

Criação de grafos verbais para o software UNITEX

*Adrielle B. Kucinskas¹, Oto A. Vale².

1. Universidade Federal de São Carlos – UFSCar/PIBITI-CNPq; *adriellekucinskas@hotmail.com

2. Depto.de Letras, UFSCar, São Carlos/SP

Palavras Chave: grafos-verbais, português (br), UNITEX

Introdução

O UNITEX é uma ferramenta linguístico-computacional que foi desenvolvida pela equipe de Linguística e informática da Université Paris-Est Marne-la-Vallée. Esta ferramenta permite a busca por expressões regulares em grandes corpora e possui dicionários eletrônicos de grande cobertura gerados a partir de regras de flexão que permitem a busca por categoria gramatical ou ainda pelo lema das entradas lexicais. Mas, esses dicionários, uma vez construídos, precisam ser constantemente revisados e aprimorados.

Os modelos de flexão que geram os dicionários são representados através de grafos de autômatos de estado finito e apresentam todos os modos verbais de um mesmo verbo, com sub-grafos para melhor organizar a sua estrutura visual.

Os grafos verbais foram criados pois não existiam para o português do Brasil. Criamos então 100 grafos. Estes são pertencentes a primeira, a segunda e a terceira conjugação verbal.

Resultados e Discussões

A pesquisa se dividiu em seis etapas.

Na primeira etapa foi feita a observação da conjugação completa do verbo através do dicionário Houaiss informatizado procurando por regularidades e irregularidades, e identificando o grupo ao qual aquele verbo fazia parte.

A segunda etapa da pesquisa foi a mais extensa. A partir de um verbo e seguindo um modelo de conjugação -ar, -er ou -ir, criávamos a flexão do verbo completa. Dividindo de uma forma com que a estrutura visual do grafo ficasse de fácil compreensão em sua leitura. Assim, se fez a descrição dos tempos verbais como: Presente do Indicativo (P), Pretérito Perfeito (J), Pretérito Imperfeito do Indicativo (I), Pretérito Mais Que Perfeito (Q), Futuro do Presente (F), Futuro do Pretérito (C), Presente do Subjuntivo (S), Pretérito Imperfeito do Subjuntivo (T), Futuro do Subjuntivo (U), Imperativo (Y) e Infinitivo Flexionado (W) e ainda Gerúndio (G), Particípio (K). Alguns desses modos verbais foram feitos por meio de sub-grafos, podendo assim serem utilizados para vários grafos. E as terminações iguais foram agrupadas em uma mesma linha. Informações como mesóclise, ênclise e próclise também foram adicionadas.

Na terceira etapa, se usou a ferramenta que faz as flexões automáticas, ou seja, que a partir do grafo finalizado, monta toda a conjugação em forma de lista.

Na quarta etapa, identificamos se havia algum erro na lista, como, por exemplo, algum problema de digitação ou outro erro de outra natureza, havendo se corrigia, voltando para a edição do grafo, se não, o grafo estava finalizado.

Como quinta etapa o orientador avaliava o trabalho feito e assim, podíamos ir para um próximo grafo. Este processo se deu de igual forma para todos os outros verbos.

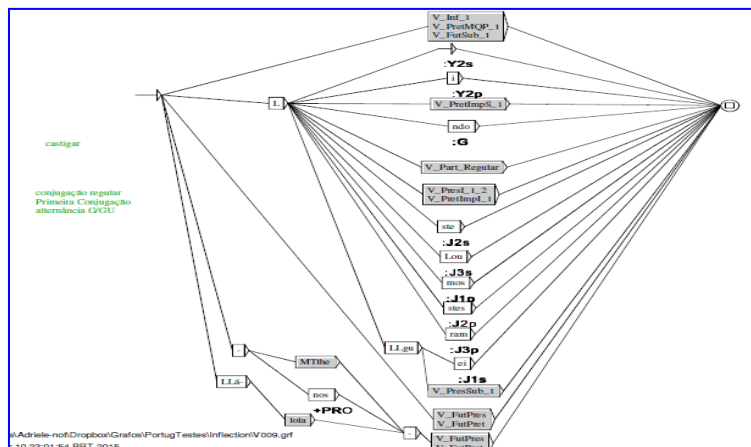


Figura 1. Grafo regular V009 castigar

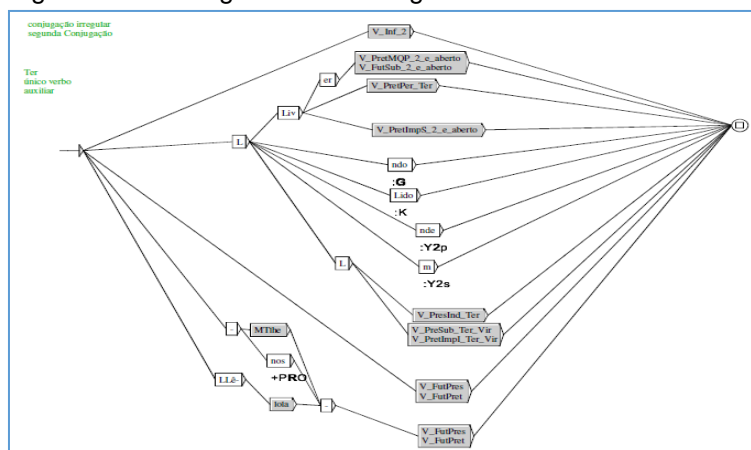


Figura 2. Grafo irregular V001 ter

Conclusões

O trabalho realizado trata-se de mais um passo importante para manter o software atualizado, visando a sua utilização por novos pesquisadores da área de PLN para a análise de fenômenos em vários níveis da língua e para a construção de novos recursos linguístico-computacionais.

MUNIZ, M. C. M. A construção de recursos linguístico-computacionais para o português do Brasil: o projeto de Unitex-PB. Dissertação de Mestrado. Instituto de Ciências Matemáticas de São Carlos, USP. 72p. 2004

PAUMIER, S. Unitex 3.0 User Manual. Université Paris-Est Marne-La-Vallée. Disponível em:

<<http://www-igm.univ-mlv.fr/~unitex/UnitexManual3.0.pdf>> Acesso em: 04 de março de 2013.

VALE, O. A. Dictionnaire électronique de conjugaison des verbes du portugais du Brésil. Mémoire de Maîtrise. Université Paris VIII, 1989