



<b>Evento</b>	Salão UFRGS 2013: SIC - XXV SALÃO DE INICIAÇÃO CIENTÍFICA DA UFRGS
<b>Ano</b>	2013
<b>Local</b>	Porto Alegre - RS
<b>Título</b>	Análise inicial de parâmetros para aquisição lexical
<b>Autor</b>	TATIANA COSTA MEISTER
<b>Orientador</b>	ALINE VILLAVICENCIO

Para melhor compreender a aprendizagem humana e seus aspectos cognitivos, um estudo sobre a aquisição de linguagem nos primeiros anos de vida se mostra um importante contribuinte. Para este estudo foi desenvolvida uma ferramenta para realizar a anotação de palavras ditas por crianças de 0 até 8 anos com informações semânticas e sintáticas (métricas) oriundas de bases lexicais e semânticas (BLS) as quais são: FrameNet [4], WordNet [2][3], Valex [6], VerbNet [5], MRC [7].

As BLS, que serão apresentadas a seguir, possuem métricas sobre palavras, aqui sobre a língua inglesa. Na FrameNet estão dispostas através de frames que descrevem um tipo de evento, como para cozinhar, temos que envolve alguém que cozinha e a comida que será cozinhada. Na WordNet há conjuntos de palavras interligadas entre si através de alguma relação léxico-semântica, tais como sinonímia e familiaridade. Na VerbNet e Valex temos métricas sobre verbos. No MRC temos dados sobre frequência, idade de aquisição, entre outros.

A ferramenta, anotador, foi desenvolvida com o objetivo de receber um conjunto de palavras e gerar um arquivo cujo conteúdo são estas palavras com suas respectivas informações de cada métrica das BLS. Por exemplo, uma métrica poderia ser o número de antônimos da WordNet ou a idade de aquisição do MRC.

A construção do anotador constituiu-se em duas etapas. Na primeira foi realizada a obtenção e reunião das BLS para um único banco de dados, com exceção da WordNet. As informações da WordNet são coletadas automaticamente pela ferramenta durante a anotação através de um programa executável disponibilizado pelos responsáveis da base em seu site. Na segunda foi desenvolvido um programa para receber as palavras, acessar o banco de dados único das informações das BLS e gerar uma saída com as palavras anotadas.

Das anotações da ferramenta desenvolvida é possível realizar análises sobre as métricas comuns às crianças em cada idade, de forma a poder compreender melhor quais fatores influenciam durante a aquisição da linguagem nos primeiros anos de vida. Foi anotado o corpus longitudinal do CHILDES [1], um corpus que contém transcrições de conversas com e entre crianças para mais de 30 línguas. Foi verificado quanto uma métrica é capaz de identificar a idade da criança, treinando um classificador J48.

#### Referências:

- [1] MACWHINNEY, B. The CHILDES Project: Tools for analyzing talk. Third Edition. Mahwah, NJ: Lawrence Erlbaum Associates, 2000.
- [2] MILLER, George A. WordNet: A Lexical Database for English. Communications of the ACM. Vol. 38, No. 11: 39-41, 1995.
- [3] FELLBAUM, Christiane. Wordnet: An Eletronic Lexical Database. Cambridge, MA: MIT Press, 1998.
- [4] BAKER, Collin F., FILMORE, Charles J., LOWE, John B. The Berkeley FrameNet Project. In Proceedings of COLING-ACL'98, pages 86–90, Montréal, Canada, 1998.
- [5] KIPPER, Karin; KORHONEN, Anna; RYANT, Neville; PALMER, Martha. A Large-scale Classification of English Verbs. In Language Resources and Evaluation Journal Vol. 42(1), pp. 21-40, Springer Netherland, 2008.
- [6] KORHONEN, Anna; KRYMOLOWSKI; BRISCOE, Ted. VALEX
- [7] WILSON, M.D. The MRC Psycholinguistic Database: Machine Readable Dictionary, Version 2. Behavioural Research Methods, Instruments and Computers, 20(1), 6-11, 1988