

# UMA DESCRIÇÃO DAS EXPRESSÕES CRISTALIZADAS E O PROCESSAMENTO AUTOMÁTICO

Alzira da Penha Costa Davel<sup>1</sup>

**RESUMO:** O presente estudo analisa o processo de construção de expressões cristalizadas, com a estrutura Verbo + Nome, observando a extensão do sentido metafórico que contribui para a compreensão do significado de textos, tendo como principal aporte teórico os pressupostos da teoria do Léxico-Gramática (GROSS, 1975, 1984). Para a identificação das estruturas, são aplicados alguns critérios formais de ordem morfossintático-semântica e de transformação, cujo propósito é o de manter o sentido inicial da sentença. Esses critérios facilitam a averiguação do linguista em relação à delimitação de determinados itens quando se depara com aspectos relacionados à composicionalidade e à ambiguidade, além da intuição linguística que pode interferir no julgamento da aceitabilidade. Tais procedimentos de averiguação oferecem maior segurança para formalização e processamento automático das expressões na elaboração de um dicionário eletrônico.

**Palavras-chave:** expressões cristalizadas; morfossintaxe; semântica; não-composicionalidade; ambiguidade.

**ABSTRACT:** The present study analyses the construction of crystallized expressions with Verb + Noun structure observing the extension of the metaphoric sense contributing to the understanding of the meaning of texts, based mainly on the Lexic-Grammar Theory (GROSS, 1975, 1984). For the identification of structures, some formal, morphosyntatic-semantic and transformation criteria are applied for the main purpose of maintaining the initial sense of the sentence. Those criteria make the linguist's inquiry about the delimitation of certain items easier, whenever facing aspects of compositionality and ambiguity, besides the linguistic intuition that can interfere on the acceptability scrutiny. Such investigation proceedings provide higher security for automatic formalization and processing of expressions in the elaboration of an electronic dictionary.

**Keywords:** crystallized expressions; morphosyntax; semantics; non-compositionality; ambiguity.

## INTRODUÇÃO

Este estudo tem como objetivo apresentar uma descrição de expressões cristalizadas do português do Brasil, para o processamento automático da linguagem natural. Deu-se ênfase ao estudo de expressões, tendo em vista que, de acordo com as divergências acerca das

---

<sup>1</sup> Mestre em Estudos Linguísticos na área de estudos analítico-descritivos da linguagem, pela Universidade Federal do Espírito Santo-UFES.

terminologias e a falta de critérios de análise adequados, essas expressões são consideradas pelos estudos tradicionais como objetos linguísticos excepcionais, não sendo integradas na gramática da língua, por não serem inseridas nas “regras gerais”.

Nas últimas décadas, mudanças significativas aconteceram no campo computacional, em especial, o estudo das línguas naturais, possibilitando a obtenção de informações a respeito do funcionamento/regularidade da língua em suas diferentes construções. Sendo assim, o objetivo da linguística computacional é construir programas, com a capacidade de interpretar informações fornecidas pela linguagem natural. Para isso, são necessários os conhecimentos referentes à língua para construção de sistemas e técnicas que possam realizar o processamento dos níveis linguísticos de uma língua, como a morfologia, a sintaxe, a semântica e também os discursos (pragmática) que podem trazer significado no uso.

Sabe-se que é comum o falante seguir determinadas convenções da língua que lhe permitem reconhecer se determinadas palavras ou expressões são válidas. É a intuição linguística relacionada à competência comunicativa dos nativos que governam essas regras. No Processamento das Línguas Naturais, essa função compete aos analisadores sintáticos, e, para isso, é preciso que a língua seja especificada por um léxico e por uma gramática.

Para o processamento da linguagem natural, além da língua escrita, a linguística computacional também tem como objeto de estudo a língua falada. Por isso, precisa de uma tecnologia especial para interpretar a fala por meio da manipulação da representação do conhecimento fonético-fonológico.

Com referência às questões do significado, expressas por meio dos sistemas do falante em relação ao ouvinte, é necessário que se recorra a mecanismos que representem o conhecimento de mundo. Além disso, esses mecanismos proporcionam relações entre os vários componentes e segmentos de um texto ou discurso que, do ponto de vista semântico, podem estar envolvidos tanto na construção de um modelo de interpretação de um texto quanto no reconhecimento de um sentido específico, dentro de um contexto, em casos de palavras ambíguas, como, por exemplo: **manga** de camisa ou **manga** como fruta, ou ainda **manga** como coletivo de porcos.

A língua natural tem sempre seu significado relacionado à situação de uso, e a semântica caracterizou-se como uma área de estudo que considera o significado das expressões linguísticas de modo dependente de quem as usa e de como são usadas. No entanto, as questões ligadas ao uso da linguagem estão associadas ao domínio da pragmática, cuja consideração para as análises centra-se no contexto linguístico, na interpretação das expressões linguísticas. O contexto linguístico é mais facilmente tratado pela linguística computacional, uma vez que se refere ao que está explicitado no texto.

### **A mente humana e a inteligência artificial**

O objetivo maior da linguística computacional é promover, cada vez mais, a interação entre o homem e o computador. A Inteligência Artificial tem como finalidade imitar, por intermédio de máquinas eletrônicas, o máximo possível, a atividade mental e, conseqüentemente, talvez, trazer alguma melhoria para a capacidade humana. Segundo Winston (1987), a “Inteligência Artificial é o estudo de conceitos que permitem aos computadores serem inteligentes”. O estudo da robótica, por exemplo, procura atender às exigências da indústria, relacionadas a aparelhos para realizar tarefas “inteligentes”, que até então exigiram a intervenção humana. Há também interesse no desenvolvimento dos sistemas especializados, nas profissões da medicina, do direito, entre outras.

Nesse sentido, surgem, inclusive, questionamentos se o conhecimento e a experiência dos seres humanos seriam substituídos pela “máquina”. É preciso considerar que esse aspecto traria implicações sociais mais abrangentes. A psicologia, e talvez a filosofia, nesse caso, trariam relevantes contribuições, pois, na tentativa de imitar o comportamento humano por meio de um recurso eletrônico, é possível que se aprenda algo sobre o funcionamento do cérebro humano. No entanto, embora a Inteligência Artificial já tenha obtido inúmeros avanços, ainda há muitas lacunas a serem preenchidas, no que diz respeito à capacidade de o computador interpretar certos comportamentos do cérebro humano. Em outros termos, pensa-se a possibilidade de o computador trabalhar à semelhança do cérebro humano e, com isso, ocorra um processo reverso de entendimento, já que o cérebro é uma “engrenagem” complexa.

## O Léxico-Gramática

Para a descrição das expressões cristalizadas, toma-se como base a teoria do Léxico-gramática (GROSS, 1975), que é também um método para descrição lexical de uma língua, cuja orientação é a aceitabilidade de frases como fonte básica para o conhecimento linguístico. A análise é feita sob o ponto de vista sintático-semântico, observando as restrições paradigmáticas, por meio das inserções e/ou substituições de elementos, que indicam maior ou menor nível de fixidez.

As abordagens do Léxico-gramática, acima mencionado, são ancoradas na teoria transformacional de Harris (1970) que, muitas vezes pode relacionar-se com a Gramática Gerativa; porém, caracteriza-se de modo diferente, pois os postulados do Léxico-Gramática baseiam-se nos procedimentos empíricos. Por esta razão, Gross (1975) tem pontos de vista bem diversos dos estudos gerativistas, uma vez que seus princípios fundamentais são os de construção de ‘modelos’, que não consideram a real produtividade na língua. A teoria do Léxico-Gramática assume uma postura taxonômica e a Gramática Gerativa caracteriza-se pelo método hipotético-dedutivo.

Os estudos sobre a fixidez de estruturas<sup>2</sup> têm sido desenvolvidos nos últimos 30 anos por esse autor, juntamente com sua equipe de linguística, tendo como base a unidade de significado das frases simples. Para tanto, o método adotado consiste em estabelecer classes, cujos elementos possuam características sintáticas semelhantes.

Gross (1986) estabelece também uma classificação que separa as expressões verbais de acordo com a estrutura interna: os continentes fixos e os constituintes livres. Pode-se, então, observar que as expressões são em número bem maior do que se costuma supor nos estudos linguísticos e que, em geral, são tratadas como ‘exceções’. De acordo com essas investigações, o autor já elaborou vários trabalhos sobre as ECs verbais em diferentes línguas.

---

<sup>2</sup> No decorrer do estudo são adotados diferentes termos: construção, estrutura, sequência, mas todas eles referem-se à “expressão”.

Nesse aspecto, embora as expressões fixas tenham sido relegadas pelos estudos gramaticais e sintáticos a uma zona de indefinição, tem-se assistido, nos últimos tempos, a um crescente interesse por esses objetos linguísticos ‘anômalos’, sobretudo na área do Processamento das Línguas Naturais. As formas fixas, ou seja, as sequências de elementos lexicais que não possuem as propriedades combinatórias tal como era suposto, são tão numerosas em qualquer tipo de texto que não podem ser ignoradas.

Bechara (2001, p. 603), por sua vez, menciona essas expressões na subsecção “Anomalia da Linguagem”, definindo-as como: “Idiotismo ou expressão idiomática é toda a maneira de dizer que, não podendo ser analisada ou estando em choque com os princípios gerais da Gramática, é aceita no falar culto”. Em geral, isso ocorre também com muitos outros itens e/ou expressões que frequentemente ocorrem no uso da língua, mas que não receberam a devida atenção dos estudiosos e que não foram sistematizados pelos estudos gramaticais.

Segundo Ranchhod (1999), baseando-se em Gross (1996), embora já existam muitos estudos relacionados às expressões fixas, na noção de fixidez ainda há lacunas para um consenso, tanto na perspectiva analítica e conceptual como do ponto de vista da terminologia. A esse respeito, a definição das ECs é, às vezes, polêmica por existirem, na literatura, vários termos para essa designação. Assim, acabam por ocupar uma posição indesejada, uma vez que não podem ser definidas por aplicação de um modelo criterial único de análise e, conseqüentemente, são subestimadas pelo interesse linguístico.

Essa última citada dota a designação de ‘Frases Fixas’ para estruturas em que existem fortes restrições lexicais e sintáticas entre um verbo e, pelo menos, um dos seus argumentos. São registros que possuem características comuns de combinações de **verbo + nome**, que não têm possibilidades produtivas, distribucionalmente, e nem podem ser interpretadas do ponto de vista composicional. Esse tipo de estrutura tem recebido diferentes denominações, tais como Fraseologias, Idiomatismos, Lexias Complexas, Expressões Fixas, Expressões Cristalizadas, entre outras.

Nesse sentido, Vale (2001) faz uma observação importante quando diz que, no aspecto sintático, as expressões cristalizadas<sup>3</sup> são frases aparentemente semelhantes às comuns, mas no aspecto semântico são, em geral, interpretadas pelos falantes com seu significado idiomático, e não pela soma dos sentidos literais das palavras que as compõem. Assim como esse autor, adota-se, neste estudo, a última denominação (ECs), conforme vem sendo mencionada.

Considerando o posicionamento de Smarsaro (2002, p. 84), do ponto de vista da reprodutibilidade, há vulnerabilidade e divergência no conceito de estruturas composicionais, porque podem ser avaliadas de acordo com nossa intuição. Ao observar essas flutuações de avaliações, pode-se dizer que existem diferentes graus de composição.

Assim, os significados, tanto do ponto de vista composicional como do uso idiomático das expressões, dependem do conhecimento de mundo dos falantes, que podem produzir sentidos múltiplos, diferenciados e até mesmo antagônicos.

### **A composicionalidade e a não-composicionalidade**

A concepção de composicionalidade relaciona-se à possibilidade de poder extrair/deduzir o significado de uma sequência a partir dos significados dos componentes. Em outros termos, significa que é possível deduzir o significado de cada elemento que compõe a expressão, o que possibilita o cálculo de um processo a ser formalizado. Para isso, é necessário que haja certa transparência semântica, bem como produtividade. Produtividade essa que pode ser aplicada a inúmeras construções sintáticas com o mesmo padrão, como no caso de sequências (palavras compostas) como **guarda-chuva, guarda-roupa**, entre outras. De acordo com Vale (2001, p. 72), “o entendimento da transparência é observado de acordo com a maior proximidade do cálculo do significado total da expressão por seus componentes, enquanto que a opacidade seria a total impossibilidade desse cálculo”.

Nesse sentido, a não-composicionalidade relaciona-se à falta de transparência semântica, ou seja, não é possível a apreensão do significado dos elementos que compõem a expressão,

---

<sup>3</sup> O autor usa a caracterização Expressões Cristalizadas para as expressões fixas.

mas da expressão como um todo, como, por exemplo, as sequências, mencionadas mais adiante, **Dar zebra, Engolir sapo, Ser galinha** etc, que, a partir de uma construção (**V+Nome**), não podem ser entendidas como construções transparentes e sim totalmente opacas, porque o significado não pode ser obtido por meio do significado de cada constituinte das expressões. Os critérios para classificação dessas expressões se assemelham aos de algumas palavras compostas porque possuem certa transparência semântica, mas não apresentam produtividade distribucional.

Dependendo do contexto, embora essas ECs possam ser intercaladas por um advérbio, um artigo, flexão de número etc (dar muita zebra), (ser um galinha), (ser uma anta), por analogia, não há possibilidade de substituição, por exemplo, do Nome (Dar macaco), (Engolir mosquito), (Ser frango). Sobretudo para os falantes da língua, elas ganhariam um novo sentido e perderiam o significado metafórico que caracterizam as Expressões Cristalizadas. Mais adiante, poderá ser percebido que algumas dessas substituições e/ou inserções são possíveis, sem que haja perda do caráter metafórico, indicando menor grau de fixidez, ou seja, menor grau de cristalização, portanto.

### **As expressões idiomáticas e suas extensões metafóricas**

Desde a antiguidade, o homem, conforme sua evolução, é definido como um ser racional; no entanto, mais recentemente, percebeu-se, com maior clareza, que ele é também emocional. Pesquisas demonstram que o ser humano se emociona por meio de ideias, quando razões e emoções se misturam no cérebro, construindo imagens que dialogam entre si em forma de metáforas.

Ao longo do tempo, várias foram as concepções discutidas e revistas sobre o entendimento da metáfora. Segundo a visão da retórica clássica, a metáfora é considerada como figura de linguagem e definida como substituição de uma palavra por outra, quando existe uma relação de similaridade. Devido à insuficiência dessa caracterização, ela passa a ser considerada como constituição de sentido, ou seja, outra possibilidade criada pelo contexto de leitura (FIORIN, 2002, p. 86).

Estudos mais recentes mostram que a metáfora trata-se de um processo cognitivo, que estabelece semelhanças entre o novo e as experiências; e ainda, que exerce a função de propiciar a extensão da capacidade de conceituar e, conseqüentemente, de tornar mais fácil a comunicação, pois institui o jogo entre o concreto e o abstrato que perpassa o processo mental na comunicação linguística. É um processo que perpassa nosso sistema conceptual, de modo que “tentamos, por meio de exemplos, dar algumas indicações do papel considerável da metáfora na maneira como agimos, falamos, vivemos” (LAKOFF; JOHNSON, 2002, p. 205).

Na verdade, percebe-se que são vários os fatores que contribuem para as construções de sentido da linguagem: as diversas relações que se estabelecem de ordem histórico-social e cultural, o lugar da interlocução, o conhecimento compartilhado dos falantes de determinada língua.

São fatores importantes a serem levados em consideração para o tratamento das expressões idiomáticas que são também estruturas compostas de várias palavras, adicionando, porém, o fato de que uma análise lexical é exigida para a representação do significado dos itens complexos. Esse entendimento aplica-se também às expressões constituídas de **V + SN** (verbo+nome de animal, inseto) que não apresentam alta produtividade do ponto de vista morfossintático, mas sim, sob o aspecto da fala cotidiana da língua portuguesa, considerando que os termos adquirem valores semânticos opostos entre si, conforme seus ambientes linguísticos, havendo alternância entre o denotativo e o metafórico nos processos comunicacionais.

Para o caso do processamento automático desse tipo de expressões para elaboração de um dicionário eletrônico, é preciso seguir alguns caminhos para a descrição:

- a) apresentar critérios linguísticos que propiciem a identificação das expressões;
- b) relacionar as expressões;
- c) descrever as relações morfossintáticas, bem como o processo de flexão;
- d) formalizar o conhecimento linguístico sobre as expressões de maneira a serem utilizadas por sistemas de processamento automático.



Seguindo esses critérios, esse estudo auxiliará, de algum modo, a posterior formalização e o processo de automatização das expressões a seguir relacionadas:

### 1) Soltar os cachorros

Flexão verbal

1.a) Mário **soltou os cachorros** na secretária.

Nessa sequência, há possibilidade de flexão do verbo **soltar**, com a manutenção do sentido metafórico da expressão, porém, dependendo do contexto de uso, pode haver ambiguidade, pois é possível imaginar que Mário soltou os ‘animais’ para atacar a secretária.

Ambiguidade

1. b) \*? Mario **soltou os cachorros** no quintal.

“**Soltar os cachorros**” sem o complemento-adjunto “no quintal” torna a informação ambígua, tendo em vista que são possíveis outras leituras. Dependendo da ambiência linguística em que ocorre, o verbo **soltar** perde o significado de libertar, e **cachorro** deixa de significar ‘um animal mamífero’. A perda do sentido literal de cada termo é que evidencia a existência de uma expressão que constrói o seu valor a partir do sentido global dos elementos **soltar os cachorros**, portanto, significa ter um comportamento explícito. Nos dois casos, a natureza dos nomes que compõe o adjunto (secretária e quintal), é relevante para a alteração do sentido.

Substituição lexical do nome

1. c) \* Mário **soltou os gatos** na secretária.

Quando uma sequência, como **soltar os cachorros**, é fixa, não se pode fazer substituição lexical do nome, no caso, **cachorros**. A substituição de **cachorros** por **gatos** não é possível, considerando que altera o sentido da expressão.

Variação de número

1. d) \* Mário **soltou o cachorro** na secretária.

A variação de número da expressão também revela fixidez, por exemplo, **soltar o cachorro no vizinho** muda o sentido metafórico da expressão.

Inserção de advérbio

1. e) \* Mário **soltou pouco/muito os cachorros** na secretária.

A inserção os advérbios de intensidade **pouco/ muito** não é permitida, pois levaria a expressão a ter um sentido literal.

Variação de gênero do nome

1. f) \* Mário **soltou as cachorras** na secretária.

Não há possibilidade de variação de gênero **as cachorras**. O índice de fixidez não permite, pois a expressão perderia o sentido metafórico.

Apassivação

1. g) \*Os cachorros **foram soltados** por Mário.

Não admite apassivação, pois o sentido passaria a ser literal, como se os animais tivessem sido ‘libertados’ por Mário. Há, nesse caso, um bloqueio distribucional.

Redução do determinante e do nome

1. h) \*Mário **soltou na secretária**.

Não há possibilidade de omitir/reduzir termos da sequência, uma vez que se tornaria sem sentido nas formas usadas por falantes da língua portuguesa.

## 2) Lavar a égua

Acréscimo do adjunto adverbial

2. a) Mário **lavou a égua** com a venda de dois cavalos.

Essa expressão permite o acréscimo de outros elementos, pois o sentido metafórico é mantido, denotando que a pessoa tirou proveito (teve ganhos) ao fazer uma negociação. Contudo, esse tipo de sequência não permite certas modificações (substituições e/ou inserções) de itens, podendo causar o comprometimento do sentido metafórico, como:

Substituição verbal

2. b) \*Mário **esfregou a égua no quintal**.

Observa-se um bloqueio distribucional, porque não admite substituição verbal **esfregar**, pois a expressão ganharia um sentido literal. Além disso, a expressão passaria a ter um caráter ambíguo devido ao emprego do adjunto adverbial de lugar **no quintal**. Poderia se imaginar o fato de **Mário** estar no quintal e lá, ter lavado o animal.

#### Apassivação

2. c) \***A égua foi lavada** por Mário.

Não admite a forma passiva porque a expressão perderia o caráter metafórico, uma vez que o nome **égua** estaria sendo empregado no sentido denotativo (um animal).

#### Substituição lexical

2. d) \***Mário lavou a burra** no jantar.

Não admite a substituição do nome por outro como **burra/cavalo/cachorra**, mesmo sendo do mesmo campo lexical.

#### Redução do artigo

2. e) \***Mário lavou égua** no jantar.

Não aceita a perda do determinante, pois, além de ficar mal estruturada, essa forma não é comumente usada.

#### Inserção de complemento

2. f) \***Mário lavou a égua** com sabonete.

Não admite o acréscimo do adjunto, caracterizado por material, que poderia também ser **com detergente, com sabão** etc., permitindo várias possibilidades de substituições, o que caracteriza uma sequência livre. Percebe-se, então, um bloqueio distribucional, cuja explicação é feita de modo inverso de uma expressão cristalizada.

#### Acréscimo/ inserção de adjunto adverbial

2. g) \***Mário lavou a égua** no lava-jato, com shampoo.

Novamente, a sequência não admite o acréscimo do adjunto adverbial de lugar devido a sua natureza, além do complemento **com shampoo** ser caracterizado por material. Isso influencia o

sentido da expressão para literalidade, em que o nome **égua** assume um sentido propriamente de animal.

2. h) \*Mário **lavou muito/ pouco a égua** no jantar.

Da mesma forma, não é admissível a intercalação dos adjuntos **muito/ pouco**, na expressão, assim como não é possível a substituição do determinante **a** por outros elementos, como:

2. i) Mário lavou (\***umas** + \***várias** + \***algumas** + \***todas** + \***estas**) éguas. Além disso, a flexão do nome **égua(s)** comprometeria o sentido.

Em todas as construções que apresentam asteriscos há um bloqueio distribucional, pois, ao se introduzir determinados elementos, perde-se o sentido metafórico e as sequências passam a ser livres.

### 3) Dar zebra

Inserção de determinantes

3. a) **Deu uma zebra** lá em casa hoje!

3. b) **Deu a zebra** no concurso.

3. c) **Deu umas zebras** no concurso de hoje.

Nessa sequência é permitida a inserção dos determinantes **uma, a, umas**. Embora possam adquirir nuances diferentes, pelo acréscimo contextual, o sentido metafórico é preservado. Em (3. a) pode-se entender que um fato desagradável aconteceu; em (3.b) remete à ideia de que já se esperava algum problema no concurso, referindo-se a um conhecimento partilhado, e em (3.c) remete à ideia de que alguns problemas aconteceram no concurso.

Inversão dos termos

3. d) O concurso **deu zebra**.

Percebe-se, nesse exemplo, a possibilidade de inversão dos termos (inserção do determinante + nome), que funciona como sujeito, mantendo o mesmo sentido da expressão cristalizada.

Inserção de adjuntos

3. e) **Deu muita/pouca zebra** no concurso.

Admite a inserção dos adjuntos **muita/pouca** porque não acontece a perda do caráter metafórico da expressão. Contudo, o sentido metafórico se desfaz se forem realizadas as alterações seguintes:

Substituição de item lexical

3. f) \***Deu girafa** no concurso.

Não há possibilidade de substituição do elemento lexical **zebra** por **girafa** (mesmo sendo do mesmo campo semântico), devido ao bloqueio distribucional.

Apassivação

3. g) \***Zebra foi dada** no concurso.

Da mesma forma não admite a formação da passiva, pois há um índice de fixidez elevado. Trata-se de uma expressão mais flexível que admite algumas inserções/ substituições e outras, não, como em (f-g).

#### 4) Engolir sapo

4. a) Às vezes, é preciso **engolir simplesmente** sapo.

Observa-se que é possível inserir o adjunto adverbial de modo **simplesmente**, sem que haja perda da extensão metafórica.

Substituição de item lexical

4. b) \***Engoli um mosquito** no trabalho hoje.

Não há aceitação da substituição lexical do mesmo campo semântico, pois a estrutura ficaria no sentido literal de realmente engolir um inseto.

Flexão verbal

4. c) **Engolimos sapo** no trabalho sempre.

Nesse exemplo, é possível a flexão verbal em número, sem a perda do sentido metafórico, que remete à ideia de se ter, normalmente, problemas que causam aborrecimento no trabalho.

Inserção de advérbio

4. d) **Engoli muito** sapo no meu trabalho.

Admite a inserção do advérbio de intensidade, sem que haja perda do sentido metafórico. Remete à ideia de que o indivíduo, no ambiente de trabalho, teve que ouvir coisas desagradáveis e manter-se calado.

Inserção de determinante

4. e) Sempre **engolimos uns sapos** no trabalho.

Esse item admite a inserção do determinante **uns** mantendo, ainda, a extensão do sentido metafórico, pois leva à ideia de que sempre passamos por aborrecimentos no trabalho, sem poder reagir.

Inserção de pronome

4. f) Mário **engoliu aquele sapo**.

Há possibilidade da inserção do pronome demonstrativo **aquela**, uma vez que a expressão mantém o seu sentido figurado. Pode-se imaginar que **Mário** já esperava pelo constrangimento pelo qual iria passar, ou ainda, para dar realce à determinada situação discursiva, de conhecimento partilhado.

Apassivação

4. g) **\*O sapo foi engolido** por Mário.

Não admite apassivação, uma vez que a expressão poderia adquirir um sentido literal. A característica do verbo remete à ideia de que realmente **Mário** ingeriu o animal.

Substituição lexical

4. h) **\*Mário engoliu jacaré**.

Não aceita a substituição do nome por outro do mesmo campo semântico, como **engolir peixe**, **engolir porco** etc., pois haveria a perda do sentido metafórico e passaria a ser literal,

constituindo-se um bloqueio distribucional da expressão, devido à baixa produtividade e ao alto índice de fixidez.

Variação de grau

4. i) \* **Engoli um sapinho** no trabalho essa semana.

Não aceita a variação de grau no diminutivo, pois perderia o sentido metafórico. Além disso, não é usada pelos falantes.

5) Ser galinha

Inserção de determinante

5. a) Mário é **um galinha**.

A sequência admite a inserção do determinante **um**, sem alteração do sentido, denotando que Mário é um promíscuo, vulgar. Nota-se que aqui não é possível eliminar o determinante **um**, como **Mário é galinha**.

Flexão do determinante

5. b) **Maria é uma galinha**.

Nesse caso, é admissível a flexão do determinante **uma** antes do nome **galinha**, mantendo o sentido metafórico, o quer dizer que **Maria** é uma pessoa vulgar.

Flexão verbal

5. c) **Somos galinha**.

Esse item admite a flexão do verbo **ser**, podendo continuar com o mesmo sentido da expressão original, **ser galinha**. Nesse caso, o nome **galinha** é, normalmente, usado no singular, pois, se for usado no plural, perde o sentido metafórico da expressão.

Substituição verbal

5. d) **Mário ficou um galinha**.

Admite a substituição verbal de **ser** por **ficar** sem modificar o sentido, significando que **Mário** tornou-se uma pessoa vulgar.

Substituição lexical

5. e) \*Mário é **um frango**.

Nesse caso, a expressão não admite a substituição do nome por outro do mesmo campo semântico. Existe aqui um índice de fixidez e um bloqueio distribucional, pois perderia o sentido figurado e passaria para o sentido literal.

Flexão de grau

5. f) \*Mário é **um galinhão**.

Esse item não admite a flexão do grau aumentativo, uma vez que perderia o sentido figurado e passaria a ter um sentido literal.

## 6) Ser anta

Inserção de determinante

6. a) Mário é **uma anta**.

Admite a inserção do determinante **uma**, sem a perda do sentido metafórico, que remete à ideia de alguém que é bobo, tolo. Da mesma forma poderíamos dizer que **Maria é uma anta**, o que possibilita a substituição do nome (sujeito) **Mário** por **Maria**.

Flexão verbal

6. b) **Nós somos umas antas**.

Permite a flexão do verbo **ser** (somos), e do determinante **umas**, mantendo o caráter metafórico da expressão, cujo significado é: **nós somos uns bobos**.

Inserção de advérbio

6. c) Mário é **muito anta**.

Essa sequência permite a inserção do advérbio de intensidade **muito**, em que o sentido metafórico se mantém, indicando a ideia de que Mário é muito bobo.

Inserção de adjetivos

6. d) \*Mário é uma **anta grande/inteligente**.



Há um bloqueio distribucional porque após o nome **anta** não é possível o acréscimo de adjetivos **grande, inteligente** etc., mostrando que a expressão **ser anta** possui maior grau de fixidez.

### 7) Acertar na mosca

Flexão verbal

7. a) Mário **acertou na mosca** com aquela compra.

Admite a flexão verbal, uma vez que se mantém o sentido metafórico, denotando que Mário agiu acertadamente. Em contrapartida, não admite as seguintes modificações: Flexão nominal

7. b) \***Mário acertou nas moscas.**

A flexão da locução **nas moscas** não é permitida, pois o sentido passaria a ser literal, como se **Mário** tivesse matado os insetos.

Substituição lexical

7. c) \***Mário acertou no mosquito.**

A substituição do nome **mosca** por **mosquito** deslocaria o sentido para a literalidade, além do bloqueio distribucional existente.

Apassivação

7. d) \***A mosca foi acertada** por Mário.

A apassivação também alteraria o sentido metafórico, como se “Mário tivesse realmente matado a mosca”.

### 8) Cantar de galo

Flexão verbal

8. a) Mário **cantou de galo** hoje.

Essa sequência admite a flexão do verbo **cantar**, mantendo o sentido metafórico da expressão, denotando que Mário se portou como ‘um valentão’. No entanto, não aceita as modificações seguintes:

Inserção de advérbios

8. b) Mário **canta** (\***bem** + \***mal** + \***muito** + \***bastante**) **de galo**.

A expressão, com a inserção desses advérbios, além de comprometer o sentido, não é usada pelos falantes.

Substituição lexical

8. c) \*Mário **canta de pavão**.

Não admite substituição do nome por outro do mesmo campo semântico, **pavão**, causando uma ruptura paradigmática. Esse item apresenta bloqueio distribucional e um alto grau de fixidez. Aliada a isso, também, a falta de transparência faz com que a expressão fique no sentido literal.

Variação de grau

8. d) \***Mário** canta de galinho.

Também não admite variação de grau no diminutivo, pois não é normalmente usada, e, além disso, não permitiria manter o sentido original da expressão cristalizada.

## 9) Cair do cavalo

Inserção do determinante

9. a) \*Mário **caiu de um cavalo**.

Não admite inserção do determinante **um**, pois a sequência se tornaria ambígua, logo, passaria a ter um sentido literal de que Mário realmente caiu de um cavalo, isto é, sofreu um acidente.

Inserção de adjunto adverbial

9. b) Mário **caiu** (\***sempre** + \***muito**) **do cavalo** com essas coisas.

A inserção dos adjuntos adverbiais **sempre**, **muito** possibilitaria também a ambiguidade. Dependendo do ambiente sintático em que se encontra, pode ser considerada sequência livre ou composta, podendo, ainda, assumir um sentido conotativo/ literal. Essa expressão apresenta, portanto, um bloqueio, não sendo possível aplicar um paradigma distribucional com o nome **cavalo**, ou seja, não poderia haver substituição lexical por outro nome, como **caiu** (\***do burro** + \***do boi** + \***do bode**) etc.

Observa-se, então, que nas sequências (7), (8) e (9) há um bloqueio distribucional, devido à baixa produtividade e maior índice de fixidez entre os elementos das expressões, entre a preposição e o nome: **na mosca**, **de galo** e **do cavalo**, denotam sentidos metafóricos usados pelos falantes de um idioma e que representam determinados comportamentos. Se em tais expressões fossem substituídos os elementos (nomes), por outros do mesmo campo lexical, como: acertar **no mosquito**, cantar **de pavão** e cair do **burro**, elas passariam a ter outro sentido.

### **Conclusão**

No decorrer deste estudo foi possível observar o comportamento de algumas expressões, sobretudo quando modificadas pela substituição e/ou pela inserção de certos itens lexicais. Esse procedimento evidencia as restrições distribucionais que ocorrem entre os verbos e os nomes e entre as preposições e os nomes, mostrando a fixidez interna entre esses elementos.

Observou-se também que algumas delas apresentam elevada fixidez, o que implica a não-composicionalidade dos seus significados, apontando para o fato de que sejam reconhecidas como um ‘bloco’, funcionando como um só item lexical. No caso daquelas que permitem a inserção e/ou a substituição de itens (sem bloqueio distribucional), sem alteração de sentido, são mais flexíveis, podendo ser consideradas menos cristalizadas.

Importante se faz ressaltar que este estudo centrou-se na análise descritiva, tendo sido apresentados somente alguns exemplos das expressões, como forma de demonstrar que elas têm alto teor de produtividade na vida cotidiana dos falantes brasileiros e é necessário que recebam um tratamento especializado. Dentre os critérios utilizados, destacam-se a verificação do caráter de fixidez e da desambiguação dos enunciados, aspectos fundamentais que devem ser observados na definição de entradas lexicais. Porém, para que constem de um dicionário eletrônico, é preciso ainda que essas expressões sejam representadas e formalizadas de modo a melhorar a qualidade dos programas computacionais do processamento da linguagem natural.

Conforme o posicionamento de alguns linguistas, as expressões cristalizadas não devem ser tratadas como anomalias linguísticas, mas sim inseridas no léxico computacional de um programa que trabalhe com o processamento da linguagem natural e implementadas em um dicionário eletrônico. Isso não só possibilita maior abrangência do léxico como também beneficia a questão da tradução de itens lexicais complexos como um todo, em vez de serem resolvidos por fragmentos sintáticos.

## Referências

BAPTISTA, J.; CORREIA, A.; FERNANDES, G. *Léxico-gramática das frases fixas do português europeu*. Universidade de Algarve. Faculdade de ciências humanas e sociais: cadernos de fraseologia galega, n. 7, 2005, p. 41-53.

BECHARA, E. *Moderna gramática portuguesa*. Rio de Janeiro: Lucerna, 2001.

FIORIN, J. L. (Org.). *Introdução à Linguística I - Objetos teóricos*. São Paulo: Contexto, 2002. p. 226.

GROSS, G. *Les expressions figées du français*. Paris: Ophrys, 1996.

GROSS, M. *Méthodes empiriques em syntaxe*. Paris: Hermann, 1975.

LAKOFF, G.; JOHNSON, M. *Metáforas da vida cotidiana*. Coordenação de tradução: Mara Sophia Zanotto. São Paulo: Mercado das Letras, 2002.

RABUSKE, R. A. *Inteligência artificial*. Florianópolis: Ed. da UFSC, 1995.

RANCHHOD, E. M. *Tratamento das línguas por computador: uma introdução à linguística computacional e suas aplicações*. Lisboa, dez. 1999.

SMARSARO, A. *Um estudo sobre composição*. Revista Fala Palavra, n. 2, Outubro, 2002.

VALE, O. *Expressões Cristalizadas do português do Brasil: uma Proposta de Tipologia*, Tese de doutoramento, Araraquara: UNESP. 2001.

VIEIRA, R.; LIMA, V. L. S. *Linguística computacional: princípios e aplicações*. In: IX Escola de informática da SBC-Sul. Passo Fundo, Maringá-São José, SBC-Sul, 2001.