

que poderá naturalmente enriquecer-se com o diálogo que se seguir às intervenções anunciadas.

A Dra. Fátima Biscaia vai iniciar o nosso trabalho.

Fátima Biscaia

Eu queria agradecer à Associação Portuguesa de Linguística e à professora Maria Helena Mateus a oportunidade que me é dada a título pessoal e, também, profissionalmente, em nome da Junta Nacional de Investigação Científica e Tecnológica, a oportunidade de falar de um projecto em que, efectivamente, temos investido muitos esforços.

A minha exposição vai ser muito breve e informal até mesmo porque não quero maçá-los com questões que são muito específicas e técnicas do domínio em que trabalho. Penso, conforme a Profa. Maria Helena Mira Mateus referiu, que a tradução científica e técnica coloca alguns problemas que não são certamente idênticos ou totalmente idênticos aos da tradução literária.

Um desses problemas tem a ver exactamente com a questão da terminologia. Para que haja uma boa comunicação da informação científica e técnica, é imprescindível a precisão e a univocidade da terminologia. É esta questão da terminologia coloca-se não só na tradução de uma língua para a outra, como também a nível da transferência de conhecimentos dentro de uma mesma língua.

Assim, e situando-me no domínio em que trabalho (o domínio normalmente designado por informação científica e técnica, ou por sistemas de informação sem ambiguidades, ou reduzindo, tanto quanto possível, as ambiguidades.

Ora acontece que a linguagem natural, dado o seu carácter linear, dada a complexidade das suas construções e até a ambiguidade frequente do seu vocabulário, se adapta mal a esta transferência unívoca dos conhecimentos científicos e técnicos, transferência esta que exige a representação unívoca dos conceitos, a sua normalização, a normalização desta representação e, por outro lado, uma estruturação semântica por forma a tornar possível a pesquisa e

a recuperação da informação. Num sistema de informação científica e técnica o que interessa é que, no tratamento de uma unidade informativa, sejam identificados os conceitos, representados numa linguagem sistema, linguagem essa que vai ser também utilizada na saída do sistema. Quando um utilizador põe uma questão a uma base de dados, essa questão que o utilizador coloca em linguagem natural é convertida numa metalinguagem, que é uma linguagem artificial mais ou menos codificada, que vai dialogar com a linguagem de entrada do sistema.

Portanto, tendo em conta o que anteriormente referi das linguagens naturais, existe uma necessidade de construção de metalinguagens, genericamente designadas por linguagens documentais, e que vão desde as classificações tradicionais aos actualmente designados thesauri (vocabulários controlados e estruturados que têm o objectivo de identificar os conceitos de informação científica e técnica e de facilitar, de otimizar a sua recuperação).

Nestas metalinguagens a irregularidade das correspondências entre significantes e significados, inerentes à linguagem natural, dá lugar à normalização dos primeiros, isto é, dos significantes, fundamentada na invariabilidade dos significados num domínio científico e técnico bem delimitado, bem determinado. Portanto, a componente mínima de uma linguagem documental é a redução normativa de léxico, isto é, a sua definição controlada por forma a que, a um conceito, vá corresponder um só termo, e que nestes sistemas de informação se chama descritor. Um descritor é um termo preferencial em casos de sinonímia ou de quase-sinonímia. É evidente, portanto, que nestas linguagens se procura, fundamentalmente, reduzir a sinonímia e a polissemia por forma a que a um significado corresponde sempre e só um conceito.

Se por um lado há um só termo considerado preferencial para descrição de informação, - o tal descritor -, por outro lado, todos os outros termos são registados nestas linguagens e constituem pistas úteis para a pesquisa. Isto como enquadramento genérico com o qual eu não gostaria de os maçar muito, porque o que me parece poder ter um interesse mais imediato e directo será dizer-lhes alguma coisa do que foi feito pela Junta Nacional de Investigação

Científica e Tecnológica em relação a esta questão. A Junta tem estado a construir, desde 1972, uma base de dados sobre política de ciência e tecnologia que é um domínio muito vasto, muito heterogéneo, e que inclui questões muito diversas, que vão desde os fundamentos científicos e metodológicos das políticas de ciência e tecnologia, aos recursos implicados nestas políticas, à elaboração, formulação, gestão e execução das mesmas, e igualmente às políticas sectoriais que devem ser coordenadas por essa mesma política. Claro que se nos pôs, desde o início, a questão de uma linguagem suficientemente eficaz para o tratamento desta documentação e desta informação, que pela sua natureza transdisciplinar e pluridisciplinar é muito heterogénea. A única linguagem existente era o SPINES que aliás apareceu em 1976, originalmente em língua inglesa. O SPINES é assim um vocabulário controlado e estruturado. Tem 11462 termos, tem 78134 relações semânticas, e já por aqui podem ver que é uma linguagem de um certo volume e sobretudo muito estruturada. A Junta pretendeu, desde o início, utilizar esta linguagem em língua portuguesa, e assim foi decidido fazer a sua adaptação. Não se pode falar de uma tradução, porque a este nível não se trata exactamente de uma tradução mas sim de uma adaptação, porque o que é importante é saber qual é o conceito que está implicado na versão original em inglês, tentar agarrá-lo com muita precisão e encontrar o termo que na nossa língua ou no português é mais exacto. Esse trabalho foi feito em colaboração com o Brasil, com o Instituto Brasileiro de Ciência e Tecnologia. Mas porque é que nós realizámos este projecto nesta fórmula difícil de cooperação? Porque, é muito difícil, com vocabulário tão extenso e um pouco complicado como este, colaborar com o Brasil que fica muito distante de nós, tendo constantemente que confrontar questões, conceitos e termos. Nós quisemos fazê-lo porque conseguimos assim uma maior garantia para a transferência de informação, através desta linguagem. Na prática, o que acontece é que neste thesaurus, na sua versão original, o conceito é sempre definido por um código numérico. Código numérico esse que não é utilizado depois para a representação dos conceitos mas só para a sua identificação. Quando para um mesmo conceito existe uma variante diferente, ou um termo diferente (como