

ELABORAÇÃO DE DICIONÁRIO ELETRÔNICO

Aluno: Mauro Ricardo Rebelo de Paiva
Orientadora: Prof. Dra. Violeta de San Tiago Dantas Barbosa Quental

Introdução

Essa pesquisa faz parte do Projeto Prepoly, desenvolvido em colaboração com o Prof. Eckhard Bick, da Aarhus University, e se situa dentro do escopo do Projeto Processamento de Linguagem Natural do Português do Brasil, PLN-BR, financiado pelo CNPq e desenvolvido em parceria com as universidades USP/São Carlos, Unisinos, Federal de São Carlos, PUCRS e Mackenzie. O objetivo geral do projeto Prepoly é a análise de expressões prepositivas do português, especialmente quando constituem expressões multivocabulares.

Quando submetidas à análise de um parsing, expressões multivocabulares são etiquetadas preferencialmente como um vocábulo único, por questões como economia de processamento e por corresponderem mais adequadamente ao fenômeno da composição lexical. A expressão “em pêlo”, por exemplo, se analisada como multivocábulo, seria classificada como sintagma preposicional com função adjetival, como em “nu ‘em pêlo’”, e não como sintagma adverbial formado de preposição e substantivo, como em “carrapato ‘em pêlo de cachorro’”.

Para a decisão de análise de sintagmas preposicionais como expressões multivocabulares, tem de se realizar a coleta dessas expressões em seus contextos de ocorrência, e analisá-las lingüística e estatisticamente quanto à sua relevância como multivocábulos, para posterior inserção no léxico computacional – no caso dessa pesquisa o léxico utilizado pelo parser Palavras (Bick 2000) -, de modo a constarem nos resultados de análise morfossintática automática da língua portuguesa.

Objetivos

Esta etapa do projeto tem como objetivo a análise de expressões multivocabulares coletadas nos corpora de língua portuguesa disponíveis na internet, bem como o estudo estatístico e descritivo dessas expressões de acordo com os princípios da composicionalidade, substitucionalidade, e modificabilidade, para a produção posterior de identificações individuais relativas a marcas contextuais e a possíveis restrições quanto ao seu uso como multivocábulos.

Metodologia

A partir de uma lista prévia de 1400 sintagmas preposicionais fornecida pelo Prof. Eckhard Bick, da Universidade de Aarhus, Dinamarca, estão sendo estudadas as características de formação e atuação desses sintagmas em seus contextos de ocorrência, de modo a identificar as situações em que devem ser analisados como expressões fixas, e as situações que devem ser analisados como multivocábulos.

Os vocábulos são analisados segundo os critérios da composicionalidade – princípio através do qual se obtém o significado da expressão a partir do significado de suas partes, da substitucionalidade, princípio que julga a possibilidade de se realizarem substituições de morfemas de uma expressão sem alterar seu significado, e a modificabilidade, princípio que aborda a inserção de palavras no corpo da expressão de modo a alterar ou não seu significado.

Assim, por exemplo, seguindo os princípios da composicionalidade, é realizada a análise da expressão “de efeito” na sentença “foi observada uma alteração de efeito inverso ao registrado anteriormente(...)”. A expressão “de efeito” é composta a partir da preposição “de” somada ao substantivo “efeito”, que exprimem, separadamente, um significado semântico que se mantém na sentença. No entanto, quando a expressão é analisada na frase “Como eu não sabia mais o que falar, comecei a soltar frases de efeito para ver se o convencia(...)”, semanticamente ela não mais exprime o significado composicional da sentença anterior, mas representa um significado distinto que abrange a expressão como um todo, e não somente suas partes separadamente. Nesses casos, “de efeito” seria classificada como uma expressão sem composicionalidade, ficando sujeita a uma análise etimológica apurada e ao estudo de seu uso como multivocábulo.

Ao analisarmos a expressão “de bolso”, por exemplo, na frase “acabo de comprar uma edição de bolso do livro de viagens (...)”, constatamos que o significado da expressão na sentença não seria alcançado se a expressão fosse, por exemplo, “para bolso”, ou “de miniatura”, mesmo que literalmente as últimas alcançassem um nível de entendimento próximo ao alcançado pela primeira. A expressão “de bolso” seria classificada como uma expressão sem substitucionalidade, pois seu significado não seria igualmente reconhecido se houvessem substituições em sua estrutura gramatical.

De acordo com o princípio da modificabilidade, ao analisarmos a expressão “com bons olhos”, por exemplo, na sentença “(...)eu vejo com bons olhos este casamento(...)”, é clara a compreensão do significado da expressão no contexto em que está inserida pois trata-se de uma expressão popular bastante utilizada. Se fosse acrescentada à sentença a palavra “castanhos”, resultando em “vejo com bons olhos castanhos esse casamento”, mesmo que a compreensão do significado da expressão se mantivesse identificável, esta nova expressão seria caracterizada como pouco usual, causando estranheza e dificuldade de entendimento.

Conclusões

Até o momento foram retirados dos diversos corpora do português os resultados das buscas pelas expressões de caráter adjetival em seus contextos de aparição. Na fase atual da pesquisa, estão sendo analisados os sintagmas preposicionais em seus contextos de ocorrência de acordo com as premissas da composicionalidade, substitucionalidade, e modificabilidade, para sua posterior classificação como multivocábulo ou não, e para a formalização de suas propriedades.

Referências

- 1- Bick, E. 2000. *The Parsing System Palavras - Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework*, Århus 2000.
- 2- Davies e Ferreira. Corpus do Português. <http://www.corpusdoportugues.org/>
- 3- Projeto PLN-BR - <http://www.ronaldomartins.pro.br/plnbr/>
- 4- LINGUATECA - <http://www.linguateca.pt/>