

Relato sobre a Parceria Linguateca-NILC

Maria das Graças Volpe Nunes
NILC – ICMC-USP
São Carlos - Brasil

O NILC (Núcleo Interinstitucional de Lingüística Computacional) tem usufruído de recursos e iniciativas da Linguateca desde que tivemos contato com seus integrantes pela primeira vez, durante o IV Propor, em Setembro de 1999, em Évora.

Já no primeiro contato, ficou evidente o quanto toda a comunidade de Processamento da Língua Portuguesa ganharia com a parceria entre os grupos de pesquisadores de Portugal (e Europa) e do Brasil. Independentemente de quanto cada parte poderia, de fato, se beneficiar dos recursos gerados e distribuídos, o que tem mantido firme essa parceria é a crença na importância do trabalho comum realizado e, principalmente, da sua divulgação e ampla disponibilização.

O corpus CETENFolha, com textos do jornal Folha de São Paulo, cedido pelo NILC, foi o primeiro e, talvez, o mais útil recurso compartilhado com a Linguateca. São inúmeros os pesquisadores brasileiros que se beneficiaram desse corpus, depois de ele ter sido processado, documentado e se tornado facilmente acessível pela Linguateca. Outros recursos, como o a Floresta Sintá(c)tica, o Repentino, entre outros, têm possibilitado diversas pesquisas no NILC. De outro lado, o NILC tem disponibilizado, via Linguateca, vários recursos ali desenvolvidos: corpora NILC, TeMário, Lacio-Ref, Mac-Morpho, sumariador, material didático, etc.

Nas atividades de avaliação, o NILC não tem podido participar num mesmo nível que os demais participantes, uma vez que a natureza dos recursos sob avaliação não coincidiu ainda com aquelas dos sistemas que desenvolvemos. Mas a parceria tem se mantido em outros níveis, sempre cercada pelo respeito mútuo ao trabalho desenvolvido.

O que cerca o planejamento, o desenvolvimento e a criação de recursos ou aplicativos, nesta e em outras áreas, muitas vezes está fora de nosso controle, o que acaba dificultando o compartilhamento que todos almejamos. A Linguateca tem servido como um alerta para que o desenvolvimento visando o uso comum e a qualidade atestada passem a fazer parte do trabalho do pesquisador em PLN.

A importância do processamento lingüístico na atual sociedade informatizada nos impõe uma grande responsabilidade. Alçar o português ao lugar que lhe cabe na Sociedade da Informação tem sido mais do que uma opção de pesquisa; passou a ser uma missão. Se, como cientistas, devemos investigar a melhor e mais eficiente solução para os problemas de processamento da língua, como peças de uma grande engrenagem nos cabe responder rapidamente às demandas geradas pelas mudanças tecnológicas e, principalmente, à variedade cada vez mais desafiadora da demanda gerada por novos e diferentes usuários.

Nesse cenário, são de singular importância o compartilhamento de recursos para rápidos novos desenvolvimentos, e a existência de mecanismos de controle de qualidade. Enquanto que a decisão de compartilhar recursos depende principalmente da disposição de quem os gera, a elaboração de um sistema de avaliação é bastante mais complexa.

Promover a avaliação de recursos, nem sempre construídos sob os mesmos parâmetros e objetivos, requer uma organização cuidadosa e demorada. Uma diferença relevante entre as avaliações conduzidas pela Linguatca e as congêneres ligadas a eventos internacionais é o número de participantes. Como a comunidade de PLN/português é relativamente pequena, é compreensível que alguns dos recursos avaliados não sejam produzidos por todos os grupos. Aliás, temos aí uma contradição involuntária, já que compartilhar recursos acaba por implicar sua criação distribuída entre os grupos. Nesse sentido, seria igualmente importante a existência de um mecanismo de prospecção de recursos necessários para o português, e conseqüente incentivo ao seu desenvolvimento por um ou mais grupos sabidamente preparados para o desafio. A Linguatca tem todas as condições para servir a esse papel.

A importância do trabalho desenvolvido pela Linguatca nos seus 10 anos de existência pode ser avaliada sob diferentes aspectos e, em todos os casos, a conclusão será positiva.