101 **AQUISIÇÃO AUTOMÁTICA BASEADA NA VARIABILIDADE SINTÁTICO-SEMÂNTICA DE EXPRESSÕES MULTIPALAVRAS.** *Leonardo Fernando dos Santos Moura, Aline Villavicencio (orient.)* (UFRGS).

Esta pesquisa trata da aquisição automática de Expressões Multipalavras (EMs), em particular de construções verbo-e-partícula (VPCs) – tais como "take off" em "The plane took off late" no inglês - combinando propriedades estatísticas e alguns padrões lingüísticos destas EMs como combinações produtivas de verbos e partículas. Investigamos também a determinação da idiomaticidade semântica de um dado VPC. Dado um conjunto de candidatos a VPCs, utilizamos árvores de decisão – uma técnica de aprendizado de máquina – e as medidas estatísticas de Mutual Information (MI), $\chi 2$ e Entropia para distinguir automaticamente VPCs de outras construções aparentemente similares. Para classificá-los automaticamente como composicionais ou idiomáticos, usamos informações sobre sinonímia: verificamos se na combinação o verbo pode ser substituído por sinônimos. Os sinônimos são obtidos de alguns recursos computacionais eletrônicos como Wordnet 3.0 e Levin's classes, e as probabilidades são calculadas baseadas em dados coletados da web (utilizando a Yahoo API). Na primeira fase do experimento dos 3078 candidatos, 2848 foram considerados VPCs genuínos, sendo 429 desses, falsos positivos, dos candidatos eliminados, 100 foram falsos negativos. Esses dados levam a um recall de 96% e uma precisão de 84.9%. Resultados esses que confirmam que o uso de informações estatísticas e linguísticas de fato é uma maneira de melhorar a cobertura dos recursos léxicos atuais. As informações obtidas sobre candidatos composicionais permitem gerar regras específicas para evitar redundâncias em dicionários e VPCs idiomáticos podem ser adicionados a esses. Além disso, nós também pretendemos investigar métodos de clustering, para agrupar os verbos automaticamente em categorias, que levem em consideração o quão bem um grupo de verbos se combina com um de partículas. (PIBIC).