

Caracterização da Aquisição Léxico-Semântica

Objetivos

- Explorar as diferenças de linguagem de três grupos de falantes da língua portuguesa de diferentes idades para melhor compreender o processo de aquisição da linguagem humana.
- Buscar métricas e padrões que permitam distinguir esses grupos etários.
- Realizar um estudo de agrupamento por similaridade permitindo agrupar falantes de acordo com o seu grupo etário.

Materiais e Métodos

- Três grupos etários estudados: C1, com 22 crianças mais novas, C2, com 29 crianças mais velhas, e um grupo composto de 57 indivíduos adultos.
- Dados provenientes de experimento em que foram exibidos 29 vídeos de ações e registrados os verbos utilizados pelos espectadores para descrevê-los.
- Análises de comparação bayesiana, correlações de Pearson e Spearman e índice de Jaccard para a comparação das formas de similaridade.
- Métricas aplicadas para identificar e descrever os comportamentos linguísticos dos grupos:

- ◆ **número de palavras distintas**, onde para cada grupo é calculada a média de respostas distintas por vídeo, normalizada pelo número de falantes;

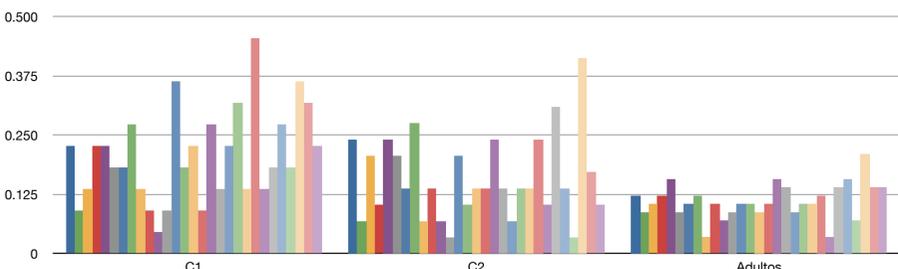


Figura 1 - Variação para a métrica de número de palavras distintas entre os grupos (cada cor representa um dos 29 filmes)

- ◆ **grau de concordância**, a média das concordâncias dos falantes de um grupo com a resposta mais frequente do seu grupo para cada filme;
- ◆ **convencionalidade**, que retrata a proporção de respostas de um grupo ditas convencionais dentro de seus contextos, de acordo com o julgamento de especialistas;

- ◆ **entropia**, uma forma de retratar a quantidade de informação de uma resposta, atribuindo valores maiores quanto menos provável for ela.

Resultados

- Observadas relações entre os grupos etários, como:
 - ◆ decrescimento na métrica de número de palavras distintas e crescimento na entropia conforme a idade;
 - ◆ maior uso de verbos considerados convencionais pelas crianças mais velhas e adultos;
 - ◆ maior correlação entre o grupo de crianças maiores com o de adultos do que com o de crianças menores ou do que do grupo mais jovem com o de adultos;
 - ◆ maior concordância com a resposta mais comum na faixa etária das crianças mais velhas.

	Similaridade por Distribuição Estatística	Correlação Pearson	Correlação Spearman	Palavras em Comum
C1-C2	77.78%	55.88%	55.21%	55.56%
C1-Adultos	66.67%	52.44%	57.02%	39.44%
C2-Adultos	74.07%	70.49%	68.42%	40.85%

Tabela 1 - Resultados das análises comparativas entre os grupos etários

Conclusões

- As análises descritivas e comparativas aplicadas identificam comportamentos nos grupos e indicam diferenças e semelhanças.
- Ajudam a estabelecer métricas e valores relevantes para a determinação do grupo linguístico de um determinado indivíduo.
- Dessa forma, podem, se estendidas, por exemplo, auxiliar no diagnóstico de doenças degenerativas que afetam a linguagem.

Agradecimentos

- Este trabalho faz parte do projeto **Coleta e Estudo Computacional da Aquisição de Linguagem**, financiado pela Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS).