

Obstáculos na aplicação do CREA como meio de obter colocações

Manuela Arcos Machado

Orientador Prof. Dr. Félix Bugueño Miranda

UFRGS, Instituto de Letras, Av. Bento Gonçalves, 9500 – Porto Alegre (RS)

Atualmente, no que diz respeito a materiais lexicográficos da língua espanhola, além dos dicionários, conta-se também com ferramentas computacionais disponíveis na rede, tais como o *Corpus de Referencia del Español Actual* (CREA), ferramenta criada e disponibilizada pela *Real Academia Española* (RAE). Embora exista uma grande variedade de obras lexicográficas, os dicionários do espanhol, com raríssimas exceções, ainda oferecem resultados aquém do esperado em relação, por exemplo, às combinações léxicas. Isso explica a grande valia que o CREA possui. O CREA é um banco de dados que registra o uso do léxico da língua espanhola compreendido na faixa cronológica de 1975 até a atualidade. O corpus possui um conjunto de opções (como filtros e seqüências lógicas) para efetuar a pesquisa, o que possibilita um resultado mais direcionado ao desejado pelo usuário.

O objetivo da presente comunicação é avaliar de que forma se pode empregar o CREA para estabelecer padrões colocacionais do espanhol.

Como metodologia empregaremos alguns filtros oferecidos pelo CREA, especificamente o filtro de frequência, que permite isolar um número de 2, 3 ou 5 elementos à esquerda ou à direita de cada concordância. A partir dos resultados oferecidos, podem-se estabelecer possíveis padrões colocacionais. Segundo Hausmann (1979), colocações são produtos da linguagem que o locutor recupera de sua memória como um todo, e são constituídas, além do critério de frequência, por um critério de afinidade semântica. Sendo assim, o filtro de frequência do CREA, apenas, não é suficiente. Em efeito, ele apenas permite que se chegue a candidatos a colocações que, segundo a proposta semântica de Hausmann, devem passar por uma espécie de filtro que possa diferenciar o que é colocação do que são combinações recorrentes de palavras. Contudo, na tarefa de extrair candidatos a colocações, o corpus tem se mostrado não completamente eficiente, por uma limitação do próprio sistema, que compromete os resultados obtidos, oferecendo resultados aleatórios no que diz respeito ao critério de frequência.

Nossos primeiros resultados indicam que a extração de colocações oferecidas pelo CREA com base apenas no critério de frequência não é suficiente. É necessária uma metodologia que possa avaliar semanticamente quais desses resultados constituem, de fato, padrões colocacionais. Ainda, concluímos que o CREA possui uma grande valia por filtrar seus resultados em função do critério de frequência e oferecer candidatos a colocações, mas sua serventia, atualmente, é parcial, devido apenas a uma deficiência de funcionamento da própria ferramenta.

Palavras-chave: colocações; corpus; CREA.