*: popa-da-bunda e b) *Informantes analfabetos*: bochecha-da-bunda: popa-da-bunda, popados-quartos e poupa-da-bunda. Observamos que nas unidades fraseológicas, as palavras mais recorrentes, para a definição de nádegas, são *bochecha*, *bunda*, *quartos* e *popa*, lexemas que fazem parte da constelação de tabuísmos de ânus. Segundo Houaiss (2009), *bunda*, datada de 1836, é uma palavra de origem angolana com etimologia “*mbunda*”, que quer dizer, “*quadris, nádegas*”. Também em Houaiss (2009), *bunda* tem acepção de “*nádegas de gente alcatreira*”, ou seja, vale dizer, “*nadeguda*”, palavra reconhecida, em Portugal, entre os usuários atuais da língua, mas não é empregada, ao certo, por razões de tabuísmo. Já a associação do lexema nádegas ao lexema *bochechas*, como no caso de informantes analfabetos, é uma das marcas do regionalismo e jocosidade dos cearenses.

4. Considerações finais

Os resultados preliminares da análise do ALECE apontam para um significativo número de compostos e locuções nominais (adjetivas, adverbiais, substantivas e verbais) evocados pelos informantes escolarizados e analfabetos. Os informantes analfabetos, com 142 dados linguísticos, produziram maior número de unidades fraseológicas contra 96 ocorrências dos escolarizados. Os lexemas que mais apresentaram produtividade fraseológica foram os relacionados com a *natureza*, como *arco-íris*, com 29 ocorrências; *temporal*, com 21 ocorrências e relacionados com o *tempo* como *transanteontem*, com 26 ocorrências e *anteontem*, com 21 ocorrências. De modo geral, os lexemas selecionados na pesquisa as confirmaram características mais marcantes da fraseologia: (a) *pluriverbalidade*; (b) *fixação*; (c) *idiomaticidade* e (d) *institucionalização*, o que revelam a regularidade destas unidades na estrutura sintagmática e suas restrições combinatórias, sejam oriundas do léxico mental de informantes escolarizados ou analfabetos.

REFERÊNCIAS BIBLIOGRÁFICAS

ARAGÃO, Maria do Socorro Silva de; MENEZES, Cleusa Palmeira Bezerra de. *Atlas linguístico da Paraíba: cartas léxicas e fonéticas*. Brasília: UFPB/CNPq, 1984a.

_____; _____. *Atlas linguístico da Paraíba: análise das formas e estruturas linguísticas encontradas*. Brasília: UFPB/CNPq, 1984b.

BENVENISTE, Émile. *Problemas de linguística geral II*. 2. ed. Campinas: Pontes, 2006.

BESSA, José Rogério Fontenele. (Coord.). *Atlas linguístico do Estado do Ceará*. Vol.1. Fortaleza, UFC, 2010a.

_____. (Coord.). *Atlas linguístico do Estado do Ceará*. Vol. 2. Fortaleza, UFC, 2010b.

_____ et ali. (Dir.). *Questionário do atlas linguístico do Estado do Ceará*. Fortaleza: UFC, 1982.

CAMARA JUNIOR, J. Matoso. *Dicionário de linguística e gramática*. 21. ed. Petrópolis: Vozes, 2000.

DELBECQUE, Nicole. *Linguística cognitiva: compreender como funciona a linguagem*. Lisboa: Instituto Piaget, 2008.

DUBOIS, Jean et ali. *Dicionário de linguística*. São Paulo: Cultrix, 2004.

FERREIRA, Aurélio Buarque de Holanda. *Novo dicionário da língua*. 2. ed. Rio de Janeiro: Nova Fronteira, 1986.

HOUAISS, Antônio; VILLAR, Mauro de Salles. *Dicionário Houaiss da língua portuguesa*. Elaboração do Instituto Antônio Houaiss de Lexicografia e Banco de Dados da Língua Portuguesa. Rio de Janeiro: Objetiva, 2001.

_____; _____. *Dicionário Houaiss da língua portuguesa*. Elaboração do Instituto Antônio Houaiss de Lexicografia e Banco de Dados da Língua Portuguesa. Rio de Janeiro: Objetiva, 2009. [com a nova ortografia da língua portuguesa]

MARTINS, Vicente. *Tratamento lexicográfico dos compostos nominais antes e depois do acordo ortográfico (ao): O caso de unidades léxicas relacionadas com o lexema pé nas edições do dicionário Houaiss (2001–2009)*. Disponível em:

http://www.filologia.org.br/xiv_cnlf/tomo_2/1670-1695.pdf Acesso em: 23/09/2010

NEVEU, Franck. *Dicionário de ciências da linguagem*. Petrópolis: Vozes, 2008.

SÁNCHEZ, Mario García-Page. *Introducción a la fraseología española: estudo de las locuciones*. Barcelona: Anthropos, 2008.