Uma Perspectiva de Progresso para a Engenharia da Linguagem

contribuição para o debate preparatório do Livro Branco

António Horta Branco
Antonio.Branco@di.fc.ul.pt
Fac. de Ciências de Lisboa, Dep. de Informática

25 de Março de 1999

Nota prévia: Este texto é um esboço duma contribuição cuja versão final estará pronta até dia 17 de Abril de 1999. Para eventuais versões posteriores à presente ver
http://www.di.fc.ul.pt/~ahb/livbranco.html

0 Introdução

Procuramos com este texto contribuir para o debate preparatório do Livro Branco do Desenvolvimento Científico e Tecnológico Português (1999-2007) no que respeita à área da engenharia da linguagem [1]. As nossas reflexões expandem-se por três secções em que abordaremos (1) os bloqueios a ultrapassar face ao actual estado de desenvolvimento da disciplina, (2) as principais necessidades da área em Portugal, e (3) as oportunidades e desafios que se colocam à política nacional de ciência e tecnologia para a engenharia da linguagem.

[1] Agradeço os comentários de Félix Costa, Helder Coelho, José Luiz Fiadeiro, Miguel Filgueiras e Palmira Marrafa.

1 Bloqueios a ultrapassar

Recursos de base

Tal como em outras áreas científicas e tecnológicas, a engenharia da linguagem, ao se ocupar com soluções tecnológicas que apoiam e potenciam a utilização dos meios naturais de comunicação dos seres humanos (http://www.linglink.lu/le/reps/p_and_p98/index.html) necessita de equipamento específico apropriado. Entre esse equipamento encontram-se colecções sistemáticas e informatizadas de documentos, conversas, inquéritos e outros registos de utilização de línguas naturais (corpora), bases de dados com informação acerca das características linguísticas de unidades lexicais (léxicos), colecções de programas para o tratamento de entidades linguísticas (etiquetadores morfológicos, alinhadores de documentos em diferentes idiomas, analisadores sintácticos, detectores de nomes próprios, resolvedores de anáfora, seleccionadores de ocorrências, expurgadores de palavras funcionais,...), colecções padronizadas de testes de avaliação de aplicações (conjuntos de frases de teste para gramáticas computacionais, pares de traduções para sistemas de tradução automática,...), etc. Estes são instrumentos específicos usados na procura de soluções tecnológicas para melhorar as condições de utilização da linguagem natural.

Tal como em outras áreas, também a existência deste equipamento específico – que é frequente encontrar referido sob a designação de recursos linguísticos – constitui uma condição *sine qua non* para a construção de diversas aplicações. É o caso, entre outros, do desenvolvimento de software educativo de apoio ao ensino de línguas (jogos para a melhoria da ortografia, correctores sintácticos com ajuda especializada para falantes não nativos,...), assistentes para comunicação multilingue (sistemas de tradução assistida por computador, processadores de texto com funcionalidades específicas para falantes não nativos da língua em que o texto está a ser escrito, bases de dados terminológicas,...), de aplicações para apoio a pessoas com incapacidades (reconhecedores de falante, conversores de texto para fala,...), de programas para a facilitação do acesso à informação (seleccionadores de documentos, sumarizadores, identificadores de assuntos, encaminhadores de documentos,...), de assistentes para a comunicação homem-máquina (interfaces em linguagem natural, conversores de fala para texto,...), etc.

Por outro lado, do mesmo modo que na química a bancada laboratorial não está apenas ao serviço da tecnologia, mas serve igualmente o propósito do estudo da estrutura

química da matéria e está ao serviço da descoberta científica, assim na engenharia da linguagem os recursos de base desempenham um papel crucial na investigação fundamental acerca das línguas naturais, da comunicação linguística e, em grau suficientemente relevante, da racionalidade e cognição humana.

Recursos humanos e materiais

Relativamente aos recursos de base, um primeiro aspecto a sublinhar é o de que, ao invés do que sucede, por exemplo, na química, não há ainda o que se pode considerar um mercado ou indústria do equipamento específico para a engenharia da linguagem, apesar dos passos iniciais estarem a ser dados a um ritmo febril (http://www.icp.inpg.fr/ ELRA/). As razões imediatas para tal circunstância são fáceis de encontrar quando se compara uma engenharia jovem de décadas, com um peso-pesado com séculos de história, como é a química.

Assim sendo, o equipamento específico para a investigação fundamental e aplicada em engenharia da linguagem tem de ser, e está a ser, construído pelos próprios investigadores que dele precisam, à medida que dele precisam e o concebem, e com os meios ao seu alcance.

E aqui encontramos um primeiro bloqueio que importar conhecer. Não havendo equipamento, não há actividade de investigação e desenvolvimento continuada e consistente. Não havendo tal actividade, não há quem a faça. Não havendo investigadores, não há equipamento, pois faltam meios humanos adequados para o construir.

Importa romper este círculo e não há que procurar muito para saber onde encontrar as chaves. Elas encontram-se no financiamento consequente e numa visão de longo prazo que regule a sua aplicação i) em termos da formação de meios humanos com competência e vocação para a concepção e validação dos recursos de base e, claro, ii) no emprego de recursos humanos com o treino adequado para a construção dos mesmos.

Numa outra vertente ainda, a construção de equipamento para a engenharia da linguagem revela também como esta não pode estar dissociada da formação de recursos humanos de alto nível. A transferência de tecnologia de ponta para a indústria não se reduz a uma transferência de objectos, ferramentas ou equipamento, inclui o envolvimento de recursos humanos capazes de os operar e deles retirar o rendimento devido.

E claro, pensar apenas na formação de competências em desenvolvimento, ou mesmo investigação aplicada, ignorando a investigação fundamental, seria fazer apenas trabalho de cigarra, sem amealhar para o futuro, sendo que sobre a vital articulação entre saberes "puros" e "aplicados" nos eximimos aqui de repetir melhores apologias e manifestos.

As teias da lei e da boa vontade

Um outro bloqueio a focar relativo ao progresso da engenharia da linguagem e ao seu equipamento específico diz respeito ao regime da sua difusão e disponibilização.

Ao contrário do que acontece em outras áreas – como a química, para voltarmos à nossa analogia, em que o tempo já trabalhou a favor de serem encontradas práticas estáveis –, na área da jovem engenharia da linguagem muita coisa parece ainda estar por definir. Em torno da questão do regime de disponibilização dos recursos base e, porque não dizê-lo com toda a frontalidade, da sua comercialização, decorre um vivo debate internacional sobre as complexas questões legais envolvidas (http://www.icp.inpg.fr/ELRA/legals.html).

Vejamos alguns aspectos a título meramente sugestivo. Para a construção de um corpus, uma editora benemérita cedeu o texto de uma obra editada por si: pode quem compila o corpus disponibilizá-lo ignorando que coloca assim em risco investimento e interesses comerciais da editora, podendo até estar em rota de colisão com preceitos legais?

Uma equipa universitária concebeu e desenvolveu um programa inovador para a etiquetagem de funções gramaticais: essa mesma equipa, que já está a receber direitos de autor pelo livro de divulgação — de grande sucesso, mas ligeiro conteúdo científico — que publicou usando os resultados da sua experiência, deve desistir de receber qualquer recompensa por este seu outro mais árduo trabalho? Deve esta recompensa limitar-se a um reconhecimento e referência da autoria onde o sistema for utilizado ou integrado? Deve isto ser regulado legalmente com preceitos específicos, ou os que já existem bastam?

Estas, e muitas outras, são questões à espera de uma regularização e de práticas estáveis (vd. *Elsnews*, 8.1, Fev. 99, p.16). Entretanto, o prolongamento desta indefinição não se deixará de fazer sentir no ritmo do progresso da área.

4 000 línguas, 4 000 equipamentos

Um último aspecto a destacar antes de encerrarmos esta secção tem a ver com o facto de que não há, nem haverá, equipamento "universal" para a engenharia da linguagem, pronto a ser importado.

Cada língua natural pressupõe, para o desenvolvimento e aplicação das tecnologias para a linguagem humana, a existência de recursos específicos. Dada a natureza das línguas naturais, não se pode, nem se pode esperar poder, importar ou adaptar a esmagadora maioria do equipamento de engenharia da linguagem construído para uma língua X para o reusar no processamento de outra língua Y. O melhor que se pode esperar neste particular é usar algumas metodologias de construção de recursos de base testadas para outras línguas e, desejavelmente, padronizar os recursos a desenvolver de acordo com práticas e exigências internacionais.

Perante isto, é compreensível que as diferentes línguas — e respectivas comunidades de falantes — se encontrem em situações de desigualdade relativamente ao desenvolvimento do equipamento respectivo para a engenharia da linguagem. Desigualdades que se ampliarão, como é previsível, como desigualdades discriminatórias — de âmbito linguístico, nacional, étnico ou cultural, ou combinações destes, consoante os casos — na fruição das potencialidades da sociedade global da informação, acerca da qual já estamos razoavelmente convencidos de que se avizinha. Por conseguinte, e dado o volume de investimento envolvido e a dimensão de capacidade organizativa requerida, quais as línguas para que está a ser desenvolvido, ou virá a ser desenvolvido em tempo útil, o equipamento de engenharia da linguagem conveniente, isso é determinado basicamente pela lógica de dois tipos de factores: de mercado, e políticos.

Para uma multinacional da área das tecnologias da informação, as línguas a tratar são escolhidas pelo potencial de negócio que o seu tratamento proporcionar. Se não o calculássemos, disso fomos publicamente informados no último Verão pelo representante, neste caso, da Microsoft presente na First International Conference on Language Resources and Evaluation (cf. *Elsnews*, 7.3, Julho 1998, p.2). O número de falantes de uma dada língua, ou o património literário nela edificado — para afastarmos de vez determinantes compassivas —, é secundário nessa escolha. O poder económico do conjunto dos falantes duma língua e/ou a utilização alargada dessa língua por falantes não nativos em situações de multilinguismo contam como factores decisivos. Não há aqui surpresas: a tecnologia só é desenvolvida se houver compradores que rentabilizem um investimento gigantesco.

Para agências de governação com capacidade de intervenção na área, a lógica será certamente diferente. O que importará será, a um tempo, intervir para contrabalançar a estrita lógica de mercado no sentido de impedir que o fosso entre os info-pobres e os info-ricos à escala planetária tenha o seu enraizamento em diferenças linguísticas, e muito

especialmente, claro, que a barreira deixe no território da exclusão os falantes da língua em questão; a outro tempo, contribuir para manter e alargar a utilização da língua em causa como veículo internacional de comunicação e interacção entre seres humanos e como base – raramente reconhecida mas nem por isso menos crucial – para a projecção cultural e económica dos falantes que a usam; e ainda a outro tempo, intervir para garantir que no panorama geral do desenvolvimento científico, não seja negligenciado o progresso do saber acerca da linguagem e comunicação humanas, a cujos avanços sob a tremenda pressão de novas capacidades e necessidades tecnológicas se assiste neste século, após quase dois milénios de estagnação.

2 Principais Necessidades: o estado das nossas coisas

O conjunto de apontamentos da secção anterior traz-nos até um ponto em que é incontornável a pergunta sobre o que se está a passar em Portugal no que respeita à engenharia da linguagem. Todos sabemos que a perguntas deste teor se responde com números, tabelas comparativas e gráficos. Porém, como no nosso caso estes elementos não existem para que uma análise abrangente e sistemática possar ser feita, limitar-nos-emos aqui àquilo que, à falta de melhor designação, poderemos chamar de análise sintomática, dirigindo a nossa atenção sobretudo para a área da linguagem, umas das duas secções clássicas da nossa engenharia, a par da fala (*language/speech*).

Investigação

Julgamos que uma amostra sugestiva do estado das coisas no que diz respeito à investigação pode ser obtida se procedermos da seguinte forma: i) sendo crucialmente interdisciplinar a actividade no campo da engenharia da linguagem, e encontrando-se as publicações relevantes repartidas por sítios tão diversos como relatórios técnicos de inteligência artitificial, actas de conferências de lógica aplicada ou revistas de linguística formal, seleccionamos as três conferências internacionais de grande prestígio e maior especificidade interdisciplinar para a área: o Annual Meeting of the Association for Computational Linguistics (ACL, que teve o primeiro evento em 1966), a Conference of the European Chapter of the Association for Computational Linguistics (EACL, que tem lugar desde 1983), e a International Conference on Computational Linguistics (COLING, que acontece desde 1964), as duas últimas de periodicidade bianual (a revista Natural Language Engineering conta apenas com cinco volumes: http://www.cup.cam.ac.uk/Journals/

JNLSCAT/nle/nle.html); ii) seleccionamos o período de oito anos, que vai de 1990 até aos mais recentes eventos, que ocorreram no ano passado, o que terá a vantagem de nos dar uma imagem actualizada, ao mesmo tempo que cobre o período mais fecundo da comunidade portuguesa da área, já que para trás deste período não se encontram sinais de actividade portuguesa nestas conferências; iii) e finalmente contamos o número de comunicações de investigadores de instituições portuguesas e o número global de comunicações constantes das actas. Se assim procedermos, obtemos aproximadamente 6 e 2 000.

Vejamos outro indicador, o número de doutoramentos nas últimas duas décadas obtidos por portugueses, em universidades portuguesas ou estrangeiras. Desde 1979, data do primeiro doutoramento, sobram muitos dedos de duas mãos para contar os doutorados. E se a seguir perguntarmos não pelos doutorados, mas pela classe supostamente mais abrangente de pessoas integradas no sistema académico que encontraram condições para fazer da investigação continuada na área o objecto central do seu trabalho, então o número em vez de aumentar, diminui.

A este propósito, evitando a inútil comparação com a realidade esmagadora dos Estados Unidos, e mantendo-nos na nossa galáxia europeia, tomemos a Alemanha como termo de comparação. De cerca de duas dezenas de departamentos com actividades na área, limitemo-nos a considerar o departamento de linguística computacional da Universidade de Saarland, em Saarbrüken. Basta-nos entrar em qualquer dos quatro pisos de gabinetes do respectivo edíficio para encontrarmos doutorados em número que duplica os nossos vinte anos de doutoramentos.

Outro indicador interessante do estado da investigação diz respeito às pessoas integradas no sistema científico em situação de emprego precário – em projectos, com bolsas de pós-graduação, etc. Dada a grande flutuação inerente a este grupo de pessoas, a estimativa aqui é susceptível de maior margem de erro. Mas arrisquemos uma apreciação em excesso. Neste caso, diríamos que cerca de duas dezenas e meia é o número de pessoas que neste momento se encontram nesta situação. O que importa perceber é que, mesmo que este número esteja mal estimado e que seja na realidade, digamos, o dobro, ele ainda assim não ultrapassa o número de estudantes de pós-graduação nesta área – tipicamente com bolsas de pós-graduação – da recém-criada Division of Informatics da Universidade de Edimburgo (que inclui o anterior Center for Cognitive Science e o Human Communication Research Center: www.cogsci.ed.ac.uk/school/people/postgrads).

Desenvolvimento e transferência de tecnologia

No que diz respeito ao desenvolvimento comercial de aplicações tecnológicas e transferência de tecnologia para fora dos meios académicos, o panorama é também simples de descrever. Evitando misturar ecumenicamente as tecnologias da linguagem humana e a mais vasta indústria de edição multimédia, encontramos – continuamos a falar do domínio da linguagem – uma editora associada a uma empresa de software, as quais mantêm em comercialização para o mercado da língua portuguesa um pacote de ferramentas de primeira geração de apoio ao processamento de texto. A par disso, temos conhecimento da criação de uma empresa dedicada à lexicografia computacional, criada por pessoas que durante vários anos se treinaram em projectos de investigação.

Novamente, a dificuldade no que concerne à comparação com realidades não portuguesas não está em encontrar termos sugestivos de comparação, mas no facto de que a desproporção que vem ao de cima ridiculariza a própria tentativa de comparação. Mas ainda assim podemos tentar abrir os relatórios do projecto EUROMAP (http://www.linglink.lu/le/reps/euromap98/index.html) ou o registo de software para o processamento da linguagem da ACL (http://www.dfki.de/lt/registry/).

Recursos de base

Voltemo-nos agora para a situação em que nos encontramos relativamente aos recursos de base para a engenharia da linguagem. Passemos para além da discussão das dificuldades associadas à disponibilização dos recursos existentes, e concentremo-nos nas questões relativas à quantidade destes, à sua avaliação e validação, ao seu estado de padronização, e à sua manutenção e actualização. Olhando à voltando, o que encontramos?

Sabemos que houve projectos que realizaram protótipos significativos de gramáticas computacionais (EUROTRA, LS-GRAM,...), construíram colecções lexicais de relevo (PAROLE, SIMPLE,...), compilaram corpora de dimensão importante (PF, CORPUS, …), desenvolveram programas úteis (Palavroso, Jspell,...). É importante, no entanto, notar que muitos destes recursos foram desenvolvidos como módulos instrumentais para a prossecução de objectivos centrais dos projectos em que se enquadraram. Em consequência, não se encontram documentados de forma útil, a sua estrutura está formatada para servir objectivos específicos, a sua manutenção foi descontinuada após o final do respectivo projecto, a sua dimensão e/ou cobertura é limitada. Por outro lado, é de notar ainda que tem existido um considerável preconceito por parte da comunidade académica relativamente ao

trabalho científico aplicado envolvido na construção de recursos de base, o qual só recentemente parece dar sinais de se desvanecer.

Perante este cenário torna-se difícil estabelecer análises comparativas com outras realidades. Mas existem talvez dois indicadores sugestivos que podem ajudar nesta tarefa.

Se abrirmos a lista de recursos escritos *(written resources)* que a ELRA-European Language Resources Association possui neste momento em condições de serem disponibilizados, ficamos a saber que das centenas de items da lista – que inclui corpora, dicionários, etc – apenas cinco dizem respeito ao português (http://www.icp.inpg.fr/ELRA/cata/tabtext.html).

O outro indicador encontramo-lo na conferência já atrás mencionada, a First International Conference on Language Resources and Evaluation, que teve lugar em Maio de 1998. Neste evento participaram cerca de 500 pessoas de 35 países diferentes (*Elsnews*, 7.3, Julho 98, p.2). Dessas, não mais de três eram portuguesas (http://ceres.ugr.es/~rubio/elra/program.html).

Investimento público

Detenhamo-nos agora por um pouco em tentar perceber o estado das coisas relativamente ao investimento público na área da investigação e desenvolvimento em engenharia da linguagem. Talvez aqui, ainda mais do que noutra dimensão, seja difícil aproximarmo-nos de uma análise produtiva por falta de elementos objectivos.

Socorrendo-nos de elementos dispersos, a primeira nota que tentativamente se poderá avançar diz respeito aos montantes envolvidos. Aqui importa assinalar que, apesar de Portugal e a língua portuguesa aparecerem sistematicamente em último lugar nas estatísticas da Comissão Europeia no que diz respeito ao investimento comunitário – menos projectos financiados, menos instituições participantes, menos verbas transferidas –, ainda assim esse investimento ultrapassa largamente o investimento público nacional.

Esta ideia baseia-se em dados imprecisos, mas cuja ordem de grandeza é significativa. No início de 1995, juntamente com alguns colegas e com os dados de que então dispunhamos e as estimativas razoáveis que arriscávamos, apercebemo-nos de que para um período de sete anos, entre 1988 e 1994, o montante transferido para Portugal em resultado do financiamento pela União Europeia de projectos de investigação em engenharia da linguagem muito facilmente ultrapassaria o meio milhão de contos. Compare-se este

número com a estimativa de 300 mil contos divulgada para todo o investimento nacional até ao presente, onde foram ecumenicamente incluídos projectos da mais variada sorte (http://www.portugues.mct.pt/branco/listagem.html).

Estes dois factos em conjunto, menos investimento europeu em Portugal (por nação ou por língua oficial) e – tudo o indica – o menor investimento nacional entre os pares europeus, em termos absolutos ou relativos, não nos deixa onde estávamos há uma década. Deixa-nos muito mais para trás ainda. Este cenário desolador completa-se com o seguinte conjunto de observações acerca do financiamento da investigaão científica em engenharia da linguagem em Portugal: raramente há aferição do mérito das candidaturas a financiamento por avaliadores independentes com competência na área e nos seus subdomínios; não há acompanhamento do desenrolar dos projectos; não há controlo do cumprimento dos objectivos; não há uma estratégia de crescimento do potencial científico humano nem de acumulação de recursos de base.

Formação e fixação de recursos humanos

Finalmente, um ângulo importante sob o qual a situação da engenharia da linguagem em Portugal deve ser também observada diz respeito à formação e fixação de meios humanos competentes na área.

Aqui, a primeira observação irá para o facto de, enquanto em outras realidades nacionais aflui à área gente com formação diversificada em lógica, psicologia, filosofia, ciência da computação, linguística, ciência cognitiva, inteligência artificial, em Portugal o recrutamento de meios humanos tem-se limitado em regra ao universo dos licenciados em informática e linguística. Porém, mesmo neste universo, a fixação de longa duração de recursos humanos com experiência continuada e acumulada de trabalho na área tem sido difícil, e sem dimensão para um crescimento sustentado.

Dada a grande procura de profissionais de informática e o consequente "aquecimento" deste mercado de trabalho, são conhecidas as dificuldades de a universidade competir nesse mercado tendo em vista a captação de recursos humanos para a actividade académica e para a investigação no próprio centro de gravidade do campo da informática. Isto dito, compreende-se as dificuldades acrescidas sentidas em segurar jovens talentos cujo espírito de risco fez chegar à área da engenharia da linguagem. Quando o que se oferece a seguir à participação num primeiro projecto é a possibilidade de, talvez, eventualmente, quem saberá, vir a haver outro projecto, e dada a inexistência de investimentos privados no

sector, não é preciso consultar a bola de cristal para calcular a dimensão do sentido de risco dos resistentes.

Quanto aos licenciados formadas na área da linguística, embora o seu mercado de trabalho seja de uma "temperatura" diferente, em termos de algo que possa ser chamada de crescimento sustentado de massa crítica, o resultado final da análise que cabe fazer é idêntico. A precaridade do emprego científico já proporcionou que assistíssemos ao abandono da área por uma primeira geração de jovens investigadores que incluiu talentos magníficos, formados em projectos irrepetíveis que proporcionaram condições de trabalho e acumulação de experiência dificilmente igualáveis. As consequências desastrosas para a acumulação e crescimento da competência na área em termos de comunidade que nela trabalha são claras: muito raramente os juniores chegam a seniores – o que aliás já se poderia depreender do que atrás ficou dito na subsecção sobre investigação.

Estas circunstâncias mostram sinais de vir a ser alteradas, uma vez que, com a chegada ao mercado de trabalho dos primeiros diplomados da nova licenciatura de Engenharia da Linguagem e do Conhecimento (promovida pelos Deps. de Informática, Filosofia, Estatística e Investigação Operacional, Matemática e Linguística, das Faculdades de Ciências e Letras de Lisboa: http://www.di.fc.ul.pt/licenciaturas/lelc/index.html), passará a existir um novo grupo de profissionais qualificados com uma forte motivação para apostar no destino desta área.

Entretanto, a decisão de lançar no próximo ano lectivo um novo mestrado interdisciplinar em Processamento e Tecnologias da Linguagem (Fac. Letras de Lisboa), envolvendo a participação de um leque amplo de especialistas nacionais e estrangeiros de áreas diversificadas, não pode deixar de ser tomado senão como um sinal de bom augúrio.

Mas aqui, tal como em outras circunstâncias humanas cujos desenvolvimento depende de parâmetros complexos, o futuro está em aberto e depende daquilo que fizermos dele.

3. Oportunidades e desafios

As reflexões e apontamentos que atrás ficam expostos levam-nos a compreender que existe no presente um conjunto de circunstâncias que justificam e reclamam uma atenção especial relativamente à área da engenharia da linguagem ou, usando, terminologias confluentes, do processamento das línguas naturais, da linguística computacional, das

tecnologias da linguagem humana. Há forte **motivação** para uma intervenção nacional na área. Há que aproveitar um momento histórico de crescimento do saber científico acerca das línguas naturais, da comunicação e da cognição humanas – o qual se encontra em curso sob uma enorme pressão das necessidades tecnológicas – para dar oportunidade a que exista uma participação portuguesa de relevância internacional na construção desse saber científico. Há que evitar a info-exclusão de raiz linguística dos falantes do português e há que ajudar a projectar a língua portuguesa como veículo internacional de comunicação na era da globalização e da sociedade da informação.

Do ponto vista das actividades de I&D, estas motivações colocam diante de nós como **objectivo** primordial fazer crescer a capacidade e o potencial científico e tecnológico do nosso país nesta área. Isto passa por estimular o rápido crescimento e acumulação de recursos humanos e não humanos formados de acordo com padrões internacionais de qualidade. Passa também por favorecer o enraizamento de competências científicas no tecido académico português e por apoiar a transferência de capacidades e soluções tecnológicas para o tecido industrial.

Colocar a tónica apenas na formação de recursos humanos seria como construir um departamento de química apenas com gabinetes e salas de reunião. Colocar a tónica apenas na construção de recursos linguísticos com financiamento público seria como instalar um arquivo histórico nacional para que a biblioteca do congresso melhor os pudesse fotocopiar. Precisamos que as multinacionais da engenharia da linguagem encontrem condições para se interessarem pela língua portuguesa, mas precisamos também que os falantes do português possam contar com capacidade científica e tecnológica independente e actualizada.

Sobre as **linhas orientadoras** que devem regular o estabelecimento dos meios para alcançar estes objectivos, e para abreviar uma história longa, sabemos que elas começam por duas palavras: financiamento e organização.

De nada valerá pensar que a solução se resume a destinar para a área um orçamento com a dimensão do orçamento do programa espacial norte-americano. O problema não seria apenas o de que não existiria capacidade de absorção completa de tal estímulo. É do senso comum que um organismo sobre-alimentado entra em graves riscos de desenvolvimento.

Ao financiamento generoso, teremos de juntar organização. Mas neste capítulo novamente o problema não é de quantidade, mas da quantidade adequada. Uma atitude excessivamente voluntarista e directiva depressa alcançaria resultados opostos àqueles a

que se proporia. A conhecida história do programa EUROTRA, um programa ambicioso que arregimentou várias equipas com métodos de trabalho, padrões de exigência e culturas diferentes e que foi conduzido pelos próprios funcionários da Comissão Europeia, é a todos os títulos eloquente e fonte de muitos ensinamentos neste particular.

Aqui, mais do que dirigir, há que estimular.

Há que estimular a auto-organização de trabalhos, equipas, departamentos, e faculdades que cultivem talvez a característica mais rara mas mais crucial na área, a *interdisciplinaridade*.

Há que favorecer trabalhos que cruzem a *investigação fundamental* e a *investigação aplicada*, a qual inclua o desenvolvimento de recursos de base e de aplicações.

Há que apoiar *todas as ciências e disciplinas confluentes*, com contributos importantes para o desenvolvimento do conhecimento e estudo da linguagem natural e da comunicação linguística.

Há que criar condições para que as *equipas existentes* se tornem naquilo que o seu potencial permitir.

Há que atrair jovens talentos e *novas equipas* para a área dando sinais claros de que há condições para cobrir o risco envolvido na aposta numa área nova.

Há que *importar saberes e competências* enviando estudantes para os melhores departamentos no estrangeiro e contratando jovens investigadores estrangeiros em início promissor de senioridade para que o crescimento do potencial científico ganhe momentum suficiente para produzir efeitos práticos a tempo.

Há que estabelecer condições para que a disponibilização de recursos de base desenvolvidos seja incontornável por parte de quem os desenvolve com financiamento público por forma a que a sua *reutilização e extensão* seja possível.

Há que procurar uma *articulação de iniciativas* entre entidades com interesses convergentes (MCT, Biblioteca Nacional, Instituto Camões,...), não perdendo de vista a cobertura de diferentes variantes da língua portuguesa

Há que informar e *incentivar o sector industrial*, nacional e internacional, do e para o interesse e potencial da engenharia da linguagem aplicada à língua portuguesa.

Mas sobretudo, há que fazer tudo isto com um apurado sentido de *eficiência, longo-prazo e progressividade*. Há que ser rápido, mas há também que caminhar por etapas, e avaliar o caminho a tomar após cada uma delas.

[*para continuar...*]