

CONCEITUANDO FRASEOLOGIA: CONCEPÇÕES E EQUÍVOCOS SOBRE EXPRESSÕES FIXAS

DEFINING PHRASEOLOGY: CONCEPTIONS AND MISCONCEPTIONS ON FIXED EXPRESSIONS

Lúcia Fulgêncio¹

luciafulgencio@hotmail.com

RESUMO: Quantas expressões fixas existem em uma língua? Qual é a sua frequência nos textos orais e escritos? Essas perguntas têm importância não apenas para a descrição da língua em questão, mas também para a avaliação do peso do componente memorizado no conhecimento de uma língua. Com efeito, a maioria dos modelos de análise correntes se concentra em aspectos regulares da estrutura da língua: regras e princípios de validade geral. E ainda que muitos linguistas tenham chamado a atenção para a primazia do léxico na estruturação do enunciado, as expressões idiomáticas ficam geralmente na sombra – às vezes tachadas de elementos marginais, excepcionais, de ocorrência quase exclusiva na língua oral coloquial, sendo tratadas como epifenômenos sem maior relevância. No entanto, um levantamento exaustivo das expressões fixas em corpus real do português brasileiro revelou alguns fatos interessantes: (a) as expressões fixas correntes no português brasileiro – aquelas que todos os falantes podem reconhecer e interpretar facilmente – são em grande número, mais de 8.000; (b) as expressões fixas são de frequência muito alta em todos os tipos de textos, não apenas orais, mas também escritos (como livros, jornais e revistas, e em especial na linguagem publicitária); (c) as expressões apresentam individualidades não somente do ponto de vista semântico, mas também sintático, apesar de tradicionalmente ser enfatizada prioritariamente a idiosincrasia semântica. Vê-se portanto que, no estudo dos fraseologismos, encontramos vários conceitos e mal-entendidos arraigados pela tradição à caracterização das expressões fixas, que merecem reexame e reavaliação. Este trabalho tem por objetivo apresentar algumas dessas concepções que precisam ser reanalisadas, a fim de chegar a uma concepção mais adequada do conjunto das expressões fixas do português brasileiro contemporâneo. Tais questões não só são relevantes para o estudo das expressões fixas em si, mas implicam repercussões para a própria teoria da linguagem.

PALAVRAS-CHAVE: expressão idiomática; expressão fixa; expressão convencional; fraseologia.

ABSTRACT: How many fixed expressions are there in a language? What is their frequency in oral and written texts? These questions are important not only for the description of a particular language, but also for the evaluation of the weight of the memorized component of language knowledge. In fact, most current analysis models focus on regular aspects of language structure: rules and general principles. And, while many linguists have called for attention to the primacy of lexicon in utterance structure, idioms and fixed expressions are often left in the shadow – sometimes called marginal, exceptional elements, occurring almost exclusively in informal speech, and treated as epiphenomenal

¹ Doutora em Linguística e Língua Portuguesa pela Pontifícia Universidade Católica de Minas Gerais – PUC-MG. Professora da Faculdade de Letras da Universidade Federal de Minas Gerais – UFMG.

features without much relevance. However, an exhaustive survey of fixed expressions in a corpus of Brazilian Portuguese has shown some interesting facts: (a) common fixed expressions – those all speakers can easily recognize and interpret – are present in great number, more than 8,000 in all; (b) fixed expressions are highly frequent in all kinds of texts, not only oral, but also written ones (books, newspapers, magazines, with specially high presence in the language of advertising); (c) fixed expressions are not only semantically idiosyncratic (as traditionally emphasized), but also show syntactic deviations. The study of phraseologisms involves several traditionally accepted concepts and misconceptions which deserve to be reexamined and reevaluated. The present work aims at presenting some of these misconceptions, in order to arrive at a better understanding of fixed expressions in current Brazilian Portuguese. These questions are relevant, not only for the study of fixed expressions per se, but also for the study of the theory of language.

KEYWORDS: idiom; fixed expression; conventional expression; phraseology.

O RENOVADO INTERESSE PELA LEXICOLOGIA

Tem-se observado nos últimos anos um intenso incremento de estudos lexicológicos. Até há pouco tempo essa era uma área pouco explorada, uma vez que era considerada como não integrante do “core” linguístico – ou seja, do foco estrutural formal no qual se concentrava o interesse dos linguistas e a maioria dos estudos sobre o funcionamento das línguas. Mais recentemente essa situação se modificou, em grande parte devido à constatação de que léxico e gramática não são áreas isoladas e sistemas independentes, e sim faces interligadas da organização da língua (GROSS, 1979, 1994; LANGACKER, 1987, 1991; GOLDBERG, 1995).

Culicover e Jackendoff (2005) afirmam que a porção idiossincrática e o léxico “constituem a maior parte da língua” e negam a possibilidade de se fazer a distinção tradicional entre léxico e gramática (p. 26). Culicover e Jackendoff são explícitos ao incluir os fraseologismos e as construções idiossincráticas de mais de uma palavra no léxico, visto por sua vez como um sistema sem limites precisos:

[...] há um contínuo uniforme do material linguístico presente léxico, que se estende desde as palavras, expressões idiomáticas, construções idiomáticas, construções mais gerais mas ainda especializadas, até os princípios mais gerais, mais nucleares (*core-like principles*). Em princípio não existe distinção entre núcleo (*core*) e periferia, mas somente uma gradação de generalidade.² (Culicover; Jackendoff, 2005, p. 39).

Observamos que o cômputo linguístico é feito concomitantemente em dois níveis: o nível estrutural e também o nível lexical, que inclui estruturas memorizadas.

² Texto original: “[...] there is a smooth continuum of linguistic material in the lexicon, ranging from words through idioms through truly idiosyncratic constructions through more general but still specialized constructions to the most general core-like principles. There is no principled distinction between core and periphery, only a gradation of generality.”

Atualmente é ponto pacífico que o conhecimento das construções de uma língua não pode ser descrito somente do ponto de vista de sua forma abstrata, mas depende essencialmente do comportamento de cada item lexical. Isso quer dizer que o conhecimento do léxico e das idiossincrasias de cada vocábulo é determinante para a boa formação dos sintagmas e sentenças. Essa nova perspectiva redirecionou os estudos linguísticos, conferindo maior relevância à pesquisa lexicológica.

Uma parte importante da investigação sobre o léxico refere-se ao estudo dos fraseologismos. No entanto, ainda se observa nessa área a permanência de alguns conceitos tradicionais que vêm sendo mantidos e repetidos, e nem sempre revelam uma análise adequada de todo o conjunto das expressões fixas. No estudo das expressões do português brasileiro (PB), algumas concepções comuns e bastante difundidas precisam ser revistas e redimensionadas. Há alguns equívocos, alguns mal-entendidos e outras falsas concepções. Algumas afirmações seguem a tradição e outras são pouco precisas, feitas sem base empírica, sem levar em conta um corpús real, sem examinar a língua em uso. Portanto, os conceitos normalmente associados à análise das expressões fixas merecem exame e reflexão. Algumas observações nesse sentido serão examinadas a seguir. É esse o objetivo deste artigo, que visa a apresentar e reavaliar algumas características das expressões fixas que nem sempre têm sido adequadamente dimensionadas.

1. O PAPEL DOS FRASEOLOGISMOS

É evidente que uma língua inclui necessariamente um extenso componente computacional e de geração criativa de sentenças, tendo em vista a grande liberdade de produção disponível aos usuários. Mas também é fato que unicamente um sistema algorítmico, associado a um léxico constituído por palavras isoladas, não consegue explicar o grande número de fraseologismos utilizados comumente pelos falantes. Fato é que uma grande porção da língua é composta de construções memorizadas.

As línguas em geral incorporam um grande número de **convencionalidades**. O falante, condicionado e moldado por essa cultura linguística, repete as fórmulas institucionalizadas, que se cristalizam. Como resultado, o léxico não é formado somente por uma lista de palavras isoladas, mas também por vários agrupamentos de mais de uma palavra, por sequências que se fixam e se consolidam com determinado formato e significado específico. Os falantes memorizam não só palavras e morfemas,

mas também sintagmas e sequências complexas de palavras que são repetidas integralmente na construção de novos enunciados, como um bloco solidário tomado da memória. Portanto, as unidades que compõem o inventário lexical são compostas não somente por palavras, mas também por sequências de palavras que vão juntas, por agrupamentos com mais de um vocábulo que ocorrem sistematicamente no mesmo formato e que aparecem reiteradamente com a mesma forma e significado fixos. Modelos de base sintática dão conta de descrever a organização de uma sentença produzida composicionalmente, mas mostram-se incapazes de explicar a preferência do falante por sintagmas lexicalmente preenchidos que se repetem consistentemente, da mesma forma.

Vejamos alguns exemplos: pode-se dizer que um ato inconveniente *pega mal*, mas não se diz que **segura mal*, apesar da semelhança semântica entre *pegar* e *segurar* e da gramaticalidade estrutural nos dois casos; e quando alguém arruma uma confusão pode-se dizer que *comprou uma briga*, mas não que **adquiriu uma briga*, mesmo sendo essa estrutura bem formada do ponto de vista sintático. O que se conclui daí é que as expressões *pegar mal* e *comprar briga* foram convencionalizadas com esse molde, e não admitem substituição lexical, possível em casos de sintagmas montados pelo falante. Os itens léxicos que compõem esses sintagmas não podem ser alterados porque as expressões são fixas, cristalizadas, não passíveis de montagem ou de modificação pelo falante com base no sistema computacional da língua. Quando o falante usa tais construções, demonstra que utilizou um conjunto lexical presente na sua memória, não construído no momento do enunciado.

O reconhecimento de estruturas “prontas”, memorizadas em grupo, é a única saída capaz de explicar a repetição *ipsis litteris* de alguns grupos lexicais e a maior frequência de certos conjuntos de palavras frente a outras sequências criadas livremente pelos falantes.

Quando pensamos em construções decoradas e expressões idiomáticas, imediatamente nos remetemos a casos como *pagar o pato*, *chutar o pau da barraca*, *matar dois coelhos com uma cajadada*, *misturar alhos com bugalhos*, *comprar gato por lebre* ou *ser cheio de nove horas*. É verdade que esses são, sim, exemplos de fraseologismos; mas, no entanto, convém salientar que não são só essas as expressões fixas da língua: esse grupo de expressões idiomáticas representa somente um dos tipos de expressões fixas. Há muitos outros casos diferentes desses, com uma estruturação interna não necessariamente do tipo SV, que possuem uma relevância

também especial e diferenciada na composição do discurso. Outro tipo de expressão fixa pode ser exemplificado por casos como os seguintes: *no entanto, de vez em quando, ou seja, à toa, quem dera, por enquanto, já que, à primeira vista, à solta, prisão perpétua, lua de mel, até mais, que tal, diga-se de passagem, e assim por diante* e tantos outros. Todos esses são exemplos de expressões fixas, mas estranhamente muitas vezes esquecidos quando se pensa em fraseologismos.

Esse último grupo de expressões cristalizadas inclui sequências imprescindíveis na construção do discurso, que usamos a todo momento, sem nem mesmo nos darmos conta de que grande parte do que falamos inclui porções significativas de palavras recuperadas da memória em bloco, como um todo único. Sem as expressões fixas seria bastante difícil (senão impossível) elaborar qualquer enunciado com naturalidade e fluência.

Uma das características importantes desse grupo é o fato de que se trata de expressões corriqueiras, que usamos com frequência, em qualquer nível diastrático, em qualquer tipo de discurso oral ou escrito, tanto quanto qualquer outra palavra da língua, compondo as sentenças e entremeando qualquer enunciado. Não se trata de elementos eventuais ou substituíveis, mas de grupos de palavras indispensáveis, radicalmente inerentes à própria construção das sentenças, às vezes necessários para promover a coesão textual, e que formam parte integrante e essencial da própria tessitura do discurso. Por exemplo, são expressões fixas os casos abaixo salientados em negrito (todos colhidos em *cópus real*). Ao final do exemplo é apresentada a indicação do tipo de fonte de onde foi extraído o dado, com o objetivo de constatar a diversidade dos meios linguísticos em que são empregadas as expressões fixas:

- (1) Argumentamos tudo isso e, como boas brasileiras, preferimos **deixar pra lá, até porque, por mais que** expliquemos, ele nunca iria entender mesmo. [jornal]

- (2) **Por incrível que pareça**, os hipocondríacos costumam **gozar de** boa saúde. **No entanto, por conta dessas** ironias da vida, os hipocondríacos mais afoitos acabam **correndo risco** ao se submeter a diversos exames invasivos. [livro]

- (3) Os chineses têm conseguido crescer com fartura. **No entanto, é líquido e certo** que não terão condições de manter esse ritmo acelerado indefinidamente. [tv]
- (4) O pai assumiu a filha, mas nunca foi meu marido. **Quer dizer, assumiu em termos**, porque nome não mata a fome, e ele só deu o nome. [revista]
- (5) Acredita-se que a serotonina seja, **de fato, um elemento-chave** nesse processo, **uma vez que** os medicamentos antidepressivos agem preferencialmente sobre o metabolismo desse neurotransmissor. [livro]

Expressões como essas são grupos cristalizados e convencionais que se mostram indispensáveis na elaboração das sentenças e na construção e encadeamento do texto. Tal como os fraseologismos do tipo *ser unha e carne* ou *botar lenha na fogueira*, expressões fixas como *até porque, antes de mais nada, em linhas gerais, de fato, o quanto antes, quer dizer, isto é, às vezes, por outro lado, tal qual* ou *que nem* (que é a fórmula normalmente usada como marca de comparativo no português brasileiro) são igualmente sequências convencionais, memorizadas em bloco, e nesse sentido são igualmente expressões fixas que constituem unidades integrantes do léxico brasileiro.

O grupo dos fraseologismos é portanto bastante grande, formado não somente pelas expressões idiomáticas reconhecidas tradicionalmente – como *bater as botas* ou *fazer das tripas coração* – mas também por outras sequências convencionais memorizadas, cujo conjunto é de alta frequência na língua e inclui sequências que se mostram indispensáveis na construção do texto.

Trata-se de um grupo bastante amplo e de tipologia variada, que inclui SNs (*ponta de estoque, ar-condicionado* ou *paraíso fiscal*), sintagmas verbais (*ganhar tempo, dar-se bem, não ter nada a ver*), preposições (*em vez de, em cima de, de acordo com*), SPreps (*a certa altura, na hora H, sem mais nem menos*), conjunções (*por conseguinte, no entanto, se bem que*), interjeições (*ora essa!, puxa vida!*), advérbios oracionais (*conforme for*), orações (*sendo assim*) e até fórmulas em aberto (como *em questão de* [+ período de tempo]: *em questão de horas / em questão de minutos* etc.) – além de vários outros casos.

Do ponto de vista semântico há todo tipo de conteúdo, como qualificativos (*espírito de porco, peso morto, azul marinho*), expressões referenciais (*ponte aérea, praça de alimentação*), expressões temporais (*hoje em dia, em cima da hora, de repente*), expressões de modo (*sem querer, à toa*), quantificadores (*por volta de, mais ou menos*), intensificadores (*sem tamanho, um monte de*) e negativas (*sei lá, não [V] patavina, qual nada*); exprimem eventos (*perder a hora, dar de cara*) ou estados (*ser a gota d'água, ficar elas por elas*); apresentam fórmulas sociais (*até logo mais, já vai!*), moduladores discursivos (*digamos assim, veja bem, tendo em vista, se não me engano, sabe como?*) e por aí vai.

Todos esses elementos fazem parte do grande conjunto de expressões fixas de que lançamos mão ao compor o discurso. Trata-se de estruturas “institucionalizadas” ou “lexicalizadas”, que Tagnin (1989) descreve como “convenções que se aplicam à língua”. As expressões fixas são então caracterizadas por se constituírem de sequências de mais de uma palavra que são memorizadas em bloco (de qualquer tamanho – desde locuções a orações inteiras), conhecidas pelos falantes como construções já prontas e acessíveis na memória como blocos lexicais coesos. São estruturas convencionais e idiossincráticas que não admitem a mesma liberdade de montagem que se verifica em sintagmas composicionais, e caracterizam assim a forma peculiar de expressão de uma língua. Possuem duas características essenciais: a convencionalidade linguística e a armazenagem em bloco na memória de longo termo dos falantes, constituindo assim uma unidade do inventário lexical. Tagnin (2013) apresenta uma longa taxonomia de classes diversificadas dessas expressões convencionais, incluindo expressões idiomáticas, coligações, colocações, binômios, marcadores conversacionais e fórmulas situacionais, às quais chamamos todas de “expressões fixas”³.

Praticamente **todo** e qualquer texto é permeado de sequências memorizadas. Até mesmo a escolha da preposição em casos de diátese verbal (como em *pensar em* e *preocupar-se com*) é um fato idiossincrático, sendo que as preposições que acompanham verbos (e nomes) por força de sua valência, chamadas “preposições funcionais”, precisam ser decoradas caso a caso pelos falantes da língua, o que comprova por si só a grande incidência de elementos memorizados na composição do enunciado.

³ A nomenclatura “expressões fixas” não é novidade: é largamente utilizada em inglês (*fixed expressions*) e em francês (*expressions figées*), sendo inclusive o nome de uma das principais obras sobre o assunto escrita por Gross (1996), intitulada *Les expressions figées en Français*.

2. O PESO DO MATERIAL MEMORIZADO: A LÍNGUA INCLUI UMA VASTA PORÇÃO MEMORIZADA

Um aspecto que nunca foi quantificado é o seguinte: qual é o peso da porção idiossincrática da língua? Quantas seriam as expressões fixas do português brasileiro? Qual é a sua frequência nos textos orais e escritos? Indagações como essas têm a sua relevância não no sentido do estabelecimento de um número, qualquer que seja ele, mas no sentido de permitir uma avaliação da amplitude da parte memorizada da língua comparativamente à porção estrutural em aberto. Uma tentativa de responder a essa questão foi feita no levantamento realizado por Fulgêncio (a sair), tendo sido encontradas mais de 8.000 expressões fixas no PB contemporâneo. Culicover & Jackendoff (2005) também dão uma estimativa a esse respeito, hipotetizando que o inglês teria possivelmente 8.000 expressões idiomáticas⁴. Quando perguntado sobre a origem desse cômputo, Jackendoff respondeu o seguinte: “[...] Não me lembro como chegamos a nosso palpite de 8.000 expressões idiomáticas, mas é realmente bem espantoso que você tenha chegado a esse número empiricamente”. (Jackendoff, mensagem de e-mail, 16.01.2014)⁵.

O cômputo das expressões fixas tem seu interesse por permitir avaliar o peso do componente memorizado frente aos aspectos de composição criativa da sentença, e também frente ao próprio componente lexical, ou seja, em comparação com a quantidade de lemas individuais que compõem o inventário lexical dos falantes. Portanto, o que se procura estabelecer não é um número em si; o que se deseja indagar ao fazer o levantamento das expressões fixas é qual a proporção das estruturas aprendidas de cor frente à atuação das regras, e qual a porcentagem das sequências de palavras memorizadas dentro do próprio inventário lexical dos falantes.

Com efeito, a maioria dos modelos de análise correntes se concentra em aspectos regulares da estrutura da língua: regras e princípios de validade geral. A amplitude da memorização de fatos individuais e a enorme quantidade de dados que decoramos e usamos normalmente na construção do texto ainda não foi adequadamente considerada na descrição das línguas. E ainda que muitos linguistas tenham chamado a atenção para a primazia do léxico na estruturação do enunciado

⁴ Texto original: “if English has, say, 8,000 idioms [...]” (Culicover & Jackendoff, 2005, p. 43).

⁵ Texto original: “I don’t remember how we arrived at our guess of 8,000 idioms, but it is indeed pretty amazing that you have come up with that number empirically”. (comunicação pessoal)

(como Levelt, 1989), os fraseologismos nem sempre são entendidos como uma porção significativa da linguagem. Muitas vezes são considerados como excepcionais e como epifenômenos de ocorrência quase exclusiva na língua oral coloquial – o que definitivamente não corresponde aos fatos, e traduz uma ideia ingênua do que vem a ser o conjunto das expressões fixas.

Ou seja, não se procurou ainda verificar empiricamente a dimensão da parte memorizada individualmente, comparativamente à parte composicional. Os dados sugerem que talvez não esteja bem dimensionado o grau de atuação das regras em relação às estruturas aprendidas de cor, uma a uma. Em outras palavras, parece não estar adequadamente avaliada a dosagem entre o componente gerativo (de regras) e o idiossincrático (de elementos aprendidos de cor caso a caso).

Falta enfatizar devidamente um aspecto importante da linguagem, que é nossa enorme capacidade de memorização de unidades – isto é, palavras, locuções, sintagmas e frases prontas. Parece haver uma concepção subjacente de que a parte idiossincrática da língua seria pouco expressiva. Fillmore (2000, p. 50) tem uma posição diferente: “[...] devo explicar que acredito que uma porção muito grande da habilidade de uma pessoa de usar uma língua consiste no domínio de expressões fixas”.⁶

Ao que tudo indica, a memorização de estruturas e a idiossincrasia de comportamentos têm um peso e uma relevância na língua bem maior do que se costuma admitir. O papel crucial da memorização individual não se concentra exclusivamente na aprendizagem de expressões fixas, mas se expande por todos os componentes da língua, seja na fonologia, na semântica ou na sintaxe. O conhecimento linguístico do falante não é composto exclusivamente por um sistema computacional, mas também por construções decoradas, por hábitos que se instalam e se repetem automaticamente. A maior evidência disso encontra-se no léxico, constituído não somente por palavras isoladas, mas também por sequências multilexicais cristalizadas.

A observação de que a língua é parcialmente sistemática e parcialmente idiossincrática aponta para uma forte participação da memória no conhecimento linguístico, e subverte em parte os modelos teóricos com que lidamos até então, que concebiam a língua essencialmente como um sistema de regras em aberto.

⁶ Texto original: “[...] I should explain that I believe a very large portion of a person’s ability to get along in a language consists in the mastery of formulaic utterances.”

3. UMA METODOLOGIA PARA O LEVANTAMENTO DAS EXPRESSÕES FIXAS E CONVENCIONAIS

Para efetuar o levantamento sincrônico das expressões fixas contemporâneas, o campo de busca tem de fazer uso indispensável de corpus real de língua viva, examinando a língua em uso. Em primeiro lugar e principalmente, o uso do corpus ajuda a nos proteger da nossa própria fantasia. Além disso, reprime a tendência comum de inflar o número de expressões, como se uma quantidade numérica grande fosse mais representativa ou tivesse maior valor. Na verdade, o importante não é um número em si, mas a coleta de casos realmente empregados, que possam demonstrar o conhecimento dos falantes da língua contemporânea. Em segundo lugar, o uso de outras fontes, como os dicionários tradicionais, aumenta o inventário com expressões desusadas, de estágios arcaicos da língua, e eventualmente inclui fraseologismos correntes em outros países onde se fala português. Por exemplo, os dicionários de Pugliesi, Cascudo, Nascentes e Aurélio registram expressões desconhecidas como *ter lobo próprio aprisco* (que teria o significado de “contar com um traidor”); *ser uma carta viva* (“falar demais”), *não chegar à craveira*, *passar ginja*, *levar buçal de couro fresco*, *às rápidas* e muitas outras expressões que, hoje e aqui, nos soam “estranhas”. Por outro lado, deixam de registrar expressões contemporâneas de uso frequente como *rápido e rasteiro*, *recarregar as baterias*, *recolha-se à sua insignificância*, *rolar a dívida* e muitas mais.

Obviamente, se o objetivo for o de apresentar uma dimensão diacrônica, ou apresentar as expressões que existem ou já existiram em outros países de língua portuguesa, a perspectiva e o material de trabalho poderão ser diferentes. Mas nesse caso convém ter ciência de que as expressões coletadas não necessariamente seriam aquelas presentes no português brasileiro contemporâneo.

Outra questão a que se deve estar atento refere-se à possibilidade de se confundir uma construção metafórica, construída no discurso, com uma expressão idiomática da língua. Para a caracterização de uma expressão fixa, de qualquer tipo, é condição *sine qua non* que a sequência seja **convencional**, conhecida pelos falantes da língua em geral, e que represente um conjunto lexical consagrado pelo uso. Não basta, certamente, que uma sequência seja semanticamente opaca para que seja considerada idiomática – a condição para a determinação de uma expressão fixa

centra-se na convencionalidade do grupo, e portanto na sua disseminação entre os falantes da língua.

Como metodologia de pesquisa de fraseologismos da língua brasileira atual foi necessário, portanto, o exame e a coleta em *córpus real*. Isso não excluiu, certamente, certa dose de introspecção. A introspecção é essencial em dois aspectos: (a) na seleção de dados, expurgando ocorrências que podem ser atribuídas a incoerências, mudanças de plano discursivo e outros “desvios” de performance; e (b) no acréscimo de material que, por questões fortuitas, pode não ter ocorrido no *córpus* levantado, mas que sabidamente está presente no português atual. Além, naturalmente, da determinação do significado de cada expressão, o que só é acessível por introspecção. A introspecção se justifica no sentido de que, em qualquer estudo linguístico, o objetivo do pesquisador não pode ser a descrição exclusiva de um *córpus* como fim em si mesmo, mas sim a descrição da língua e do conhecimento dos falantes, para o qual o *córpus* constitui somente um meio para se chegar lá.⁷

O levantamento em *córpus real* (incluindo todos os tipos de sequências lexicais convencionais memorizadas) permitiu evidenciar o que é realmente empregado na linguagem atual do dia a dia. Com isso, foram descartadas as expressões arcaicas, desusadas, regionais, empregadas fora do Brasil ou simplesmente fictícias – talvez usadas ocasionalmente por um único falante ou um pequeno grupo, não apresentando os traços necessários para a caracterização de uma sequência como convencional. A pergunta que se fez é: qual é realmente o conjunto das expressões fixas do português brasileiro contemporâneo? E quais são os significados atribuídos a essas expressões pelos falantes do século XXI?

Os dados foram colhidos em *córpus misto* (escrito e oral) de ocorrência livre: no discurso oral de conversação espontânea ou em TV; e em textos escritos – como jornal, revista, livro, e-mail ou internet –, incluindo assim diversos registros e meios linguísticos. A coleta em *córpus livre* foi relevante para garantir dois requisitos fundamentais: (a) a contemporaneidade do uso da expressão e (b) o conhecimento da expressão pelos falantes do PB em geral. A coleta das expressões em jornais e revistas foi especialmente relevante, tendo em vista que esses veículos de comunicação se destinam à difusão em todo o território nacional, indicando com isso a expectativa de compreensão do texto por falantes de qualquer região. Além disso, a checagem em

⁷ Sobre a importância da introspecção na pesquisa linguística, ver Talmy (2007); e também Perini & Othero (2011).

cópus livre permitiu a detecção de expressões fixas cuja sequência é interrompida por inserções (como em abrir o berreiro / abrir o **maior** berreiro; a cada passo / a cada **novo** passo) ou expressões que contêm variantes (como *história da carochinha/conto da carochinha; botar pra quebrar/pôr pra quebrar; a custo/depois de muito custo; com relação a/em relação a*).

De grande auxílio foi igualmente a coleta em cópus concebido justamente com o objetivo de funcionar como um centro de recursos para o processamento computacional da língua, sobretudo o “Linguateca”, que serviu para fornecer novas abonações e confirmar o uso de expressões já coletadas.

Os cópus orais e escritos utilizados, além de servirem de fonte para a coleta das expressões convencionais, e além de permitirem a comprovação do uso das expressões na linguagem atual, também serviram como fonte de abonação e exemplificação das diversas acepções das expressões coletadas. Foi assim composto o “DEIC – Dicionário de expressões idiomáticas e convencionais do português brasileiro contemporâneo” (Fulgêncio, a sair). O objetivo da composição desse dicionário foi o de apresentar dados da língua viva, moderna, realmente usada pelos falantes do PB atual.

Certamente, nem todas as expressões coletadas têm a mesma frequência de uso; e também é de se esperar que nem todas sejam reconhecidas por todos os falantes – pelo fato de ser implausível que um falante seja detentor de todo o inventário léxico de uma língua. Porém, não há dúvida quanto à existência das expressões elencadas, uma vez que foram todas coletadas em cópus real contemporâneo. Esse elenco permite, em última análise, o desenvolvimento de estudos mais rigorosos dos fraseologismos do português, uma vez que oferece dados empíricos que permitem revelar fatos mais aderentes à realidade linguística moderna do Brasil.

Assim, a coleta das expressões permitiu a análise da estrutura formal, bem como do comportamento semântico e discursivo dessas sequências. Surpreendentemente, revelaram-se certos comportamentos diferentes daqueles esperados, ou daqueles que tradicionalmente são apresentados na literatura sobre os fraseologismos. Foram evidenciados conceitos e mal-entendidos arraigados pela tradição quanto à caracterização das expressões fixas, que merecem reexame e reavaliação.

Alguns dos fatos observados foram os seguintes: (a) as expressões correntes no português brasileiro – aquelas que todos os falantes podem reconhecer e interpretar facilmente – são em grande número, em torno de 8.000⁸; (b) nem todas as expressões têm significado opaco; (c) as expressões apresentam individualidades não somente do ponto de vista semântico, mas também sintático, apesar de tradicionalmente ser enfatizada a idiosincrasia semântica.

Para exemplificar, vamos examinar nas próximas seções fatos que revelam traços ainda pouco salientados ou até mesmo inesperados na caracterização e descrição das expressões fixas do PB.

4. AS EXPRESSÕES FIXAS NÃO SÃO TÍPICAS DA LINGUAGEM COLOQUIAL

Entre os fatos revelados no levantamento em *cópus* real das expressões fixas do PB, o que mais salta aos olhos é o enorme uso de expressões fixas em textos escritos. Diferentemente do que se costumava pensar, as expressões fixas não são de uso quase exclusivo da linguagem oral coloquial, pouco cuidada, mas são de frequência muito alta em todos os tipos de textos, não apenas orais, mas também escritos, de qualquer nível de coloquialidade – ou seja, tanto em jornais e revistas quanto em livros técnicos ou de cunho literário. Ressalta-se ainda uma ocorrência significativa de fraseologismos na linguagem publicitária, muitas vezes com emprego de jogos de palavras que levam à construção do humor. Vejam-se, por exemplo, os títulos de artigos de revista abaixo mencionados, onde a desconstrução da expressão idiomática promove o humor e o interesse pelo texto:

(6) Acabou em pinga! (título de artigo sobre festival de cachaça)

(7) Divagar e sempre (título de artigo sobre movimento filosófico que vai na contramão da pressa)

As expressões fixas são, na verdade, fatos comuns da língua, disseminados em qualquer meio escrito ou oral, em qualquer nível de formalidade.

⁸ Não estão computados neste número os casos de colocação preposicional nas valências verbais e nominais.

5. AS EXPRESSÕES FIXAS PODEM SER TRANSPARENTES

Como vimos, uma expressão fixa é uma sequência linguística multilexical, convencional e memorizada. Trata-se não apenas de uma expressão não computável composicionalmente e cujo significado não pode ser depreendido a partir da soma do significado individual dos itens componentes, mas sim de qualquer grupo de palavras convencional na língua, guardado como um todo na mente dos falantes. Podemos entender as expressões fixas como o hiperônimo das expressões idiomáticas, ou seja, o termo mais genérico, a classe mais alta que engloba também as expressões idiomáticas tradicionais.

Para que se analise um grupo de palavras como uma expressão fixa basta que a sequência seja convencional e memorizada, ou seja, basta que faça parte do inventário lexical dos falantes da língua como uma **unidade da memória**. O conceito de expressão fixa não está ligado à semântica, mas à memória.

Como se sabe, há expressões opacas cujo significado é indecomponível e idiossincrático, como *a três por dois*, *mandar às favas*, *da pá virada* ou *gato-pingado*. O significado do grupo inteiro, em bloco, não corresponde à soma do significado das palavras integrantes do grupo. Quando uma expressão fixa é opaca, a fórmula [conhecimento da estrutura da língua + conhecimento do significado individual de cada item léxico envolvido na expressão] não garante a compreensão do significado global do grupo. Esse é o caso também das expressões *falar cobras e lagartos*, *cruz-credo*, *casa da mãe joana*, *sem eira nem beira* ou *fazer de conta*, em que o significado não é transparente, isto é, não é possível chegar ao significado da expressão através do cômputo do significado individual de cada item integrante do grupo.

Mas nem sempre o significado da expressão fixa é incompreensível através de composição. Existem também expressões fixas transparentes, cujo significado pode ser montado pelo ouvinte a partir da semântica individual de cada elemento lexical. Exemplos de expressões fixas transparentes são os seguintes: *dia sim, dia não*; *ir de mal a pior*, *precisar ver para crer*, *gosto não se discute*, *amor-próprio*⁹, *guarda-volumes*, *achados e perdidos*, *quem tudo quer tudo perde*, *a união faz a força*, *por*

⁹ Convém observar que a presença ou não de hífen não pode ser tomada como traço definidor do tipo de lexia, tendo em vista que as normas ortográficas em vigor não apresentam um parâmetro coerente para a inclusão ou não de hífen entre palavras (cf. Fulgêncio e Perini, 2012). Por exemplo, temos *alta-roda* e *alta-tensão* (mas *alta classe* e *altas horas*); *barra-pesada* (mas *barra suja*); *arco-da-velha* (mas *pé de moleque*); *chá-mate* (mas *chá preto*).

incrível que pareça, prisão perpétua, salvo engano, até prova em contrário, sobe e desce, de alto a baixo, assim espero, e tantas outras mais. A opacidade semântica da expressão **não** é condição para a caracterização das expressões fixas, mas sim o seu caráter convencional dentro da organização da língua.

Apesar de transparentes e compreensíveis mesmo por ouvintes que nunca tenham entrado em contato anteriormente com essas sequências, ainda assim os exemplos citados são analisados como expressões fixas porque se trata de blocos cristalizados, de unidades lexicais. Expressões como *achados e perdidos, de lá pra cá, sofá-cama* e outras são interpretáveis palavra por palavra, através da integração dos elementos componentes; mas ainda assim continuam pertencendo ao conjunto das expressões fixas porque caracterizam uma estrutura rígida, armazenada em bloco na memória: não se costuma dizer **perdidos e achados*, nem **de lá para aqui*, nem **poltrona-cama*, o que evidencia que não são composições livres, uma vez que não aceitam mudança de ordem sintática nem apresentam liberdade de substituição lexical. São sempre repetidos do mesmo modo, na mesma ordem, com o mesmo formato, com os mesmos itens, não admitindo substituíbilidade dos componentes nem alteração da forma e da ordem linear¹⁰. Não são admissíveis variantes como **precisar ver para acreditar* (com substituição lexical da expressão convencional *precisar ver para crer*) ou **de baixo a alto* (em vez de *de alto a baixo*, com alteração na ordem das palavras). Os itens integrantes de expressões como essas são insubstituíveis por outros de mesma carga semântica e potencial sintático. A expressão é cristalizada, e não um sintagma composto livremente – apesar de seu significado ser transparente.

Como se vê, nem todos os fraseologismos apresentam idiosincrasias semânticas: alguns são regulares do ponto de vista semântico, mas idiosincráticos sob outros pontos de vista.

6. A CARACTERIZAÇÃO DAS EXPRESSÕES FIXAS DEVE CONSIDERAR O PAPEL DO ENUNCIADOR, E NÃO SOMENTE O DO DECODIFICADOR

Quando o ouvinte está diante de uma expressão fixa transparente, é capaz de interpretá-la mesmo se não a conhece antecipadamente, operando na elaboração do

¹⁰ A respeito de testes e critérios para a detecção de expressões fixas veja-se Fulgêncio (2008, Cap. 7).

significado a partir da computação semântica dos termos que compõem o grupo, assim como faz para gerar o significado de qualquer sintagma.

Apesar de expressões transparentes poderem ser decodificadas e compreendidas por montagem de significado, é pouco provável que, no aspecto da produção, possam ser empregadas por quem não as conheça por antecedência. As expressões fixas só podem ser produzidas adequadamente, da forma convencionalizada pela língua, pelos falantes proficientes que já têm o grupo lexical guardado na memória. A recuperação do sentido (do ponto de vista do ouvinte) pode ser possível em alguns casos, mas não a produção (do ponto de vista do enunciador).

A enunciação de uma expressão fixa por parte de um falante que não a conhece de antemão fica comprometida, porque seria difícil (senão impossível) imaginar exatamente qual seria a estrutura fixada e cristalizada dentre as tantas possibilidades que a língua oferece. Seria improvável a enunciação de uma expressão fixa sem que se soubesse *a priori* qual é essa estrutura determinada pela convenção linguística. Existiriam chances ínfimas de que um indivíduo que não conhecesse a expressão, qualquer que seja ela, pudesse construí-la exatamente da forma consagrada na língua. Se o falante montasse o sintagma criativamente, poderia compor construções do tipo **prisão eterna* (em vez de *prisão perpétua*) ou **exceto engano* (em vez de *salvo engano*) ou **perdidos e achados* (com mudança na ordem dos elementos coordenados). Aliás, a ordem **perdidos e achados* refletiria uma lógica cronológica desfeita na expressão original, e estaria mais de acordo com a cronologia dos acontecimentos, refletindo a ocorrência dos fatos, do ponto de vista pragmático – só que não foi essa ordem que a língua escolheu.

Em caso de desconhecimento da forma da expressão transparente, o processamento pelo ouvinte pode ser viável, mas não o emprego da expressão por parte do enunciador. A recuperação do significado pode ser possível (pela decodificação do sintagma ou por ancoramento pragmático em expectativas ou em dicas do discurso), mas é improvável a produção com o exato formato estabelecido pela língua e consagrado pelo uso.

Portanto, não é somente a impossibilidade de computação semântica que caracteriza o grupo como idiomático, mas igualmente a implausibilidade de **produção** por quem não conheça de antemão a forma codificada pela língua. Essa constatação aponta para o fato de que a definição de uma expressão fixa não pode estar apoiada unicamente no ponto de vista do decodificador (como se costuma

fazer), mas também deve levar em conta o ponto de vista do enunciador. A definição de um fraseologismo não deve considerar somente a impossibilidade de decodificação do grupo ou a impossibilidade de atribuição do significado pelo ouvinte, mas sobretudo a baixa possibilidade de montagem do grupo pelo falante, naquele formato consagrado e convencional (caso o conjunto não esteja previamente armazenado no seu léxico mental).

Portanto, no âmbito das expressões fixas é importante diferenciar as atividades de recepção e produção linguística, que podem determinar operações linguísticas diferentes e envolver gramáticas distintas.

7. AS EXPRESSÕES FIXAS NÃO SÃO PRODUTIVAS

Ainda tratando de “produção”, convém ter presente que as expressões fixas de uma língua não são produtivas – não no sentido de produtividade lexical empregado usualmente no campo da morfologia, que lida com morfemas produtivos que podem gerar novos itens. Na verdade, as noções de produtividade lexical e de convencionalidade são até mesmo antagônicas.

É importante e necessário estabelecer uma diferença clara entre as noções de frequência e de produtividade. As expressões fixas são frequentes, mas não são produtivas: não podem ser usadas para gerar novos itens e elas próprias não podem ser criadas a partir de formações neológicas construídas por um único falante. Não é possível inventar ou criar um novo fraseologismo numa perspectiva sincrônica. Portanto, apesar de ocorrerem com frequência no enunciado, os fraseologismos não são produtivos.

A razão é simples: para classificar um grupo lexical como fraseológico é necessário que esse conjunto seja conhecido por vários falantes e que seja consagrado pelo uso. Lembremos que o que caracteriza uma expressão fixa é a sua convencionalidade, de modo que a criação de uma expressão fixa é um processo extensivo no tempo: para ser criada uma nova expressão é preciso primeiro que a sequência lexical “contamine” vários falantes e se espraie *ipsis litteris* pela língua. Sendo assim, a formação de uma nova expressão é um processo que demanda tempo, e não pode ser feito por decisão de um único falante, como acontece no caso da formação de neologismos. O que se pode construir neologicamente é um sintagma

novo, não uma expressão convencional que, para ser caracterizada como tal, precisa ser disseminada e conhecida pelos falantes em geral.

Claro, diacronicamente podem ser criadas expressões, que depois de serem repetidas por vários falantes acabam por se tornarem conhecidas e se incorporarem ao inventário lexical da língua. Mas do ponto de vista sincrônico não tem sentido falar em produtividade fraseológica ou em construção neológica de expressões fixas, o que pode soar até mesmo como uma contradição em termos.

8. AS EXPRESSÕES FIXAS PODEM INCLUIR FORMAS LEXICAIS QUE NÃO EXISTEM FORA DA EXPRESSÃO

A idiosincrasia lexical também se evidencia pelo fato de que, em alguns casos, o fraseologismo inclui itens que só existem dentro daquela expressão, isto é, que nunca podem aparecer desacompanhados do restante do grupo. Isso acontece, por exemplo, com as expressões *à paisana*, *de chofre* ou *não obstante*, em que aparecem as palavras *paisana*, *chofre* e *obstante*, que não existem sozinhas, ou seja, não ocorrem fora do grupo idiomático. São formas grafadas como palavras, mas que não são itens independentes, uma vez que ocorrem exclusivamente nesse conjunto.

Alguns outros exemplos que ilustram palavras que só são empregadas na expressão fixa são os seguintes (negritados em cada expressão): *à **beça***, *à **toa***, *às **pampas***, *a **esmo***, *por um **triz***, *meter o **bedelho***, *mentira **deslavada***, *terreno **baldio***, *sem **eira** nem beira*, *perder as **estribeiras***, *em **frangalhos***, *no que **tange***, *ao **léu***, *ao **bel** prazer*, *desta **feita***, *em **riste***, *pé **ante** pé* e muitos outros.

Os idiomatismos acima possuem certamente um significado global para o grupo como um todo, mas os termos que os compõem não têm significado independente, uma vez que não têm liberdade de se destacarem do conjunto idiomático. Não ocorrem na língua atual, fora da expressão, os substantivos *beça*, *esmo*, *eira* ou *feita* (no feminino), por exemplo. Os falantes do português sabem o que significa *por um triz* ou *meter o bedelho*, mas só como um conjunto idiomático cristalizado. Não seria possível isolar os componentes desses grupos e dizer coisas como **o triz foi pequeno* ou **esse bedelho foi inconveniente*. Apesar de presentes nas expressões, essas pretensas “palavras” não configuram vocábulos da língua atual em uso; e, portanto, isoladas do grupo idiomático são desprovidas de significado.

Quando cada um desses grupos foi criado, as palavras que os compunham tinham provavelmente existência autônoma e algum significado reconhecível – tanto é que naquele momento essas formas foram usadas para a construção de sintagmas que posteriormente constituíram os idiomatismos usados hoje em dia. Portanto, é concebível e provável que, um dia, palavras como *triz* ou *bedelho* tenham tido um significado independente. Do ponto de vista sincrônico, porém, as palavras salientadas acima não podem ser usadas independentemente, fora da expressão; portanto, na língua contemporânea já perderam a carga semântica que tiveram no passado e a sua condição de item lexical autônomo. Recuperar, hoje em dia, o significado de formas como *triz*, *léu* ou *toa*, por exemplo, é trazer para o presente fatos semânticos que só existiram no passado e que não são conhecidos atualmente pelos falantes nativos contemporâneos.

Vê-se, mais uma vez, a importância e a necessidade da distinção clara entre diacronia e sincronia no estudo dos fraseologismos.

9. AS EXPRESSÕES FIXAS PODEM INCLUIR IDIOSSINCRASIAS SINTÁTICAS

Sabe-se que muitas vezes o significado de um fraseologismo não pode ser obtido através da somatória do significado das partes componentes. As expressões idiomáticas têm um significado para o conjunto que muitas vezes não tem nada a ver com o significado das palavras que integram o idiomatismo. Esse é o caso exemplificado por *rasgar seda*, *de uma figa*, *pintar o sete*, *de afogadilho*, *sem eira nem beira*, *pagar mico* e tantas outras expressões idiomáticas. Temos aí uma idiossincrasia semântica.

Mas se uma sequência é transparente, como vimos anteriormente (em casos como *de lá pra cá* ou *até prova em contrário*), se é inteligível e se é possível compreendê-la composicionalmente, como se pode saber que se trata de um fraseologismo? A resposta é que às vezes a idiossincrasia está em outro lugar, muitas vezes na parte formal (diferentemente do que se costuma pensar).

As idiossincrasias não se limitam exclusivamente a questões semânticas de composição de significado, mas também podem envolver questões sintáticas ou lexicais, rompendo padrões gramaticais, e assim revelando uma estrutura idiomática também quanto à estrutura formal. Tanto do ponto de vista semântico quanto do ponto de vista sintático e também pragmático, as expressões fixas podem apresentar

idiosincrasias sob diversos aspectos, como ordem fixa ou diferente da habitual, verbos com diátese incomum, nomes não referenciais e muitas outras idiosincrasias que não se conformam aos padrões linguísticos esperados. Veremos a seguir alguns desses casos.

9.1 IRREVERSIBILIDADE

Um dos casos em que as expressões não seguem o mesmo comportamento dos sintagmas composicionais refere-se ao imobilismo da ordem sintática. Frequentemente não é possível mudar a ordem das palavras de uma expressão.

Em estruturas construídas (que não são fraseologismos), a conjunção *e* coordena palavras ou sintagmas de mesma classe, e normalmente esses dois elementos são permutáveis: [A e B] ou [B e A]. Na língua em geral, em sintagmas coordenados pode-se alternar a ordem, de forma que o primeiro elemento passa para a segunda posição e vice-versa. Por exemplo, pode-se comprar *laranja e banana* ou *banana e laranja* – a ordem é irrelevante.

Porém, em casos de expressões fixas que incluem uma estrutura coordenada, a ordem pode ser rígida. Quando há uma coordenação aditiva ou alternativa, isto é, quando os itens são unidos por conjunções como *e*, *ou*, ou por vírgula (que são os principais casos em que se esperaria a possibilidade de permuta de ordem), muitas vezes a ordem dos termos é irreversível e não se pode trocar a posição dos elementos dentro da expressão. Também em outros casos a ordem das palavras pode ser cristalizada, mesmo se as regras da língua permitiriam em princípio o movimento dos integrantes da sequência. Vejam-se alguns exemplos:

<i>arco e flecha</i>	<i>(*flecha e arco)</i>
<i>são e salvo</i>	<i>(*salvo e são)</i>
<i>pele e osso</i>	<i>(*osso e pele)</i>
<i>prós e contras</i>	<i>(*contras e prós)</i>
<i>ser unha e carne</i>	<i>(*ser carne e unha)</i>
<i>a ferro e fogo</i>	<i>(*a fogo e ferro)</i>
<i>a torto e a direito</i>	<i>(*a direito e a torto)</i>
<i>cara ou coroa</i>	<i>(*coroa ou cara)</i>
<i>par ou ímpar</i>	<i>(*ímpar ou par)</i>

<i>mais ou menos</i>	(* <i>menos ou mais</i>)
<i>de cabo a rabo</i>	(* <i>de rabo a cabo</i>)
<i>de cima a baixo</i>	(* <i>de baixo a cima</i>)
<i>sem dó nem piedade</i>	(* <i>sem piedade nem dó</i>)
<i>sem pé nem cabeça</i>	(* <i>sem cabeça nem pé</i>)
<i>mais dia, menos dia</i>	(* <i>menos dia, mais dia</i>)
<i>ter em alta conta</i>	(* <i>ter em conta alta</i>)
<i>por assim dizer</i>	(* <i>por dizer assim</i>)
<i>azar seu!</i>	(* <i>seu azar!</i>)

Vê-se então que a cristalização da expressão implica também um rompimento de regras sintáticas, como no caso da irreversibilidade de ordem dentro da expressão, em casos em que o movimento seria permitido em princípio pela gramática da língua.

9.2 EM EXPRESSÕES FIXAS O NOME NEM SEMPRE É O CONTROLADOR MORFOSSINTÁTICO

Como explicam Perini *et al.* (1996), um SN é geralmente estruturado em torno de um nome, que tem potencial referencial e constitui o núcleo do sintagma, uma vez que é esse o elemento que define o gênero e número dos demais elementos integrantes do SN (como determinantes, possessivos, qualificativos etc). O núcleo é não só o controlador da referência, como também o controlador morfossintático (segundo Zwicky, 1993, o “*morphosyntactic locus*”). Isso quer dizer que se o núcleo é masculino, o SN inteiro também é; e se é feminino, o SN também é. Em um sintagma composicional se diz, portanto, *o livro amarelo*, mas *a revista amarela*.

Pois bem, quando o SN é uma expressão fixa, essa condição pode ser subvertida, de modo que nem sempre o nome interno à expressão é o controlador morfossintático. Esse fato se verifica, por exemplo, nas expressões *o caixa dois*, *o pele-vermelha*, *o longa-metragem* e *o boia-fria*. Apesar de o nome aí presente ser feminino (*a caixa*, *a pele*, *a metragem* e *a boia*), o sintagma, como um todo, é masculino. Isso mostra que em casos como esses o nome interno à expressão não é o núcleo do SN idiomático e não atua como controlador morfossintático, diferentemente das regras que funcionam em sintagmas composicionais.

9.3 O ARTIGO PODE SER DEFINIDOR DO SIGNIFICADO

Em sintagmas construídos, o artigo tem a função de indicar características semânticas do nome com o qual entra em construção (como genérico/específico) ou características discursivas (como definido/indefinido, ou dado/novo). Porém, quando se trata de uma expressão fixa o artigo pode ter funções não previstas em seu repertório regular.

Normalmente, entre duas construções iguais, a presença do artigo ou o tipo do artigo usado não é capaz de atribuir significados diferentes às construções, além dos já previstos dentro da própria semântica do artigo. Mas há casos de expressões fixas em que a mudança de artigo é capaz de atribuir significado radicalmente diferente ao grupo, fazendo com que duas expressões se distingam unicamente pelo tipo de artigo que entra na formação da estrutura idiomática. Ou seja, ao se comparar “pares mínimos” de expressões idênticas, cuja única diferença é o tipo de artigo usado ou a presença/ausência do artigo, observa-se que exclusivamente a mudança de artigo é capaz de demarcar um significado diferente para cada expressão. São exemplos desse caso os apresentados no quadro abaixo:

<i>sair da linha</i> (perder a compostura) / <i>sair de linha</i> (parar de ser fabricado)
<i>pegar um fogo</i> (ficar bêbado) / <i>pegar fogo</i> (incendiar-se)
<i>ter a palavra</i> (ter a vez de falar) / <i>ter palavra</i> (cumprir o que diz)

Essa característica das expressões fixas apresenta um ponto que ainda não havia sido considerado, que é a possibilidade de o artigo ser o elemento definidor de um campo semântico.

9.4 PODEM APARECER ANÁFORAS SEM REFERENTE

Em construções livres podem ser usadas anáforas sem antecedente, cuja interpretação depende de conhecimentos partilhados entre falante e ouvinte (Fulgêncio, 1983). O referente não é recuperado nem por relação com o texto, nem por relação exofórica (ou seja, com o ambiente físico em que se desenrola a cena), mas é evocado inferencialmente, com base em conhecimentos pragmáticos. Mas nos idiomatismos não ocorre o mesmo fenômeno que se pode observar em construções

livres: em expressões fixas é possível fazer menção a um referente que não é expresso de nenhuma forma, mas que já está implícito, por convenção, na própria expressão.

Um desses casos é o de expressões dêiticas como *estar por aqui [com alguém]* e *entrar por aqui e sair por ali* (indicando “entrar por um ouvido e sair pelo outro”). Essas expressões se referem ao contexto físico (o pescoço e o ouvido), mas não são obrigatoriamente acompanhadas do gesto que relaciona a dêixis ao seu referente; isso porque o significado já é conhecido. Dessa forma, as dêixis acabam sendo usadas sem serem obrigatoriamente acompanhadas do ancoramento gestual que marcaria ou apontaria o seu referente.

Em expressões fixas é também possível o uso de anáfora sem referente, o que nunca ocorre em construções livres. Exemplos de anáfora sem referente são os seguintes (onde o símbolo \emptyset indica uma elipse): *sem mais \emptyset nem menos \emptyset* , *dar uma \emptyset de, não estar nem **aí**, ora **essa**, fazer das **suas** ou sem mais **aquela**.*

Se isso ocorresse numa construção nova (não fixa), não seria possível determinar o significado, e a comunicação seria afetada negativamente. Mas no fraseologismo, como o significado é fixo e já conhecido para o grupo, já convencionalizado e já determinado para o bloco inteiro, o fato de haver anáforas com referente não identificável não interfere na compreensão, uma vez que o significado interno não é computado. Em situação composicional, não idiomática, o ouvinte precisa recuperar o referente das anáforas para montar o significado; mas quando se trata de uma expressão, o significado já está pronto na sua memória lexical. Não interessa qual é o referente de um item anafórico que possa aparecer dentro de uma expressão – só interessa o significado do grupo idiomático inteiro, que já está determinado e sabido *a priori*.

9.5 A ANÁFORA PODE APARECER NA FORMA MARCADA

Como se viu acima, no grupo das expressões fixas encontram-se expressões como *ora **essa**, fazer das **suas** e sem mais **aquela***, em que a anáfora aparece no feminino. O mesmo acontece em muitos outros casos, como também em *cair **nessa**, corta **essa**, estar **na mesma** e ficar **elas** por **elas***. Se não há antecedente de gênero feminino nem referente de sexo feminino, seria de se esperar que a anáfora migrasse para a forma não-marcada, que é a do masculino singular. Estranhamente, a anáfora aparece no feminino, que é uma forma marcada no português.

A situação *default* na língua é a de as formas tenderem para o masculino singular. Uma situação inesperada é a ocorrência de casos de escolha da forma feminina, ou plural, ou feminina plural, sem que haja relação com outro termo que condicione a flexão. Quando se trata de expressões fixas, observa-se uma tendência acentuada ao uso de formas marcadas no feminino e no plural, mesmo quando não há concordância com um termo regente ou quando não há relação conceitual com o número plural. Exemplos desse caso são os seguintes: *levar a pior, às cegas, voltar às boas, cair na real, qual é a sua? e essa não!* (dentre outros).

É surpreendente a constatação de que muitas vezes a flexão é feita no feminino e no plural, que são justamente as duas formas marcadas na língua. A pergunta é: porque as expressões fixas se comportam assim, na contramão da tendência que funciona para os sintagmas construídos?

Para explicar a presença do feminino e/ou do plural seria possível imaginar a presença de um referente anafórico (elíptico) com o qual seria feita a concordância. Por exemplo, em *estar na mesma* alguém poderia teorizar que a concordância seria feita no feminino devido a uma suposta elipse relativa à palavra *situação* (como em *estar na mesma “situação”*). Mas essa seria uma explicação *ad hoc*, uma vez que não há nenhuma evidência independente de que a suposta elipse seria *situação* (palavra feminina) e não, por exemplo, *estado* ou *jeito* (palavras masculinas). A explicitação da elipse é feita retrospectivamente: uma vez que já se sabe que a expressão aparece no feminino, imagina-se a presença de uma elipse de palavra feminina que imporá a concordância. Ora, esse tipo de explicação é evidentemente circular e não serve para explicar a preferência por formas marcadas.

Para complicar mais a situação, há ainda casos não anafóricos, onde a expressão está no feminino, mas não há ausência de substantivo – como em *estar com a macaca, na calada da noite, desta feita* – e também no plural, como em *dar as caras, às pressas* ou *às cegas*. Nesses casos o substantivo aparece explicitamente no feminino e/ou no plural, sem motivação sintática para tal – ou seja, fica novamente sem explicação o porquê da escolha da forma marcada na composição da expressão.

Não há até o momento nenhuma explicação plausível que possa justificar a escolha pela forma marcada em muitas expressões fixas. Portanto, a pergunta sobre a motivação para o emprego do feminino e do plural em fraseologismos continua em aberto.

CONSIDERAÇÕES FINAIS

Vimos então que, ao lado das expressões idiomáticas tradicionalmente consideradas, como *botar a boca no trombone*, *comer mosca* ou *falar cobras e lagartos*, há um enorme número de outras expressões fixas, como *a fim de que*, *às vezes*, *quem dera*, *levar a melhor*, *muito embora*, *tão logo* ou *não há de quê*, igualmente idiossincráticas e convencionais, que compõem o grupo das expressões fixas guardadas na memória, amplamente utilizadas na composição dos enunciados.

As expressões fixas não se caracterizam exclusivamente por idiossincrasias do ponto de vista semântico, como se costuma pensar. Ao lado da opacidade semântica e de especificidades lexicais, aparecem também várias características inesperadas quanto à composição estrutural, indicando que os fraseologismos se comportam idiossincraticamente também do ponto de vista sintático. Foram aqui pinçadas algumas dessas idiossincrasias, dentre um rol de várias outras.

A não-conformidade das expressões fixas aos parâmetros da língua acrescenta mais a indicação de que essas estruturas só podem ser entendidas como um bloco único, que o usuário emprega sem passar pela composição, nem para produzi-lo nem para decodificá-lo. Caso fosse feito algum tipo de processamento, muitas expressões teriam de ser rejeitadas por serem consideradas agramaticais – o que evidentemente não é o que acontece. Os dados indicam que a estrutura interna das expressões fixas nem sempre se conforma aos padrões gramaticais que regem a montagem de sintagmas. As expressões aceitam a violação de condições estruturais e semânticas de natureza diversa, infringem muitas vezes padrões lexicais, semânticos ou formais que funcionam para o restante da língua, sendo que tal fato nem mesmo é percebido pelos falantes. A não conformidade às regras da língua (semânticas e sintáticas) é inclusive um dos critérios para a identificação de expressões fixas.

Diferentemente do que geralmente se pensa, as expressões fixas não são típicas da linguagem coloquial. Esse tipo de confusão aparece em muitos autores, que insistem em afirmar que as expressões idiomáticas seriam criações de origem popular, sobretudo da linguagem oral. Ora, as expressões fixas fazem parte da competência do falante e não são exclusivas de textos orais, uma vez que aparecem comumente em textos escritos, de qualquer nível de formalidade, sem que tal fato cause espanto ou impacto por um hipotético choque de registros. Há inclusive certas

expressões que são usadas exclusivamente em língua escrita formal (como *por conseguinte, em tela, à guisa de, até mais ver, com o fito de, dar de ombros, de mais a mais, tão-somente, em tempo, fazer mister* ou *isto posto*). As expressões fixas são simples fatos da língua.

Não é possível evitar o emprego de seqüências memorizadas em textos orais ou escritos. Autores normativistas tratam certas expressões como “chavões”, e dão a elas o carimbo de “construções inadequadas a evitar”. Se a ideia de evitar sintagmas prontos e construções padronizadas fosse posta em prática, chegaríamos a uma língua bem diferente da que é usada hoje em dia. O fato é que os enunciados, sejam orais sejam escritos, são permeados de constantes e inúmeros usos de expressões fixas.

A grande quantidade, a frequência e a constância de uso das expressões fixas na composição do discurso indicam que não se pode tratar o grupo como marginal. Essas características sugerem ainda que não é possível imaginar uma teoria linguística adequada para o português que não incorpore os fraseologismos como um traço proeminente da língua.

REFERÊNCIAS

1. CASCUDO, Luís da Câmara. *Locuções tradicionais no Brasil*. 2ª edição, Rio de Janeiro: FUNARTE, 1977.
2. CULICOVER, Peter W.; JACKENDOFF, Ray, *Simpler Syntax*. Oxford: Oxford University Press, 2005.
3. FERREIRA, Aurélio Buarque de Holanda. *Novo dicionário eletrônico Aurélio versão 6.0*. Editora Positivo, 4ª edição, 2009.
4. FILLMORE, Charles J. On Fluency. In: RIGGENBACH, Heidi (org.), *Perspectives on Fluency*. Ann Arbor: The University of Michigan Press, 2000.
5. FULGÊNCIO, Lúcia. DEIC - *Dicionário de expressões fixas e idiomatismos do português brasileiro*. (inédito)
6. FULGÊNCIO, Lúcia. *Expressões fixas e idiomatismos do português brasileiro*. Tese de Doutorado – Pontifícia Universidade Católica de Minas Gerais. Belo Horizonte, 2008, 502 p.

7. FULGÊNCIO, Lúcia. *O problema da interpretação dos elementos anafóricos*. Dissertação de Mestrado – UFMG. Belo Horizonte, 1983.
8. FULGÊNCIO, Lúcia; PERINI, Mário A. A dança da ortografia. *Revista do Centro de Estudos Portugueses*, v. 32, n. 27, 2012
9. GOLDBERG, Adele. *Constructions: A construction grammar approach to argument structure*. Chicago, IL: University of Chicago Press, 1995.
10. GROSS, Gaston. *Les expressions figées en français*. Paris: Ophrys, 1996.
11. GROSS, Maurice. Constructing lexicon-grammars. In: ATKINS, B. T. S.; ZAMPOLLI, A. *Computational approaches to the lexicon*. Oxford University Press, 1994.
12. GROSS, Maurice. On the failure of generative grammar, *Language*, 55, 4, p. 859-885, 1979.
13. LANGACKER, Ronald W. *Concept, image and symbol – the cognitive basis of grammar*. Berlin: Mouton de Gruyter, 1991.
14. LANGACKER, Ronald W. *Foundations of cognitive grammar – theoretical prerequisites*. Stanford: Stanford University Press, 1987.
15. LEVELT, W. J. M. *Speaking: from intention to articulation*, Massachusetts: MIT Press, 1989.
16. LINGUATECA – Córpus Brasileiro - acessível em <http://www.linguateca.pt/acesso/corpus.php?corpus=CBRAS>
17. NASCENTES, Antenor. *Tesouro da fraseologia brasileira*. São Paulo: Livraria Freitas Bastos S.A., 2ª edição, 1966.
18. PERINI, Mário A. *et al. O sintagma nominal em português: estrutura, significado e função*. Número especial da Revista de estudos da linguagem – Faculdade de Letras da UFMG. jul./dez. 1996.
19. PERINI, Mário A.; OTHERO, Gabriel de Ávila. 2011. Córpus, introspecção e o objeto da descrição gramatical. *Signo*, Universidade de Santa Cruz do Sul, RS, vol. 35, nº 59, p. 2-12.
20. PUGLIESI, Márcio. *Dicionário de expressões idiomáticas*. São Paulo: Editora Parma, 1981.
21. TALMY, Leonard. 2007. Foreword. In GONZÁLEZ-MÁRQUEZ, M.; MITTELBERG, Irene; COULSON, Seana; SPIVEY, Michael J. (eds). 2007. *Methods in cognitive linguistics*. Amsterdam: John Benjamins, p. XI-XXI.

22.ZWICKY, Arnold. Heads, bases and functors. In: CORBETT *et al* (orgs), *Heads in grammatical theory*. Cambridge: Cambridge University Press, 1993.

Recebido em: 17/06/2017

Aceito em: 25/07/2017