

As TIC e a Nova Metodologia de Análise Textual

*Maria de Lurdes Abrantes Garcia
(Universidade Nova de Lisboa)*

As Tecnologias de Informação e Comunicação, suporte de estudos linguísticos, vieram impulsionar a criação e introdução de novos métodos na investigação lexicológica e lexicográfica e trouxeram enormes benefícios tanto à qualidade como à quantidade dos produtos obtidos.

A «democratização» da informática, o desenvolvimento das Tecnologias da Informação, os inúmeros instrumentos que põe à disposição do utilizador e a sua operacionalidade permitem aos linguistas e a todos os profissionais da «língua» utilizar directamente este meio nas suas pesquisas, nos seus trabalhos, adaptando-o eficazmente às suas necessidades.

A revolução operada veio trazer novos desafios não só à investigação linguística, mas também aos meios electrónicos cuja evolução tem sido também determinada pela necessidade de dar resposta às solicitações dos linguistas.

A investigação linguística, com a análise e tratamento automático da informação, tem evoluído bastante com uma aproximação estreita à informática e aos seus recursos. Deste modo, a Engenharia Linguística tem vindo a afirmar-se como uma área produtora de inúmeros materiais electrónicos muito úteis tanto na pesquisa linguística, como na realização de trabalhos de tradução.

Na Engenharia Linguística destacamos assim a Terminótica como área auxiliar da investigação linguística, da elaboração e edição de materiais dicionarísticos e resultante da intervenção participativa das Tecnologias de Informação na Terminologia Aplicada. Tem impulsionado a criação de diversos programas informáticos que têm também contribuído para o seu desenvolvimento e, consequentemente, para a implantação das Indústrias da Língua. A utilização de instrumentos informáticos adequados facilita e torna a análise textual mais rápida e

eficiente.

As Tecnologias de Informação, proporcionando a utilização de programas específicos, facilitam a informatização de textos e a preparação de *corpora*. A partir destes é possível a elaboração de Bases Textuais, que armazenam e tratam grande volume de informação. Possibilitam a realização de análise textual e a execução de diversos tipos de pesquisa e gestão de recursos linguísticos, nomeadamente, tratamento automático extensivo e intensivo da informação textual, observação do funcionamento real das unidades linguísticas em contexto, identificação de termos, repertório de termos, pesquisa de formantes, inventário de formantes, extracção de contextos de termos e de formantes, pesquisa de definições, estudo de equivalentes, estudo comparativo de diversos aspectos linguísticos... Permitem ainda a realização de muito mais trabalhos, entre outros, a constituição de bancos de termos, glossários, dicionários informatizados, acesso multilingue a bases de dados, transferência de dados (importação/exportação de ficheiros), tradução assistida e comunicação em tempo real.

Sistematizando, as Tecnologias de Informação oferecem o suporte de produção de materiais dicionarísticos (base de dados linguísticos, glossários, dicionários informatizados), que disponibilizam recursos de gestão terminográfica: estruturação e armazenamento de dados, tratamento da informação: copiar, modificar, acrescentar, suprimir dados, importação e exportação de ficheiros, localização de dados...

Pelas características do seu suporte informático, os materiais dicionarísticos proporcionam por sua vez recursos de pesquisa, que apoiam a nova metodologia de análise textual: pesquisas interactivas diversas, determinadas pela investigação que se pretenda realizar, como atrás referimos, pesquisa de termos em texto livre, pesquisa de termos a partir de um formante, extracção de listagens de termos a partir de um formante, pesquisas diversas, dirigidas individualmente a cada uma das entradas e a cada um dos campos, ou abrangendo a totalidade de entradas e a informação respectiva em cada um dos campos, pesquisa via consulta simples ou complexa. Por consulta simples, entende-se a que possibilita a visualização e edição da informação contida num só campo, por exemplo a listagem de entradas. Consulta complexa é a que pressupõe o estabelecimento de um relacionamento entre vários campos, permitindo obter diversos tipos de relatórios. Permitem portanto realizar consultas em virtude da programação estabelecida, esta última obviamente dependente do objectivo da pesquisa a empreender.

O suporte informático em que os produtos dicionarísticos são elaborados disponibiliza também recursos editoriais, isto é a edição programável de registos de

diversos tipos e em formatos diferentes, por exemplo: a edição de glossários, vocabulários bilingues ou multilingues, dicionário monolingue, bilingue ou multilingue.

O suporte electrónico viabiliza também a difusão de dados realizada em disquete, CD-ROM ou pela Internet, meios que, em virtude das características do *software*, permitem uma actualização permanente de dados.

Em síntese, as novas tecnologias optimizam a análise textual, a realização de diversos tipos de pesquisa e gestão de recursos linguísticos através da utilização de programas adequados que conferem maior rapidez à investigação, menor possibilidade de erro e maior rigor em relação à pesquisa manual. Constituem um auxiliar extremamente importante na investigação linguística e no trabalho desenvolvido pelo Tradutor.

Concluindo, realça-se ainda a importância do acesso generalizado à Internet que veio introduzir alterações profundas, revolucionando a metodologia de trabalho em Lexicologia e Terminologia na actual sociedade da informação.

Através de motores potentes de pesquisa (Excite, Yahoo, Google, Altavista, etc.), em poucos segundos, a Internet faculta o acesso rápido e, por vezes gratuito, a localização de *sites* que dão resposta a interrogações e que esclarecem dúvidas e que fornecem também materiais indispensáveis à investigação.

O internauta tem à sua disposição grande variedade de programas de extracção, gestão e tratamento de dados, inúmeros bancos terminológicos e dicionários *on line*.

O trabalho terminológico torna-se um trabalho colectivo, interactivo em que participam investigadores, especialistas e utilizadores das terminologias.

A Internet possibilita a criação de redes informatizadas de produção e de troca de dados, a realização de trabalho à distância, oferece *e-learning*, permite uma actualização constante dos dados face à evolução dos conhecimentos e faculta a difusão electrónica dos produtos lexicográficos e terminográficos.

Estas potencialidades que as TIC oferecem têm assim contribuído imenso para o progresso da Engenharia Linguística. Esta tem facultado a produção e utilização de *software* especializado de apoio à investigação linguística, ADN de uma nova metodologia de análise textual e respectiva constituição e difusão de dados que

disponibiliza recursos valiosos aos Tradutores.

Bibliografia

- AUGER, P. (1994): "Les outils de la terminotique: typologie des logiciels d'aide à la terminologie et/ou d'automatisation de la chaîne de travail en terminographie", Terminologies Nouvelles, n° 11, RINT, Québec.
- BLANCO, J.(1999): Cómo "Leer" un diccionario. Algunos aportes de la Lingüística informática, (comunicação), Universidade Autònoma de Barcelona.
- BOGURAEV, B. et alii, (1990). "Database models for Computational Lexicography", in Euralex ' 90 Proceedings, Barcelona.
- BOURIGAULT, D., Slodzian, M. (1999): "Pour une terminologie textuelle", in Actes du colloque de Nantes: Terminologie et intelligence artificielle, Terminologies Nouvelles, n° 19, RINT, Bruxelles, Belgique.
- GARCIA, M. de Lurdes, (1993): "Base de données textuelles - Corpora trilingue", in Revue Cumfid, Université de Nice, CNRS- InaLF.
- GARCIA, M. de Lurdes, (1997): "Projet de dictionnaire interactif multilingue de termes médicaux", in META, Journal des traducteurs, vol. 42, n°1, mars 1997, Les Presses de l'Université de Montréal, Canada.