

Desenvolvimento de um tradutor automático Português Brasileiro-Inglês

Lucia Helena Rozario da Silva (USP / PG)

Marcello Modesto (USP)

Alon Lavie (CMU)

Tradução Automática é a aplicação computacional ao processo de tradução de uma língua natural para outra. Existem muitas razões para a sua existência, dentre elas a vantagem potencial de ser menos demorado e menos dispendioso que os tradutores técnicos, podendo traduzir grandes quantidades de textos rapidamente e com um menor custo (Lavie et al. 2004).

O objetivo geral deste trabalho é mostrar o desenvolvimento de um tradutor automático utilizando métodos estatísticos e de transferência estruturais e lexicais entre o Português Brasileiro e o Inglês, a primeira como língua-alvo e a segunda como língua-fonte, e investigar as dificuldades da tradução automática sob uma perspectiva lingüística. Para investigar tais dificuldades, identificamos alguns fenômenos lingüísticos que afetam a eficácia do processo de tradução automática envolvendo as duas línguas, para assim podermos melhorar a eficiência do sistema em desenvolvimento.

O tradutor Português – Inglês em desenvolvimento é composto basicamente pelos seguintes componentes: léxico, analisador morfológico, gramática, mecanismo de transferência, modelo de linguagem e decodificador, conforme Figura 1 abaixo.

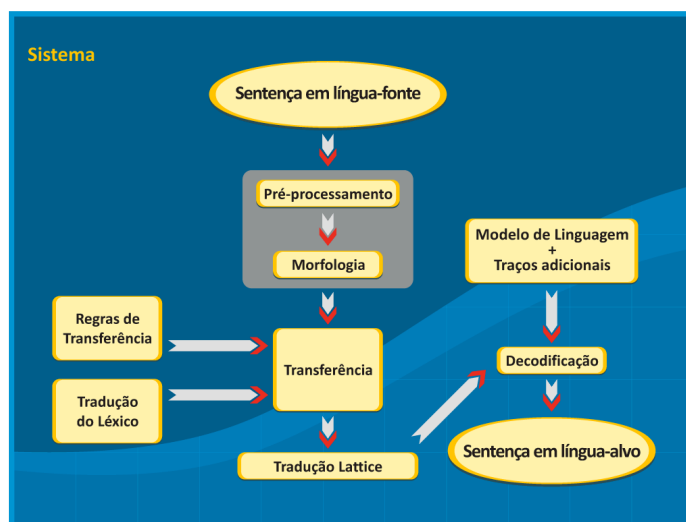


Figura 1. Arquitetura do sistema

A partir de análises e avaliações iniciais do *corpus* construído para a pesquisa, observamos que algumas das questões levantadas pertencem a diferentes níveis de análise lingüística, mas nos limitaremos aqui a discutir questões morfológicas e

sintáticas, esta decorrente de diferenças entre os parâmetros da língua portuguesa e inglesa.

Tomamos como referência qualitativa o tradutor Google Translate (disponível em translate.google.com). A partir do uso do avaliador automático Meteor (Lavie 2007), comparamos os resultados do Google Translate com os resultados do tradutor em desenvolvimento para verificarmos a ocorrência dos mesmos erros em ambos os tradutores. Após a verificação da coincidência dos erros e avaliação dos resultados, identificamos com base em um *corpus* inicial que o parâmetro do sujeito nulo, o emprego dos tempos verbais e a construção de sentenças interrogativas foram alguns dos fenômenos lingüísticos que apresentaram uma baixa pontuação em ambos sistemas.

Diante de dar conta das questões mencionadas, foram necessárias algumas modificações no sistema. Novas regras foram adicionadas e regras existentes foram ajustadas para que pudessem tratar dos problemas apresentados. Após tais modificações, uma nova avaliação do *corpus*, utilizando o Meteor, foi feita para se identificar melhorias no desenvolvimento do sistema, de onde obtivemos uma pontuação de *0.4538228* em contraste com a avaliação inicial, que era de *0.4379914*.

Apesar de se constituir uma análise preliminar, os resultados iniciais são satisfatórios e indicam uma diretriz baseada em análises lingüísticas a se trabalhar para o aprimoramento do estado da arte dos tradutores automáticos em uso e em desenvolvimento, principalmente em relação ao Português-Inglês. Assim, através deste trabalho, buscamos contribuir para o desenvolvimento de recursos computacionais para o Português Brasileiro.

Referências

- LAVIE, A., WINTNER, S., EYTANI, Y., PETERSON, E., PROBST, K. "Rapid prototyping of a transfer-based Hebrew-to-English machine translation system". In *Proceedings of the 10th International Conference on Theoretical and Methodological Issues in Machine Translation (TMI-2004)*, Baltimore, MD. 2004.
- LAVIE, A., AGARWAL, A. "METEOR: An Automatic Metric for MT Evaluation with High Levels of Correlation with Human Judgments". In *Proceedings of the 2nd Workshop on Statistical Machine Translation, 45th Meeting of the ACL (ACL-2007)*. Prague, 2007. pp 228-231.