

## **Field – Dicionário de Expressões do Futebol: um recurso lexicográfico baseado no aporte teórico-metodológico da Semântica de Frames e da Linguística de Corpus**

*Field – Football Expressions Dictionary: a lexicographic resource based on the theoretical-methodological approach of Frame Semantics and Corpus Linguistics*

**Rove Luiza de Oliveira Chishman**

**Aline Nardes dos Santos**

**Diego Spader de Souza**

**João Gabriel Padilha**

Universidade do Vale do Rio dos Sinos – Unissinos – São Leopoldo – Rio Grande do Sul – Brasil



**Resumo:** O presente artigo tem por objetivo realizar um estudo acerca da relevância do aporte teórico-metodológico da Semântica de Frames (FILLMORE, 1982) no desenvolvimento do *Field – Dicionário de Expressões do Futebol*, cuja organização permite a consulta à linguagem do futebol a partir de expressões e de cenários – ou frames semânticos. A Semântica de Frames, desenvolvida no contexto da Linguística Cognitiva, opera através de dados empíricos retirados através da análise de corpora eletrônicos. A extração e o tratamento dos dados que possibilitaram o estudo apresentado neste artigo aconteceram através do concordanceador Sketch Engine, enquanto a análise de tais dados ocorreu a partir de critérios presentes no arcabouço da Semântica de Frames. Dentre os resultados, é possível apontar para a forma como a teoria de Fillmore (1982) contribui para a análise da polissemia, apresentando os diferentes sentidos de uma unidade lexical a partir das diferentes situações, diferentes frames, em que aparecem. A partir deste artigo, destaca-se a pertinência da Linguística de Corpus e do processamento de corpora a partir de softwares que permitem a análise minuciosa das propriedades linguísticas encontradas nos textos. Ademais, é importante também ressaltar o caráter valoroso da Semântica de Frames na construção de um recurso lexicográfico com vistas a um público não especializado, uma vez que a teoria permite a contextualização da linguagem a partir de nossas experiências. Dessa forma, cita-se também o fato de que o Field, ao adaptar conceitos presentes na Semântica de Frames, representa uma importante contribuição para a prática lexicográfica.

**Palavras-chave:** Linguística Cognitiva. Semântica de Frames. Lexicografia. Linguística de Corpus.

**Abstract:** The present article aims at problematizing the relevance of Frame Semantics (Fillmore, 1982) in the development of *Field – Dictionary of Football Expressions* – which the configuration allows the access to football language through expressions or through scenarios – or semantic frames. Frame Semantics, a theory developed in the realm of Cognitive Linguistics, is based on empirical data collected from the analysis of electronic corpora. The extraction of the data presented in this study was done with the Sketch Engine concordance, while their analysis was relegated to Frame Semantics. Among the results, it is possible to point out at the manner in which Fillmore's theory contributes to the analysis of polysemy, presenting the different senses of a lexical unit considering different situations – or different frames – in which they appear. This article also emphasizes the pertinence of corpus linguistics and the processing of corpora as resources that allow the analysis of linguistic constructs present in the texts. It is also important to emphasize the applicability of Frame Semantics to a resource devoted to a non-specialized public, once the theory makes the contextualization of language possible through the everyday routine of the speakers.

**Keywords:** Working Memory. Learning. Psycholinguistics. Language Acquisition.

## 1 Introdução

A Semântica de Frames, desenvolvida por Fillmore (1982), é uma das teorias da Linguística Cognitiva que se pauta na análise de *corpora* processáveis por computador, procedimento que constitui apenas uma entre tantas formas de análise dentre as abordagens sociocognitivas existentes. Consoante Talmy (2006), outras metodologias utilizadas incluem introspecção, análise comparativa entre línguas, estudos pautados em gravações audiovisuais e técnicas experimentais. Como bem ressalta o autor, todos os aportes metodológicos possuem vantagens e limitações conforme a perspectiva utilizada.

No caso da teoria de Fillmore, os estudos baseados em *corpora* foram consolidados por meio da criação da plataforma FrameNet (RUPPENHOFER et al., 2010), recurso que evidencia o potencial da abordagem na busca de relevância lexicográfica. A FrameNet, como primeiro recurso lexicográfico baseado em *frames*, ao propor um programa de pesquisa baseado em *corpus*, através do qual a seleção de unidades lexicais e exemplos deixa de ser intuitiva e passa a ser mais sistematizada, contribuiu significativamente para a prática lexicográfica, que até então ainda carecia de abordagens empíricas as quais direcionassem o trabalho do lexicógrafo. (FILLMORE; ATKINS, 2000).

Nesse contexto, o *Field – Dicionário de Expressões do Futebol* – é um recurso lexicográfico trilingue organizado a partir da teoria da Semântica de Frames (FILLMORE, 1982), abordagem que se insere no contexto da Linguística Cognitiva. Sua construção envolveu a compilação de três *corpora* comparáveis, que foram processados a partir da ferramenta Sketch Engine (KILGARRIF, 2004). Considerando-se os preceitos da Linguística de Corpus, utilizada neste projeto como recurso metodológico, essa ferramenta foi fundamental para o processo de identificação dos *frames* e também para o tratamento de fenômenos como a polissemia.

Este trabalho visa a demonstrar a pertinência do aporte teórico-metodológico da Semântica de

Frames na construção do *Field*, um dicionário de futebol direcionado ao público não especializado, no qual as unidades lexicais organizam-se em torno da noção de *frame* semântico, que, no contexto desse recurso, é chamado de cenário. A partir do uso da ferramenta Sketch Engine, os resultados mostram que a análise de *corpora* foi relevante não apenas para a identificação de unidades lexicais, mas também para o reconhecimento dos *frames* que constituem os eventos do futebol, os quais aparecem, no dicionário *Field*, acompanhados de uma ilustração, indicando todas as unidades lexicais que pertencem ao respectivo cenário. Os estudos aqui descritos referem-se à análise da polissemia do lexema *tocar* e à identificação dos *frames* evocados pelo verbo *marcar*.

O artigo se organiza da seguinte forma: na primeira seção, contextualizaremos a Semântica de Frames enquanto abordagem cognitiva do significado; a seguir, abordaremos o empirismo metodológico da teoria, consolidado por meio da criação da plataforma FrameNet. A terceira seção concerne à Linguística de Corpus e à ferramenta Sketch Engine, utilizada em nosso projeto para o processamento dos *corpora* de pesquisa. Na sequência, são trazidos as análises e os resultados obtidos a partir do estudo aqui descrito, para então procedermos com as considerações finais.

## 2 Semântica de frames

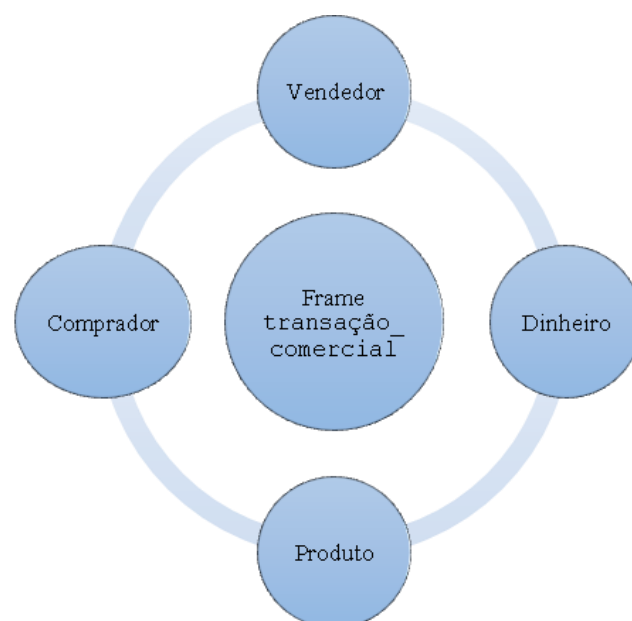
A Semântica de Frames é uma teoria desenvolvida pelo linguista americano Charles Fillmore (1982, 1985) no escopo da Linguística Cognitiva. Tendo surgido no cenário linguístico nos fins da década de 1970 e princípios da de 1980, a Linguística Cognitiva se estabelece como um paradigma que se desvincula de aspectos-chave presentes em abordagens formais em vigor na época. Nesse sentido, a Linguística Cognitiva propõe para os estudos linguísticos uma abordagem que rompe com o tratamento periférico dispensado a questões relacionadas ao sentido e ao uso. Assim, descontentes com o tratamento dispensado à semântica e à pragmática, um grupo de linguistas –

em que se destacam George Lakoff, Ronald Langacker, Leonard Talmy, Charles Fillmore e Gilles Fauconnier – se empenham em firmar um novo modelo para os estudos da linguagem, capaz de dar conta daquilo que era antes posto de lado, adotando o termo *Linguística Cognitiva*.

Vale ressaltar que a consciência de uma relação entre a cognição humana e a linguagem já circulava na Linguística através da Gramática Gerativa de Chomsky. Contudo, como se sabe, o Gerativismo se baseia em uma perspectiva *modular* da mente, defendendo a autonomia da linguagem em relação a outros módulos da nossa cognição, enquanto a Linguística Cognitiva defende o pressuposto de que a linguagem está intrinsecamente relacionada com os demais módulos cognitivos, assumindo a existência de aspectos cognitivos compartilhados pela linguagem e outras capacidades. Segundo Fauconnier (2003), capacidades cognitivas que possuem papel central na linguagem não são específicas a ela. Dentre essas capacidades, podemos incluir analogia, recursão, ponto de vista e perspectiva, alinhamento figura e fundo e integração conceptual. O fato de a Linguística Cognitiva promover o encontro de todos os níveis linguísticos, não mais negligenciando a semântica e a pragmática, possibilita compreendermos que o conhecimento é *enciclopédico*, noção que, na Linguística Cognitiva, se opõe a de “conhecimento de dicionário”. Por exemplo, o significado de dicionário de “cachorro” se limita a incluir as propriedades de um cachorro que nos possibilitam distingui-lo de outros seres, enquanto o conhecimento enciclopédico carrega uma série de outras informações, como o fato de cachorros serem conhecidos, em determinados contextos socioculturais, como os melhores amigos do homem (CRUSE, 2006). Sendo assim, a Linguística Cognitiva relaciona a linguagem (e a forma como nos expressamos através dela), com as nossas experiências. É nesse contexto que surge a Semântica de Frames de Charles Fillmore.

A teoria, de acordo com Evans e Green (2006, p.222, tradução nossa), que “tenta descobrir as propriedades do inventário estruturado da linguagem

associado com as palavras, e considerar quais consequências as propriedades desse sistema de conhecimento podem ter para um modelo semântico.”<sup>1</sup> Um *frame* pode ser descrito como uma esquematização da experiência depositada em nossa memória de longo-prazo (EVANS; GREEN, 2006), elencando os elementos participantes de uma determinada experiência. De acordo com Fillmore (1982, p.11), um *frame* é “qualquer sistema de conceitos relacionados de tal maneira que para entender qualquer um deles é preciso entender a estrutura que os comporta como um todo [...]”. No intuito de demonstrar esse conceito, podemos utilizar o exemplo clássico do *frame* transação\_comercial. Como é possível ver na figura a seguir, o *frame*, no centro, é rodeado por elementos: o comprador, o vendedor, o dinheiro e o produto. Tais elementos (e a forma como eles se relacionam) caracterizam os conceitos pertencentes a uma cena de transação comercial, isto é, são os participantes e objetos necessários para que a situação retratada no *frame* se concretize.



Fillmore (1982) caracteriza a Semântica de Frames como empírica. Nas palavras do linguista “A Semântica de Frames oferece um modo particular de

<sup>1</sup>No original: “[...] attempts to uncover the properties of the structured inventory of knowledge associated with words, and to consider what consequences the properties of this knowledge system might have for a model of semantics.”

se olhar para o significado das palavras [...]” (FILLMORE, 1982, p.11). Tal modo é caracterizado pela observação e análise das nossas experiências, resultantes da nossa interação com o ambiente. As palavras, para Fillmore, são *categorizações de experiência*, sustentadas por uma situação motivacional que ocorre em um plano de fundo de conhecimento decorrente da forma como nos posicionamos e percebemos o mundo ao nosso redor. Desse modo, os *frames* podem caracterizar um modo para se entender a razão pela qual determinada comunidade cria certas categorias de palavras, buscando explicar o significado de cada unidade lexical através do esclarecimento de tal motivação. Os *frames*, nesse sentido, estão intrinsecamente relacionados à cultura.

Tal relação existente entre o *frame* semântico de Fillmore e a cultura de cada um pode ser representada a partir de um *frame* de café da manhã (FILLMORE, 1982). Segundo o linguista,

Entender essa palavra é entender a prática em nossa cultura de ter três refeições diárias, em horários mais ou menos convencionalmente estabelecidos do dia, e sendo que uma dessas refeições é aquela a ser feita no início do dia, após um período de sono, consistindo de um menu um tanto quanto único (cujos detalhes podem variar de comunidade para comunidade) (FILLMORE, 1982, p.118, tradução nossa).<sup>2</sup>

De início, temos o fato de que um *frame* de café da manhã carrega mais do que apenas informações referentes ao que se consome. Compreender “café da manhã” significa compreender o papel dessa refeição na nossa sociedade. Podemos ir mais fundo: um café da manhã tradicionalmente americano contém ovos mexidos, bacon, panquecas. É possível que a ausência de algum desses elementos cause estranhamento em uma pessoa acostumada a essa realidade. No entanto, no Brasil, é igualmente possível que a presença dos mesmos elementos cause tal estranhamento. É importante

<sup>2</sup> No original: *To understand this word is to understand the practice in our culture of having three meals a day, at more or less conventionally established times of the day, and for one of these meals to be the one which is eaten early in the day, after a period of sleep, and for it to consist of a somewhat unique menu (the details of which can vary from community to community).*

ressaltar que o *frame* é o mesmo, uma vez que ele está ainda retratando a primeira refeição do dia, feita no início da manhã, após um período de sono, dentre outros aspectos que o caracterizam.

O conceito de *frame* proposto por Fillmore no fim da década de sessenta passa a subsidiar os chamados *papeis semânticos*, descrevendo as estruturas formadas pelos verbos da língua inglesa, com base na valência semântica ao invés de *elementos distribucionais* tais quais os *traços desubcategorização estrita* e as *restrições seletivas*. A valência semântica, por sua vez, considerará o papel semântico dos argumentos do verbo – a começar pelo sujeito, considerado argumento *externo* – ao contrário da valência sintática, que se baseia na bipartição *sujeito-predicado*.

Seguindo o pressuposto de que o estudo da Semântica de Frames é realizado a partir de dados reais de uso, veremos agora de que forma se dá a orientação empírica da teoria de Fillmore.

### 3 O empirismo metodológico da teoria de Fillmore

Conforme vimos na seção anterior, a Semântica de Frames, enquanto teoria pertencente ao escopo da Linguística Cognitiva, está comprometida com o empirismo no sentido experiencial, considerando-se, contudo, a experiência num sentido lato, que envolve aspectos culturais, históricos e sociais. Além disso, é possível verificar que essa teoria se utiliza de metodologia empírico-descritiva, visto que, em seu viés aplicado, assume um compromisso com a investigação baseada em *corpora* para descrever a linguagem – aspecto que será tema de nossa próxima seção. Desse modo, este capítulo visa a tratar da orientação empírica da Semântica de Frames no que tange aos procedimentos metodológicos adotados a partir desse aporte teórico.

Os textos fundadores publicados por Fillmore entre os anos 1970 e 1980 ainda não chegam a delinear propostas metodológicas relativas à

Semântica de Frames. Em *Analternativetochecklisttheoriesofmeaning* (FILLMORE, 1975), o teórico traz os primeiros exemplos que ilustram a pertinência da teoria para os estudos linguísticos em diferentes contextos, tais como coerência textual, aquisição de linguagem e fronteiras de categorização. Ainda na década de 1970, o artigo *Frame semanticsandthenatureoflanguage* (FILLMORE, 1976), com o objetivo de abordar aspectos da Semântica de Frames que poderiam contribuir para o modo como se compreende a linguagem humana, introduz um aspecto importante para compreendermos a visão do teórico quanto à forma de se investigar fenômenos linguísticos – a noção de contexto, entendido como o conjunto de circunstâncias que embasam uma sentença ou uma experiência que permite o entendimento de uma palavra. Nessa passagem, Fillmore rejeita enfaticamente a abordagem gerativista por desconsiderar as relações entre o falante e as questões contextuais atinentes ao processo comunicativo. Nas palavras do autor,

O linguista tem bons motivos para se preocupar com a descrição do contexto [...] como sendo mais uma complexidade do seu trabalho ao descrever a língua, mas infelizmente tem sido fácil para os linguistas pensar no conhecimento do contexto como apenas mais uma complexidade adicional para o usuário da língua<sup>3</sup>. (FILLMORE, 1976, p. 23).

Dessa forma, temos claramente uma refutação de abordagens intuitivas que desconsiderem o real uso linguístico dos falantes. Esses preceitos são retomados e aperfeiçoados na década seguinte, quando o autor consolida a Semântica de Frames como um programa de análise do significado a partir de situações reais de comunicação (FILLMORE, 1982; 1985).

É no trabalho de 1992, o qual descreve um estudo realizado por Fillmore juntamente com Atkins, que aparece um primeiro delineamento quanto à

orientação metodológica da Semântica de Frames. Em *Toward a Frame-BasedLexicon: The Semanticsof RISK and its Neighbors*, os autores indicam seu objetivo de criar um recurso lexicográfico através da aplicação da teoria, realizado posteriormente através do projeto FrameNet. Pela primeira vez, traz-se uma análise mais minuciosa e prática quanto à proposta de Fillmore, na qual se levam em conta as descrições valenciais que estruturam os *frames*. Esse experimento é conduzido a partir de concordâncias retiradas de um *corpus* de 25 milhões de palavras, compilado pela American PublishingHouse for theBlind. Para a análise do lexema *risk* e das respectivas unidades lexicais, foram extraídas 1770 sentenças<sup>4</sup>. Através desse experimento, Fillmore e Atkins evidenciam que o fato de um dicionário separar os sentidos de acordo com as paráfrases possíveis, baseados em diferenças de padrões gramaticais, não permite que o usuário verifique processos metafóricos e metonímicos relacionados a determinadas unidades lexicais (FILLMORE; ATKINS, 1992), daí a necessidade de se criar um recurso organizado em torno de *frames*.

A Plataforma FrameNet entra em operação em 1997, consolidando-se como contraparte aplicada da teoria de Fillmore e, assim, delineando minuciosamente os procedimentos metodológicos atinentes à abordagem, pautados na análise a partir de evidências empíricas fornecidas por *corpora* de estudo. Desse modo, conforme sintetizam Atkins, Fillmore e Johnson (2003), a FrameNet é um recurso lexicográfico computacional que descreve propriedades semânticas e sintáticas de palavras em língua inglesa, a partir de *corpora* eletrônicos, apresentando essa informação através de uma plataforma online. Assim, o desenvolvimento do projeto embasa-se na análise de um vasto conjunto de textos autênticos de língua inglesa, do qual são extraídas as sentenças que ilustram as unidades lexicais.

<sup>3</sup>The linguist has good reasons for regarding the description of context [...] as an added complexity in his job of describing a language; but unfortunately it has been easy for linguists to think of knowledge of context as an added complexity for the language-user as well.

<sup>4</sup> Os autores ressaltam que, em virtude de a metodologia ainda não estar consolidada, informações concernentes a cada uma das concordâncias foram sendo registradas a partir de uma base de dados fornecida pela Universidade de Berkeley.

Fillmore e Baker (2010) explicam que o método de pesquisa adotado pelo projeto

[...] consiste em encontrar grupos de palavras cujas estruturas de *frames* podem ser conjuntamente descritas, devido ao fato de partilharem padrões e contextos esquemáticos comuns de expressões que podem se combinar com elas para formar frases ou sentenças maiores. Tipicamente, as palavras que partilham de um mesmo *frame* podem ser usadas como paráfrases umas das outras. As propostas gerais do projeto buscam oferecer descrições confiáveis de propriedades combinatórias sintáticas e semânticas de cada palavra do léxico e reunir informações sobre modos alternativos de se expressar conceitos dentro de um mesmo domínio conceptual. (FILLMORE; BAKER, 2010, p. 321, grifo nosso).<sup>5</sup>

Assim, o trabalho da FrameNet inclui desenvolvimento do *frame* semântico, a extração de *corpus* e anotação de sentenças exemplo. O desenvolvimento do *frame* tem como etapas a caracterização informal do tipo de entidade ou situação representada pelo *frame*, seleção dos elementos de *frame* e a construção de listas de palavras que pertencem ao *frame*. A extração de *corpus* tem por objetivo a verificação dos contextos sintáticos e semânticos dos constituintes das sentenças, anotando-se os exemplos de acordo com os Elementos de Frame estabelecidos (FILLMORE; JOHNSON; PETRUCK, 2003).

Através do maquinário da FrameNet, a Semântica de Frames estabelece-se como teoria que parte de evidência proveniente de *corpora* para conduzir seus procedimentos de análise. Conforme explicam Ruppenhoffer et al. (2010), são os dados do *corpus* que indicam agrupamentos de unidades lexicais, as quais constituem evocadores de um mesmo *frame*, estabelecendo-se a rejeição de métodos introspectivos de análise linguística. Fillmore e Baker (2010) reiteram que a teoria implica uma pesquisa *empírica, cognitiva e etnográfica* por

<sup>5</sup> "The method of inquiry is to find groups of words whose frame structures can be described together, by virtue of their sharing common schematic backgrounds and patterns of expressions that can combine with them to form larger phrases or sentences. In the typical case, words that share a frame can be used in paraphrases of each other. The general purposes of the project are both to provide reliable descriptions of the syntactic and semantic combinatorial properties of each word in the lexicon, and to assemble information about alternative ways of expressing concepts in the same conceptual domain."

natureza, destacando que a exploração de *corpora*, além de rejeitar análises baseadas na intuição, potencializa o uso da abordagem no que tange a técnicas computacionais.

A seção a seguir aborda a Linguística de Corpus, aporte metodológico que embasa o presente trabalho.

#### 4 Linguística de Corpus

A *Linguística de Corpus (LC)*<sup>6</sup> é uma subárea da Linguística que surge na segunda metade do século XX, sendo seu marco inicial a compilação do *corpus* Brown, em 1964 (SARDINHA, 2000). O contexto que propiciou seu surgimento foram insatisfações relacionadas à linguística chomskyana, mais precisamente à natureza intuitiva dos dados usados pelos seguidores dessa corrente teórica. Em outras palavras, Chomsky e seus seguidores desprezavam dados empíricos porque seriam irregulares (CHOMSKY, 1957), além da crença que a intuição de um falante nativo *ideal* seria o suficiente para descrever as regras de *boa formação* das sentenças de uma língua. Nesse contexto, o Brown surgiu da necessidade de estudar a língua inglesa em situações reais de uso, ao invés de situações hipotéticas. Vale lembrar que a quantia de um milhão de palavras que integravam esse *corpus* haviam sido computadas através de cartões de papel furados, que eram lidos por um computador *main frame*, que ocupava uma sala inteira (SARDINHA, 2000, p. 324).

Além de seu caráter empirista, é válido explicitar o que se entende pela palavra *corpus*, através de uma definição de LC proposta por Sardinha (2000):

A LC ocupa-se da coleta e exploração de *corpora*, ou conjunto de dados linguísticos textuais que foram coletados criteriosamente com o propósito de servirem para a pesquisa de uma determinada língua ou variedade linguística. Como tal, dedica-se à exploração da linguagem através de evidências empíricas, extraídas por meio de computador" (p. 325).

<sup>6</sup>Doravante "LC".



Essa definição torna clara a crítica endereçada aos gerativistas, que não se valiam de situações reais de uso da língua. Também é válida a conceituação proposta por Santos (2008), a qual se refere à área como *Linguística com corpos* (p. 46), alegando que um corpo não é um objeto de estudo, mas uma *ferramenta* de estudo. Discordâncias conceituais à parte, cabe ressaltar aquilo que é mais notável nas definições: o uso de coleções de textos no estudo para estudo de línguas e suas variedades por meio de ferramentas eletrônicas.

Há duas diferenças fundamentais entre os primeiros *corpora* e os *corpora* atuais: não eram eletrônicos, pois não havia computadores para processá-los, e seu uso enfatizava o ensino de línguas (SARDINHA, 2000).

O computador foi uma ferramenta fundamental para o estabelecimento da LC. Até então, o manuseio de textos era feito por seres humanos, o que tornava as análises imprecisas ou impraticáveis, devido à enorme quantidade de textos (SARDINHA, 2000). A invenção dos já citados computadores *main frame*, bem como a invenção de novas mídias, principalmente a fita magnética, mudaram esse quadro, facilitando o tratamento e a manutenção dos *corpora*.

Em título de curiosidade, vale a pena citar alguns *corpora* de referência, comumente citados quando o assunto é LC. Sardinha (2000), em âmbito internacional, além do já citado, o pioneiro Brown, também cita o Birmingham (360 milhões de palavras, do tipo aberto, ou seja, é constantemente alimentado) e o BNC (British National Corpus, o primeiro a atingir 100 milhões de palavras e que pode ser adquirido).

Na seção seguinte, traremos alguns conceitos fundamentais sobre o fenômeno linguístico que este trabalho propõe-se a estudar, a polissemia.

#### 4.1 SketchEngine

O *Sketch Engine*<sup>7</sup> é uma ferramenta que permite a criação, a manipulação e o estudo de

*corpora* em diversos idiomas, como o inglês, o italiano, o português, o alemão, e o japonês. Para usar essa ferramenta, o usuário pode optar por uma conta gratuita, que expira em trinta dias, ou pode comprar licenças que variam de seis meses a um ano de duração.

Através das opções de busca, o usuário é levado às *concordâncias*, que consistem em linhas baseadas em frações de texto em que a palavra ou expressão pesquisada (chamada de *node word*, ou, *palavra nó*) aparece em destaque, bem como seu cotexto (porções de texto que rodeiam a palavra nó), como se observa na imagem abaixo, em que as concordâncias mostram o comportamento do verbo *marcar* em nosso *corpus* de estudo:



O *Sketch Engine* diferencia-se de outros concordanceadores como o *AntConc* e o *WordSmithTools* em alguns aspectos: primeiramente, oferece *corpora* já compilados para seus usuários – nas outras ferramentas, a adição dos textos do estudo será feita pelo próprio usuário, o que nem sempre é uma tarefa simples, já que, para o uso efetivo dos recursos das ferramentas, os textos devem obedecer a certos requisitos de formalização, conforme nos referimos na descrição do *corpus*. Em segundo lugar, além das concordâncias e da lista de palavras (*wordlist*, recurso que enumera as palavras mais frequentes nos *corpora*), presentes nas demais ferramentas, o *Sketch Engine* possui recursos como *word sketch*,

<sup>7</sup> SKETCH ENGINE. *Sketch Engine features*. EastSussex, 2014. Disponível em:

<<http://www.sketchengine.co.uk/documentation/wiki/Website/Features>>. Acesso em: 25 ago. 2014.

que consiste em uma tabela em que são detalhados os padrões valenciais da palavra *nó*, isto é, esse recurso explicita todas as combinações sintáticas nas quais a palavra *nó* se realiza no *corpus* pesquisado.

## 5 Análise e resultados

### 5.1 A identificação de unidades lexicais polissêmicas a partir do *Sketch Engine*

O Sketch Engine possibilitou a identificação de unidades lexicais (ULs) polissêmicas. Através de uma consulta simples ao recurso *concordance*, percebemos que, em relação ao verbo *tocar*, por exemplo, era possível notar seu comportamento polissêmico com base no contexto em que esse verbo ocorre:

o , sem chances para Diogo=Silva , que ainda **tocou** na bola ! Golaco ! Segunda bucha do Pirata ! Grêmio edor. Com o placar garantido , o time soube **tocar** a bola , administrar a vantagem e garantir os três no marcador e deu mais tranquilidade para o **tocar** a bola e administrar a vitória . Aos 36 minutos , para Herrera pelo=meio da zaga. O argentino **tocou** com perfeição no canto do goleiro do Juventude . defensores e a bola veio para Ricardinho , que **tocou** com categoria na saída do goleiro do time do interior

Nos itens destacados acima, é possível ver dois sentidos do verbo *tocar*: em [...] *o time soube tocar a bola* [...], é evocado o *frame* Passe, uma vez que se trata da transferência da posse de bola entre jogadores da mesma equipe. No segundo caso, [...] *o argentino tocou com perfeição no canto do goleiro do Juventude*, trata-se do *frame* Chute, pois a bola é projetada em direção ao gol adversário.

Destacamos também o recurso Word Sketch, que se mostrou muito útil no estudo das ULs polissêmicas, uma vez que permite, conforme já referimos, o acesso a todas as realizações sintáticas da palavra pesquisada:

tocar		Futebol freq = 926 (873.1 per million)	
<b>modifier</b> 90 0.9	<b>pp_para</b> 106 8.2	<b>sujeito</b> 103 1.5	<b>pp_de</b> 59 2.0
nem 3 9.52	fundo 17 10.67	pisano 2 9.29	calcanhar 13 11.75
ainda 28 9.31	meio 13 9.77	artilheiro 2 8.12	letra 7 11.21
de=pé 2 9.17	rede 22 9.46	livre 2 8.06	cabeça 30 10.27
à=direita 2 9.01	rei 2 8.86	meia 3 7.95	bico 4 10.22
por=cima 2 8.63	centroavante 3 8.32	volante 4 7.88	
bem 10 8.36	<b>gol</b> 25 7.23	centroavante 2 7.75	<b>pp_com</b> 32 2.3
só 7 8.13	linha 4 7.17	atacante 10 7.66	categoria 11 11.02
bastante 3 8.09	brasileiro 2 6.25	atleta 2 7.6	classe 2 10.54
apenas 4 7.95	atacante 3 5.92	argentino 2 7.42	mão 7 9.89
mal 2 7.29	escanteio 2 5.92	seleção 2 7.28	precisão 2 9.75
mais 7 7.08	área 4 5.13	bola 21 7.14	pé 3 9.97
também 3 6.86		goleiro 6 6.41	ponta 2 7.23
não 2 3.82		grêmio 4 6.32	
			e ou 24 1.0

Na imagem acima, temos o resultado obtido para a *sketch* do verbo *tocar*. Em destaque, percebe-se a realização desse verbo complementada pela preposição *para*, totalizando 106 ocorrências, dentre as quais 25 envolvem o objeto *gol* – *tocar para o gol*. Outros sentidos de *tocar* emergem através do objeto *bola*:

objeto	100	1.1
<b>bola</b>	<b>75</b>	<b>8.98</b>
<b>mão</b>	<b>3</b>	<b>8.31</b>
10	2	8.09
travessão	3	7.92
trave	2	6.4
camisa	2	6.36

*Tocar a bola* pode remeter tanto a um passe como a um chute, mas *tocar a mão* envolve outro sentido e, consequentemente, outro *frame*:

Word sketch item 3 (2.8 per million)

aproveitando infração cometida por Müller -R o atleta **tocou a mão** na bola na entrada da área e , de uma forma mas a jogada foi anulada . Mais=uma=vez , Aloísio **tocou a mão** em cruzamento para a área . A bola entrou no chutou muito perto=de gol . Aos 16 min , Aguilár **tocou a mão** na bola dentro=de área e o árbitro marcou

A consulta às três ocorrências em que *mão* ocorre como objeto do verbo *tocar* revela que o sentido evocado remete a uma infração, visto que envolve a ideia de tocar a bola com as mãos, conforme mostram as sentenças acima. Este é mais um ponto a ressaltar em relação ao Sketch Engine: seus recursos trabalham de forma integrada, já que, enquanto se utiliza o Word Sketch, é possível acessar as concordâncias clicando nos *links* dispostos na frequência dos itens no *corpus*, conforme demonstramos acima.

### 5.2 O recurso *Word sketch* e a construção de frames: identificando eventos a partir da unidade lexical marcar

O Word Sketch, como pudemos ver, oferece a possibilidade de visualizar as diferentes combinações



de uma determinada palavra com outras, apresentando mapeamento no nível sintático-semântico. De forma a identificar os *frames* instanciados a partir das orientações da FrameNet, é pertinente olharmos para os chamados elementos de *frame* nucleares de cada evento em potencial, visto que estes “instanciam um elemento conceptualmente necessário de um *frame* tornando o *frame* único e diferente de outros *frames*”. (RUPPENHOFER, 2010, p. 19). Tais elementos devem ser sempre especificados de alguma forma nas sentenças, evidenciando, por exemplo, a estrutura argumental de um verbo e diferenciando-se de itens acessórios, como os elementos de *frame* que designam noções de tempo e espaço.

Assim, para verificarmos possíveis diferenças entre os eventos, atentamos primeiramente para os sujeitos instanciados com o verbo *marcar*, conforme a respectiva coluna do Word Sketch:

**marcar** Futebol freq = 2,680 (2,527.0 per million)

modifier	543	1.6	objeto	840	2.7	sujeito	459	2.0	pp_a+o	141	2.7	pp_para	106	2.4
quase	137	11.86	gol	441	11.19	árbitro	87	11.45	min	32	10.21	19h30	6	10.61
contra	16	9.84	pênalti	68	10.6	juiz	27	10.52	41	5	9.7	18h30	4	10.07
bem	35	9.63	impedimento	39	10.3	arbitragem	19	9.97	22	6	9.59	16h20	3	9.75
novamente	19	9.59	golado	34	10.18	terceiro	9	8.9	18	5	9.45	21h	3	9.75
também	30	9.56	segundo	16	8.87	assistente	7	8.83	26	4	9.27	dia	11	9.72
forte	21	9.12	estrela	14	8.87	atacante	25	8.7	28	4	9.2	21h50	3	9.57
não	80	9.01	falta	35	8.76	centroavante	6	8.22	min.	6	9.19	terça-feira	4	8.74
por-pouco-não	9	8.93	retorno	9	8.35	relógio	4	8.14	36	3	8.83	domingo	8	8.4
logo	10	8.86	presença	8	8.19	zagueiro	11	8.1	24	3	8.79	sábado	5	7.97
mais	30	8.84	tento	8	8.11	auxiliar	3	7.65	12	3	8.56	quarta-feira	4	7.93
já	18	8.72	reencontro	7	8.07	meia-atacante	3	7.57	minuto	23	8.51	time	13	6.28
ainda	20	8.43	quarto	7	7.95	brasileiro	7	7.56	equipe	6	5.85	equipe	6	5.85
mais-uma-vez	7	8.39	penalidade	6	7.73	tipo	3	7.5	pp_por-o	111	3.7			
quando	10	8.28	um	9	7.69	x	3	7.47	equilíbrio	5	10.02	pp_em-o	73	0.6
apenas	9	8.23	dois	12	7.69	partida	16	7.44	vez	5	9.28	acréscimo	6	9.82
de-novo	5	8.07	volta	6	7.68	francês	3	7.41	meia	7	9.15	fim	5	8.92
cedo	4	7.87	saída	7	7.64	impedimento	3	7.28	francês	3	8.73	inicio	3	8.08
mesmo	5	7.82	vez	5	7.37	Jogador	7	7.19	chileno	3	8.67	tempo	11	7.19
primeiro	5	7.62	o	4	7.26	camisa	5	7.15	artilheiro	3	8.66	lance	3	6.88
respectivamente	3	7.48	despedida	4	7.25	time	24	7.07	atacante	13	8.03	final	6	6.85
pela-primeira-vez	3	7.38	toque	4	7.03	grêmio	8	7.03	zagueiro	6	7.66	competição	3	6.56
finalmente	3	7.38	tiro	3	6.63	flamengo	4	6.77	brasileiro	4	7.24	minuto	5	6.36
pelo-menos	3	7.37	três	3	6.29	jogo	14	6.67	árbitro	3	7.2	vitória	4	6.1
muito	7	7.35	min	3	5.96	equipe	8	6.12						
que	3	7.28	minuto	3	5.15	placar	3	5.67						

A partir disso, verificamos que é possível distinguir duas grandes categorias com o verbo marcar. A primeira refere-se à arbitragem (sujeitos *árbitro*, *juiz*, *arbitragem*), indicando um *frame* que abrange as decisões tomadas por esses participantes ao longo da partida, conforme mostram as sentenças

caiu após carrinho de Abidal . O *árbitro marcou* pênalti e o veterano Blanco converteu a individual pela direita , caiu , mas *árbitro marcou* apenas tiro de meta para o Ajax . Aos 30 que girou e bateu forte , mas o *árbitro marcou* impedimento do Demolidor . Aos 43 , Montillo

cujo sujeito é *árbitro*:

A segunda grande categoria mostrou que os eventos relacionados a jogadores subdividem-se em dois *frames*. O primeiro deles refere-se ao ato de se marcar um jogador ou um grupo de jogadores, consoante ilustrado pelo exemplo a seguir, com o sujeito *atacante*:

por morder Ivanovic , o *atacante* uruguaio *marcou* duas vezes e foi o grande nome da vitória e 60 mil no domingo . O *atacante* Hernane *marcou* o segundo gol do Flamengo na derrota por no=meio=de semana . O *atacante* brasileiro *marcou* um gol , mas perdeu um pênalti e viu o

O segundo *frame* dessa categoria denota a atividade de marcação em campo, quando os jogadores buscam controlar os movimentos dos adversários, conforme o exemplo com o sujeito *time*:

de defesa . O *time* do técnico Caio=Júnior *marcou* os celestes de forma compactada , mas com espalmado a bola para escanteio . O *time* carioca *marcava* a saída de bola do Inter e a equipe dirigida Fernando=Prass fez defesa tranquila . O *time* da casa *marcava* forte para seguir pressionado . Aos dois

A partir dessa análise, foi possível efetuar um mapeamento das realizações sintático-semânticas referentes a cada um dos três *frames* instanciados, identificando-se, posteriormente, os demais elementos que compõem cada evento. Esse estudo resultou na criação de três *frames* relacionados ao verbo *marcar*(*Marcação*, *Marcar\_Falta* e *Marcar\_Gol*), que compõem os cenários do Dicionário Field:

### Marcar Falta

bola parada  
falta

**Definição:**  
Ação em que o árbitro penaliza o jogador diante de uma falta cometida.



**Cenários relacionados**  
 Cobrança de Falta  
 Cobrança de Pênalti  
 Funções do juiz  
 Reações da torcida

### Marcação

▶

**Definição:**  
Ato de um ou mais jogadores de um time vigiarem e procurarem controlar os movimentos de jogadores do time adversário.



**Palavras do cenário**

apertar a marcação
cercar
dividida
escapar
fechar a marcação
marcação homem a homem
jogo de corpo
marcação
marcar

**Cenários relacionados**

Ataque
Defesa
Táticas

### Marcar Gol

▶

**Definição:**  
Evento em que um jogador acerta a bola entre as traves e abaixo do travessão da meta do time adversário. Após a validação do árbitro, o gol resulta em um ponto no placar.



**Palavras do cenário**

abrir o placar
descontar
desempatar
desencantar
fazer gol
gol
gol de cabeça
gol de falta
gol de letra
gol de pênalti
gol de placa
gol de primeira
gol olímpico
golapço
golpear
levar gol
marcar gol
sofifer gol

## 6 Considerações finais

O presente trabalho teve por objetivo mostrar de que forma a abordagem teórico-metodológica da Semântica de Frames foi pertinente à construção do dicionário Field. É de extrema pertinência ressaltar o tratamento dos *corpora* a partir do software, considerando-se não só o tamanho dessas coletâneas, mas também a possibilidade que o programa oferece de explorar propriedades linguísticas desses textos. Com isso, torna-se possível não só a identificação das palavras com maior incidência nos *corpora*, mas também a forma como essas palavras se inter-relacionam, demonstrando-se também a importância da Linguística de Corpus como metodologia de pesquisa. Além da pertinência da ferramenta para o tratamento

de casos polissêmicos, destacamos a relevância do Sketch Engine para a construção dos *frames*, haja vista que as *word sketches* mostram o comportamento sintático-semântico dos evocadores de *frame* através de evidências empíricas.

Ressaltamos também que o projeto Field, em virtude do compromisso com a criação de um dicionário para o público leigo, implicou a adaptação dos procedimentos metodológicos em voga na plataforma FrameNet, visto que a organização das informações deveria consistir em uma interface amigável, mostrando apenas aquilo que é relevante ao consulente não especializado. Nesse sentido, constatamos que tal apropriação é consideravelmente promissora no que tange à prática lexicográfica, considerando-se também que a Semântica de Frames, até então, havia sido predominantemente utilizada para a construção de recursos lexicográficos direcionados a especialistas. Essa adaptação reforça, portanto, a pertinência da teoria de Fillmore para a prática lexicográfica também para o consulente não especializado.

Finalmente, insta salientar que a Semântica de Frames, inserida no escopo das teorias da Linguística Cognitiva que se utilizam de métodos empíricos, expressa-se empiricamente de duas formas distintas, ambas igualmente relevantes à descrição lexicográfica. A primeira delas diz respeito à metodologia propriamente dita, pautada no maquinário da Linguística de Corpus e nas possibilidades que os estudos embasados na extração de *corpora* trazem ao fazer lexicográfico, tornando a prática mais sistematizada e permitindo, portanto, que recursos como o dicionário Field constituam dicionários confiáveis, que disponibilizam dados lexicograficamente relevantes. O segundo aspecto perpassa todas as teorias inseridas no escopo da Linguística Cognitiva, independentemente do aporte metodológico utilizado: o empirismo concernente à valorização da experiência, em suas múltiplas dimensões, dos sujeitos em interação com outros membros de sua comunidade. Assim, um dicionário baseado em *frames* como o Field, ao primar por uma organização que exhibe os *frames*

evocados pelos verbetes, separando as unidades lexicais consoante essas distinções, não apenas possibilita aos usuários consultas mais contextualizadas, mas também mostra como o léxico, conforme o conhecimento dos falantes e suas experiências socioculturais, serve como guia para o significado enquanto produto de seus processos cognitivos.

## Referências

- ATKINS, S.; FILLMORE, C. J.; JOHNSON, C. R. Lexicographic Relevance: Selecting Information from Corpus Evidence. *International Journal of Lexicography*, Oxford, v. 16, n. 3, p. 251-281, 2003.
- BERBER SARDINHA, T. Linguística de Corpus: histórico e problemática. *DELTA* [online]. (2000) Vol.16, n.2, p. 323-367. Disponível em <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0102-44502000000200005&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-44502000000200005&lng=en&nrm=iso)> Acesso em: 4 jul. de 2013.
- CRUSE, Alan. *A glossary of semantics and pragmatics*. Edimburgo: Edinburgh University Press, 2006.
- FAUCONNIER, Gilles. Cognitive linguistics. In: NADEL, Lynn. *Encyclopedia of cognitive science*. Londres: Macmillan, 2003.
- FILLMORE, C. J. An alternative to checklist theories of meaning. In: Cogen, C. et al. (Eds.). *Proceedings of the First Annual Meeting of the Berkeley Linguistics Society*. Berkeley: Berkeley Linguistics Society, 1975. p. 123-31.
- \_\_\_\_\_. Frame semantics and the nature of language. In: *Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech*, Vol. 280, 1976. p. 20-32.
- \_\_\_\_\_. *Frame Semantics*. In: *Linguistics in the Morning Calm*. Seoul, Hansinh Publishing Co., 1982. p.111-137.
- \_\_\_\_\_. Frames and the semantics of understanding. In: *QuadernidiSemantica*, vol. 6, n. 2, 1985. p.222-254.
- FILLMORE, C. J.; ATKINS, S. Describing Polysemy: the Case of “Crawl”. In: RAVIN, Y.; LEACOCK, C. *Polysemy: theoretical and computational approaches*. Oxford: Oxford University Press, 2000. p. 91-110.
- FILLMORE, C. J.; BAKER, C.A frames approach to semantic analysis. In: HEINE, B.; NARROG, H. (Eds.). *The Oxford Handbook of Linguistic Analysis*. New York: Oxford University Press, 2010. p. 313-339.
- FILLMORE, C. J.; JOHNSON, C. R.; PETRUCK, M. R. L. Background to FrameNet. *International Journal of Lexicography*, Oxford, v.16, n.3, p. 235-250, 2003.
- RUPPENHOFER, J. et al. *FrameNet II: Extended Theory and Practice*. Berkeley, California: International Computer Science Institute, 2010. Disponível em: <<https://framenet2.icsi.berkeley.edu/docs/r1.5/book.pdf>>. Acesso em: 20 ago. 2013.
- TALMY, L. Foreword. In: GONZALEZ-MARQUEZ, M. et al. (Orgs.) *Methods in Cognitive Linguistics*. Amsterdam: John Benjamins, 2003.