

Classes de objeto no futebol: resultados

Maria Cristina A. dos Santos¹, Oto Araújo Vale²

¹Programa de Pós-Graduação em Linguística – Universidade Federal de São Carlos

²Departamento de Letras – Universidade Federal de São Carlos
Caixa Postal 676 – 13.565-905 – São Carlos – SP – Brasil

{andrade.cristina,otovale}@gmail.com

Abstract. *The results presented in this article are the product of a research that aimed to apply the model of classes of object proposed by Gaston Gross (1994) at the construction of ontologies of domain. 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 purpose of this study was to arrive at semantic classes through the application of syntactic criteria which that may be reproduced by other researchers. The classes of objects are semantic classes selected from the argumental restrictions of each predicate.*

Resumo. *Os resultados apresentados neste artigo são o produto de uma pesquisa que visou aplicar o modelo de classes de objeto de Gaston Gross (1994) à construção de ontologias de domínio. A construção de ontologias de um modo geral segue uma metodologia baseada na tradição semântica, pela qual são estabelecidas classes conceituais a partir do julgamento do criador. A proposta desse estudo foi chegar às classes semânticas por meio da aplicação de critérios sintáticos que possam ser reproduzidos por outros pesquisadores. As classes de objetos são classes semânticas selecionadas a partir das restrições argumentais de cada predicado.*

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, definida assim como a unidade mínima de significação. A frase simples é composta por um predicado acompanhado de seus argumentos. Assim, cada item lexical deve ser analisado em função de sua vizinhança e pelo ambiente sintático em que ele se encontra, pois as relações de co-ocorrência das palavras determinam seu significado. A partir dos postulados de Harris, Gaston Gross (1994) desenvolveu o modelo de *Classes de Objeto*, que visa, principalmente, análise e utilização de critérios sintáticos que possibilitem a distribuição das unidades lexicais em classes.

As classes de objeto podem ser definidas como classes semânticas criadas a partir de critérios sintáticos. Cada classe de objetos 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 e Mathieu-Colas 1998, p. 6].

Tendo em vista esses pressupostos, a pesquisa intitulada “Proposta de aplicação do modelo de classes de objeto para a construção de uma ontologia do domínio do futebol” teve como objetivo aplicar o modelo de classes de objeto de Gaston Gross à construção de ontologias de domínio. Para atingir o objetivo proposto utilizamos uma gramática de operadores e argumentos, tendo como unidade mínima de significação a frase, para a qual cada item lexical deve ser analisado em função do ambiente sintático em que se encontra, pois as relações de co-ocorrência dos itens lexicais determinam seu significado. Para isso, foi construído um corpus do domínio do futebol, que serviu de subsídio para verificar a co-ocorrência dos itens lexicais.

2. Classes de Objeto de Gaston Gross

Como referido, classes de objeto são classes semânticas criadas a partir de critérios sintáticos: “As classes de objeto, consideradas como classes de argumentos, são definidas pelo relacionamento com os predicados que são específicos.” [Le Pesant e Mathieu-Colas 1998].

Lerat (2000) afirma que as classes de objetos não são apenas classes extensionais de objetos, mas sim a validação lógica de classes lexicais (substantivos), através 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, observemos o exemplo abaixo, retirado de Le Pesant e Mathieu-Colas (1998):

redigir <texto>

Verbos como *redigir* possuem critérios necessários para estabelecer a classe de argumentos que o acompanha. Assim, as unidades lexicais que farão parte da classe <texto> 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

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 e Mathieu-Colas 1998, p. 12].

Como dito anteriormente, a frase simples é composta por um predicado acompanhado de seus argumentos. Podemos dizer que ela também é construída em torno de um 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 núcleo predicativo requer somente o argumento sujeito essencial para compor a frase:

1) Ronaldo brilha.

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

2) O Corinthians vence o Palmeiras.

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

3) Chicão lança a bola para Ronaldo.

Le Pesant e 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 acompanham esse predicado. Aqui apresentaremos dois tipos de situações:

Situação 1: Ausência de conflito entre os argumentos.

Alguns predicados podem não impor qualquer tipo de restrição à semântica de seus argumentos, N (N=substantivo) nos exemplos abaixo pode pertencer a qualquer grupo nominal:

Penso em N.

Lembro-me de N.

Sonhei com N.

Isto diz respeito a N.

N é importante.

Situação 2: apesar dessa ausência de conflito, a maioria dos predicados exige a especificação de sua semântica nos 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 acompanham (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:

- 4) Corinthians contratou Tevez.
- 5) O clube inglês contratou o jogador do Grêmio.
- 6) Ronaldo ganhou um chute de Douglas.
- 7) Douglas ganhou o cartão do juiz.

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

- 8) *O Corinthians contratou a bola.
- 9) *A camisa contratou Tevez.
- 10) *Ronaldo ganhou cartão do gol.
- 11) *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 pelas 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+animado, N1 inanimado+concreto, N2 hum+animado)

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.

Até aqui percebemos que, para a aplicação do modelo de classes de objeto, a análise exaustiva dos dados torna-se 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" definicional. [Le Pesant e Mathieus-Colas 1998 , p.13].

Observemos os exemplos:

- 12) O Corinthians perdeu para (o Palmeiras, o São Paulo, o vice-líder do campeonato).

- 13) Corinthians venceu (o Palmeiras, o São Paulo, o vice-líder do campeonato).
- 14) O Corinthians contratou Ronaldo.

Em se tratando de língua geral, nenhum dos verbos acima, por si só, é específico para uma classe, como podemos demonstrar nos exemplos abaixo:

- 15) Marcos perdeu seus óculos.
- 16) Douglas venceu a aposta.
- 17) 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 uma classe de objeto. Porém, 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>.

3. Metodologia

A metodologia utilizada no trabalho foi dividida em 3 etapas. A primeira etapa trata-se da constituição do corpus. O corpus deste trabalho foi constituído de dois corpora: o corpus de língua falada, que é formado de quatro narrações transcritas provenientes da transmissão de partidas de futebol [Turtelli 2002], e o corpus de língua escrita, formado pelo subcorpus do projeto COMET (Corpus Multilíngue para ensino e tradução) e por textos disponíveis online que foram recolhidos e convertidos para o formato txt.

Na segunda etapa foi realizada a extração de itens lexicais do domínio definido para a pesquisa, para que esses fossem trabalhados na constituição das classes na etapa 3.

A partir das primeiras observações, começamos a delinear classes de argumentos e classes de predicados, a saber: argumentos substantivos, argumentos adjetivos, predicados adjetivais, predicados nominais, predicados verbais. Cinco tabelas foram criadas para fins de registro das ocorrências encontradas no corpus. Assim, os argumentos selecionados pelos operadores (predicados) foram distribuídos nas duas primeiras tabelas (argumento adjetivo e substantivo), já as últimas três (predicados adjetivais, nominais e verbais), possuem os padrões sintáticos para as ocorrências dos operadores.

3. Resultados

O presente trabalho chegou ao resultado de 20 classes de objeto do domínio do futebol, listadas no quadro abaixo:

<acessório>	Acessórios utilizados pelos árbitros em uma partida de futebol.
<bola>	Formas de nomear a bola.
<campo>	Formas de nomear o campo.
<clube>	Formas de nomear os clubes.
<divisão>	Divisões de um campeonato.
<estádio>	Estádios de futebol.
<faltas>	Faltas cometidas pelo jogador.
<função_jogador>	Funções dos jogadores em campo.
<jogador>	Formas de nomear o jogador.
<local do gol>	Local onde o gol pode ser marcado.
<parte do corpo>	Partes do corpo humano.
<parte do campo>	Locais demarcados no campo.
<parte do gol>	Partes integrantes do gol.
<penalidade/equipe>	Penalidades cometidas pela equipe.
<penalidade/jogador>	Penalidades cometidas pelo jogador.
<período do jogo>	Períodos que compõem um jogo.
<recompensa>	Recompensa adquirida pelo resultado de um jogo ou campeonato.
<tática>	Táticas de jogo.
<torcedor>	Formas de nomear um torcedor.
<uniforme>	Peças integrantes do uniforme dos jogadores.

Quadro 1. Classes de objeto do domínio do futebol

No entanto, para o francês, o número de classes é um pouco maior. Sabatier (1998) delineou de 32 classes de objeto. 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 atestadas através de corpus.

Observando os resultados alcançados, verificamos que as classes de objeto obedecem a 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 de seus itens lexicais. Além da possibilidade de ancoragem ontológica, as classes de objeto do domínio do futebol são mais uma opção de descrição do português, já que o modelo abre portas para, por exemplo, a partir de uma comparação com classes do francês, a realização de testes de tradução automática.

Agradecimento

A pesquisa contou com o apoio financeiro da Fundação de Apoio à Pesquisa do Estado de São Paulo – FAPESP.

Referências

Anastassiadis-Symeonidis. (2000) 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.

- Gross, Gaston. (1994) "Classes d'objet et description des verbs", *Langages* n.115 p.15-30.
- Harris, Zellig. (1976) *Notes du cours de syntaxe*. Paris: Le Seuil.
- Le Pesant, Denis; Mathieu-Colas, Michel. (1998) "Introduction aux classes d'objet", *Langages* n.131 p. 6-33.
- Lerat, Pierre. (2000) *Les emplois juridiques das 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.
- Sabatier, Paul. (1997) "Un lexique-grammaire du football", *Lingvisticae investigationes* n.21/1 p. 163-197.
- Turtelli, Sandra R. (2002) "Estudo da linguagem de um evento esportivo numa abordagem sócio-léxico-computacional." Tese (doutorado). Universidade de São Paulo. São Paulo.