



Programa de
Pós-Graduação em
Linguística

Maria Cristina Andrade dos Santos

**PROPOSTA DE APLICAÇÃO DO MODELO DE CLASSES DE OBJETO PARA A
CONSTRUÇÃO DE UMA ONTOLOGIA DO DOMÍNIO DO FUTEBOL**

São Carlos

2010



Universidade Federal de São Carlos

PROPOSTA DE APLICAÇÃO DO MODELO DE CLASSES DE OBJETO PARA A
CONSTRUÇÃO DE UMA ONTOLOGIA DO DOMÍNIO DO FUTEBOL

UNIVERSIDADE FEDERAL DE SÃO CARLOS
Centro de Educação e Ciências Humanas
Programa de Pós-Graduação em Linguística
Linha de Pesquisa: Linguagem Humana e Tecnologia

**PROPOSTA DE APLICAÇÃO DO MODELO DE CLASSES DE OBJETO PARA A
CONSTRUÇÃO DE UMA ONTOLOGIA DO DOMÍNIO DO FUTEBOL**

Dissertação apresentada ao Programa de Pós-Graduação em Linguística da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do Título de Mestre em Linguística.

Orientador: Prof. Dr. Oto Araújo Vale.

**Ficha catalográfica elaborada pelo DePT da
Biblioteca Comunitária da UFSCar**

S237pa

Santos, Maria Cristina Andrade dos.

Proposta de aplicação do modelo de classes de objeto para a construção de uma ontologia do domínio do futebol / Maria Cristina Andrade dos Santos. -- São Carlos : UFSCar, 2013.
89 f.

Dissertação (Mestrado) -- Universidade Federal de São Carlos, 2010.

1. Linguística – processamento de dados. 2. Linguística de *corpus*. 3. Linguagem humana e tecnologia. 4. Análise sintática. I. Título.

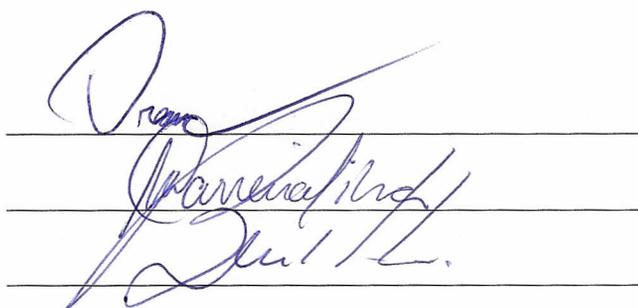
CDD: 410.285 (20^a)

BANCA EXAMINADORA

Prof. Dr. Oto Araújo Vale

Profª. Dra. Maria Cristina Parreira da Silva

Prof. Dr. Bento Carlos Dias da Silva



Three horizontal lines with handwritten signatures in blue ink. The first signature is 'Oto Araújo Vale', the second is 'Maria Cristina Parreira da Silva', and the third is 'Bento Carlos Dias da Silva'.

À minha mãe, Ana.
(*In memoriam*)

AGRADECIMENTOS

Ao meu orientador, mestre e amigo Prof. Dr. Oto Araújo Vale pelo constante incentivo, sempre indicando a direção a ser tomada nos momentos de maior dificuldade. Agradeço principalmente pela confiança em mim depositada.

Ao Prof. Dr. Bento Carlos Dias-da-Silva e à Profa. Dra. Lucia Rino, pela leitura, apontamentos e sugestões durante o exame de qualificação.

Ao Prof. Dr. Bento Carlos Dias-da-Silva e à Profa. Dra. Maria Cristina Parreira da Silva, pela leitura, apontamentos e sugestões durante o exame de defesa.

À Profa. Dra. Gladis Maria B. de Almeida, à Profa. Dra. Flávia B. M. Hirata-Vale e ao Prof. Dr. Bento Carlos Dias-da-Silva que, em suas aulas, permitiram espaços para discussão de algumas questões relevantes e importantes desenvolvidas neste trabalho.

À Profa. Dra. Margarita Correia pelas sugestões, apoio e companheirismo.

Ao meu pai, por acreditar sempre em minha vitória, independente da escolha.

Aos meus Irmãos Cláudia e Gabriel, pela alegria que me proporcionaram todas as (poucas) vezes que eu voltei pra casa. Obrigada pela admiração e por acreditarem nesta conquista.

Aos meus amigos Mc' Hanne, Dayse, Cláudia, Renata, Jacque, Débora, Cacilda, Rose, Paty, Edgar, Thali e Barbie pela ajuda atribuída sempre que necessário. Obrigada pelo carinho, incentivo e pelos momentos de descontração.

Durante boa parte do curso de mestrado contei com os financiamentos concedidos pela FAPESP, que facilitaram a realização deste trabalho.

A todas as pessoas que contribuíram direta ou indiretamente para a realização desta pesquisa.

RESUMO

Este trabalho tem como objetivo aplicar o modelo de classes de objeto de Gaston Gross (1994) à construção de ontologias de domínio, no caso uma ontologia do domínio do futebol. As classes de objetos são classes semânticas selecionadas a partir das restrições argumentais de cada predicado. A construção de ontologias, de modo geral, segue uma metodologia baseada na tradição semântica, pela qual são estabelecidas classes conceituais a partir do julgamento do criador. A proposta deste estudo foi chegar às classes semânticas por meio da aplicação de critérios sintáticos que possam ser reproduzidos por outros pesquisadores. Para isso, utilizamos uma gramática de operadores e argumentos, tendo, como unidade mínima de significação, a frase. Assim, cada item lexical deve ser analisado em função do ambiente sintático em que se encontra, pois as relações de co-ocorrência das palavras determinam seu significado.

Palavras-chave: Classes de objeto. Ontologia. Futebol.

ABSTRACT

This work aimed at developing an ontology of the football domain based on the model of classes of objects proposed by Gaston Gross (1994). The construction of an ontology in general follows a methodology based on the semantic tradition through which conceptual classes are established according to the conceiver's judgment. The proposal of this study was to arrive at semantic classes through the application of syntactic criteria which may be reproduced by other researchers. In order to do that, we use the theory proposed on Harris (1976), who considers the sentence as the minimal unit of meaning and a grammar of operators and arguments. The classes of objects are semantic classes selected from the argumental restrictions of each predicate.

Key words: Classes of object. Ontology. Soccer.

LISTA DE FIGURAS

Figura 1 - Triângulo de Ogden e Richards adaptado	30
Figura 2 - Dinâmica do processo de construção de um SPLN.....	31
Figura 3 - Triângulo de Ogden e Richards adaptado II	41
Figura 4 - Classificação do corpus de acordo com o tamanho.....	45
Figura 5 - Concordância de <i>bola</i> seguida de um verbo.....	48
Figura 6 - Word Lists do Unitex.....	49
Figura 7 - Conceito BOLA	71

LISTA DE QUADROS

Quadro 1 - Traços semânticos	21
Quadro 2 - Lista de códigos dos dicionários do Unitex	47
Quadro 3 - Unidades Lexicais mais recorrentes.....	50
Quadro 4 - Classe de objeto <bola>.....	55
Quadro 5 - Exemplo classe de objeto <clube>.....	55
Quadro 6 - Classe de objeto <jogador>	55
Quadro 7 - Classe de objeto <jogador>	56
Quadro 8 - Classe de objeto <acessório>	56
Quadro 9 - Classe de objeto <campo>.....	57
Quadro 10 - Classe de objeto <divisão>	57
Quadro 11 - Exemplo classe de objeto estádio	58
Quadro 12 - Classe de objeto <falta>.....	58
Quadro 13 - Classe de objeto <parte do corpo>	59
Quadro 14 - Classe de objeto <partes do campo>.....	59
Quadro 15 - Classe de objeto parte do gol.....	60
Quadro 16 - Classe de objeto <local do gol>	60
Quadro 17 - Classe de objeto <penalidade/equipe>	61
Quadro 18 - Classe de objeto <penalidade/jogador>	61
Quadro 19 - Classe de objeto <período/jogo>.....	62
Quadro 20 - Classe de objeto <recompensa>.....	62
Quadro 21 - Classe de objeto <tática>.....	63
Quadro 22 - Classe de objeto <uniforme>	63
Quadro 23 - Classe de objeto <função/jogador>.....	64
Quadro 24 - Classe de objeto <torcedor>	65
Quadro 25 - Classes de Objeto Futebol	65
Quadro 26 - Exemplo tabela predicados adjetivais	66
Quadro 27 - Exemplo tabela predicados nominais.....	66
Quadro 28 - Exemplo tabela predicados verbais.....	66

Abreviaturas, símbolos e notações

hum – humano
-hum, – não-humano
humcoll – humano coletivo
inc – inanimado concreto
loc – local

N – Substantivo
N0 – Grupo nominal em posição argumental sujeito.
N1 – Grupo nominal em posição argumental primeiro complemento.
nn – Substantivo neutro
nm – Substantivo masculino
nf – Substantivo feminino
Prep – Preposição
Adv – Advérbio
V – Verbo
Vaux – Verbo auxiliar
Vsup – Verbo suporte

Colchetes angulares e fonte `courier new` – classes de objeto.

Fonte *itálico* – palavras mencionadas.

* – Marca a inaceitabilidade.

.

SUMÁRIO

1	Introdução	11
1.1	Ontologia e ciência da computação	12
1.2	Escolha do domínio.....	13
1.3	Classes de objeto aplicadas à ontologia	14
2	Fundamentos teóricos.....	15
2.1	Classes de Objeto de Gaston Gross.....	15
2.1.1	Definição de classes de objeto	15
2.1.2	Classes de argumentos	16
2.1.3	Classes de predicados	26
2.1.4	Restrições de seleção dos argumentos.....	27
2.1.5	Outros trabalhos	28
2.2	Ontologias e organização do conhecimento	29
2.2.1	Metodologias para a Construção de ontologias.....	32
2.3	Classes Ontológicas versus Classes de objeto.....	38
2.3.1	Linguagem figurada	40
2.4	Linguística de corpus	43
3	Metodologia.....	46
3.1	Etapa 1: Constituição e manipulação do corpus	46
3.2	Etapa 2: Extração das ocorrências utilizadas na descrição	48
3.3	Etapa 3: Aplicação do modelo de Gaston Gross.....	50
4	Resultados	52
4.1	Constituição das classes de objeto	52
4.2	Hierarquia das classes de objeto e ancoragem ontológica	67
	Considerações finais	73
	Referências Bibliográficas	75
	Apêndices.....	79

1 Introdução

Há muito a estrutura de predicado/argumentos tem sido alvo de estudos descritivos. Podemos destacar variações dessa estrutura em Tesnière (noções de actante e valência), Fillmore (modelos de casos) e Harris (noção de operador). Harris (1976) propõe uma gramática baseada em operadores e argumentos. Sua hipótese de base é a de que a informação é transmitida na linguagem por meio de frases simples, definidas assim como a unidade mínima de significação. A frase simples é composta por um predicado acompanhado de seus argumentos. Para o autor, cada item lexical deve ser analisado em função de sua vizinhança, i.e. o ambiente sintático em que ele se encontra, pois as relações de co-ocorrência das palavras determinam seu significado.

A proposta a ser apresentada é a utilização de critérios sintáticos que possibilitem a distribuição das unidades lexicais em classes – as Classes de Objeto (GROSS, 1994; LE PESANT; MATHIEU-COLAS, 1998).

As classes de objeto podem ser definidas como classes semânticas criadas a partir de critérios sintáticos. Cada classe de objeto é definida a partir de uma classe de predicados que selecionam de forma apropriada os argumentos que a compõem, ou seja, cada classe de predicados possui restrições combinatórias que permitem selecionar apropriadamente o argumento que a acompanha (LE PESANT & MATHIEU-COLAS, 1998).

Para Gross (1994) definir o sentido de um predicado implica descrever a totalidade de suas partes. Assim, como hipótese inicial, acreditamos que a delimitação das classes de objeto possibilita a formação de hierarquias que podem compor uma ontologia, pois a partir da criação de classes bem definidas seria possível estabelecer uma relação consistente entre elas. Partindo dessa hipótese, nosso objetivo foi estabelecer as classes de objeto de um domínio específico e demonstrar como as classes podem ser organizadas hierarquicamente, possibilitando assim a ancoragem linguística em uma ontologia. Optamos por desenvolver esta pesquisa com o léxico do domínio futebol, por ser este um domínio em que se poderia colocar à prova a abordagem proposta no trabalho.

1.1 Ontologia e ciência da computação

Para trabalharmos com ontologia, é necessário primeiramente especificar de qual *ontologia* falamos. Na ciência da Computação, as classes ontológicas são definidas a partir de um conjunto de propriedades. Borst (1997) define o termo ontologia como uma especificação formal de uma conceitualização compartilhada. Essa especificação é formal porque deve ser feita por meio de uma representação formal e compartilhada porque deve ter uma visão universal sobre o conhecimento trabalhado.

Sowa (2003) define ontologia formal como:

(...) um conjunto de nomes atribuídos para um conceito e os tipos de relações existentes organizadas em uma ordenação parcial pelo tipo e subtipo de relação. Ontologias formais são ainda distinguidas pela forma como os subtipos são distintos de seus supertipos: uma ontologia axiomatizada distingue subtipos de axiomas e de definições apresentadas em uma linguagem formal, através da lógica ou alguma notação orientada por computador que pode ser traduzida para a lógica, uma ontologia baseada em protótipos distingue subtipos por uma comparação com um membro típico ou protótipo para cada subtipo¹. (SOWA, 2003, [s/p]).

Uma ontologia formal é definida principalmente por sua aplicação: “ontologia formal” refere-se ao grupo de ontologias que são legíveis por máquina. Já Gruber (1996, [s/p]) define ontologia como

(...) uma especificação explícita de uma conceitualização. [...]. Quando o conhecimento de um domínio é representado em um formalismo declarativo, o conjunto de objetos que podem ser representados é chamado de universo de discurso. [...]. Em tal ontologia, definições associam nomes de entidades no universo do discurso (por ex. classes, relações, funções, e outros objetos) com textos que descrevem o que os nomes significam e os axiomas formais que restringem a interpretação e o uso desses termos².

¹ Texto original Sowa (2003, [s/p]): “A formal ontology is specified by a collection of names for concept and relation types organized in a partial ordering by the type-subtype relation. Formal ontologies are further distinguished by the way the subtypes are distinguished from their supertypes: an axiomatized ontology distinguishes subtypes by axioms and definitions stated in a formal language, such as logic or some computer-oriented notation that can be translated to logic; a prototype-based ontology distinguishes subtypes by a comparison with a typical member or prototype for each subtype”.

² Texto original Gruber (1996, s/p) “An **ontology** is an explicit specification of a conceptualization. [] When the knowledge of a domain is represented in a declarative formalism, the set of objects that can be represented is called the universe of discourse. []. In such an ontology, definitions associate the names of entities in the universe of discourse (e.g., classes, relations, functions, or other objects) with human-readable text describing what the names mean, and formal axioms that constrain the interpretation and well-formed use of these terms”.

Segundo Gruber, os conceitos são explicitamente definidos por meio da aplicação de restrições que dizem respeito ao seu uso. Entendemos por axioma as regras/leis/princípios que regem a organização das classes em uma ontologia. Como podemos perceber, a definição de ontologia varia conforme a sua aplicação. Neste trabalho adotaremos a definição de Gruber (1996) por considerá-la mais apropriada para os objetivos do trabalho.

No que diz respeito às metodologias existentes para a construção de ontologias e suas aplicações, deixaremos para uma discussão posterior, que tratará de ontologias e organização do conhecimento (item 3.1).

1.2 Escolha do domínio

O número de pesquisas linguísticas acerca do léxico do domínio futebol que já foram desenvolvidas no Brasil não reflete a importância desse esporte na vida do país. Apesar da proposta de Carvalho (1996) para uma organização semântica do domínio, não temos conhecimento de nenhum trabalho que apontasse para uma abordagem linguística na construção das classes.

A escolha do domínio futebol não se deu aleatoriamente: esse esporte é bastante popular e difundido no Brasil e está ligado de tal maneira à cultura brasileira a ponto de seu vocabulário ser utilizado não somente no esporte, mas também em outros domínios (p.ex.: *pisar na bola* = fazer algo errado: *bater na trave* = chegar perto do objetivo sem conseguir atingi-lo). Por outro lado, boa parte das expressões empregadas no futebol parece vir de metáforas da língua corrente (*tapete verde* = campo bem cuidado / *caixa de fósforo* = estádio muito pequeno). A diversidade de expressões coloca uma dificuldade suplementar no desenvolvimento de uma ontologia de domínio: a delimitação de seus componentes. Para tentar sanar essa dificuldade utilizamos um corpus de domínio para subsidiar a análise empírica com os métodos da Linguística de Corpus (SINCLAIR, 1991, 2005; BERBER SARDINHA, 2004).

1.3 Classes de objeto aplicadas à ontologia

As ontologias servem como um meio para a representação de informações que contêm um caráter semântico comum, que podem ser aplicadas em situações diversas do mundo real.

Em geral, os estudos voltados para a construção de ontologias têm como base o julgamento de seus criadores, que definem as classes semânticas em que os conceitos se inserem. Como exemplo, podemos citar o método usado por Uschold (1996), guiado por “cenários de motivação”. Esses cenários são descrições de problemas que não são solucionados pelas ontologias existentes e fornecem soluções “intuitivas” para esses problemas descritos.

Conforme o objetivo da pesquisa, pretendemos demonstrar como o modelo de classes de objeto pode ser útil para delimitar a hierarquia presente em uma ontologia. Para isso, mantemos em foco a pergunta que motivou seu estudo: seria possível utilizar um método mais fiável para a construção de ontologias? Acreditamos que a aproximação classes de objeto/classes ontológicas é positiva, pois uma vez criados critérios sintáticos reproduzíveis, qualquer estudioso pode chegar a resultados semelhantes no trabalho de desenvolvimento de uma ontologia por meio da aplicação dos mesmos testes sintáticos.

O trabalho foi dividido em *Fundamentos teóricos*, que traz uma descrição mais ampla sobre: a noção de classes de objeto (2.1), ontologias e organização (2.2), relação entre classes ontológicas e classes de objeto (2.3), e uma explanação do método da linguística de corpus (2.4). Em *Metodologia* trataremos da constituição e manipulação do corpus construído para o trabalho (seção 3), da utilização de programas computacionais que auxiliaram no trabalho de extração de informações do corpus. Na quarta seção trataremos da constituição das classes de objeto para o domínio futebol (4.1) e a hierarquia das classes de objeto e ancoragem ontológica (4.2). Por fim, levantaremos considerações a cerca do trabalho, apontando possíveis utilizações do objeto final.

2 Fundamentos teóricos

Nesta seção serão tratadas as três perspectivas que serviram de base para o presente trabalho: as Classes de Objeto, a construção de Ontologias e a Linguística de Corpus.

2.1 Classes de Objeto de Gaston Gross

Gaston Gross, em seu artigo *Classes d'objet et description des verbes* de 1994, apresenta a noção de Classes de Objeto. Essa abordagem tem sido aplicada sobretudo para o francês, no Laboratório Lexiques Dictionnaires et Informatique da Universidade de Paris XIII³.

Gross (1994) utiliza a noção de frase simples de Zellig Harris para apoiar a teoria sobre classes de objeto. Para ele, a frase canônica é formada por um operador acompanhado de seus argumentos. Um operador determina a presença de um tipo de argumento que lhe é específico. O operador é o responsável pela predicação na frase, é comumente chamado de predicado.

2.1.1 Definição de classes de objeto

Como já mencionado anteriormente, classes de objeto são classes semânticas criadas a partir de critérios sintáticos. Essas classes são definidas a partir de uma classe de operadores que selecionam de forma apropriada os argumentos que a compõem, ou seja, cada classe de operadores possui restrições combinatórias que permitem selecionar apropriadamente o argumento que a acompanha: “As classes de objeto, consideradas como classes de argumentos, são definidas pelo relacionamento com os predicados que são específicos” (LEPESANT E MATIEU-COLAS, 1998, p.6). Assim, classes são constituídas por argumentos que podem ser selecionados por operadores.

³ <http://www-ldi.univ-paris13.fr/ldi.php>

Lerat (2000) afirma que as classes de objetos não são apenas classes extensionais de objetos, mas, sim, a validação lógica de classes lexicais (substantivos), por meio de uma relação semântica (traços definitórios compartilhados) e sintática (traços combinatórios: predicados verbais e adjetivais apropriados).

Para Anastassiadis-Symeonidis (2000) o modelo de classes de objetos não só tenta compilar uma lista de operadores gerais/apropriados, dividindo os substantivos em classes gerais e subclasses homogêneas, mas também tenta identificar verbos suporte que acompanham os substantivos predicativos, adjetivos e determinantes apropriados aplicados aos substantivos de cada classe. Assim, a teoria procura estabelecer uma lista de expressões que incluem nomes pertencentes a cada classe estudada.

Em alguns casos, a especificidade de apenas um predicado é suficiente para definir uma classe inteira:

`redigir <texto>`

Verbos como *redigir* possuem critérios necessários para estabelecer a classe de argumentos que os acompanha. Assim, as unidades lexicais que farão parte da classe `<texto>`⁴ são todas aquelas que podem aparecer como complemento do verbo *redigir*: *redigir carta, artigo, romance, etc.* A criação de uma classe não é baseada em relações de hiperonímia e hiponímia, ou seja, não se trata de uma relação como no exemplo *ofício é um tipo de texto*, mas de uma relação sintática entre o predicado que define a classe e os elementos que nela figuram. Desse modo, o que define a classe `<texto>` deve ser a possibilidade de um de seus itens aparecer na posição complemento do verbo *redigir* (LE PESANT & MATHIEU-COLAS, 1998 p.12).

2.1.2 Classes de argumentos

Como dito anteriormente, a frase simples é composta por um predicado acompanhado de seus argumentos. Ela também é construída em torno de um

⁴ Colchetes angulares e fonte `courier new`: notação para as classes de objeto.

núcleo predicativo que é complementado com um ou mais argumentos nominais, sendo o predicado determinante para o número de posições que compõem a frase:

P (x)	brilhar	x brilha
P (x, y)	vencer	x vence y
P (x,y,z)	lançar	x lança y à z

Em *x brilha*, o predicador requer somente o argumento sujeito para compor a frase:

- 1) Flávio brilha.

Em *x vence y*, o núcleo predicativo seleciona dois argumentos para acompanhá-lo.

- 2) Flu vence o Corinthians.

Já em *x lança y à z*, o predicador seleciona três argumentos para compor a frase:

- 3) Paulo Miranda lança a bola para Juninho.

Harris (1976) também aponta que esse núcleo é capaz de múltiplas realizações e não pode ser identificado apenas como uma única categoria morfológica. Essas múltiplas realizações não ocorrem somente em forma de verbos, mas também em forma de adjetivos, como mostram os exemplos abaixo:

Inocente (x)	x é inocente.
Habilidoso (x)	x é habilidoso.

- 4) O atacante é inocente.
- 5) (Ele) é habilidoso.

O mesmo ocorre com os substantivos predicativos, que estão associados a verbos ou adjetivos:

Dar carrinho (x,y)	x deu um carrinho em y.
--------------------	-------------------------

Ter habilidade (x) x tem habilidade.

- 6) Viola deu um carrinho na bola.
- 7) (Ele) tem uma habilidade fantástica.

Enquanto o verbo carrega suas marcas próprias (tempo, pessoa, aspecto), o adjetivo e o substantivo são acompanhados pelos verbos-suporte. Neves (2000, p.53) define os verbos-suporte como verbos “bastante esvaziados do ponto de vista semântico e formam com o seu complemento (objeto direto) um significado global, geralmente correspondente ao que tem outro verbo na língua”.

Le Pesant & Mathieu-Colas (1998) propõem, nessa mesma linha, uma análise de restrições de seleção que recupere a semântica na descrição da frase, uma vez que um predicado nem sempre é suficiente para que a frase seja interpretada no sentido de descrever os argumentos que o acompanham. Aqui apresentaremos dois tipos de situações:

Situação 1: Ausência de conflito entre os argumentos

Há predicados que podem não impor restrições à semântica de seus argumentos, N (N=substantivo) nos exemplos abaixo pode ser de qualquer grupo nominal:

- Penso em N.
- Lembro-me de N.
- Sonhei com N.
- Isto diz respeito a N.
- N é importante.
- N não tem qualquer interesse.

Situação 2: Especificação semântica dos argumentos

Apesar dessa ausência de conflito, a maioria dos predicados exige a especificação da semântica de seus argumentos. Para uma descrição adequada devemos caracterizar os itens lexicais com rigor, ou seja, dizer, para um dado conjunto de predicados, que tipo de nome pode estar em posição de reivindicá-lo. Observemos o padrão sintático dos verbos *contratar* e *ganhar*, sem nenhuma

referência sobre a natureza dos argumentos que os acompanha (N0 indica o argumento sujeito, N1 e N2 primeiro e segundo argumentos objeto):

N0 contratou N1
N0 ganhou N1 de N2.

A partir dessa estrutura temos:

- 8) Sport contrata Giba.
- 9) Clube contrata outro volante.
- 10) O elenco ganha folga.
- 11) (Ele) já ganhou três faltas.

Mas é improvável que encontremos, no discurso padrão, ocorrências como:

- 12)*Sport contrata a bola.
- 13)*A camisa contrata outro volante.
- 14)*Ronaldo ganhou cartão do gol.
- 15)*Apito ganhou Ronaldo da camisa.

O recurso que pode ser utilizado para situações como esta é a caracterização dos argumentos por meio de traços semânticos aliados às restrições de seleção dos predicados. Utilizando traços como o humano / não humano, animado / inanimado, concreto / abstrato na estrutura padrão dos verbos *contratar* e *ganhar* temos:

Contratar (N0hum, N1hum)

Ganhar (N0 hum, N1inanimado+concreto, N2 hum)

- 16) Fla contrata jovem.
- 17)(Eu) ganhei uma camisa do Zé Roberto.

Podemos dizer que o verbo *contratar* seleciona como argumentos, tanto no argumento sujeito como no argumento objeto, substantivos humanos. Já o verbo *ganhar* seleciona como argumento sujeito um substantivo humano, e como primeiro e segundo argumentos objeto um substantivo inanimado e um substantivo humano, respectivamente.

Observemos agora o adjetivo “agressivo”:

18) D’Alessandro é agressivo.

19) O time é agressivo.

VS

20) *A trave é agressiva.

Nos exemplos 18 e 19, o adjetivo *agressivo* ocorre após um argumento sujeito humano animado, já em 20, o adjetivo parece rejeitar substantivo não humano (N-hum). Podemos dizer que, como restrição de seleção, *agressivo* seleciona como argumento o Substantivo humano (hum+animado). Ao observarmos esse mesmo predicador em outro domínio, a seleção de traços não é válida:

21) O câncer é agressivo.

22) A cirurgia é agressiva.

Os mesmos traços atribuídos aos argumentos do predicador em 18 e 19 não podem ser aplicados com eficiência em 21 e 22, uma vez que *câncer* e *cirurgia* não possuem traços humano animado.

Os traços semânticos utilizados neste trabalho estão listados na tabela a seguir:

Sigla	Traço semântico
hum	humano
humcoll	humano coletivo
inc	inanimado concreto

loc	local
-hum	não- humano

Quadro 1 - Traços semânticos

Até aqui percebemos que para a aplicação do modelo de classes de objeto, a análise exaustiva dos dados é fundamental para se chegar às restrições combinatórias selecionadas por cada predicado. Porém, em alguns casos são necessários apenas um ou dois verbos para delinear uma classe:

Para delimitar uma classe de objetos, é necessário usar a combinação de vários critérios (como é feito para os fonemas): dois ou três verbos são muitas vezes suficientes para constituir uma espécie de "feixe" definicional⁵. (LE PESANT & MATHIEU-COLAS, 1998, p.13).

Observemos os exemplos:

- 23) O Náutico perdeu para o Grêmio. N0 V prep N1
 24) O Vitória venceu o Atlético-Ba. N0 V N1
 25) O Corinthians contratou jogadores. N0 V N1

Em se tratando de língua geral, nenhum dos verbos acima, por si só, é específico para uma classe:

- 26) Marcos perdeu seus óculos.
 27) Douglas venceu a aposta.
 28) Miguel contratou um motorista.

O verbo *contratar* exige que os dois argumentos que o acompanham sejam sujeito humano e objeto humano. Somente com essas restrições não conseguimos delinear a classe de objeto que o acompanha. No entanto, a reunião de apenas dois desses predicados (*perder, contratar / vencer, contratar*), no domínio do futebol, nos permite delimitar uma classe de nomes de clubes <clubes>. Alguns

⁵ Il faut, pour délimiter une classe d'objets, recourir à la conjonction de plusieurs critères (comme on le fait pour les phonèmes) : deux ou trois verbes suffisent souvent pour constituer une sorte de « gerbe » ou de « faisceau » définitionnel. (LE PESANT & MATHIEU-COLAS, 1998, p.13)

operadores podem selecionar seus argumentos conforme a necessidade do domínio. *Vencer* e *contratar* selecionam o argumento clube como N0 no domínio do futebol. Para verificar se essa relação é válida em outros domínios, é necessário aplicar os mesmos testes sintáticos.

Como visto anteriormente, o verbo *redigir* possui restrições de seleção suficientes para delimitar a classe de argumentos <texto>. A partir dessa primeira análise, podemos ir mais além, e pensar na delimitação de uma classe de objeto formada por argumentos que podem servir de suporte para se redigir um texto:

redigir (N0 hum+animado, N1inanimado, N2 inanimado+concreto)

29)Ronaldo redigiu o texto no bloco de notas.

30)Ronaldo redigiu o a carta no caderno.

31)Ronaldo redigiu o romance no notebook.

Nos exemplos acima, *bloco de notas*, *caderno* e *notebook* servem de suporte para que sejam redigidos o *texto*, a *carta* e o *romance*. O que caracteriza carta como um tipo de <texto> é que ela pode ser redigida; do mesmo modo, o que caracteriza caderno como suporte para se redigir é o fato de que nele podemos registrar um <texto>. Assim, qualquer objeto que sirva de suporte para a redação de um texto faria parte da classe de objeto <texto>, uma de suas características é o traço semântico inanimado+concreto (inc) (agenda, diário, touchpad, etc.).

Similar ao verbo *redigir*, o verbo *ler* também seleciona os argumentos que o acompanha. Este verbo foi objeto de uma análise realizada por Le Pesant (1994, p. 31) a título de ilustração do modelo de Gross (1994):

O padrão sintático

ler (N0, N1)

permite construções como:

32) A atriz leu um romance.

33) Ana Maria Braga leu um poema.

34) Sandra leu uma carta.

No entanto, seria difícil encontrarmos, no discurso padrão, construções como:

35) * Ronaldo leu uma casa.

36) * Maria leu um ministro.

37) * Felipe leu uma caixa.

As frases acima têm pouca probabilidade de ocorrer devido as restrições de seleção do operador *ler*, ou seja, *ler* seleciona como argumento sujeito um substantivo que possua o traço humano (hum), e como argumento objeto um substantivo com traço não humano (-hum):

Ler (N0hum + N1-hum)

No entanto, uma observação é feita por Le Pesant (1994):

38) Eu li Paulo Coelho.

Na ocorrência acima, o argumento objeto é Nhum e possui um N -hum apagado. Podemos ler um N hum desde que ele seja autor do <texto>. Por ser um complemento apropriado, <texto> pode ser apagado, desde que seja seguida a ressalva acima.

Ademais, não são todos os substantivos que possuem traço -hum que podem ocorrer na posição N1, mas somente o que pode ser “lido” por um sujeito humano: carta, romance, contrato, etc. Para encontrar os substantivos não-humanos que podem ser “lidos” foi necessário repertoriar todos os nomes do léxico, o resultado foi uma lista de 1500 itens lexicais (LE PESANT, 1994). Essa análise empírica revelou que esses substantivos não ocorrem somente como argumentos do verbo *ler*, mas também dos verbos *decifrar*, *decodificar* e *reler*:

N0 hum	ler reler decifrar decodificar	N1-hum
--------	---	--------

Também foi constatado que na lista de itens lexicais, que ocorrem como argumentos do verbo *ler*, não continha apenas *o que* pode ser lido, mas também o *local* onde os <tipos de texto> que podem ser lidos estão registrados:

- 39) Ronaldo leu um soneto sobre um caderno.
- 40) Ronaldo leu um soneto em um notebook.
- 41) Ronaldo leu um soneto sobre uma cadeira.
- 42) Ronaldo leu um soneto sobre a areia.

A partir destas frases é extraída a seguinte estrutura:

N0+hum	ler reler decifrar decodificar	N1-hum (<i>em / sobre</i>) N2 inc
--------	---	-------------------------------------

O argumento sujeito da estrutura acima deve ser um substantivo que possui o traço humano (Nhum), o primeiro argumento objeto um substantivo com traço não humano (N-hum), e o segundo argumento objeto um substantivo inanimado/concreto.

No entanto, é possível encontrar ocorrências como:

- 43) Ronaldo leu um livro.
- 44) Ronaldo leu uma revista.

Sendo que N1 (*livro, revista*) não é o mesmo que ocorre nos exemplos 32, 33 e 34 (*romance, poema, carta*), mas possui o mesmo traço semântico (inanimado concreto) de N2 em 39, 40, 41 e 42.

Para diferenciar os nomes que designam coisas que podem ser lidas do local onde elas estão registradas, o autor propõe uma nomenclatura:

<s-> = ardósia, livro, revista, etc.

<é-> = endereço, história, decreto, ideograma, etc.

As formalizações abaixo foram propostas pelo autor:

ler/N0hum/N1:<s->/

ler/N0hum/N1:<é->/

decifrar/N0hum/N1:<s->/

decifrar/N0hum/N1:<é->/

decodificar/N0hum/N1:<s->/

decodificar/N0hum/N1:<é->/

reler/N0hum/N1:<s->/

reler/N0hum/N1:<é->/

e

ler (em / sobre)/N0hum/N1:<é->/N2:<s->/

decifrar (em / sobre)/N0hum/N1:<é->/N2:<s->/

decodificar (em / sobre)/N0hum/N1:<é->/N2:<s->/

reler (em / sobre)/N0hum/N1:<é->/N2:<s->/

A classe de objeto <s-> é definida por ser selecionada para a posição N1, pelos operadores *ler*, *decifrar*, *decodificar* e *reler*. Já na posição N2, esta mesma classe é selecionada pelos operadores *ler*, *decifrar*, *decodificar* e *reler* sucedidos das preposições *em* ou *sobre*.

Ao longo de seu artigo, Le Pesant (1994) apresenta as classes criadas a partir do verbo *ler*, que vão desde <suporte de publicação> à <escritos periódicos e não-periódicos>. Ao concluir, o autor coloca que a ideia de classes de objeto parece fazer avançar a compreensão da relação entre semântica e sintaxe.

Como veremos adiante, a descrição por meio do modelo de classes de objeto não se restringe apenas às classes de argumentos. As classes de operadores/predicados também têm sua importância, uma vez que são elas as responsáveis pela seleção dos argumentos.

2.1.3 Classes de predicados

No modelo de classes de objeto, os mesmos princípios utilizados para a descrição dos argumentos são utilizados para uma classificação semântica dos predicados.

Alguns adjetivos possuem a propriedade de ocupar a posição de epíteto de um substantivo que lhe é apropriado para qualificá-lo, como mostram os exemplos abaixo:

Edmundo tem um comportamento agressivo = Edmundo é agressivo.

O gol tem a forma retangular = O gol é retangular.

Esta relação apresentada nos exemplos acima serve de base para a criação de classes de adjetivos de <forma>, *retangular*, *quadrado*, *redondo* etc, e de <comportamento> *agressivo*, *individualista*, *violento*.

Podemos destacar também, em se tratando do futebol, as qualidades humanas que podem ser encontradas em um jogador:

NO	V suporte	<qualidade humana>
O Jogador	é	ágil
		dinâmico
		criativo
		brilhante

Assim, qualquer adjetivo que possa ocorrer em <qualidade humana> na estrutura acima, fará parte desta classe de predicados.

2.1.4 Restrições de seleção dos argumentos

Ao abordar a relação operador/argumento por meio do modelo de classes de objeto, constatamos que o operador determina qual argumento deve lhe acompanhar. Em se tratando de realização linguística, podemos dizer que as ocorrências de uma língua são sequências de itens lexicais que obedecem certas restrições combinatórias.

A primeira restrição que apresentaremos consiste na ordem parcial desses itens. A restrição operador-argumento(s) impõe uma ordem parcial sobre os itens lexicais de uma língua em relação à possibilidade de esses co-ocorrerem nas frases. Assim, em:

45) O Jogador chutou a bola. N0 chutar N1

Chutar ocorre como operador de *Ronaldo* e *bola*. Todas as outras restrições que porventura aparecerem (os traços semânticos de cada argumento, por exemplo) são consequência desta primeira, pois é nela que temos a capacidade de expressar relações semânticas em uma sequência de itens lexicais.

A segunda restrição a ser apontada trata do significado específico de um item lexical em uma frase, ou seja, cada argumento possui maior probabilidade de ocorrer acompanhado por determinados operadores.

No domínio do futebol, é mais provável que para argumentos *Ronaldo* e *bola* apareça o operador *chutar* no lugar de *comer*, sendo este último mais provável que *respirar*.

46) Ronaldo chutou a bola.

47) Ronaldo comeu a bola.

48) Ronaldo respirou a bola.

Como podemos observar, o modelo de classes de objeto é uma teoria que permite abordar fenômenos de natureza lexical, semântica e sintática de uma forma relativamente eficaz. A partir dessas constatações, aplicamos esse mesmo modelo para o português. Trabalhos semelhantes já foram feitos nas línguas francesa e grega (SABATIER; 1997; MOUSTAKI; DIMITRIADI, 2006).

2.1.5 Outros trabalhos

Neste item serão apresentados alguns trabalhos que abordaram a teoria de classes de objeto de Gaston Gross.

Sabatier (1997) propõe, para o francês, um léxico-gramática do futebol feito a partir do modelo de classes de objeto. Seu trabalho consistiu em desenvolver duas tabelas. A primeira traz classes de construções frasais, sendo que essas construções podem ser aplicadas às construções sintáticas inerentes a cada verbo. Na segunda tabela o autor indica 32 classes conceituais (classes de objeto), encontradas a partir da análise. Essas classes vão desde *árbitros*, *jogadores*, *cartões* até *partes do corpo* e *gestos técnicos*:

<arbitre> formas de nomear um árbitro.

<buts> formas de nomear o tiro-de-meta.

<carton-sanction> cartões que o árbitro pode atribuir ao jogador sancionado.

Moustaki e Dimitriadi (2006) realizaram um estudo contrastivo entre as línguas francesa e grega utilizando o léxico-gramática desenvolvido por Sabatier (1998). As autoras partiram de uma seleção de 600 frases simples para estabelecer não somente o léxico-gramática do futebol para o grego, mas também uma lista das classes dos objetos e de substantivos compostos desse léxico-gramática. Os objetivos desse trabalho foram descrever similaridades morfossintáticas e as diferenças entre construções de frases simples e substantivos compostos em ambas as línguas, visando aplicações futuras em tradução automática.

Observando os textos teóricos e os trabalhos feitos a partir do modelo de classes de objeto de Gross, o objetivo deste trabalho foi traçado de modo que essa teoria fosse testada no delineamento de classes e atributos de uma ontologia. Para isso, diferentemente dos autores acima citados, que trabalharam com introspecção – quando os exemplos são criados a partir do conhecimento de mundo do pesquisador –, optamos por uma pesquisa baseada em linguística de corpus.

2.2 Ontologias e organização do conhecimento

As ontologias servem como um meio para a representação de informações que contêm um caráter semântico comum, que podem ser aplicadas em situações diversas do mundo real. Uma ontologia assemelha-se à classificação que os biólogos utilizam para classificar os seres vivos e serve como base para uma comunicação livre de ambiguidades. Sua hierarquia é definida pelos critérios semânticos, permitindo que usuários relacionem conceitos utilizando relacionamentos do tipo *parte-de*, *tipo-de*, *membro-conjunto*, *lugar-região*, etc.

Ontologia é definida no dicionário Oxford de Filosofia como “o termo derivado da palavra grega que significa ‘ser’, mas usado desde o século XVII para denominar o ramo da metafísica que diz respeito àquilo que existe” (BLACKBURN, 1997). Na ciência da Computação, as classes são definidas a partir de um conjunto de propriedades.

Di Fellipo (2008) destaca que as ontologias podem ser representadas por meio de várias metalinguagens formais ou formalismos; a informação relevante, no entanto, não é a forma, mas o conteúdo, ou seja, o conjunto de noções oferecidas como um modo de “ver” o conhecimento.

Como dito anteriormente, neste trabalho adotamos a definição de Gruber (1996, [s/p]) para ontologia:

Uma ontologia é uma especificação explícita de uma conceitualização. [...]. Em tal ontologia, definições associam nomes de entidades no universo do discurso (por ex. classes, relações, funções, e outros objetos) com textos que descrevem o que os nomes significam e os axiomas formais que restringem a interpretação e o uso desses termos.

Antes de prosseguir, cabe abrir um espaço para falarmos de *conceito*. Não pretendemos aprofundar na discussão sobre sua definição, mas cremos, assim como Di Fellipo (2008, p.2), que o conceito pode visto como uma “‘descrição mental’, uma ideia (compartilhada pelos falantes) de um tipo de coisa”, que permite a categorização por parte do falante, assim este pode discriminar diferentes tipos de entidades. O conceito pode, em primeira instância, ser lexicalizado em todas as

línguas, mas sua lexicalização depende da interação entre convenção e motivação (TAYLOR, 1985; LAKOFF, 1987 *apud* DI FELIPPO, 2008).

A construção de uma ontologia apóia o desenvolvimento do Processamento de Língua natural (PLN), na medida em que pensamos a sistematização feita como forma de apoio para que o computador possa simular o uso de uma língua natural. Isso pode ser alcançado por um módulo ontológico (conceitos) que faça a ponte entre o mundo (extralinguístico) com o qual se pretende trabalhar – no caso do trabalho apresentado, o domínio do futebol – e as formas presentes no léxico (signo/significante). Uma forma adaptada do triângulo de Ogden e Richards (Figura 1) pode nos ajudar a explicitar melhor essa ideia:

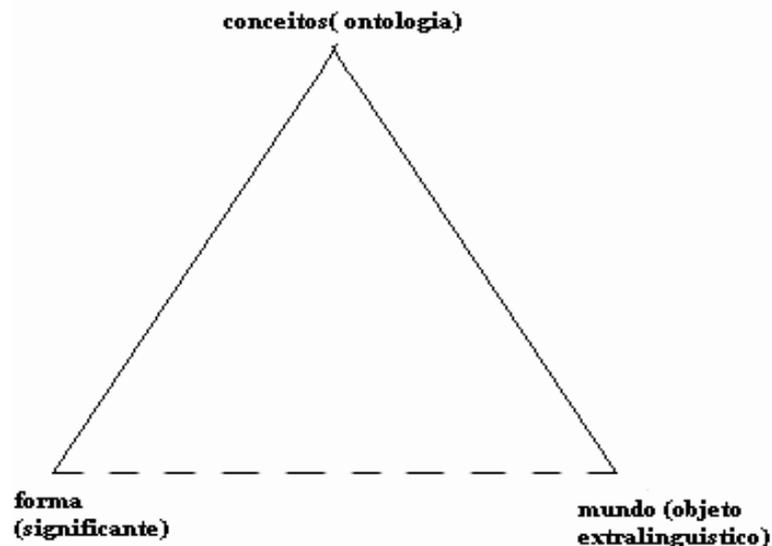


Figura 1 - Triângulo de Ogden e Richards adaptado

Além disso, é importante destacar os níveis de representação feitos por Jackendoff (1983). O autor divide a organização do pensamento em quatro domínios: *mundo real*, *mundo projetado*, *estrutura conceitual* (conceitos) e *expressões linguísticas* (formas presentes no léxico). Para esse autor, as entidades às quais nos referimos se encontram em um mundo projetado, e não no mundo real, já que os conceitos são constituídos pela estrutura da mente. Em decorrência disso, as formas presentes no léxico não denotam os objetos reais, mas sim as representações mentais que são resultantes das informações que chegam à mente. Como exemplo, podemos estabelecer que a expressão linguística *bola* (forma) é um

tipo de BOLA (conceito) e que, em uma realização concreta desse conceito em *Eu chutei a bola*, o referente é BOLA (DIAS-DA-SILVA, 1996).

Em se tratando de PLN, também cabe ressaltar a dinâmica do processo de construção de um SPLN (Sistema de Processamento de Língua Natural) adotada por Dias-da-Silva (2006) (Figura 2). Para o autor, a montagem de SPLNs exige o desenvolvimento de, no mínimo, três etapas: “extração do solo’ (explicitação dos conhecimentos e habilidades linguísticas), ‘lapidação’ (representação formal desses conhecimentos e habilidades) e ‘incrustação’ (o programa de computador que codifica essa representação)” (DIAS-DA-SILVA *et al* 2007, p.15). Estas três etapas foram chamadas de: *Fase Linguística*, em que o trabalho consiste em construir um “corpo de conhecimentos sobre a própria linguagem, dissecando e compreendendo os fenômenos linguísticos necessários para o desenvolvimento do sistema”; *Fase Linguístico-Computacional*, que é caracterizada pela “construção conceitual do sistema, envolvendo a seleção e/ou proposição de sistemas formais de representação para os resultados propostos pela fase anterior”; e *Fase Computacional*, cujo trabalho consiste na “codificação das representações elaboradas durante a fase anterior em termos de linguagens de programação e planejamento global do sistema.” (DIAS-DA-SILVA, 2006, p.125).

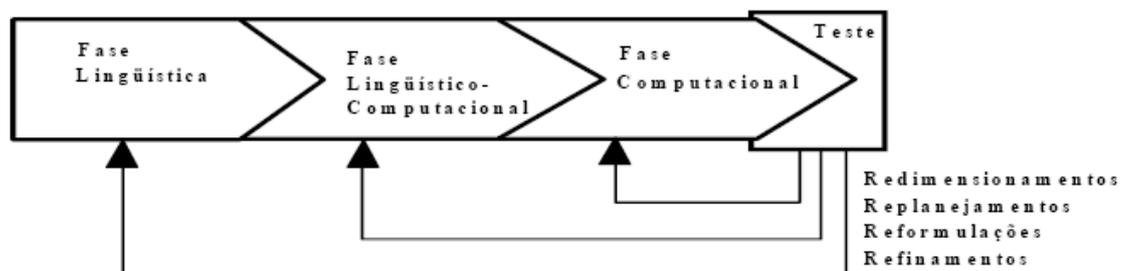


Figura 2 - Dinâmica do processo de construção de um SPLN
Fonte: (DIAS-DA-SILVA, 2006)

O presente trabalho constitui a fase linguística, pois explicitamos o conhecimento e uso linguístico do domínio do futebol por meio das Classes de objeto de Gross (1994).

2.2.1 Metodologias para a Construção de ontologias

Atualmente, pesquisadores de várias áreas computacionais estão reconhecendo que é necessário um certo tipo de semântica para superar dados estatísticos de PLN. Hovy (2005) coloca que o mais problemático aspecto no que diz respeito ao conteúdo semântico é a ausência de debate metodológico e teorias emergentes, que pudessem fornecer o rigor necessário, sistematicidade e métodos de verificação que facilitariam o trabalho de construção de ontologias.

Com crescimento progressivo e constante de trabalhos que envolvem a construção de ontologias, crescem também diferentes formas de desenvolvimento de metodologias, o que torna difícil a escolha de qual metodologia adotar. Hovy (2005) destaca que a ausência desse debate metodológico tem como consequências as seguintes questões: quem constrói, não sabe dizer se a ontologia construída hoje tem a mesma consistência da que foi construída ontem e da que será construída amanhã. Em consequência disso tem-se um usuário que não sabe escolher entre as diversas teorias semânticas existentes, sendo a escolha baseada, na maioria das vezes, na reputação do pesquisador.

2.2.1.1 Estilos de abordagem ontológica

Hovy (2005) ainda identifica cinco estilos de abordagem ontológica decorrentes da: Filosofia, Ciência Cognitiva, Linguística, Inteligência Artificial (AI)/linguística Computacional, Raciocínio de Domínio. O autor ainda complementa que cada uma das abordagens funciona de uma maneira distinta, resolvendo questões com argumentos e recursos de diversos padrões de raciocínio, o que os conduzem a resultados diferentes.

Tipo 1: recombinação abstrata (filósofos):

Estes pesquisadores fazem o processo de criação pelo conceito de especificação, ou seja, usam o recurso sistemático adicionando novos *differentiae*. Essa forma permite uma combinação mais ou menos mecânica, isto faz com que a abordagem infelizmente não funcione para além dos níveis abstratos, e,

consequentemente, não é muito útil para, por exemplo, a prática de construção de ontologias de domínio.

Tipo 2: distinções ontológicas intuitivas

A mais antiga e natural razão para a criação de um novo conceito é a sensação intuitiva de que este conceito não é o mesmo que qualquer outro já definido. Infelizmente as pessoas não são coerentes em fazê-lo, dividindo e criando subgrupos de conceitos já existentes, que formam variantes distintas com qualquer tipo de recurso.

Tipo 3: fenômenos linguísticos

Para algumas pessoas, os conceitos podem ser motivados pura e simplesmente porque palavras ou expressões para eles aparecem em várias línguas. Para Vigotsky, Sapir, Whorf ou Piaget a interligação entre linguagem e mente é bastante aceita. Um exemplo disso é a EuroWordNet (VOSSEN, 1998), em que a análise cruzada das diferenças linguísticas revela complexas inter-relações entre significados. Hovy (2005) coloca que existem muitos conceitos que não possuem a palavra correspondente em uma determinada língua, enfatizando que a utilização desse tipo de pensamento permite resultados satisfatórios em ontologias gerais, o que não acontece com as ontologias de domínio, que perdem a expressiva utilidade.

Tipo 4: Inferência de generalizações (AI/ Linguística Computacional)

Para os sistemas computacionais a ontologia é uma instalação que garante que os itens do domínio serão tratados (agrupados) do mesmo modo. Os metadados devem ser definidos em detalhe suficiente com um apoio razoavelmente poderoso de um conjunto de regras. O autor aponta essa metodologia como relativamente limpa, dependendo da solução para as regras computacionais. Porém, pesquisadores de áreas diversas (filósofos, linguistas, etc.) não possuem conhecimento computacional para desenvolver essas regras.

Tipo 5: Especialistas do domínio

A necessidade de se criar ontologias para determinadas áreas surge por parte dos próprios especialistas do domínio. No entanto, o ponto de vista que o especialista tem do domínio reflete na ontologia construída, ou seja, cada especialista obterá um resultado diferente.

Pode-se considerar as cinco abordagens diferentes apontadas por Hovy (2005) como complementares, no entanto, essa mistura de práticas tende a exigir que as partes saibam um pouco sobre cada abordagem. Para criar uma ontologia de domínio, a metodologia geralmente adotada é a seguinte:

1. Determinação das características gerais da ontologia que será construída
2. Reunião dos conhecimentos sobre o domínio
3. Delimitação dos grandes fenômenos que deverão ser considerados
4. Levantamento da lista de ocorrências
5. Construção da definição, com as principais características (*differentiae*)
6. Gravação em parâmetros essenciais

Hovy (2005) completa afirmando que trabalhar com todos os detalhes as fases dessa metodologia leva tempo e esforço, que nem todos os aspectos são aplicáveis a todos os casos, e que, porém, um cuidadoso estudo sobre esse procedimento poderá ajudar no aperfeiçoamento de uma metodologia sistemática para a construção de ontologias.

Alguns trabalhos já foram realizados a fim de avaliar metodologias existentes para a construção de ontologias. Como exemplo podemos citar o trabalho *Ontologias e vocabulários controlados: comparação de metodologias para construção* de Silva, Sousa e Almeida (2008). Após a análise de vários documentos das áreas da computação e da ciência da informação, os autores selecionaram as metodologias para comparação e definiram categorias para a análise das metodologias.

Foram selecionadas sete metodologias para comparação: a metodologia de Gruninger e Fox (GRUNINGER; FOX, 1995 *apud* SILVA, *et al*, 2008); o método de Uschold e King (USCHOLD; KING, 1995 *apud* SILVA, *et al*,

2008); a metodologia Methontology (GÓMEZ-PÉREZ; FERNANDEZ-LOPES; VICENTE, 1996 *apud* SILVA; SOUSA; ALMEIDA, 2008); o método Cyc (REED; LENAT, 2002 *apud* SILVA; SOUSA; ALMEIDA, 2008); o método Kactus (BERNARAS; LARESGOITI; CORERA, 1996 *apud* SILVA; SOUSA; ALMEIDA, 2008); o método Sensus (VALENTE *et al*, 1999 *apud* SILVA; SOUSA; ALMEIDA, 2008); o método 101 (NOY; GUINNESS, 2001 *apud* SILVA; SOUSA; ALMEIDA, 2008). Para aplicar o método comparativo foram estabelecidas onze categorias para análise: o gerenciamento do projeto, o pré-desenvolvimento, a especificação de requisitos, a modelagem conceitual, a formalização, a implementação, a forma de apresentação, a manutenção, a integração, a avaliação e a documentação.

Diante da análise feita, Silva, Sousa e Almeida (2008) levantaram algumas considerações acerca dos resultados da análise, como a variedade de estratégias adotadas para a construção de ontologias, a falta de um padrão para construção de ontologias, e a “falta de explicações sistemáticas de como, onde e sob quais limites podem ser utilizadas as abordagens teóricas dentro do processo de elaboração”. Os autores destacam, ainda, que a solução para esses problemas estaria centrada em “uma proposta metodológica fundamentada em princípios teóricos e metodológicos que dessem sustentação científica no processo de construção de ontologias”.

Acreditamos que os resultados de Silva, Sousa e Almeida (2008) vêm somar-se à ideia central deste trabalho: a construção de uma ontologia por meio de uma metodologia que utiliza critérios sintáticos reproduzíveis. Assim, poder-se-ia chegar a resultados semelhantes aplicando os mesmos testes sintáticos em diferentes trabalhos.

A seguir serão apresentadas algumas metodologias para desenvolvimento de ontologias.

2.2.1.2 Alguns modelos metodológicos

MIKROKOSMOS (BEALE; NIRENGURG; MAHESH 1995; MAHESH; NIRENGURG 1995) foi construída no Laboratório de Investigação da Computação do Novo México (NMSU) e contém 4790 conceitos. O projeto foi desenvolvido para apoiar a tradução automática entre variedade de línguas. Assim, definem uma

metodologia para representar o significado da linguagem natural de textos em uma linguagem neutra, uma interlíngua formal, nomeada *text meaning representation* (TMR). Uma das principais aplicações da ontologia foi o processamento de textos sobre aquisições de empresas, que não são padronizados e abrange um universo amplo de conceitos, isso significa que a ontologia possui um grande número de conceitos associados.

A ontologia Mikrokosmos é baseada nos seguintes princípios descritos abaixo:

1. A ontologia é a língua independente. Uma ontologia não é um dicionário de transferência entre duas línguas, pois ela não é específica de uma determinada língua natural, como Inglês ou Espanhol.

2. Uma ontologia deve ser motivada independentemente.

3. Uma ontologia deve ser formada de acordo com a necessidade, especificação axiomática e internamente consistente em termos de estrutura, nomes, e de conteúdo em todos os conceitos.

4. As ontologias devem ser coerentes e compatíveis com outros conhecimentos e transformação dos recursos, como o léxico, o analisador semântico, e qualquer microteoria.

5. Para PLN, tradução automática e outras tarefas, a estrutura da ontologia é necessariamente rica, não se trata apenas de uma hierarquia de conceitos. Cada conceito não é apenas um rótulo, mas tem rica estrutura interna, resultando em um alto grau de conectividade com outros conceitos da ontologia.

6. Uma ontologia deve ser fácil de compreender e simples para pesquisar conceitos. Ela deve ser fácil de navegar, fácil de treinar adquirentes, apresentável, e assim por diante.

7. A ontologia deve ter uma grande utilidade para a tarefa para a qual é destinada. A linguagem na ontologia é independente, porém é dependente de tratamento da língua:

a. deve dispor de todos os conhecimentos necessários para condicionar suficientemente a geração e interpretação da linguagem.

b. deve ter uma rica rede de conceitos ligados para permitir uma variedade de inferências.

A somatória de todos esses princípios exprimem a definição de ontologia utilizada nesta pesquisa. Em especial, destacaremos dois princípios adotados pela Mikrokosmos: o primeiro, que trata da independência dos conceitos em relação às línguas; e o quinto, sobre a conectividade de um determinado conceito com outros pertencentes à ontologia.

Cabe abrir um espaço para comentarmos a ontologia CYC (REED, LENAT, 2002) que tem sido desenvolvida por CYCorp, Texas, desde meados da década de 80 . É uma tentativa de criar uma única e grande ontologia como um possível modelo para o mundo inteiro. A primeira versão da OpenCyc foi lançada em 2002 e continha apenas 6.000 conceitos e 60.000 axiomas. Atualmente, a Ontologia CYC contém cerca de 40.000 conceitos e mais de 300.000 axiomas.

O problema que motivou o desenvolvimento da CYC está no fato de que muitos analistas, pesquisadores, políticos no Governo, militares, e representantes de setores empresariais trabalham em meio um mar de informações de fontes díspares de formatos incompatíveis. O analista é responsável por identificar, extrair e integrar as informações necessárias. Deste modo, o analista tem que gastar um tempo valioso na triagem das fontes para obter a informação verdadeiramente relevante.

Para solucionar o problema, os pesquisadores apresentaram a SKSI (Semantic Knowledge Source Integration), que permite uma base de conhecimentos para atuar como uma fonte de dados inteligente e mediadora.

Suas classes superiores – em torno de 3.000 – estão organizadas em 43 categorias. Segundo seus criadores, os 3.000 termos obedecem a critérios de universalidade (qualquer conceito pode ser corretamente relacionado com um desses 3.000 termos) e articulação (para cada distinção feita, existem justificações pragmáticas e teóricas).

Os conceitos da CYC estão organizados de acordo com suas relações, que expressam dois diferentes tipos de submissão: *elemento essencial* e *sub-conjunto de*.

Este método foi dividido em 3 processos:

Processo I: extração manual do conhecimento;

Processo II: extração do conhecimento apoiada por computadores;

Processo III: extração do conhecimento gerenciada por computadores.

A sua base de conhecimento é continuamente atualizada. Também foram desenvolvidas microteorias para apoiar a exposição de informação em domínios específicos.

Como podemos constatar, o desenvolvimento de uma ontologia requer planejamento e sistematicidade no tratamento dos dados. Tendo isso em mente, o presente trabalho aplica um modelo formal de descrição (GROSS, 1994) para a observação da hierarquia existente entre as classes de uma ontologia.

2.3 Classes Ontológicas versus Classes de objeto

As classes de objeto são classes semânticas construídas a partir de critérios sintáticos. Cada classe é composta a partir de predicados que selecionam uma classe de argumentos. Uma classe relativamente homogênea de argumentos selecionados por uma classe de predicados constitui uma classe de objeto. Além disso, para cada classe de objeto são também definidos os demais predicados que a ela se aplicam.

Para Gross (1994) essas classes são definidas a partir da noção de emprego de predicado. Para o autor, definir um sentido de um predicado consiste em descrever a totalidade de suas propriedades. Com a aplicação do modelo de classes de objeto, a noção de hiperonímia ganha um novo sentido, já que as heranças das classes devem se enquadrar no tipo de relação sintática existente entre o predicado e seus argumentos. Para a aplicação do modelo de classes de objeto, podemos recorrer tanto a um método introspectivo, em que as frases são criadas a partir do conhecimento de mundo do pesquisador, quanto a um método baseado em lingüística de corpus, utilizando corpus computadorizados: coletâneas de textos, transcrições de fala, etc. Isto implica que o objeto tratado no modelo de classes de objeto é a língua, seja por meio de um método introspectivo, em que as orações são criadas para fins de explicitação, seja por meio do método baseado em lingüística de corpus, com corpus computadorizado.

Neste trabalho tentamos fazer uma aproximação de classes de objeto e classes ontológicas. As ontologias têm sido utilizadas como um meio para a representação de informações que contém um caráter semântico comum, que

podem ser aplicadas em situações diversas do mundo real. As ontologias oferecem para a Web uma base de informações comum e padronizada. Essa base engloba conceitos que, quando requisitados, podem ser usados para cada situação particular.

Uma ontologia de domínio, por exemplo, é o parâmetro segundo o qual o vocabulário específico é interpretado. A definição desse vocabulário é feita de forma rigorosa para que não aja ambiguidade ao aplicarmos essas definições na realidade daquele domínio.

França (2009) discute exaustivamente a definição de ontologia postulada por vários autores (GUARINO, 1995:1998, GRUBER, 1996, SMITH *et al.*, 2005), sendo que, por unanimidade, os autores relacionam ontologia com conceitualização: “Uma ontologia é definida em relação directa com o termo conceptualização, partindo do pressuposto que é de conceitos que uma ontologia trata.” (FRANÇA 2009, p.111).

Para compreender melhor a definição de conceito, podemos observar algumas interpretações para o triângulo semiótico de Ogden e Richards apresentadas por Lyons (1980) na figura a seguir:

<i>(i)</i> <i>símbolo</i>	<i>(ii) pensamento/ referência</i>	<i>(iii)</i> <i>referente</i>
signo	conceito	significatum
signo	intensão	extensão
palavra / lexema	conceito	coisa
signo	significatum	denotatum
signo	pensamento	objecto

Figura 3 – Interpretações do triângulo semiótico de Ogden e Richards apresentadas por John Lyons

Fonte: França (2009).

As três colunas apresentadas por Lyons (1980 apud FRANÇA, 2009) são o signo, o significado e o objeto real ao qual ambos se referem (referente/ realidade extralinguística).

As classes de objeto estão ligadas à primeira coluna, *símbolo*, uma vez que trabalhamos com as formas presentes no léxico em situações reais de uso. Já os conceitos de uma ontologia estão representados na segunda coluna, *referência*, pois os mesmos tratam da representação do conhecimento/ pensamento.

No domínio do futebol o uso expressões idiomáticas/cristalizadas é bastante significativo. As expressões cristalizadas são interpretadas pelo falante com seu significado idiomático, pois não podemos atribuir um sentido à construção apenas com a soma dos significados de suas partes (VALE, 2001). Como então incluir expressões cristalizadas na descrição das classes de objeto? Esta questão será tratada no item a seguir.

2.3.1 Linguagem figurada

Neste subitem avaliaremos a importância da linguagem figurada para o trabalho proposto. Diante de algumas leituras (ORTONY, 1993; OLIVEIRA 2006) constatou-se a necessidade de discorrer sobre o tema em duas vertentes. Linguagem figurada nas classes ontológicas e linguagem figurada nas classes de objeto.

Linguagem figurada nas classes ontológicas

As classes ontológicas são classes constituídas de conceitos. Esses conceitos estão ligados à representação do conhecimento, que antecede à realização linguística. Uma vez que ouvimos alguém pronunciar *A bola rolou no tapete verde* ligamos o referente, “a coisa” extralinguística, ao seu conceito. Assim, ao ouvirmos:

- 49) A bola rolou no tapete verde.
- 50) A bola rolou no gramado.
- 51) A bola rolou no campo de futebol.

ativamos a ligação do objeto extralinguístico *campo de futebol* de maneiras diversas a um único conceito campo de futebol. A escolha de qual signo usar para representar o conceito é do falante, e este levará em conta o contexto e as

circunstâncias para utilizar um ou outro signo. A figura a seguir, baseada no triângulo de Ogden e Richards, pode explicitar melhor essa ideia:

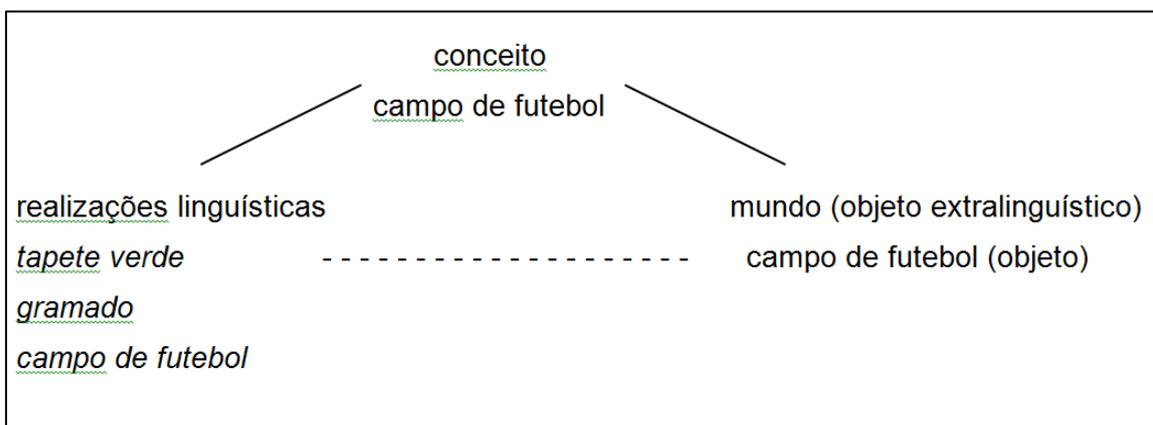


Figura 3 - Triângulo de Ogden e Richards adaptado II

As classes ontológicas e as classes de objeto não pertencem ao mesmo nível, pois a primeira não é específica de uma determinada língua natural e a segunda trata de realizações linguísticas. A relação entre as duas classes pode ser estabelecida pela ancoragem ontológica⁶ das classes de objeto a uma ontologia (conceitos) previamente montada.

Linguagem figurada nas classes de objeto

As classes de objeto são compostas por resultados de uma análise exaustiva da co-ocorrência das unidades lexicais na língua. Por tratar-se de realização linguística, as classes de objeto não podem ser confundidas com os conceitos de uma classe ontológica, por motivos já mencionados anteriormente. No entanto, as classes de objeto podem ser ancoradas a um conceito.

Ortony (1993) destaca que o significado figurado, enquanto um significado não-composicional, indireto, e pragmático (extralinguístico), por violar os critérios empíricos do significado (literal) é originalmente definido como um uso linguístico não significativo.

Algumas características são atribuídas aos significados literais e figurados: (i) o significado literal é automático e obrigatório, enquanto o significado

⁶ Ancoragem ontológica foi abordada no trabalho de Tosqui (2007): Construção e ancoragem ontológica do vocabulário básico bilíngue do Turismo para fins didáticos.

não-litera é não-automático e opcional; (ii) o significado litera é o significado não-marcado, ou a norma, e é o significado frequentemente usado; (iii) o significado não litera é gerado apenas quando o significado litera é não plausível (ARIEL, 2002, p. 364).

Alguns trabalhos já abordaram a importância dos significados figurados para a descrição do léxico. Como exemplo, podemos citar Vale (2001) que propõe uma tipologia para as expressões cristalizadas do português. O que caracteriza uma expressão cristalizada é o fato de seu significado não poder ser calculado a partir de seus componentes (GROSS, 1982).

No que diz respeito à linguagem figurada e o domínio escolhido, o domínio do futebol carrega particularidades no que se refere ao léxico especializado, pois a terminologia usada no futebol muitas vezes se confunde com os usos de língua geral. Observemos os exemplos das ocorrências a seguir:

52) Nem com a *caixa de ferramentas aberta*, o São José conseguiu segurar a pressão

53) O importante é *acumular gordura* para distanciar do quinto colocado.

No exemplo 52, a expressão *caixa de ferramentas* é recorrente no mundo do futebol. Ela é usada para expressar que os jogadores jogaram com agressividade, cometendo muitas faltas a fim de barrarem o ataque do time adversário.

Do mesmo modo, a expressão *acumular mais gordura*, do exemplo 53, não está relacionada ao ganho de peso. Na esfera do futebol essa expressão passa a ter um novo sentido: a acumulação de pontos no campeonato de futebol que tem por fórmula a disputa por pontos corridos.

Essas expressões cristalizadas são interpretadas pelos falantes do domínio com seu significado idiomático, e não pela soma dos sentidos literais das palavras que as compõem. O caráter polissêmico dessas expressões faz com que elas não possam ser analisadas fora do contexto em que se encontram.

Diante disso, constatamos que a linguagem figurada é essencial no que diz respeito à realização linguística no domínio do futebol. Por isso cremos que a linguagem figurada, enquanto realização linguística de um domínio, deve fazer parte de uma descrição sistemática do mesmo.

2.4 Linguística de corpus

O desenvolvimento de pesquisas utilizando corpus como fonte não é uma prática exclusiva dos trabalhos atuais. Há muito os corpora são utilizados na elaboração de dicionários e podemos citar, como exemplo, o dicionário feito no século XVIII, *Vocabulário Portuguez e Latino*, elaborado pelo Padre Rafael Bluteau e publicado entre 1712-1728. O Vocabulário Bluteau possui oito volumes e foi o primeiro dicionário a ser desenvolvido com base em corpus. Esse corpus foi constituído de 406 obras, aproximadamente, com autores dos séculos XV a XVII, (ALUISIO; ALMEIDA, 2006).

Com o advento do computador, pesquisadores começaram a optar por corpus computadorizado, pois com o auxílio de programas computacionais tornou-se possível processar um número bem maior de textos em busca de dados morfológicos, sintáticos e semânticos, a fim de se encontrar regularidades em uma língua.

Aluísio e Almeida (2006) fazem uma discussão minuciosa sobre a concepção de corpus para a Linguística e para a Linguística de Corpus trazendo as definições de corpus para autores como Galisson e Coste, Dubois, Ducrot e Todorov, Trask e Sinclair. Ao fim as autoras constataram que a maior diferença de corpus para a Linguística e a Linguística de Corpus é o formato do corpus, pois para a última, o corpus deve estar necessariamente em formato eletrônico. Neste trabalho optamos pela definição de Linguística de Corpus apresentada por Berber Sardinha (2004):

A linguística de Corpus ocupa-se da coleta e da exploração de corpora, ou conjuntos de dados linguísticos textuais coletados criteriosamente, como propósito de servirem para a pesquisa de uma língua ou variedade linguística. Como tal, dedica-se à exploração da linguagem por meio de evidências empíricas, extraídas por computador. (BERBER SARDINHA, 2004, p.3).

Optamos pelo método baseado em linguística de corpus para delinear as classes de objeto do futebol por acreditar que o mesmo proporciona maior contato com a realidade linguística, com ocorrências reais de uso. Foi criado um corpus computadorizado para uso em ferramentas computacionais. Eventualmente também utilizamos exemplos retirados da internet para apoiar a descrição realizada

neste trabalho. Nestes casos, utilizados os motores de busca *Google* e *Yahoo Search* para efetuar pesquisas específicas na *web*.

Inicialmente tentamos restringir os gêneros textuais que poderiam ser usados na construção do corpus, no entanto, deparamo-nos com a seguinte questão: apesar de ser um domínio muito difundido no país, a oferta de textos científicos sobre o futebol é escassa, além de não ser comum o uso expressões cristalizadas nos textos encontrados na internet que concernem o domínio do futebol, no que diz respeito à atividade esportiva.

Por isso, utilizamos para a construção do corpus textos eletrônicos de diversos gêneros disponíveis on-line, como notícias, crônicas e manuais de normas do futebol e de campeonatos. No início cogitamos a utilização de fontes como bate-papos, porém chegamos à conclusão de que esse gênero seria de pouca produtividade para a montagem do corpus. Os textos selecionados obedeceram a critérios de legitimidade e confiabilidade, de modo que estes viessem a constituir uma representação fiel do uso da língua dentro do domínio pesquisado. Para isso, utilizamos principalmente *sites* especializados com notícias atualizadas do domínio, criados, em sua maioria, única e exclusivamente para a divulgação e discussão do esporte, como é o caso do site *Globo Esporte* (<http://globoesporte.globo.com/Esportes/Futebol/0,,9825,00.html>), que trabalha com a divulgação de resultados e notícias do mundo do futebol.

A opção por textos eletrônicos se justifica na medida em que levamos em conta o critério de praticidade, já que esses se encontram em larga escala na Web, além de já estarem em formato eletrônico.

Para McEnery e Wilson (*apud* ALUÍSIO e ALMEIDA, 2006), um corpus constitui uma referência padrão para a língua que ele representa, neste caso, os autores defendem o reuso de corpora, ou seja, é de boa aplicabilidade que um corpus esteja disponível para outros pesquisadores.

Visando o reuso de corpora existentes representativos do domínio futebol, obtivemos dois corpora - um de língua falada e outro de língua escrita - utilizados em pesquisas diversas. O corpus de língua falada é formado de quatro narrações transcritas provenientes da transmissão de partidas de futebol, sendo duas delas provenientes de emissoras de rádio e duas de emissoras de televisão, das cidades do Rio de Janeiro e de São Paulo (TURTELLI, 2002). O corpus de

língua escrita é um subcorpus do projeto COMET⁷ (Corpus Multilíngue para ensino e tradução), e foi nomeado de Copa-COMET, por ser formado de notícias da Copa do Mundo de 2006.

Os corpora adquiridos somados às notícias recolhidas nos sites sobre o futebol, manuais e normas somam um total de 460 mil ocorrências/tokens (tokens correspondem ao número total de ocorrências de itens lexicais, repetidos ou não). Conforme tabela apresentada por Berber Sardinha (2004) (figura 4), o nosso corpus possui tamanho médio.

Tamanho em palavras	Classificação
Menos de 80 mil	Pequeno
80 a 250 mil	Pequeno-médio
250 mil a 1 milhão	Médio
1 milhão a 10 milhões	Médio-grande
10 milhões ou mais	Grande

Figura 4 - Classificação do corpus de acordo com o tamanho

Fonte: Berber Sardinha (2004)

Uma vez que o objetivo do trabalho não é fazer uma descrição minuciosa do domínio, o que seria necessário se estivéssemos, por exemplo, desenvolvendo um dicionário do futebol ou se a proposta do trabalho fosse gerar uma ontologia de domínio utilizando as classes de objeto criadas, acreditamos, inicialmente, que este tamanho de corpus supriria as necessidades da pesquisa.

⁷ O projeto COMET – Corpus Multilíngue para Ensino e Tradução, coordenado pela professora Stella Esther Ortweiler Tagnin, em construção junto ao CITRAT (Centro Interdepartamental de Tradução e Terminologia) da Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo, é um corpus eletrônico que tem por objetivo servir de suporte a pesquisas linguísticas, principalmente nas áreas de tradução, terminologia e ensino de línguas. Disponível em: <http://www.fflch.usp.br/dlm/comet/>.

3 Metodologia

A metodologia utilizada no trabalho foi dividida em 3 etapas. A primeira etapa trata da constituição do corpus, que se fez sob a luz da Linguística de Corpus. Na segunda etapa foi realizado o levantamento das ocorrências das unidades lexicais no corpus para que as mesmas fossem trabalhadas na constituição das classes de objeto (etapa 3), após o término da construção das classes tentamos mostrar como é possível relacionar as classes de objeto com a hierarquia de uma ontologia. Apresentamos uma descrição detalhada desses passos a seguir.

3.1 Etapa 1: Constituição e manipulação do corpus

O corpus compreende 460 mil tokens distribuídos entre narrações transcritas, notícias e normas do futebol.

Constituição do corpus:

1° língua falada:

a) quatro narrações transcritas provenientes da transmissão de partidas de futebol, sendo duas delas provenientes de emissoras de rádio e duas de emissoras de televisão, das cidades do Rio de Janeiro e de São Paulo (TURTELLI, 2002). As transcrições possuem 74.600 tokens.

2° língua escrita:

a) Copa-COMET – Subcorpus do projeto COMET (Corpus Multilíngue para ensino e tradução), com 13.700 tokens.

b) Normas de futebol colhidas em sites na internet, com 13.500 tokens.

c) Textos eletrônicos disponíveis on-line e colhidos semi-automaticamente com o auxílio da ferramenta BootCat⁸ 107.700 tokens.

d) Notícias sobre o domínio do futebol colhidas manualmente no site Globo esporte.com. 251.300 tokens.

⁸ O BootCat é uma ferramenta que compila corpus semi-automaticamente a partir de uma lista de palavras-chave oferecida pelo pesquisador. Site: <http://bootcat.sslmit.unibo.it/>

Os textos de narrações transcritas utilizados no corpus são provenientes do estudo de doutoramento realizado na Universidade de São Paulo (USP) por Sandra R. Turtelli, intitulado *Estudo da linguagem de um evento esportivo numa abordagem sócio-léxico-computacional*. Além destes textos de língua falada, foram recolhidos e convertidos para o formato txt os textos de língua escrita disponíveis *online* para constituir o corpus deste trabalho. O formato txt foi escolhido por não ter opções de formatação dos textos, facilitando o processamento do corpus por ferramentas computacionais como o *WordSmith Tools*⁹ e *Unitex*¹⁰.

A manipulação do corpus foi feita com o auxílio da ferramenta Unitex. O corpus foi pré-processado pela ferramenta, o mesmo consiste em aplicar as seguintes operações: normalização de separadores, segmentação em unidades lexicais, normalização de formas não ambíguas, segmentação em frases e aplicação dos dicionários.

Ao pré-processar um texto, são aplicados dicionários já existentes no programa: um subconjunto dos dicionários que contêm somente as formas presentes no texto pré-processado pode ser visualizado na tela. Também é criada uma "Token List" que pode ser ordenada pela frequência ou pela ordem alfabética.

Utilizamos algumas ocorrências obtidas na Token List em outra ferramenta do programa, o concordanciador. O mesmo permite buscas de expressões regulares no corpus por meio de filtros morfológicos:

A: Adjetivo	ADV: Advérbio
DET: Determinante	INTJ: Interjeição
N: Substantivo	PREP: Preposição
PRO: Pronome	V: verbo
M: Masculino	F: Feminino
N: Neutro	s: Singular
P: Plural	1,2,3: 1ª, 2ª, 3ª pessoa
P: Presente do indicativo	I: Pretérito imperfeito do indicativo
S: Presente do subjuntivo	T: Imperfeito do subjuntivo
Y: Presente do imperativo	J: Pretérito perfeito
W: Infinitivo	G: Gerúndio
K: Particípio	F: Futuro

Quadro 2 - Lista de códigos dos dicionários do Unitex

Uma expressão regular pode ser:

⁹ WordSmith Tools disponível em: <http://www.lexically.net/wordsmith/>

¹⁰ Unitex disponível em: <http://www-igm.univ-mlv.fr/~unitex/>

- ✓ uma unidade lexical (livro) ou um paradigma lexical (<comer>);
- ✓ a concatenação de duas expressões racionais (eu como);
- ✓ a união de duas expressões racionais (Pedro+Paulo);
- ✓ o asterisco de Kleene de uma expressão regular (muito*).

Ao buscar uma determinada expressão, a ferramenta permite visualizar as ocorrências dentro de seu contexto. Por exemplo, podemos observar o termo *bola* e a co-ocorrência de verbos em sua vizinhança (Figura 5).

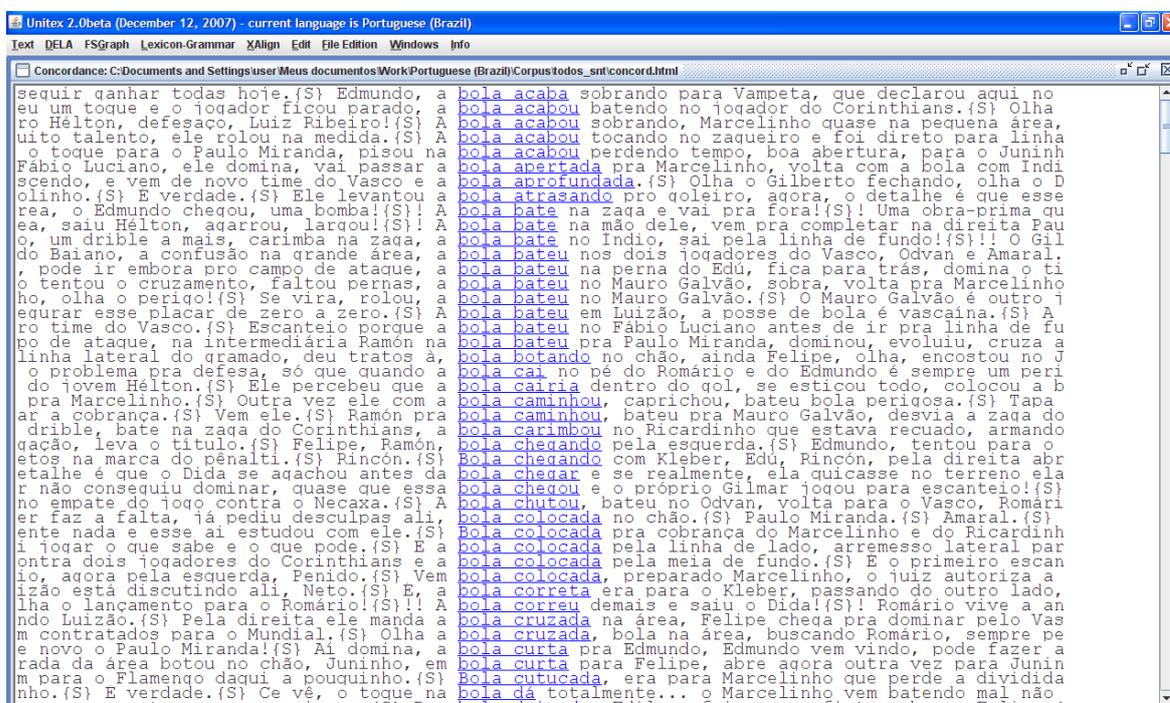


Figura 5 - Concordância de *bola* seguida de um verbo

As classes são delineadas a partir das possibilidades de co-ocorrência das expressões linguísticas.

3.2 Etapa 2: Extração das ocorrências utilizadas na descrição

Nesta etapa utilizamos a ferramenta *Unitex*, que é um conjunto de programas que possibilitam o tratamento de textos em linguagem natural, permitindo análises de *corpus* nos níveis da morfologia, do léxico e da sintaxe por meio de

recursos linguísticos como: *Word List*, *Concordanciador* e *Dicionários Eletrônicos*, que podem auxiliar na busca por termos. Utilizaremos a ferramenta *Word List* para extrair as ocorrências do corpus (Figura 6):

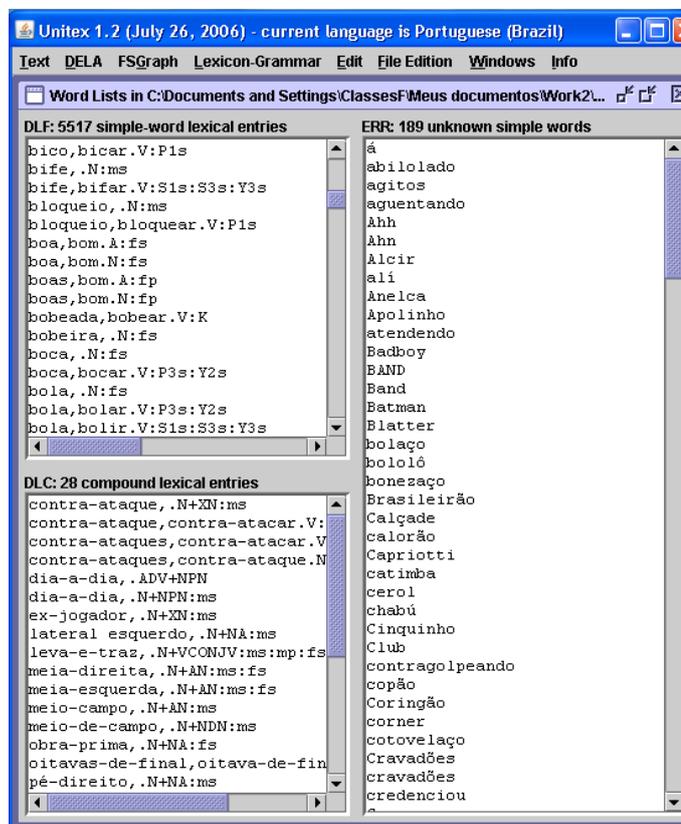


Figura 6 - Word Lists do Unitex

Para a análise de frases simples foram aproveitados os itens lexicais mais frequentes presentes nas fontes pesquisadas. O critério de frequência não foi o único a ser utilizado, os itens lexicais menos frequentes foram aproveitados segundo o critério de relevância para o domínio em questão. Nesta etapa obtivemos uma lista de nomes (N) mais recorrentes:

1 bola	12 goleiro	23 domingo
2 jogo	13 área	24 árbitro
3 time	14 técnico	25 atacante
4 jogador	15 primeiro	26 Copa
5 contra	16 Campeonato	27 decisão
6 gol	17 futebol	28 treinador
7 partida	18 clube	29 fora
8 campo	19 minutos	30 quarta
9 equipe	20 tiro	31 torcida
10 tempo	21 falta	32 lado
11 final	22 título	33 linha

Quadro 3 - Unidades Lexicais mais recorrentes

As ocorrências de nomes próprios foram excluídas dessa lista. A mesma foi utilizada na etapa de aplicação do modelo de Gross.

3.3 Etapa 3: Aplicação do modelo de Gaston Gross

Nesta fase foram analisadas as ocorrências obtidas na etapa 2. A atividade foi desenvolvida com o auxílio do concordanciador do programa Unitex.

Para construirmos cada classe, foi utilizado o modelo de classes de objeto, por meio da análise as possibilidades de co-ocorrência dos itens lexicais:

N0 V N1:

- 54) Alex leva o cartão vermelho.
- 55) Canindé Recebeu cartão amarelo.

N0 V N1 Prep N2:

- 56) (O árbitro) mostra cartão vermelho ao jogador.
- 57) O árbitro aplicou o cartão vermelho no jogador.

Nos exemplos 58 e 59 o argumento *jogador* é selecionado como sujeito pelos operadores *levar cartão* e *receber cartão*, e como complemento do predicado *aplicar cartão* (exemplo 60 e 61) já que, dentro de campo, quem pode *mostrar cartão* e *aplicar cartão* é o árbitro.

- 58) Saiu a bola pela linha de fundo.

- 59) O chute a gol caracteriza-se quando a bola cruza a linha de fundo.
- 60) Lance ordenado desde que a bola tocada pela equipe defensora ultrapasse a linha de fundo.
- 61) Branquinho correu livre até a linha de fundo e tocou para Paulo Roberto marcar.
- 62) Alex chegou até a linha de fundo e cruzou rasteiro.

Nas ocorrências acima, nota-se que a bola *sai pela / cruza / ultrapassa a linha de fundo* (exemplos 62, 63 e 64), enquanto os jogadores *chegam / correm até a linha de fundo* (exemplos 65 e 66). Os operadores *sair, cruzar e passar* acompanhados do locativo *linha de fundo* selecionam o argumento *bola*, enquanto os operadores *chegar e correr* acompanhados do mesmo locativo selecionam como argumento *jogador*. Diante disso as ocorrências abaixo soariam estranhas:

- 63) *A bola chegou até a linha de fundo e cruzou rasteiro.
- 64) *O jogador realmente cruzou a linha de fundo ao tocar no solo.

Podemos pensar então em duas classes de operadores/predicados: uma de predicados que designam ações relacionadas à bola e outra de predicados que designam ações relacionadas ao jogador.

As expressões cristalizadas tiveram tratamento especial nesta etapa, pois aqui consistiu o trabalho de definir quais classes elas evocam:

- 65) O Araújo tem integridade e é Coxa-Branca. Sempre *suou a camisa* em campo.
- 66) Romário suou a camisa com a bola nos pés durante um minuto.

Nos exemplos acima, a expressão *suar a camisa* indica que o jogador se empenhou dentro de campo (mesmo que, no caso do exemplo 66, isso tenha ocorrido por apenas um minuto). Pode-se então adicionar a expressão *suar a camisa* na classe de predicados que designa as *ações* dos jogadores.

4 Resultados

Nesta seção apresentaremos os resultados do trabalho: as classes de objeto do domínio do futebol delineadas pela pesquisa por meio de corpus e análise de co-ocorrência dos itens lexicais. Por fim, demonstraremos como as classes de objeto obedecem à certa hierarquia, organizada pelos predicadores de selecionam de forma apropriada seus argumentos.

4.1 Constituição das classes de objeto

As classes de objeto do domínio do futebol foram constituídas, primeiramente, a partir dos argumentos mais recorrentes no corpus analisado. Inicialmente verificamos as possibilidades de co-ocorrência de *Bola* e *jogador*.

Algumas classes conceituais são relativamente fáceis de serem delineadas, como os casos abaixo:

<função jogador> Funções dos jogadores, com o padrão sintático N0 ser N1 Prep N2, conforme os exemplos:

67) Zanetti é o *goleiro* do Juventude.

68)Tiago foi o *meio-campo* do Botafogo.

<bola> referências à bola, sendo que esta pode *bater*, *explodir* e *desviar* com padrão sintático N0 V Prep N1, conforme os exemplos:

69)A *bola* bate na zaga.

70)A *bola* explodiu no corpo do vascaíno.

71)A *bola* desviou na zaga.

<clube> referências aos clubes, com padrão sintático N0 V N1:

72) O Corinthians venceu o Juventude.

73) Corinthians contrata Everton e Moradei.

<parte do corpo> Partes do corpo de um jogador que podem ser atingidas por outro jogador ou pela bola, com padrão N0 V prep N1 prep N2:

74) A Bola bateu no *peito* do zagueiro.

75) A bola bateu na *mão* de Teco.

76) (A bola) bateu na *canela* do Ramón.

A partir da definição de algumas classes mais simples, tentamos delimitar a vizinhança em que essas classes ocorrem. A princípio, trabalhamos com o verbo bater. Alguns traços são curiosos para este verbo:

77) Corinthians bate o Juventude na estreia.

78) Os chilenos começaram a *bater*, algo que fazem razoavelmente bem.

79) Gilsinho aproveitou a bobeira da zaga adversária, que parou para pedir impedimento, e *bateu* forte no canto direito do goleiro.

80) Márcio *bateu* para o gol.

No exemplo 77, *bater* indica a ação de *vencer/derrotar*. Em 78, o verbo é aplicado no sentido de *agredir/atacar/investir* contra, já em 79 e 80, o verbo pede necessariamente um complemento de lugar. Podemos então pensar nas seguintes formalizações:

<clube>	bater/vencer	<clube>
<clube>+<jogador>	bater/agredir	<clube>+<jogador>
<jogador>	bater	canto esquerdo
		canto direito
		trave
		gaveta
		<parte do corpo>

Após observar as primeiras ocorrências, começamos a delinear classes de argumentos e classes de predicados:

- a) Argumentos substantivos;
- b) Argumentos adjetivos;

- c) Predicados adjetivais;
- d) Predicados nominais;
- e) Predicados verbais.

Criamos tabelas para fins de registro das ocorrências encontradas durante a pesquisa. Assim, os argumentos selecionados pelos operadores (predicados) foram distribuídos em duas tabelas: argumento adjetivo e substantivo, (apêndices A e B). Esses argumentos constituem as classes de objeto. Já as últimas três tabelas (apêndices C, D e E), predicados adjetivais, nominais e verbais, possuem os padrões sintáticos para as ocorrências dos operadores:

N0 V	Flávio brilha.
N0 V N1	O Corinthians venceu o Juventude.
N0 V Prep N1	A bola explodiu no Amaral.
N0 Vsup Adj	A bola é esférica.
N0 Vsup N1	Marcos é o goleiro.
N0 V N1 Prep N2	O atacante recebeu a bola de Jorge.
N0 V prep N1 prep N2	A bola bateu no peito do zagueiro.

4.1.1 Argumentos substantivos

Conforme tabela 1 (apêndice 1) os argumentos foram listados e classificados por categoria gramatical (nn, nm e nf); traço semântico (hum, -hum, inc, humcoll) e a classe em que o mesmo se encaixa. As classes são representadas por colchetes angulares <>. Os nomes das classes foram escolhidos de maneira a facilitar a identificação da classe.

- **Classe <bola>**

Formas de nomear a bola, argumento inanimado+concreto. Para delinear esta classe foram selecionados os operadores *rolar*, *desviar* e *bater*.

Ex.:

81) A bola rolou, rolou e gol.

82) A bola desvia na defesa do Náutico.

83) A bola bate na trave.

Ocorrência	C. G.	Traços	Classe
bola	nf	inc	<bola>

Quadro 4 - Classe de objeto <bola>

▪ **Classe <clube>**

Formas de nomear os clubes. Já que os clubes são formados não somente por jogadores, mas também por dirigentes e comissão técnica, esta classe recebe o traço humcoll (humano coletivo), e é formada por substantivos próprios. Os operadores que utilizamos para delinear esta classe foram *vencer/perder* e *contratar*.

Ex.:

84) *Figueirense vence o gama.*

85) O time perdeu para o Atlético.

86) Corinthians contrata Everton e Moradei.

Ocorrência	C. G.	Traços	Classe
ABC	nm	hum + humcoll	<clube>

Quadro 5 - Exemplo classe de objeto <clube>

▪ **Classe <jogador>**

Formas de nomear o jogador. Esta classe é formada por nomes próprios de Jogadores.

Ex.:

87) *Osmar* fez dois gols ainda no primeiro tempo.

Ocorrência	C. G.	Traços	Classe
NP		hum	<jogador>

Quadro 6 - Classe de objeto <jogador>

▪ **Classe <árbitro>**

Formas de nomear o árbitro. Os operadores *mostrar* e *aplicar* foram utilizados para delinear esta classe.

Ex.:

88) Ele gritou e o árbitro apitou.

89) O árbitro acompanha de perto, marca, aplica bem os cartões.

Ocorrência	C. G.	Traços	Classe
árbitro	nm	hum	<árbitro>
juiz	nm	hum	<árbitro>

Quadro 7 - Classe de objeto <jogador>

▪ **Classe <instrumento>**

Instrumentos utilizados pelo árbitro e árbitro auxiliar em uma partida de futebol.

Para esta classe foram selecionados argumentos que possuem o traço semântico inc (inanimado+concreto) que são utilizados pelos árbitros dentro de campo. Os operadores que utilizamos para delinear esta classe foram *levantar*, *apitar*, *mostrar* e *aplicar*.

Ex.:

90) O Juiz levanta o cartão vermelho.

91) O juiz aplica cartão amarelo em Lugano.

92) O árbitro mostra Cartão Amarelo para ele.

Ocorrência	C. G.	Traços	Classe
apito	nm	inc	<acessório>
bandeirinha	nf	inc	<acessório>
cartão amarelo	nm	inc	<acessório>
cartão vermelho	nm	inc	<acessório>

Quadro 8 - Classe de objeto <acessório>

Apesar de o árbitro e o árbitro auxiliar utilizarem acessórios diferentes, optamos por não separar os instrumentos dos mesmos em duas classes distintas, uma vez que a classe que conteria acessórios utilizados pelo árbitro auxiliar teria somente um componente, a *bandeira*. Algo que nos chamou atenção foi o fato de o diminutivo de *bandeira* ocorrer também para referenciar árbitro-auxiliar:

93)O *bandeirinha* uruguaio marcou a posição irregular

94) Em seu primeiro lance, Jonas empatou em posição legal, mas a *bandeirinha* Ana Paula anulou.

▪ **Classe <campo>**

Formas de nomear o campo. Esta classe de argumentos possuem traço semântico locativo (loc). Utilizamos os operadores *sair de*, *deixar* e *entrar em* para estabelecer a classe.

Ex.:

95) O número 9, Ramón, está deixando o *gramado* neste momento.

96) O Jogador sairá do campo de jogo.

97) O substituto entra no campo

Ocorrência	C. G.	Traços	Classe
campo	nm	loc	<campo>
gramado	nm	loc	<campo>
tapete verde	nm	loc	<campo>

Quadro 9 - Classe de objeto <campo>

▪ **Classe <divisão>**

Divisões de um campeonato. Os nomes desta classe possuem o traço semântico loc (Local). Os operadores que utilizamos para delinear esta classe foram *subir para*, *cair para* e *disputar*.

Ex.:

98) O Juventus disputa a série B.

99) O Corinthians caiu para a segunda divisão.

100) O Goiás subiu para a série A.

Ocorrência	C. G.	Traços	Classe
primeira divisão	nf	inc	<divisão>
segunda divisão	nf	inc	<divisão>
série a1	nf	inc	<divisão>
série a2	nf	inc	<divisão>
série a3	nf	inc	<divisão>
série B	nf	inc	<divisão>

Quadro 10 - Classe de objeto <divisão>

▪ **Classe <estádio>**

Estádios de futebol. Os nomes pertencentes a esta classe também possuem o traço semântico loc. Utilizamos os operadores *lotar* e *comparecer ao* para estabelecer a classe.

Ex.:

- 101) Os torcedores tentam lotar o Olímpico.
102) Os torcedores compareceram ao estádio.

Ocorrência	C. G.	Traços	Classe
Beira-Rio	nm	loc	<estádio>

Quadro 11 – Exemplo classe de objeto estádio

▪ **Classe <faltas>**

Faltas cometidas pelo jogador. Para esta classe foram selecionados argumentos que possuem o traço semântico inc (inanimado+concreto). Os operadores que utilizamos para delinear esta classe foram *dar carrinho*, *cometer*, *fazer falta*.

Ex.:

- 103) Viola deu um carrinho na bola.
104) Irineu cometeu sua terceira falta.

Ocorrência	C. G.	Traços	Classe
carrinho	nm	inc	<falta>
falta	nf	inc	<falta>
insulto	nm	inc	<falta>

Quadro 12 - Classe de objeto <falta>

▪ **Classe <parte do corpo>**

Partes do corpo humano. Para esta classe foram selecionados argumentos que possuem o traço semântico inc (inanimado+concreto). Utilizamos os operadores *dominar*, *matar* e *tirar de* para estabelecer a classe.

Ex.:

- 105) Marcelo dominou a bola no peito.
106) Kléber matou a bola no peito.
107) Vampeta tirou de cabeça.

Ocorrência	C. G.	Traços	Classe
Braço	Nm	Inc	<parte do corpo>
Cabeça	Nf	Inc	<parte do corpo>
Costas	Nf	Inc	<parte do corpo>
Cotovelo	Nm	Inc	<parte do corpo>
Joelho	Nm	Inc	<parte do corpo>
Mão	Nf	Inc	<parte do corpo>
Nariz	Nm	Inc	<parte do corpo>
Ombro	Nm	Inc	<parte do corpo>
Pé	Nm	Inc	<parte do corpo>
Peito	nm	inc	<parte do corpo>
Perna	Nf	Inc	<parte do corpo>
Tronco	Nm	Inc	<parte do corpo>

Quadro 13 - Classe de objeto <parte do corpo>

▪ **Classe <parte_do_campo>**

Locais demarcados no campo. Para esta classe foram selecionados argumentos que possuem o traço semântico loc (local). Utilizamos os operadores *cruzar*, *ultrapassar* e *ajeitar de* para estabelecer a classe.

Ex.:

- 108) A bola cruza a linha e balança a rede.
- 109) A bola ultrapassou ou não a linha de gol?
- 110) Ademar ajeita a bola na marca do pênalti para a cobrança.

Ocorrência	C. G.	Traços	Classe
linha de fundo	nf	loc	<parte do campo>
linha de gol	nf	loc	<parte do campo>
linha de meta	nf	loc	<parte do campo>
linha da meia lua	nf	loc	<parte do campo>
linha dos 18m	nf	loc	<parte do campo>
linha dos 6m	nf	loc	<parte do campo>
linha lateral	nf	loc	<parte do campo>
marca de pênalti	nf	loc	<parte do campo>

Quadro 14 - Classe de objeto <partes do campo>

▪ **Classe <parte do gol>**

Partes integrantes do gol. Para esta classe foram selecionados argumentos que possuem o traço semântico loc (local). Utilizamos os operadores *bater em*, *cabecear em* e *acertar* para estabelecer a classe.

Ex.:

111) A bola bateu na rede pelo lado de fora.

112) Alan Kardek acertou o travessão.

Ocorrência	C. G.	Traços	Classe
primeiro poste	nm	loc	<parte do gol>
segundo poste	nm	loc	<parte do gol>
trave	nf	loc	<parte do gol>
barra transversal	nf	loc	<parte do gol>
Rede	nf	loc	<parte do gol>

Quadro 15 - Classe de objeto parte do gol

▪ **Classe <local do gol>**

Local onde o gol pode ser marcado. Para esta classe foram selecionados argumentos que possuem o traço semântico loc. Utilizamos os operadores *acertar* e *entrar* para estabelecer a classe.

113) A bola pegou efeito, fez curva e entrou no canto esquerdo de Bruno.

114) Ari acertou a gaveta do goleiro Fábio.

Ocorrência	C. G.	Traços	Classe
canto direito	nm	loc	<local do gol>
canto esquerdo	nm	loc	<local do gol>
gaveta	nf	loc	<local do gol>
meio do gol	nm	loc	<local do gol>
ângulo	nm	loc	<local do gol>

Quadro 16 - Classe de objeto <local do gol>

▪ **Classe <penalidade/equipe>**

Penalidades atribuídas à equipe. Para esta classe foram selecionados argumentos que possuem o traço semântico inc (inanimado+concreto). Utilizamos os operadores *marcar*, *cobrar* e *conceder* para estabelecer a classe.

Ex.:

- 115) O árbitro marcou *pênalti* em Osmar.
 116) O árbitro concedeu o *pênalti*.
 117) Wellington Monteiro cobrou falta na área.

Ocorrência	C. G.	Traços	Classe
pênalti	nm	inc	<penalidade/equipe>
tiro livre direto	nm	inc	<penalidade/equipe>
tiro livre indireto	nm	inc	<penalidade/equipe>

Quadro 17 - Classe de objeto <penalidade/equipe>

▪ **Classe <penalidade/jogador>**

Penalidades atribuídas ao jogador. Para esta classe foram selecionados argumentos que possuem o traço semântico inc (inanimado+concreto). Utilizamos os operadores *levar*, *receber*, *aplicar* e *mostrar* para estabelecer a classe.

Ex.:

- 118) Adilson leva cartão.
 119) Canindé recebeu cartão amarelo.
 120) (Ele) aplicou o cartão vermelho no jogador.
 121) O árbitro apitou e depois mostrou o cartão amarelo para o lateral.

Ocorrência	C. G.	Traços	Classe
advertência	nf	inc	<penalidade/jogador>
cartão amarelo	nm	inc	<penalidade/jogador>
cartão vermelho	nm	inc	<penalidade/jogador>
expulsão	nf	inc	<penalidade/jogador>

Quadro 18 - Classe de objeto <penalidade/jogador>

- **Classe <período/jogo>**

Períodos que compõem um jogo. Utilizamos os operadores *terminar*, *acabar* e *iniciar* para estabelecer a classe.

-

Ex.

- 122) Termina o primeiro tempo.
- 123) O árbitro encerra o primeiro tempo.

Ocorrência	C. G.	Traços	Classe
primeiro tempo	nf	Inc	<período/jogo>
segundo tempo	nm	Inc	<período/jogo>
prorrogação	nm	Inc	<período/jogo>
etapa inicial	nf	Inc	<período/jogo>
etapa final	nm	Inc	<período/jogo>

Quadro 19 - Classe de objeto <período/jogo>

- **Classe <recompensa>**

Recompensa adquirida pelo resultado de um jogo ou campeonato. Utilizamos os operadores *exibir*, *ganhar* e *erguer* para estabelecer a classe.

Ex.:

- 124) Fla supera pressão e ergue a taça.
- 125) Messi exibe o troféu de melhor da Europa.

Ocorrência	C. G.	Traços	Classe
medalha	nf	inc	<recompensa>
taça	nf	inc	<recompensa>
troféu	nm	inc	<recompensa>

Quadro 20 - Classe de objeto <recompensa>

- **Classe <tática>**

Táticas de jogo.

Ex.:

- 126) O time é escalado no 4-4-2.
- 127) O Sergipe jogou no 3-5-2.

Ocorrência	C. G.	Traços	Classe
3-5-2	nm	inc	<tática>
4-2-4	nm	inc	<tática>
4-3-3	nm	inc	<tática>
5-3-2	nm	inc	<tática>
5-4-1	nm	inc	<tática>

Quadro 21 - Classe de objeto <tática>

▪ **Classe <uniforme>**

Peças integrantes do uniforme dos jogadores. Utilizamos os operadores *colocar*, *vestir* e *calçar* para estabelecer a classe.

Ex.:

128) O jogador vestiu a camisa 3 alvinegra.

129) Pibe colocou o uniforme, calçou a chuteira e participou do treino da seleção argentina.

Ocorrência	C. G.	Traços	Classe
camiseta	nf	inc	<uniforme>
caneleira	nf	inc	<uniforme>
chuteira	nf	inc	<uniforme>
meia	nf	inc	<uniforme>
moletom	nm	inc	<uniforme>
shorts	nm	inc	<uniforme>

Quadro 22 - Classe de objeto <uniforme>

4.1.2 Argumentos adjetivos

Assim como os argumentos substantivos, os argumentos adjetivos foram classificados por *categoria gramatical*, *traço semântico* e *classe* (apêndice 2):

▪ **Classe <posição jogador>**

Posições dos jogadores em campo.

Ex.:

130) O *atacante* Borges brilhou ao lado de Dagoberto.

131) A outra baixa é o *lateral-esquerdo* João Victor.

Ocorrência	C. G.	Traços	Classe
goleiro	nm	hum	<posição jogador>
zagueiro	nm	hum	<posição jogador>
lateral direito	nm	hum	<posição jogador>
lateral esquerdo	nm	hum	<posição jogador>
ala esquerdo	nm	hum	<posição jogador>
ala direito	nm	hum	<posição jogador>
volante	nm	hum	<posição jogador>
meio-de-campo	nm	hum	<posição jogador>
meia-armador	nm	hum	<posição jogador>
meia-atacante	nm	hum	<posição jogador>
atacante	nm	hum	<posição jogador>
ponta-de-lança	nm	hum	<posição jogador>
centro-avante	nm	hum	<posição jogador>
capitão	nm	hum	<posição jogador>
titular	nm	hum	<posição jogador>
batedor de pênalti	nm	hum	<posição jogador>
batedor de escanteio	nm	hum	<posição jogador>

Quadro 23 - Classe de objeto <função/jogador>

▪ **Classe <torcedor>**

Formas de nomear um torcedor. Utilizamos os operadores *acompanhar*, *sofrer*, *lotar* e *vaiar* para estabelecer a classe.

Ex.:

- 132) Os corintianos acompanharam o jogo.
 133) A torcida tricolor lotou o Olímpico.

Ocorrência	C. G.	Traços	Classe
verdão	nm	hum + humcoll	<torcedor>
gavião	nm	hum + humcoll	<torcedor>
santista	nm	hum + humcoll	<torcedor>
são paulino	nm	hum + humcoll	<torcedor>
palmeirense	nm	hum + humcoll	<torcedor>
tricolor	nm	hum + humcoll	<torcedor>
tricolor das laranjeiras	nm	hum + humcoll	<torcedor>
flamenguista	nm	hum + humcoll	<torcedor>
botafoguense	nm	hum + humcoll	<torcedor>

vascaíno	nm	hum + humcoll	<torcedor>
cruzeirense	nm	hum + humcoll	<torcedor>
atleticano	nm	hum + humcoll	<torcedor>
peixe	nm	hum + humcoll	<torcedor>
gremista	nm	hum + humcoll	<torcedor>
colorado	nm	hum + humcoll	<torcedor>
tigrão	nm	hum + humcoll	<torcedor>
coxa	nm	hum + humcoll	<torcedor>
corinthiano	nm	hum + humcoll	<torcedor>
ponte pretano	nm	hum + humcoll	<torcedor>

Quadro 24 - Classe de objeto <torcedor>

Ao fim da análise foram delineadas 21 classes de objeto:

01	<árbitro>	Formas de nomear o árbitro.
02	<bola>	Formas de nomear a bola.
03	<campo>	Formas de nomear o campo.
04	<clube>	Formas de nomear os clubes.
05	<divisão>	Divisões de um campeonato.
06	<estádio>	Estádios de futebol.
07	<faltas>	Faltas cometidas pelo jogador.
08	<jogador>	Formas de nomear o jogador.
09	<instrumento>	Instrumentos utilizados pelos árbitros em uma partida de futebol.
10	<local do gol>	Local onde o gol pode ser marcado.
11	<parte do corpo>	Partes do corpo humano.
12	<parte do campo>	Locais demarcados no campo.
13	<parte do gol>	Partes integrantes do gol.
14	<penalidade/equipe>	Penalidades cometidas pela equipe.
15	<penalidade/jogador>	Penalidades cometidas pelo jogador.
16	<período do jogo>	Períodos que compõem um jogo.
17	<posição jogador>	Posições dos jogadores em campo.
18	<recompensa>	Recompensa adquirida pelo resultado de um jogo ou campeonato.
19	<tática>	Táticas de jogo.
20	<torcida>	Formas de nomear um torcedor.
21	<uniforme>	Peças integrantes do uniforme dos jogadores.

Quadro 25 - Classes de Objeto Futebol

4.1.3 Predicados adjetivais

A tabela para predicados adjetivais (apêndice C) foi dividida nos seguintes campos: *N0*, *Vsup* (*verbo-suporte*), *Npred* (*nome predicativo*) e *N1*(quando necessário).

N0	Vsup/ Vaux	N	Prep	N2
<jogador>	Ser	Hábil	+	<bola>
<clube>	Ser	escalado	+	<tática>

Quadro 26 - Exemplo tabela predicados adjetivais

4.1.4 Predicados nominais

Assim como a tabela de predicados adjetivais, a tabela de predicados nominais (apêndice D) foi separada nos seguintes campos: *Operador*, *N0*, *N1*(quando necessário), *V auxiliar/suporte* e *classe*.

N0	Vsup/ Vaux	N	Prep	N2
<jogador>	cometer	falta	-	-
<jogador>	dar	carrinho	+	<jogador>

Quadro 27 - Exemplo tabela predicados nominais

4.1.5 Predicados verbais

Os predicados verbais (apêndice E) ocorreram acompanhados de *N1* e *N2* (quando necessário).

Algumas ocorrências são listadas abaixo:

- 134) Ricardinho *chutou* duas bolas para fora em faltas da entrada da área.
- 135) O Palmeiras *venceu* o botafogo.
- 136) Alan Delon *lançou* a bola para Rafael Toledo.

N0	v	Prep	N1	Prep	N2
<bola>	entrar	+	<local do gol>	-	-
<jogador>	acertar	-	<bola>	+	<local do gol>

Quadro 28 - Exemplo tabela predicados verbais

Também estão listadas, no apêndice F, as expressões cristalizadas verbais que encontradas no corpus e nas pesquisas nos motores de busca *Google e Yahoo Search*.

O presente trabalho chegou ao resultado de 20 classes de objeto do domínio do futebol, porém, para o francês, este número é um pouco maior. Sabatier (1998) delineou 32 classes de objeto. Acreditamos que a diferença entre o número de classes se deu por conta da metodologia utilizada no trabalho. Enquanto nosso trabalho se fez sobre a luz da linguística de corpus, com ocorrências reais de uso, Sabatier optou por um método baseado em introspecção, em que são listadas todas as formas possíveis de ocorrência, sem que essas ocorrências sejam testadas por meio de corpus. No entanto, não identificamos esse item como um problema, uma vez que podemos identificar que existe, sim, uma hierarquia entre as classes de objeto. É o que passaremos agora a discutir: a hierarquia nas classes de objeto e a ancoragem destas em uma ontologia.

4.2 Hierarquia das classes de objeto e ancoragem ontológica

Já vimos anteriormente que classes ontológicas e as classes de objeto não pertencem ao mesmo nível, pois a primeira trata de conceitos e a segunda trata de realizações linguísticas. Uma ontologia é constituída de uma organização hierárquica de classes/conceitos. Essa organização é regida por axiomas/princípios que determinam a hierarquia ontológica. O que tentamos demonstrar com esta pesquisa é a possibilidade de se obter classes semânticas por meio de critérios sintáticos, bem como a possibilidade de estabelecer uma hierarquia entre essas classes.

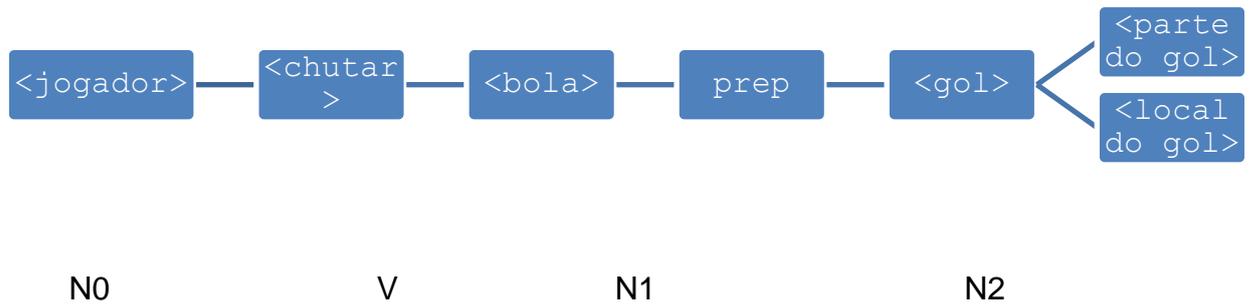
Para isso, apresentamos os exemplos abaixo

- 137) João chutou a bola no gol.
- 138) João chutou a bola na gaveta.
- 139) João chutou a bola na trave.

140) João chutou a bola no primeiro pau.

Podemos observar que o predicado *chutar* pode selecionar como argumentos a classe <jogador> (N0), a classe <bola> (N1), e as classes <parte do gol> e <local do gol>, que são especificações do gol.

Assim temos:



No entanto, não é somente a *bola* que pode ser atingida por um *chute* do *jogador*. O jogador pode atingir também outro jogador. Assim, além da formalização:

Chutar (N0hum+N1hum)

que pode realizar-se em:

141) João chutou Marcos.

pode ser acrescentado um complemento circunstancial de lugar (N2):

Chutar (N0hum+N1hum+prep+N2inc)

142) João chutou marcos na canela.

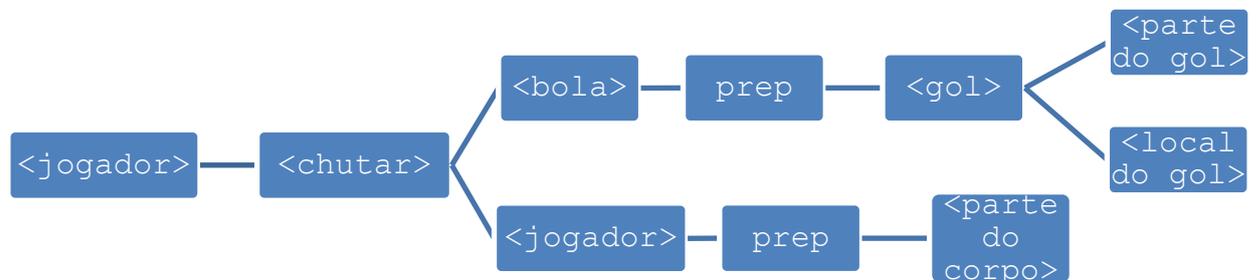
143) João chutou Marcos na mão.

144) João chutou Marcos na cabeça.

Dessas observações podemos extrair a seguinte estrutura:



Aproximando as duas estruturas, temos:



Ao observar também a classe de objeto <uniforme> dos jogadores, podemos extrair as seguintes construções:

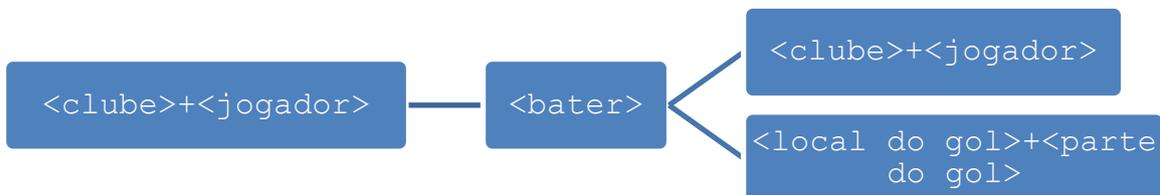
- 145) João vestiu a camisa.
- 146) João vestiu o calção.
- 147) João vestiu a meia
- 148) João calçou a chuteira.

Para delinear a classe <uniforme> precisaríamos de uma classe de predicados que a selecionaria de maneira apropriada. Porém, o predicado *calçar* não se aplica para os argumentos camisa, meia e calção, assim como o predicado *vestir* não se aplica para o argumento chuteira. Para solucionar este problema, é necessário acrescentar uma regra de restrição entre os membros da classe <uniforme> diferenciando o que pode ser *vestido* e o que pode ser *calçado*:

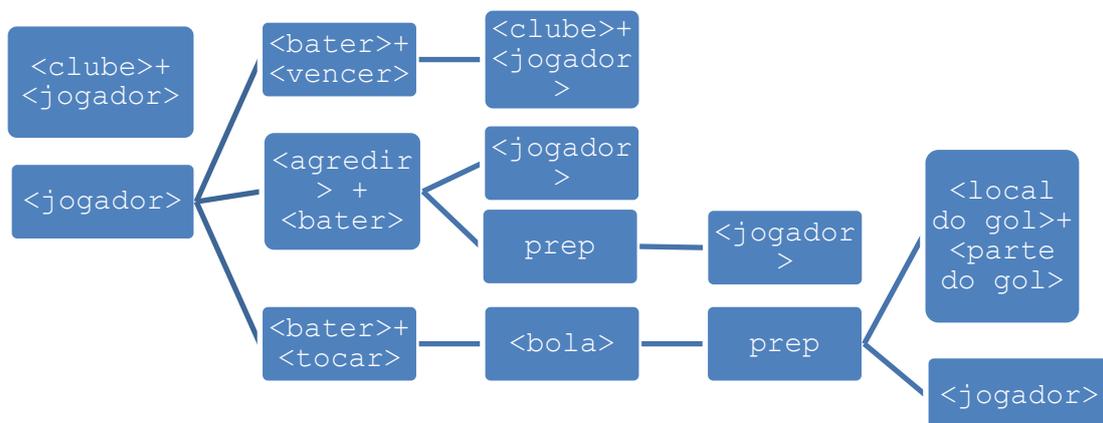


Os operadores são os responsáveis pelo estabelecimento de relação entre seus argumentos. No exemplo acima, *vestir* é responsável por estabelecer a relação entre o argumento *jogador* ao argumento *uniforme*.

Verifiquemos agora as relações do operador *bater*.



Apenas com a representação acima não conseguimos verificar quais são os argumentos selecionados pelo operador *bater*. O mesmo, por si só, não é capaz de selecionar um feixe definicional de argumentos. Para a identificação desses argumentos, é necessária a utilização de um conjunto de operadores:



Por meio da representação acima, conseguimos gerar as frases *O Vasco bateu/venceu o Fluminense*, *O zagueiro bateu no atacante*, *O zagueiro*

agrediu o atacante, O goleiro tocou a bola para o zagueiro, O atacante bateu a bola na trave.

Podemos observar que as classes de objeto obedecem certa hierarquia, relacionada pelos operadores que selecionam de forma apropriada seus argumentos. Tendo essa hierarquia em mente, acreditamos que o modelo de classes de objeto pode ser utilizado para ancoragem ontológica das ocorrências, o que também poderia ser feito para as classes de objeto construídas para o francês e para o grego. Para isso, deveria ser construída, à parte, uma ontologia de conceitos aos quais as classes de objeto remeteriam:

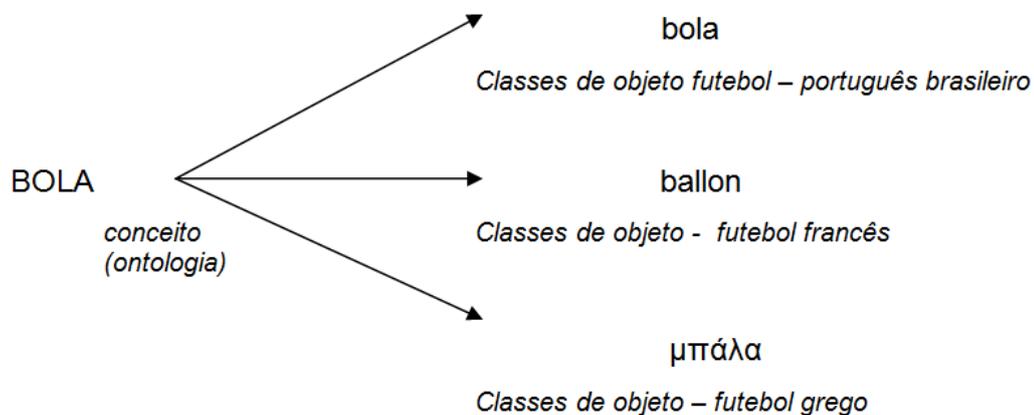


Figura 7 - Conceito BOLA

BOLA, enquanto conceito da ontologia, é a referência que os falantes possuem do objeto do mundo real. A referência é ativada em nosso pensamento quando há a realização linguística do conceito: *João chutou a bola para José.* A ancoragem ontológica facilitaria, por exemplo, pesquisas que envolvem tradução automática.

O número relativamente pequeno de classes de objeto delineadas é insuficiente para dar conta de toda a hierarquia de conceitos do domínio do futebol. No entanto, cremos que os exemplos aqui mostrados comprovam a possibilidade de delineamento de classes semânticas por meio de critérios sintáticos. É necessário destacar que a construção de um módulo ontológico demandaria um novo trabalho, o que acarretaria um debate metodológico acerca da construção de ontologias.

Para chegarmos aos resultados deste trabalho foram realizadas várias discussões acerca da relação entre *classes ontológicas* / *classes de objeto*. O trabalho teve início com a montagem do corpus. Em seguida o corpus foi observado com o auxílio de ferramentas computacionais, foi feita a delimitação das classes, repertoriando as ocorrências do corpus, e na última fase foi verificada a possibilidade de estabelecer hierarquias entre as classes.

Além das classes apresentadas nesta seção, nos apêndices apresentamos o resultado do trabalho: as tabelas de argumentos substantivos, argumentos adjetivos, predicados adjetivais, predicados nominais e predicados verbais. As tabelas de predicados nominais, adjetivais e verbais possuem os padrões sintáticos das classes de objeto. Estas últimas estão representadas nas tabelas de argumentos substantivos e adjetivos.

Considerações finais

Desde o levantamento do corpus, buscando vários gêneros textuais do domínio do futebol até o início do delineamento das classes de argumentos e classes de predicados (argumentos substantivos, argumentos adjetivos, predicados adjetivais, predicados nominais, predicados verbais), levantamos a hipótese de que o modelo de Gross pode ser aplicado à construção de uma ontologia. Nossa hipótese inicial foi confirmada, uma vez que constatamos a hierarquia existente entre os itens lexicais das classes.

Quando falamos de ontologia, tratamos de um assunto amplamente discutido no meio acadêmico, mas entendemos que ainda se faz necessário um método que normatize sua construção. O que temos hoje é uma série de métodos que divergem entre si, e quando um método é eficaz em uma determinada fase da construção, deixa a desejar em outras fases.

Em se tratando de classes de objeto, consideramos que o trabalho de repertoriar frases para a constituição das classes foi fundamental, uma vez que somente testando o modelo de Gross chegaríamos à conclusão que alcançamos no término do trabalho: é possível a ancoragem das classes de objeto aos conceitos de uma ontologia, desde que seja delineado um número de classes de objeto relativamente grande para dar conta de grande parte das hierarquias presentes em uma ontologia.

Além disso, novas teorias surgem a cada dia, buscando formas diferentes de descrição da língua. Quando falamos de descrição, o modelo testado atende à expectativa inicial proposta por Gross, que é a criação de classes semânticas a partir de critérios sintáticos.

Ademais, é necessário verificar até que ponto o método baseado em linguística de corpus foi eficiente para o desenvolvimento deste trabalho. Um corpus de 460 mil ocorrências, inicialmente caracterizado como médio, não foi suficientemente representativo para verificar todas as relações existentes entre operadores e argumentos do domínio do futebol. Acreditamos que a baixa frequência de expressões cristalizadas se deu por conta do tamanho reduzido do corpus aliado ao fato de que as mesmas não ocorrem com frequência em textos escritos, sendo mais comumente utilizadas na linguagem oral.

Dois obstáculos que dificultaram o desenvolvimento do modelo de classes de objeto foram: a escassa literatura encontrada acerca do modelo proposto por Gross (1994); e o pouco ou nenhum conhecimento sobre o método por parte dos demais pesquisadores. No entanto, discussões acerca desta pesquisa em congressos e seminários foram fatores importantes para a sua conclusão, pois nestes encontros obtivemos a oportunidade de dialogar com pesquisadores sobre a abordagem adotada no trabalho.

Como última observação podemos apontar a abertura que este trabalho dá a novas pesquisas, a partir da constatação da possibilidade de se obter classes semânticas a partir de critérios sintáticos, bem como a possibilidade de se estabelecer uma hierarquia entre as classes de objeto e ancoragem em uma ontologia. O resultado final obtido abre portas, por exemplo, para a ampliação das classes de objeto obtidas e para a realização de testes de tradução automática a partir de uma comparação com classes de objeto estabelecidas para o francês.

Referências Bibliográficas

ALUÍSIO, S. M.; ALMEIDA, G. M. B. O que é e como se constrói um corpus? Lições aprendidas na compilação de vários corpora para pesquisa linguística. **Calidoscópico** (UNISINOS), v. 4, p. 156-178, 2006.

ANASTASSIADIS-SYMEONIDIS. La classe des défauts humains en grec moderne. **BULAG : Lexique, Syntaxe et Semantique – melanges offertes à Gaston Gross à l'occasion de son 60 anniversaire**. Centre Lucien Tesniere. Numéro Hors Série, 2000.

ARIEL, M. The Demise of a Unique Concept of Literal Meaning. **Journal of Pragmatics**, v.34, n.4, p. 361-402, 2002.

BEALE, NIRENBURG; MAHESH. Semantic Analysis in the Mikrokosmos Machine Translation Project", *in Proceedings of the Second Symposium on Natural Language Processing*. Kaser Sart University, Bangkok, Tailandia., 1995.

BLACKBURN, S. **Dicionário Oxford de Filosofia**. Jorge Zahar editor. Rio de Janeiro, 1997.

BERBER SARDINHA, T. **Linguística de corpus**. São Paulo: Manole, 2004.

BORST, W. N. **Construction of engineering ontologies**. Holanda, 1997. Tese (Doutorado). Disponível em: <http://www.ub.utwente.nl/webdocs/inf/1/t0000004.pdf>
Acesso em: 10/maio/2009.

CARVALHO, N. F. Estruturas semânticas no Léxico do Futebol. **Alfa**, 40: 75-102, 1996.

DIAS-DA-SILVA, B. C. **A face tecnológica dos estudos da linguagem: o processamento automático das línguas naturais**. Tese (Doutorado em Letras) - Faculdade de Ciências e Letras, Universidade Estadual Paulista, Araraquara, 1996.

DIAS-DA-SILVA, B.C. O estudo Linguístico-Computacional da Linguagem. **Letras de Hoje**, Vol. 41, N.2, pp. 103-138, 2006.

Dias-da-Silva, B.C. *et al.* **Introdução ao Processamento das Línguas Naturais e Algumas Aplicações**. Série de Relatórios Técnicos do NILC, NILC-TR-07-10. São Carlos-SP, 2007, 121p.

Di Felippo, A. **Delimitação e alinhamento de conceitos lexicalizados no inglês norte-americano e no português brasileiro**. Tese de Doutorado em Linguística e Língua Portuguesa. Faculdade de Ciências e Letras (FCL), Universidade Estadual Paulista (UNESP). Araraquara, São Paulo, Agosto, 2008.

FRANÇA, P. C. Conceitos, classes e/ou universais: como que é que se constrói uma ontologia?. **Linguamática**. n. 01. p.105-121, 2009.

GROSS, G. Classes d'objet et description des verbes. **Langages**, n.115 p.15-30, 1994.

GROSS, M. Une classification des phrases "figées" du français. **Revue Québécoise de Linguistique**, Montréal: UQAM, v. 11, n. 2, p. 151-185, 1982.

GUARINO, N. **Formal Ontology and Information Systems**. 1998. <http://www.loacnr.it/Papers/FOIS98.pdf>.

GUARINO, N. GIARETTA, Pierdaniele. **Ontologies and Knowledge Bases. Towards Terminological Clarification**. 1995. <http://www.loacnr.it/Papers/KBKS95.pdf>.

GRUBER, T. 1993. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. *in International Journal Human-Computer Studies* Vol. 43, Issues 5-6, Novembre, 1995. pp.907-928.

HARRIS, Z. **Notes du cours de syntaxe**. Paris: Le Seuil, 1976.

HOVY, E. Methodologies for the Reliable Construction of Ontological Knowledge. **Proceedings of the 13th Annual International Conference on Conceptual Structures**, 2005.

JACKENDOFF, R. S. **Semantics and cognition**. Cambridge: MIT Press, 1983.

LYONS, J. **Semântica I**. Lisboa: Presença, 1980.

LERAT, P. Les emplois juridiques dans le lexique français. **BULAG : Lexique, Syntaxe et Semantique – melanes offerts à Gaston Gross à l'occasion de son 60 anniversaire**. Lucern Tesniere. Numéro Hors Série, 2000.

LE PESANT, D. Les compléments nominaux du verbe lire, une illustration de la notion de «classe d'objets». **Langages**, n.115, p. 31-46, 1994.

LE PESANT, D.; MATHIEU-COLAS, M.. Introduction aux classes d'objet. **Langages** n.131 p. 6-33. 1998.

MAHESH, K; NIRENBURG, S. Semantic Classification for Practical Natural Language Processing. **Proceedings of the Sixth ASIS SIG/CR Classification Research Workshop: An Interdisciplinary Meeting**. Chicago, Illinois, 1995a.

MAHESH, K; NIRENBURG, S. A Situated Ontology for Practical NLP. **Proceedings of the Workshop on Basic Ontological Issues in Knowledge Sharing**. Montreal, Canada: International Joint Conference on Artificial Intelligence. Montreal, Canada, 1995b.

MOUSTAKI, A.; DIMITRIADI, A. Un lexique-grammaire du football: Étude contrastive du grec moderne et du français . **Lingvisticæ Investigationes** 29(2), p. 275–311, 2006.

NEVES, M. H. M. *Gramática de usos do português*. São Paulo: Ed. Unesp, 2000.

OLIVEIRA, A. E. B. **A metáfora e a sua representação em sistemas de processamento automático de línguas naturais**. Dissertação (Mestrado em Linguística e Língua Portuguesa) - Faculdade de Ciências e Letras, Universidade Estadual Paulista, Araraquara, 2006.

ORTONY, A. **Metaphor and Thought**. New York: Cambridge University Press, 1993.

REED, S.L.; LENAT, D.B. **Mapping Ontologies into Cyc**. 2002. Disponível em: <http://www.cyc.com/doc/white_papers/mapping-ontologies-into-cyc_v31.pdf>. Acesso em: 23/jun/2009.

SABATIER, P. Un lexique-grammaire du footbal. **Lingvisticae investigationes** n.21/1 p. 163-197.1997.

SILVA, D.L.; Sousa,R.R.; ALMEIDA, M.B. **Ontologias e vocabulários controlados: comparação de metodologias para construção**. Revista Ciência da Informação. v. 37, n. 3, p. 60-75, 2008

SINCLAIR, J. **Corpus, Concordance, Collocation**. Oxford University Press,1991.

SINCLAIR, J. 2005. Corpus and Text - Basic Principles. *In*: M. WYNNE, **Developing Linguistic Corpora: a Guide to Good Practice**. Oxford, Oxbow Books, 2005 p. 1-16. Disponível em: <http://ahds.ac.uk/linguistic-corpora/>. Acesso em: 25/Ago/2008.

SMITH. B; CEUSTERS, W; KLAGGES, B; KOHLER, J; KUMAR, A.;LOMAX, J; MUNGALL, C; NEUHAUS, F; RECTOR, A; ROSSE, C. **Relations in biomedical ontologies**. 2005. Disponível em: <http://genomebiology.com/content/pdf/gb-2005-6-5-r46.pdf> . Acesso em: 13/ago/2009.

SOWA, J. F. **Principles of ontology**. 2003. Disponível em: <http://www.jfsowa.com/ontology/index.htm>. Acesso em: 10/abr/2008.

TOSQUI, P. **Construção e Ancoragem Ontológica do Vocabulário Básico Bilíngüe do Turismo para Fins Didáticos**. Tese (Doutorado em Letras) - Faculdade de Ciências e Letras, Universidade Estadual Paulista, Araraquara, 2007.

TURTELLI, S. R. **Estudo da linguagem de um evento esportivo numa abordagem sócio-léxico-computacional**. Tese (doutorado). Universidade de São Paulo. São Paulo, 2002.

PAUMIER, S. **Manual Unitex**. Disponível em: <http://www-igm.univ-mlv.fr/~unitex/manuel.html> . Acesso em 14/Ago/2007.

USCHOLD, M; GRUNINGER, M. **Ontologies: principles, methods and applications**. 1996. Disponível em: <http://citeseer.ist.psu.edu/uschold96ontologie.html>.) Acesso em 26/Ago/2007.

VALE, O. A. **Expressões Cristalizadas do Português do Brasil: uma proposta de tipologia**, Tese (doutorado), Universidade Estadual Paulista, Araraquara. 2001.

VOSSSEN, P. EuroWordNet: Linguistic Ontologies in a Multilingual Database. **Towards advanced document management: the creation and maintenance of electronic thesauri**, **CC AI, Communication and Cognition for Artificial Intelligence**, Special Issue, Volume 15, Numbers 1-2, 1998. 37-80.

Apêndices

Apêndice A – Tabela argumentos substantivos

Ocorrência	C. G.	Traços	Classe
bola	nf	inc	<bola>
árbitro	nm	hum	<árbitro>
juiz	nf	hum	<árbitro>
apito	nm	inc	<instrumento>
bandeira	nf	inc	<instrumento>
cartão amarelo	nm	inc	<instrumento>
cartão vermelho	nm	inc	<instrumento>
campo	nm	loc	<campo>
gramado	nm	loc	<campo>
tapete verde	nm	loc	<campo>
Abc	nm	hum + humcoll	<clube>
América	nm	hum + humcoll	<clube>
Atlético-Mg	nm	hum + humcoll	<clube>
Atlético-Pr	nm	hum + humcoll	<clube>
Bahia	nm	hum + humcoll	<clube>
Botafogo	nm	hum + humcoll	<clube>
Corinthians	nm	hum + humcoll	<clube>
Coritiba	nm	hum + humcoll	<clube>
Cruzeiro	nm	hum + humcoll	<clube>
Figueirense	nm	hum + humcoll	<clube>
Flamengo	nm	hum + humcoll	<clube>
Fluminense	nm	hum + humcoll	<clube>
Grêmio	nm	hum + humcoll	<clube>
Internacional	nm	hum + humcoll	<clube>
Palmeiras	nm	hum + humcoll	<clube>
Paraná	nm	hum + humcoll	<clube>
Ponte Preta	nf	hum + humcoll	<clube>
Portuguesa	nf	hum + humcoll	<clube>

Santos	nm	hum + humcoll	<clube>
São Paulo	nm	hum + humcoll	<clube>
Sergipe	nm	hum + humcoll	<clube>
Sport	nm	hum + humcoll	<clube>
Vasco	nm	hum + humcoll	<clube>
Vitória	nm	hum + humcoll	<clube>
primeira divisão	nf	inc	<divisão>
segunda divisão	nf	inc	<divisão>
série a1	nf	inc	<divisão>
série a2	nf	inc	<divisão>
série a3	nf	inc	<divisão>
albertão	nm	loc	<estádio>
arruda	nm	loc	<estádio>
beira-rio	nm	loc	<estádio>
castelão	nm	loc	<estádio>
cruzeirão	nm	loc	<estádio>
fonte nova	nf	loc	<estádio>
joão avelange	nm	loc	<estádio>
maracanã	nm	loc	<estádio>
mineirão	nm	loc	<estádio>
morumbi	nm	loc	<estádio>
olímpico	nm	loc	<estádio>
serra dourada	nm	loc	<estádio>
carrinho	nm	inc	<falta>
falta	nf	inc	<falta>
insulto	nm	inc	<falta>
barra transversal	nf	loc	<parte do gol>
rede	nf	loc	<parte do gol>
canto direito do gol	nm	loc	<local do gol>
canto esquerdo do gol	nm	loc	<local do gol>
gaveta	nf	loc	<local do gol>
meio do gol	nm	loc	<local do gol>
primeiro poste	nm	loc	<parte do gol>
segundo poste	nm	loc	<parte do gol>
trave	nf	loc	<parte do gol>
ângulo	nm	loc	<local do gol>
np		hum	<jogador>
braço	nm	inc	<parte do corpo>

braço	nm	inc	<parte do corpo>
cabeça	nf	inc	<parte do corpo>
costas	nf	inc	<parte do corpo>
cotovelo	nm	inc	<parte do corpo>
joelho	nm	inc	<parte do corpo>
mão	nf	inc	<parte do corpo>
nariz	nm	inc	<parte do corpo>
ombro	nm	inc	<parte do corpo>
pé	nm	inc	<parte do corpo>
perna	nf	inc	<parte do corpo>
tronco	nm	inc	<parte do corpo>
linha de fundo	nf	loc	<parte do campo>
linha de gol	nf	loc	<parte do campo>
linha de meta	nf	loc	<parte do campo>
linha da meia lua	nf	loc	<parte do campo>
linha dos 18m	nf	loc	<parte do campo>
linha dos 6m	nf	loc	<parte do campo>
linha lateral	nf	loc	<parte do campo>
marca de pênalti	nf	loc	<parte do campo>
pênalti	nm	inc	<penalidade/equipe>
tiro livre direto	nm	inc	<penalidade/equipe>
tiro livre indireto	nm	inc	<penalidade/equipe>
advertência	nf	inc	<penalidade/jogador>
cartão amarelo	nm	inc	<penalidade/jogador>
cartão vermelho	nm	inc	<penalidade/jogador>
primeiro tempo	nf	inc	<período/jogo>
segundo tempo	nm	inc	<período/jogo>
prorrogação	nm	inc	<período/jogo>
etapa inicial	nf	inc	<período/jogo>
etapa final	nm	inc	<período/jogo>
medalha	nf	inc	<recompensa>
taça	nf	inc	<recompensa>
troféu	nm	inc	<recompensa>
4-2-4	nm	inc	<tática>
4-3-3	nm	inc	<tática>
5-3-2	nm	inc	<tática>
5-4-1	nm	inc	<tática>
camiseta	nf	inc	<uniforme>
caneleira	nf	inc	<uniforme>
chuteira	nf	inc	<uniforme>
meia	nf	inc	<uniforme>
moletom	nm	inc	<uniforme>
shorts	nm	inc	<uniforme>

Apêndice B - Tabela argumentos adjetivos

Ocorrência	C. G.	Traços	Classe
Verdão	nm	hum + humcoll	<torcedor>
Gavião	nm	hum + humcoll	<torcedor>
Santista	nm	hum + humcoll	<torcedor>
São paulino	nm	hum + humcoll	<torcedor>
Palmeirense	nm	hum + humcoll	<torcedor>
Tricolor	nm	hum + humcoll	<torcedor>
tricolor das Laranjeiras	nm	hum + humcoll	<torcedor>
Flamenguista	nm	hum + humcoll	<torcedor>
botafoguense	nm	hum + humcoll	<torcedor>
Vascaíno	nm	hum + humcoll	<torcedor>
Cruzeirense	nm	hum + humcoll	<torcedor>
Atleticano	nm	hum + humcoll	<torcedor>
Peixe	nm	hum + humcoll	<torcedor>
Gremista	nm	hum + humcoll	<torcedor>
Colorado	nm	hum + humcoll	<torcedor>
Tigrão	nm	hum + humcoll	<torcedor>
Coxa	nm	hum + humcoll	<torcedor>
Corinthiano	nm	hum + humcoll	<torcedor>
Ponte pretano	nm	hum + humcoll	<torcedor>
goleiro	nm	hum	<posição jogador>
zagueiro	nm	hum	<posição jogador>
lateral direito	nm	hum	<posição jogador>
lateral esquerdo	nm	hum	<posição jogador>
ala esquerdo	nm	hum	<posição

			jogador>
ala direito	nm	hum	<posição jogador>
volante	nm	hum	<posição jogador>
meio-de-campo	nm	hum	<posição jogador>
meia-armador	nm	hum	<posição jogador>
meia-atacante	nm	hum	<posição jogador>
atacante	nm	hum	<posição jogador>
ponta-de-lança	nm	hum	<posição jogador>
centro-avante	nm	hum	<posição jogador>
capitão	nm	hum	<posição jogador>
titular	nm	hum	<posição jogador>
batedor de pênalti	nm	hum	<posição jogador>
batedor de escanteio	nm	hum	<posição jogador>

Apêndice C – Tabela predicados adjetivais

N0	Vsup/ Vaux	N	Prep	N2
<jogador>	ser	<posição jogador>	-	-
<jogador>	ser	arrogante	-	-
<jogador>	ser	imparcial	-	-
<jogador>	ser	individualista	-	-
<jogador>	ser	agressivo	-	-
<jogador>	ser	violento	-	-
<arbitro>	ser	tolerante	-	-
<arbitro>	ser	ladrão	-	-
<jogador>	ser	entrosado	-	-
<jogador>	ser	desentrosado	-	-
<jogador>	ser	ágil	-	-
<jogador>	ser	ativo	-	-
<jogador>	ser	dinâmico	-	-
<jogador>	ser	criativo	-	-
<jogador>	ser	brilhante	-	-
<jogador>	ser	malandro	-	-
<jogador>	ser	craque	-	-
<jogador>	ser	eficaz	-	-
<jogador>	ser	arisco	-	-
<jogador>	ser	bichado	-	-
<jogador>	ser	catimbeiro	-	-
<jogador>	ser	artilheiro	-	-
<jogador>	ser	robusto	-	-
<jogador>	ser	inteligente	-	-
<jogador>	ser	ágil	+	<bola>
<jogador>	ser	hábil	+	<bola>
<clube>	ser	escalado	+	<tática>

Apêndice D – Tabela predicados nominais

N0	V sup	N1	Prep	N2
<jogador>	fazer	falta	-	-
<jogador>	cometer	falta	-	-
<jogador>	dar	carrinho	+	<jogador>
<jogador>	dar	carrinho	+	<bola>
<jogador>	dar	cabeceio	+	<bola>
<clube>	dar	motivação	+	<jogador>
<clube>	dar	força	+	<jogador>
<árbitro>	dar	impedimento	-	-
<jogador>	fazer	gol	-	-
<jogador>	fazer	agressão	-	-
<jogador>	ter	agilidade	-	-
<jogador>	ter	dimanicidade	-	-
<jogador>	ter	criatividade	-	-
<jogador>	ter	estrutura	-	-
<jogador>	ter	elasticidade	-	-
<jogador>	ter	malandragem	-	-
<jogador>	ter	habilidade	-	-
<jogador>	ter	contrato	+	<clube>

Apêndice E – Tabela predicados verbais

N0	V	Prep	N1	Prep	N2
<árbitro>	levantar	-	bandeira	-	-
<árbitro>	apitar	-	-	-	-
<árbitro>	mostrar	-	<cartão>	+	<jogador>
<árbitro>	aplicar	-	<cartão>	+	<jogador>
<bola>	rolar	-	-	-	-
<bola>	bater	+	<jogador> + <parte do gol>	-	-
<bola>	desviar	+	<jogador>	-	-
<jogador>	sair	+	<campo>	-	-
<jogador>	deixar	+	<campo>	-	-
<jogador>	entrar	+	<campo>	-	-
<jogador> + <clube>	perder	+	<jogador> + <clube>	-	-
<jogador> + <clube>	vencer	-	<jogador> + <clube>	-	-
<clube>	contratar	-	<jogador>	-	-
<clube>	disputar	-	<divisão>	-	-
<clube>	cair	+	<divisão>	-	-
<clube>	subir	+	<divisão>	-	-
<torcedor>	lotar	-	<estádio>	-	-
<torcedor>	comparecer	+	<estádio>	-	-
<jogador>	dominar	-	<bola>	+	<parte do corpo>
<jogador>	matar	-	<bola>	+	<parte do corpo>
<jogador>	tirar	+	<parte do corpo>	-	-
<bola>	cruzar	-	<parte do campo>	-	-
<bola>	ultrapassar	-	<parte do campo>	-	-
<jogador>	acertar	-	<bola>	+	<parte do gol>

<bola>	bater	+	<parte do gol>	-	-
<bola>	entrar	+	<local do gol>	-	-
<jogador>	acertar	-	<bola>	+	<local do gol>
<árbitro>	conceder	-	<penalidade/ equipe>	-	-
<árbitro>	marcar	-	<penalidade/ equipe>	-	-
<jogador>	cobrar	-	<penalidade/ equipe>	-	-
<jogador>	levar	-	<penalidade/ jogador>	-	-
<jogador>	receber	-	<penalidade/ jogador>	-	-
<árbitro>	mostrar	-	<penalidade/ jogador>	+	<jogador>
<árbitro>	aplicar	-	<penalidade/ jogador>	+	<jogador>
<jogador> + <clube>	erguer	-	<recompensa>	-	-
<jogador> + <clube>	exibir	-	<recompensa>	-	-
<jogador> + <clube>	conquistar	-	<recompensa>	-	-
<clube>	atuar	+	<tática>	-	-
<clube>	treinar	+	<tática>	-	-
<jogador>	vestir	-	<uniforme>	-	-
<jogador>	calçar	-	<uniforme>	-	-
<jogador>	colocar	-	<uniforme>	-	-
<torcida>	acompanhar		jogo		
<torcida>	sofrer	-	-	-	-
<torcida>	vaiar	-	<jogador> + <clube> + <árbitro>	-	-
<torcida>	aplaudir	-	<jogador> + <clube> + <árbitro>	-	-

Apêndice F – Tabela ECs verbais

EC	N0	N1
dar um baile	<jogador> + <clube>	<jogador> + <clube>
dar o drible da vaca	<jogador>	<jogador>
entregar o ouro	<jogador>	<jogador>
ganhar no grito	<jogador>	<falta>
suar a camisa	<jogador>	-
abrir a caixa de ferramentas	<jogador> + <clube>	<jogador> + <clube>
acumular gordura	<clube>	-