

# IDENTIFICAÇÃO DE PREDICADOS NOMINAIS EM CORPUS: NOMES DE ESPORTE

Autor: Cláudia Dias de Barros

Orientador: Prof. Dr. Oto Araújo Vale

Universidade Federal de São Carlos

Linha de pesquisa: Linguagem Humana e Tecnologia

Início da pesquisa: março de 2010

Conclusão da pesquisa: março de 2014

Neste resumo é apresentada a descrição do estudo sobre a identificação de predicados nominais formados por substantivos que denominam esportes/atividades esportivas e o verbo-suporte ‘fazer’, do português do Brasil, tendo como arcabouço teórico a teoria do Léxico-Gramática (GROSS, 1975), baseada na teoria dos operadores de Harris (1961). A Léxico-Gramática possui como unidade mínima de análise a frase simples (o predicador e seus argumentos) e formaliza seus dados em tabelas binárias (tábuas) que apresentam as entradas nas linhas e as propriedades sintático-semânticas de cada entrada nas colunas. Os verbos-suporte podem ser definidos como verbos esvaziados semanticamente, que apenas suportam as marcas de tempo, modo, número e pessoa, sendo que os substantivos que os acompanham apresentam o conteúdo semântico da estrutura e são os predicadores. Para a realização do estudo, são utilizados os textos dos anos de 2003 e 2004 do *corpus* PLN.Br FULL (BRUCKSCHEIN *et al.*, 2008), que contém 103.080 textos do jornal Folha de São Paulo dos anos de 1994 a 2005 e 29.014.089 *tokens* (ocorrências). Esse *corpus* é utilizado para a identificação e extração das ocorrências de predicados nominais (verbo-suporte ‘fazer’ + nomes de esporte). Foram identificados e extraídos do *corpus* 26 predicados nominais com os nomes de esporte/atividades esportivas, como ‘fazer caminhada’, ‘fazer arvorismo’. A ferramenta utilizada para essa extração foi o processador de corpus Unitex (PAUMIER, 2002). Notou-se que alguns substantivos designam esportes propriamente ditos, como ‘atletismo’ e outros nomeiam atividades esportivas, como ‘academia’. Foram observadas, então, algumas propriedades sintático-semânticas subjacentes a esses predicados nominais, como o tipo de sujeito (se é humano ou não), de determinante (definido, indefinido ou inexistente), a possibilidade de haver um modificador (um adjetivo, por exemplo), que preposições podem acompanhá-los (de, em, a ou com), se podem ser utilizados na voz passiva, entre outras. O predicado nominal ‘fazer

atletismo’, por exemplo, apresenta as seguintes propriedades sintático-semânticas: sujeito humano (NHum), sem determinante, sem modificador, ‘atletismo’ só é usado no singular, sem preposição e somente na voz ativa (“João faz atletismo”). Os predicados nominais com suas propriedades foram inseridos em uma tabela (tábua) que poderá ser utilizada, em primeiro lugar, para a identificação das estruturas que contenham nomes predicativos de esportes que ocorrem com o verbo-suporte ‘fazer’, e em um trabalho futuro, para o treinamento de sistemas de PLN que utilizem aprendizado de máquina para a identificação automática de nomes predicativos e também para a criação do recurso léxico NomBank.Br, uma base de dados que apresentará nomes predicativos do Português do Brasil e sua estrutura argumental.

GROSS, M. **Méthodes en syntaxe**. Paris: Hermann, 1975.

HARRIS, Z. S. Strings and Transformations in Language Description. **Papers on formal linguistics 1**. Philadelphia: University of Pennsylvania Department of linguistics, 1961.

PAUMIER, S. **Unitex**: manuel d'utilisation, research report. França: University of Marne-la-Vallée, 2002. 200 p.