

Introdução ao processamento de linguagem natural através das aplicações

Diana Santos

Este texto tem como primeiro objectivo abrir os horizontes àqueles que estudam informática para depois tratar a língua e àqueles que estudam a língua para depois poderem fazer programas que a “respeitem” e “compreendam”.

Como tal, pretendo realçar a importância das aplicações na linguística computacional (ou processamento de linguagem natural), produzir uma visão de conjunto e enunciar uma série de recomendações que me parecem úteis a esse respeito.

1 Qual a relação entre aplicações e o processamento de linguagem natural?

Ao anunciar uma conferência de processamento de linguagem natural ou linguística computacional, é costume solicitar contribuições numa categoria chamada “aplicações”, ao lado de categorias mais “científicas”, por área de investigação (como morfologia, sintaxe, semântica, etc.)¹. Isto é o primeiro passo para o nascimento de conferências sobre processamento de linguagem natural aplicado² (ou melhor, aplicações) e para conferências específicas sobre várias dessas aplicações³.

O mesmo se passa com as revistas científicas. Surgem cada vez com mais frequência revistas dedicadas a sub-áreas, que levam a um progressivo distanciamento das comunidades que se dedicam a tarefas e problemas da mesma índole. Temos pois *Computational Linguistics*, mas também *Machine Translation*, *International Journal of Corpus Linguistics*, *Natural Language Engineering*, *Literary and Linguistic Computing*, *Grammars*, *Computers and the Humanities*, etc. (de facto, mais de uma centena de revistas especializadas, a grande maioria em inglês).

A secundarização das aplicações parece-me, contudo, no caso do processamento de linguagem natural, uma situação no mínimo descabida. O processamento de linguagem natural (PLN, ou o tratamento das línguas por computador) é uma disciplina que, na minha opinião, se define como a utilização de conhecimentos sobre a língua e a comunicação humana, tanto para a comunicação com sistemas computacionais como para melhorar a comunicação entre os seres humanos.⁴

Se esses conhecimentos provêm sobretudo de disciplinas como a semântica, o processamento de sinais ou a teoria da comunicação ou se, pelo contrário, têm de ser obtidos de maneiras não canónicas segundo as áreas teóricas a que pertencem, essa é uma opção dependente da aplicação considerada, ou melhor, do problema que se tenta resolver. Por outras palavras, estas disciplinas são subsidiárias do processamento da linguagem natural, e não sub-áreas.

Talvez para contrariar um pouco esta tendência, foi cunhado o termo “engenharia da linguagem” (ou melhor, o termo inglês “language engineering”), tentando dirigir a atenção para a questão do “fazer coisas” com a língua, resolver problemas específicos da linguagem natural, produzir ferramentas que comuniquem (ou possam pelo menos tolerar, senão compreender, a língua), avaliar o impacto da

utilização de tecnologias relacionadas com a língua, etc. Porém, uma mudança terminológica, embora associada a programas de financiamento europeu onde tais objectivos se encontravam expressos explicitamente, não implica necessariamente uma mudança nas mentalidades ou na organização curricular. De facto, a estruturação de vários livros de introdução ao PLN parece perpetuar a ideia de que primeiro se deve estudar o léxico, os formalismos, e as teorias, e só depois pensar em aplicações. Embora haja uma certa dose de bom senso nessa forma de apresentação, gostava de poder contrariar um pouco essa tendência no presente artigo, chamando a atenção para a perspectiva contrária: é ao tentar resolver um dado problema (isto é, ao tentar construir um dado programa que manipula a língua) que surge o momento de nos debruçarmos sobre (algumas características) do léxico ou das teorias que o tentam resolver.

Neste capítulo, vou lutar contra a tendência de ver as aplicações como uma sub-área do processamento de linguagem natural. Assim, pretendo chamar a atenção para o facto de que é igualmente válida uma perspectiva do PLN através da descrição das tarefas que os investigadores tentam resolver e que, em vez de serem denominadas "aplicações"⁵, me parece mais correcto referir como domínios de investigação e desenvolvimento. (Ou seja, não me pareceria incorrecto considerar como sub-áreas de investigação a correcção ortográfica, a tradução automática ou o garimpo do texto ("text mining").)

2 Algumas precisões

Antes de prosseguir, tentarei clarificar o que entendo por aplicações em processamento de linguagem natural, contrastando-as eventualmente com conceitos relacionados. Aproveito também para elucidar algumas opções feitas no texto.

2.1 *Manipular ou processar?*

Em primeiro lugar, é preciso distinguir entre a classe dos programas que manipulam as línguas, e a classe dos programas que tomam em consideração as características destas. Embora à primeira vista se possa pensar que as duas categorias identificam os mesmos programas, basta notar que os sistemas mais usados para lidar com a língua (escrita ou falada) são:

- editores ou formatadores de texto
- impressoras e máquinas de fax
- telefones, emissores de rádio ou televisão
- gravadores e leitores de cassetes
- folheadores ("browsers") na Internet

para reconhecer de imediato que pertencem à primeira classe descrita mas não à segunda. De facto, não são sistemas que tomem principalmente em consideração as características da linguagem tradicionalmente consideradas "linguísticas".

É possível identificar quatro tipos de sistemas informáticos em relação a uma língua. Em relação ao português, podemos pois distinguir:

1. sistemas que funcionem com a nossa língua
2. sistemas que reajam à nossa língua
3. sistemas que interajam utilizando a nossa língua
4. sistemas que produzam resultados na nossa língua (por exemplo, após consulta a informação noutras línguas)

Em relação aos programas ou dispositivos mencionados acima, constatamos que a maioria se limita ao primeiro caso. No entanto, é importante não esquecer que

é dentro de ambientes de utilização deste tipo que poderão germinar a maioria das aplicações.

2.2 *Aplicação, ferramenta, tecnologia e funcionalidade*

Uma distinção terminológica que parece útil precisar é entre **aplicação** e **ferramenta** (computacional). Uma aplicação pode ser considerada como uma (ou mais) ferramentas em contexto, para resolver um dado problema ou ajudar numa dada actividade.

Por exemplo, a aplicação "sistema de apoio à tradução" incluirá correctores ortográficos e estilísticos, interfaces com dicionários bilingues e monolingues, eventualmente um folheador de traduções anteriores, acesso a terminologias várias, assim como a um dicionário de sinónimos na língua de destino, além de um conjunto de outras ferramentas para editar facilmente o texto e procurar padrões parecidos em textos da língua de origem. Poderá também incluir formatadores e descodificadores de formatos especiais usados pelo tradutor.

Embora haja ferramentas que pareçam ter aplicação óbvia (por exemplo um sintetizador de voz ou um corrector ortográfico), a realidade é que, na maioria dos casos, é preciso fornecer – e preocuparmo-nos com isso – muito mais do que a ferramenta em si.

Por outro lado, convém também distinguir entre **ferramenta** e **tecnologia**: uma tecnologia será um certo tipo de abordagem – análise sintáctica por tabela ("chart parsing"), síntese por concatenação, linguagens orientadas por objectos, morfologia de dois níveis ("two-level morphology"), redes neuronais, etc. Uma ferramenta usará uma ou mais tecnologias, uma aplicação recorrerá a uma ou mais ferramentas na resolução de um problema concreto.

Finalmente, não deixa de ser interessante reflectir que muitas das aplicações da linguagem natural são por sua vez integradas como **funcionalidades** (opções que tornam um produto mais eficiente ou utilizável) em sistemas maiores. Assim, os correctores sintácticos vêm em geral associados a um editor de texto, os sistemas de reconhecimento ou síntese de voz são comercializados associados a ambientes de trabalho com características próprias, as perguntas em linguagem natural na rede constituem uma possibilidade de interacção associada a motores de procura com outras interfaces já com sólida implantação no mercado ou com grande audiência, etc.

2.3 *Ajudar ou substituir o ser humano?*

Na panorâmica que se segue é possível identificar (com uma transição não abrupta) dois tipos de sistemas:

1. áreas de trabalho cujo objectivo é ajudar o ser humano a desenvolver tarefas linguísticas
2. áreas de trabalho em que o objectivo é pôr o computador a realizar tarefas linguísticas

O adjectivo "linguístico" vai de uma interpretação estrita "referente à linguística" até a uma interpretação tão lata quanto possível "referente à língua", ou seja, o ponto 1 inclui aplicações cujo objectivo é a ajuda estrita à investigação linguística (como varredores de corpora) até programas cujo objectivo é melhorar o desempenho das actividades de escrita (e fala) do ser humano, tais como ferramentas de ajuda à redacção e à tradução. E, da mesma forma, o grupo 2 abrange desde a construção automática de tesouros ou dicionários (recursos estritamente linguísticos)

até à geração de cartas comerciais ou a síntese de notícias em várias línguas pelo telefone.

Parece-me importante, neste contexto, realçar que várias vezes se tem verificado que quanto menos ambicioso é um sistema – permitindo um controlo humano, ou apresentando apenas funcionalidades sem margem de erro apreciável –, mais facilmente se torna útil e é, portanto, utilizado. Parece, assim, ser uma estratégia realista tentar primeiro ajudar o ser humano numa dada tarefa em vez de principiar por tentar substituí-lo de todo.

2.4 A questão da língua no desenvolvimento de aplicações

A que ponto é que há um processamento de linguagem natural, ou um processamento do português, ou do inglês? Na minha opinião (Santos, no prelo), pode partilhar-se tecnologias, mas não se deve adaptar sem sentido crítico aplicações pensadas (ou implementadas) para o inglês.

Não só porque cada língua tem problemas específicos (áreas que causam dificuldades diferentes a um processamento automático), mas porque soluções apropriadas a uma língua podem não ajudar muito no processamento de outra.

2.5 Avaliação

Não posso deixar de mencionar que, intimamente relacionada com qualquer aplicação, está a questão da avaliação. A avaliação é a tarefa mais difícil – mas também a mais necessária – no desenvolvimento de aplicações. É preciso saber quantificar o problema, identificar as vantagens que o uso de uma dada ferramenta poderá trazer e produzir uma integração no ambiente em que a aplicação irá funcionar.

Apesar do tema me estar vedado no presente capítulo, considero que é preciso sublinhar que a descrição de uma aplicação sem a sua avaliação não deve sequer ser considerada aceitável.

2.6 Referência ou divulgação?

Finalmente, no que se refere ao presente artigo, convém esclarecer que pode haver três atitudes distintas ao escrever uma introdução aos diversos tipos de aplicações em processamento de linguagem natural:

1. produzir uma relação “realista” (daquilo que se encontra neste momento no mercado, e/ou que foi visto/experimentado pela própria autora)
2. fazer uma lista tipo “lavo daí as minhas mãos” (se alguém diz que faz, indico, e aponto para a referência)
3. produzir um texto empenhado, evangelizador: vejam as tantas coisas que se poderiam fazer, apesar de ainda não estarem feitas...

A minha atitude é essencialmente a primeira, complementada com a segunda nos casos em que a descrição me parece suficientemente convincente. Não posso, contudo, fornecer qualquer garantia de que os sistemas mencionados correspondam a mais do que tentativas de resolver um dado problema, que conduziram a uma descrição do mesmo. É possível que muitas das descrições que cito digam respeito a tentativas ainda em fase embrionária, ou que mencionem uma dada aplicação apenas para descrever um problema específico num contexto mais vasto.

Por outro lado, mesmo quando o objectivo de um artigo é descrever uma dada aplicação, não é necessariamente evidente que se esteja em presença de um sistema

que funciona, de um sistema em fase de protótipo ou mesmo de um sistema apenas planeado.

Este texto tem um fito essencialmente pedagógico (o que não significa que tenha um público fixo e delimitado à partida). Como tal, não deve ser considerado como uma obra de referência, no sentido de conter os elementos bibliográficos mais adequados a cada aplicação (tais como as primeiras referências no campo, as mais avançadas à data da escrita, e/ou aquelas que são consideradas canónicas por cada comunidade). Embora me esforce por dar prioridade ao processamento do português, não posso reclamar quer exaustividade quer representatividade.⁶ Os trabalhos mencionados são-no a título meramente ilustrativo.

De facto, o presente artigo cabe melhor dentro do discurso de divulgação⁷, mas com a esperança de que o conjunto de trabalhos a que faço referência permita ao leitor chegar a uma visão equilibrada das aplicações possíveis.

3 Panorâmicas possíveis

Seria possível estruturar uma introdução às aplicações de várias maneiras distintas. Seleciono aqui três grandes vectores, explicando para cada um deles porque não foi utilizado como elemento estruturante:

- um **modal**: a divisão entre o processamento da língua escrita e o da língua falada. No entanto, cada vez se realça mais a necessidade de colaboração entre os investigadores e as comunidades de cada pólo. Além disso, várias aplicações combinando os dois tipos existem já, como se verá em seguida. Por isso, parece-me mais didáctico juntar o processamento dos dois modos no presente capítulo.
- outro **metodológico**: o foco nos meios ou nos fins. Mesmo que alguns investigadores (ou escolas de investigação) se concentrem nos meios de melhorar o próprio trabalho na área, tal actividade só fará sentido se for subordinada ao objectivo de construir sistemas que processem robustamente alguns aspectos da linguagem natural. Por outras palavras, estou convencida de que não vale a pena construir recursos (dicionários, *corpora*, etc.), formalismos para a escrita de gramáticas computacionais, ou ambientes de tratamento de *corpora*, por exemplo, se não se tiver sempre em conta a adequação desses mesmos recursos ou programas aos seus objectivos finais.
- e outro **socio-económico**: produtos comerciais, ou apenas de I&D. Acontece que as relações entre as duas esferas (empresa/universidade) são muito variadas (por vezes são até completamente inexistentes). Assim, uma panorâmica cobrindo apenas um dos domínios arriscava-se a ter mais valor sociológico do que científico. Os produtos e as descrições técnicas e científicas de sistemas não comercializados serão, pois, referidos a par no texto, sinalizando tipograficamente os primeiros simplesmente através de maiúsculas reduzidas.⁸

Decidi, portanto, classificar as aplicações apenas segundo o tipo de tarefas a que se destinam, a um nível muito intuitivo e não especializado. Como qualquer classificação do nosso quotidiano, casos haverá que não cabem apenas numa categoria, nem as categorias são mutuamente exclusivas. Estando o leitor advertido em relação a este ponto, segue a panorâmica.

4 Tipos de aplicações

Há várias tarefas comuns na sociedade de informação actual que incluem em maior ou menor grau capacidades linguísticas e para as quais o computador, através de programas apropriados, pode ser um poderoso auxiliar:

- Escrita e produção de um texto
- Leitura e folheamento
- Tradução
- Aprendizagem e ensino
- Sistemas de informação
- Sistemas interactivos
- Indexação
- Entrada de dados
- Segurança e identificação

4.1 *Escrita (ou produção de um texto)*

No campo da ajuda à redacção incluem-se programas que detectam erros (detectores), sugerem alternativas (correctores), disponibilizam recursos – como tesouros e dicionários (monolíngues, bilingues, de sinónimos) – e fornecem ajuda de gramática. Veja-se, por exemplo, o FLIP, que oferece correcção ortográfica e sintáctica para o português de Portugal (embora esta última ainda um pouco rudimentar), e as capacidades de "grammar checking" do inglês no WORD, ou o sistema para o português do Brasil descrito em (Martins et al., 1998). Correctores estilísticos existem também que, além de problemas sintácticos, analisam fenómenos como a coesão, as repetições e o uso de um nível de língua não apropriado (Ravin, 1993; Richardson e Braden-Harder, 1993).

Os formatadores, que dispõem o texto segundo certos parâmetros, incluem em geral também conhecimento sobre a língua de forma a poderem efectuar uma translineação aceitável. Veja-se (Guerreiro, 1983), ou o hifenizador do FLIP.⁹

Existem, além disso, aplicações que tentam ir mais longe, criando semi-automaticamente documentos de tipo especializado. Por exemplo, partindo de informação presente em bases de dados ou codificada noutras formas distintas da linguagem natural, foram implementados programas para ajudar a escrever anúncios em várias línguas (Somers et al., 1997), na elaboração de boletins meteorológicos (Bourbeau et al., 1990), e na concepção de páginas da Web de conteúdo variado (DiMarco e Foster, 1997), entre muitos outros.

De facto, quanto mais especializado for o texto que se pretende escrever, mais fácil é desenvolver programas para ajudar à sua criação. Ainda que a forma de sofisticação em relação ao PLN seja muito variável, penso que se deve mencionar programas de apoio à escrita de dicionários monolíngues (Colaço, 1994; Novais e Vilela, 1997) e bilingues (Machado et al., 96), sistemas de ajuda ao trabalho terminológico (Ranchhod e Mota, 1999), sistemas especializados em cartas comerciais (Coch, 1996), além de, obviamente, sistemas de apoio à publicação na Web (DREAMWEAVER), na feitura de acetatos (POWERPOINT), e na criação de bibliografias (ENDNOTE).

A maior parte do que as pessoas escrevem é texto na sua própria língua (ou noutra), contudo podemos alargar o conceito de escrita à escrita de programas, e dentro destes referir até os sistemas de desenvolvimento de gramáticas computacionais. Estes últimos podem por sua vez incorporar processamento de

linguagem natural, através de sistemas de explicação interactivos (Tong, 1990). É também possível – e mesmo comum – ter sistemas de apoio à escrita (e à documentação) dos programas (Paris e Vander Linden, 1996) ou mesmo à documentação em geral (Ramalho et al., 1996).

Outro género de auxiliar na escrita são os sistemas de ditado, que permitem converter um discurso oral num texto escrito (VIA VOICE, DRAGON NATURALLY ORGANIZED).

Generalizando o termo "escrita" à produção de um texto, escrito ou falado, pode também mencionar-se a ajuda a deficientes motores sem capacidade de produção oral na comunicação mediada por computador (McCoy et al., 1990).

Limitando por outro lado a "escrita" à criação de um documento com certas restrições, podemos referir a existência de sistemas – como é o caso da Boeing (Wojcik e Holmback, 1996) – que, através do uso de analisadores de uma linguagem simplificada ("controlled language"), forçam uma consistência no estilo e na linguagem de milhares de produtores de manuais em inglês, simplificando significativamente a tradução automática subsequente.

Finalmente, no caso das línguas com um número muito elevado de ideogramas e combinações possíveis, uma ajuda que não é de desprezar é a dos sistemas inteligentes de digitação, associando imediatamente uma desambiguação (veja-se, por exemplo, para o chinês (Chen e Lee, 1996) e para o japonês (Takahashi et al., 1996)). Sistemas que reduzem a digitação foram também gizados para línguas europeias (Foster et al., 1997) e mesmo para o português (Calejo et al., 1986).

Outras formas de "apoio à escrita" em sentido lato serão mencionadas mais adiante nas rubricas "Entrada de dados" e "Aprendizagem e ensino de línguas estrangeiras"; resta referir a acepção de "escrita" como criação literária, para a qual também existem programas que facilitam a escrita colaborativa (Neale e Stanton, 1999).

4.2 *Leitura e folheamento*

Não é só para escrever que os programas de processamento das línguas são úteis. Também para a ajuda à leitura (em sentido lato), podemos considerar os sistemas de procura como uma ferramenta importantíssima. Essa procura pode ter como domínio uma base de documentos relativamente homogénea (por exemplo textos jurídicos, veja-se LEXONLINE, ou médicos (Quaresma e Lopes, 1993) ou uma rede que corresponde a um conjunto heterogéneo de fontes de informação como a Web (ALTAVISTA), uma intranet, ou mesmo apenas um conjunto de vários CDs diferentes (Sakai et al., 1996).

Um dos grandes problemas é como obter os documentos mais relevantes (tipicamente o problema principal da disciplina de recuperação da informação, "information retrieval" em inglês (IR), que usa tradicionalmente pouca informação linguística (Jones, 1999)) e que vários utilizadores têm tentado melhorar com o processamento de linguagem natural, veja-se (Klavans et al., 1997). A outra questão complicada é a de como apresentar de forma condensada os resultados. Este caso, em contraposição com a IR, é mais naturalmente considerado do âmbito da engenharia da linguagem.

Duas aplicações com grande interesse no nosso tempo são

- a extracção de informação (IE, "information extraction"), por exemplo de parágrafos relevantes: sistemas que seleccionam um subconjunto

relevante de um texto em vez de mostrarem o texto todo (Watanabe, 1996)

- e os sumarizadores: sistemas que criam um resumo, ou sumário, de um texto (veja-se o capítulo 3 de (Hovy et al., 1998)).

Em ambos os casos, a chamada tradução acoplada à rede dá a possibilidade de folhear (e mostrar) milhões de documentos que possam estar escritos em línguas diferentes – veja-se a rubrica "Tradução" no presente artigo – e um dos domínios de investigação mais em voga neste momento é o chamado "Multilingual Information Retrieval" (MLIR) (Hovy et al., 1998).

Também a recém-chamada actividade de "data mining" (garimpo de dados), que pretende extrair, de grandes concentrações de dados, conhecimento que apenas se encontra implícito, já inclui uma sub-área chamada "text mining", em que uma das principais actividades parece ser a criação de ambientes para manipular grandes conjuntos de textos e especificar dinamicamente procuras complexas e sua visualização (Landau et al., 1998).

Num ambiente de trabalho, a leitura de correio electrónico em andamento (e outras mensagens, em sistemas móveis) é uma aplicação muito útil dos sintetizadores de voz, que podem aliás proceder à leitura de todo o tipo de textos quando o utilizador tem deficiências visuais. As características destas duas aplicações de síntese de fala são, contudo, bem diferentes: Enquanto no primeiro caso o fundamental é o programa sintetizar a mensagem e não se perder com as convenções (assinaturas, cabeçalhos, marcas ideográficas, etc.) do correio electrónico, no segundo caso a ênfase deverá ser posta na produção de informação inteligível e variada, a partir, por exemplo, de uma multidão de janelas com efeitos gráficos intermitentes (a síntese do resultado de um leitor de écran, "screen reader", veja-se HOME PAGE READER).

As chamadas visitas guiadas (Guinan e Smeaton, 1992), criadas para iniciar um leitor num dado assunto ou local (electrónico) constituem outra forma de ajuda à leitura que permite um certo grau de liberdade por parte do utilizador, e cujas fronteiras com as actividades de aprendizagem e ensino se esbatem.

Finalmente, a consulta às últimas notícias através do telefone é uma aplicação a considerar (Boitet, 1990), assim como a procura de informação em material falado (AISA). Em geral, existe grande interesse pelas chamadas aplicações telefónicas (Gagnoulet et al., 1991), ou seja, processamento de fala integrado em serviços de telecomunicações, área em que já existem vários sistemas suficientemente robustos para serem usados no quotidiano das empresas (como é o caso da Lucas Arts).

Veja-se, para a comunicação em português falado com o computador, (Oliveira, 1996) e (Neto, 1998) respectivamente para a síntese e para o reconhecimento de fala.

4.3 Tradução

A tradução é uma das actividades que envolve mais conhecimento linguístico, visto que codifica a informação presente no texto de uma língua num texto de outra língua. Não é pois de estranhar que tenha sido a primeira área em que se trabalhou em PLN.

No âmbito da tradução encontramos aplicações que cobrem todo o espectro:

Sistemas de tradução (completamente) automática, tais como o TAUM-METEO (Isabelle, 1987), utilizado pelas autoridades meteorológicas canadianas para produzir boletins meteorológicos bilingues, ou como o sistema SYSTRAN cuja

utilização pelo motor de procura ALTAVISTA é facultada a qualquer utilizador.¹⁰ Ou como sistemas como os descritos por (Resnik, 1997) ou (Seligman, 1997), que servem para folhear a Web em línguas completamente desconhecidas sem atraso significativo para o navegante.

Sistemas de tradução semi-automática, por seu lado, podem tratar das partes fáceis, interagindo com o utilizador quando não conseguem decidir (Melby, 1987, Boitet e Tomokoyo, 1996), enquanto outros sistemas tentam minimizar o trabalho do tradutor completando as palavras ou frases ditadas após um mínimo de digitação (Brousseau et al, 1995), ou entrando em conta com as repetições em manuais técnicos (Merkel et al., 1994). Outros programas tentam mesmo oferecer caminhos já traçados, baseados em traduções prévias (Larsson e Merkel, 1993).

Entre as aplicações na área do apoio à tradução (humana), as estações de trabalho para tradutores são já uma realidade (TRADOS WORKBENCH, IBM TRANSLATION MANAGER, DÉJAVU, Picchi et al., 1992, Isabelle et al., 1993), facultando aos utilizadores o acesso a traduções prévias (memórias de tradução) e a recursos bilingues e terminológicos. Dentro dessas estações ou em separado, encontram-se programas que fazem a detecção de possíveis erros de tradução (verificadores de tradução, Macklovitch, 1995), e outros que permitem verificar a consistência terminológica de uma tradução (Macklovitch, 1996, TRANSLXIS).

Por outro lado, e apesar da sua complexidade, a tradução da fala parece ter-se imposto como área de desenvolvimento devido à sua grande utilidade, encontrando-se quer sistemas entre duas línguas (Lavie et al., 1996), quer sistemas híbridos que recorrem a uma língua comum como o caso do sistema Verbmobil, que usa o inglês como língua franca entre o japonês e o alemão (Kay et al., 1993).

Não devem ser ignoradas, finalmente, as recentes tentativas de analisar sistematicamente o comportamento dos intérpretes (Loehr, 1998) ou as ligações entre anotação linguística e sistemas visuais, por exemplo a chamada "tradução com restrições de tempo" (Toole et al., 1998).

4.4 *Aprendizagem e ensino*

Entre as actividades em que usamos mais a nossa língua encontram-se o aprender e ensinar.

Sistemas que respondem a perguntas ou juntam, de uma forma "apresentável", informação sobre um dado tema são, ou podem ser, auxiliares preciosos no processo de aprendizagem, delegando no computador uma parte da actividade de ensinar. Existem assim tutores inteligentes que tentam uma interacção com o aluno que não seja desmotivante para este, integrando capacidades de interacção em linguagem natural (Fonseca, 1993). Estes programas apresentam uma progressão dinâmica conforme as capacidades demonstradas até ali pelo utilizador, tentando por exemplo corrigir mal-entendidos (Costa et al., 1986).

Uma das primeiras aplicações da Compreensão de Linguagem Natural ("Natural Language Understanding") foi a da geração de problemas de matemática e, em geral, sistemas de ICAI ("intelligent computer-aided instruction"), como por exemplo o SCHOLAR (Carbonell, 1970).

Por razões meramente expositivas, separo aqui os sistemas com intuítos pedagógicos ou didácticos explícitos dos sistemas de informação em geral e dos sistemas interactivos em particular, que serão tratados em secções posteriores.

4.5 *Aprendizagem e ensino de línguas*

Um caso particular de ensino e aprendizagem merece especial destaque: aquele que usa o processamento da língua para ensinar a própria língua (que pode ser a língua materna do utilizador, ou uma língua estrangeira).¹¹

Existem programas concebidos para melhorar a língua materna do aluno, através da crítica das próprias redacções e da sugestão, baseada em textos literários consagrados, de formas mais bonitas de escrever, a que poderemos chamar "auxiliares estilísticos". Uma outra ajuda na compreensão de textos na língua materna (ou outra) é o acoplamento de uma enciclopédia com desambiguação automática, de forma a que o aluno leia apenas a definição apropriada (Miller e Teibel, 1991).

Podemos também considerar como ferramentas essenciais para a investigação em linguística os sistemas de trabalho com *corpora*, ou seja, programas que permitem aos investigadores um acesso aos dados da língua que de outra forma lhes estaria vedado: por exemplo, o IMS Corpus Workbench (Christ, 1994), SARA (Aston e Burnard, 1996), as ferramentas baseadas no analisador morfológico Palavroso (Medeiros, 1992), ou ambientes como o INTEX (Silberstein, 1993), integrando mesmo gramáticas. Da mesma forma, os sistemas tutores de linguagem natural (Pardo e Rino, 1999) permitem a experimentação com várias teorias de formalização da descrição de uma dada língua. Estamos aqui a falar, claro, de um outro "aprender a língua" que não o saber exprimir-se nela.

Voltando à interpretação usual de "ensino de línguas", existem alguns programas que ensinam recorrendo ao confronto com outros modos (abordagem multimodal):

- formas de onda e sons gravados, para melhorar a pronúncia (FONOGRAFE), (também para o ensino de crianças com dificuldades cerebrais, Ponte e Azevedo (1993))
- jogos didácticos com imagens, som e texto, para habituar a criança ao som e à grafia de certas palavras em língua estrangeira (CURSO DE INGLÊS - WHO IS OSCAR LAKE)
- conjuntos fala / transcrição ou texto / leitura em voz alta, para familiarizar o aluno com diversos dialectos (Lingua) ou com características fonológico-gramaticais específicas de uma dada língua

A situação mais frequente no ensino das línguas é, contudo, a concentração nas capacidades de escrita e redacção, existindo vários programas didácticos para uso dos professores na aula (Tribble e Jones, 1990). É também frequente a criação (ou adaptação) de ferramentas de ajuda à redacção numa língua estrangeira (Granger, 1998). Neste âmbito pode referir-se a criação semiautomática de exercícios baseados em corpora paralelos (Frankenberg-Garcia, 1999), ou a utilização de jogos didácticos na Web (Ludilangue).

Finalmente, conjugadores automáticos de verbos (e/ou outras formas) são considerados por muitos como um auxiliar relevante na aprendizagem de línguas com morfologia complicada (VERBOTECA), ainda que existam também programas com um nível de sofisticação do ensino da gramática muito mais desenvolvido (Bick, 1997).¹²

Ainda que, por enquanto, produtos como os materiais didácticos para professores da disciplina de Português, como o BANCO DE QUESTÕES (LÍNGUA PORTUGUESA 9º ANO), não possam ser considerados aplicações de processamento de linguagem natural, já que as respostas (e as perguntas) se encontram armazenadas,

não será contudo de desprezar a possibilidade de tais aplicações virem a evoluir para uma maior flexibilidade e interacção com o utilizador (o professor). Por exemplo, poderiam vir a sugerir exercícios sobre textos não incluídos no CD, validados *a posteriori* pelo docente.

Da mesma forma, iniciativas como a Internet Grammar of English (Aarts et al., 1998), que usa a Internet como meio de publicar uma gramática do inglês tradicional, ou aplicações que fazem uso da WWW para ilustrar e levar o aluno a identificar padrões de uso numa dada língua, tais como o Grammar Safari (Grammar Safari) para o inglês, não se podem considerar como PLN mas é de prever o enriquecimento destas iniciativas a curto prazo.

É, no entanto, interessante notar que, do ponto de vista do utilizador, o ideal seria ter uma máquina com a qual se pudesse conversar, praticando assim a língua estrangeira (Atwell, 1999).

4.6 Sistemas de informação (ou de pergunta e resposta)

O acesso a informação factual é uma das actividades mais bem cotadas do nosso tempo, e a possibilidade de interrogar um sistema em linguagem natural (independentemente da forma como essa informação está codificada, ou simplesmente organizada) um dos objectivos mais compreensivelmente desejados pelo público em geral.

Se bem que no campo da linguagem escrita tais sistemas não tenham até agora provado de forma irrefutável serem mais úteis ou mesmo mais amigáveis do que sistemas "tradicionais" de menus ou comandos, a interacção através de fala (em geral pelo telefone) corresponde a um dos maiores sucessos do processamento computacional da língua (conjugando reconhecimento de fala, processamento e geração de língua escrita, e síntese de fala). Existem neste momento sistemas de informação sobre viagens aéreas que podem ser interrogados em inglês (Ward, 1990) e mesmo através da Web (Julia et al., 1997).

Existem também bases de dados (MICROSOFT ENGLISH QUERY), páginas amarelas (LEXIQUEST, QUIQUOIÒÙ) e enciclopédias com interacção em linguagem natural (edição CD-ROM ou WWW da Enciclopédia Britânica (BRITANNICA)) assim como jornais que permitem interacção electrónica com algumas capacidades linguísticas. Tal funcionalidade é possível em inglês em motores de procura na rede como o ALTAVISTA. Além disso, cada vez são mais sofisticados os métodos de apresentação de notícias nas edições electrónicas dos jornais (veja-se, por exemplo, o PÚBLICONLINE). Assim, o serviço prestado pelo Financial Times Electronic Publishing é baseado num tesouro actualizado dinamicamente (Collier, 1998). Veja-se também (Maria et al., 1998) para o português. Não é, além disso, de excluir em breve uma catalogação por perfil de leitor, constantemente renovada (Pereira et al., 1997).

O estabelecimento de resumos linguísticos, ou melhor, resumos em linguagem natural, do conteúdo de bases de dados é outra área de investigação neste momento (Bosc et al., 1998).

Finalmente, não é possível deixar de mencionar sistemas de diálogo em linguagem natural cujo objectivo é o aconselhamento do utilizador, por exemplo conselho financeiro (Bronisz et al., 1990), orientação curricular (Garcia e Lopes, 1999), ou simplesmente ajuda na interacção com um sistema operativo (Heyer et al., 1990).

4.7 *Sistemas interactivos*

Dar comandos a uma máquina através da fala é já possível, embora de forma rudimentar, em vários sistemas sofisticados, como carros, aviões, ou casas inteligentes (AVACCM); assim como na interacção com programas no trabalho ou em casa, como editores de texto e organizadores pessoais (DRAGON NATURALLYORGANIZED).

Além disso, com os novos paradigmas de objectos e agentes na computação, foi possível pensar e implementar agentes interactivos que não tenham como único objectivo transmitir informação. É já possível interagir em linguagem natural, falada, com assistentes (por exemplo um controlador de um leitor de CDs (Ball e Ling, 1993)), com peritos (Ball et al., 1997) ou com personagens de um mundo virtual (Lester et al., 1998). (Bates et al., 1991) descrevem a interacção por voz com um sistema de planeamento de transporte militar, enquanto (Hirschman et al., 1991) usam a fala para interagir com um sistema que procura o melhor caminho para uma dada deslocação.

É também possível obter ajuda contextual quer na procura de caminhos (Fraczac et al., 1998), quer na utilização de equipamentos (Ansari e Hirst, 1998).

Outra área em que a linguagem natural é enriquecida com descrições analógicas apenas possíveis no campo visual é a simulação animada de instruções (Webber e Di Eugenio, 1990); em contrapartida, também se procura obter uma narração automática em linguagem natural das acções efectuadas por esses agentes (ibidem), ou a partir de filmes em domínios restritos, tais como o futebol (Andre et al., 1988; GoalGetter).

Os sistemas designados por MOOs ("MUD, Object Oriented") em que vários participantes comunicam entre si através de um mundo virtual, podem também incorporar capacidades linguísticas, veja-se por exemplo o sistema MOOsaico em Portugal (MOOsaico).

Finalmente, destaque-se o potencial do processamento de linguagem natural para os jogos de computador, não só na criação automática de charadas, palavras cruzadas e jogos tipo "scrabble" (EU APRENDO PORTUGUÊS, 9.º ANO), como na própria interacção com as personagens, à semelhança do célebre programa Eliza (Weizenbaum, 1966).

4.8 *Indexação*

Para a posterior identificação de livros e outros materiais de colecção, assim como de textos e revistas electrónicas, é preciso uma incessante actividade de indexação, independentemente de estarmos a falar de bibliotecas digitais (Isaías, 1995), tradicionais, ou mistas (Lewis, 1991).

A actividade de indexação pode ser em alguns casos semi-automatizada (através de técnicas afins às das de recuperação de documentos). (Feldman et al., 1998) apresentam a construção semi-automatizada de uma taxonomia relativa a grandes conjuntos de documentos, enquanto (Quaresma et al., 1999) exemplificam o uso da expansão de uma procura com o auxílio de um tesouro e de um modelo de conhecimento jurídico.

Além disso, a indexação pode também ser tornada mais fácil através do próprio uso da linguagem natural, seja através de meta-comentários (como os existentes no protocolo HTTP (Zarri, 1997)), seja através do processamento das próprias referências presentes na Web (Amitay, 1998) ou da criação de uma rede lexical em hipertexto (Pustejovsky et al., 1997).

Note-se que a indexação de material textual pode ser considerada como um caso extremo da sumarização, tendo contudo usos bem precisos e diferentes da redução de material fornecido ao leitor. Veja-se a construção de índices, de tesauros, ou de bancos de neologismos (NW).

É também possível facilitar a catalogação de outro tipo de dados, como imagens, reconhecendo e processando os comentários falados do anotador (Srihari et al., 1997).

Finalmente, o próprio "baptismo" de produtos farmacêuticos, por exemplo, pode ser melhorado através do estudo dos nomes actuais e possíveis dos produtos (ver recente discussão na lista electrónica *corpora*), o mesmo se passando com a escolha de nomes de marcas internacionais (que é uma das actividades da empresa sueca Skriptor).

4.9 Entrada de dados

Para a maior parte das utilizações da informação presente em linguagem natural é preciso uma maior estruturação dos dados, de forma a dar atenção apenas aos factores considerados relevantes. Isso significa que é preciso proceder à extracção de dados sobre um determinado assunto para depois alimentar, por exemplo, um sistema de base de dados, ou qualquer outra forma de armazenar a informação sobre um dado assunto, em vez de criar manualmente repositórios de informação.

Este é um dos objectivos mais antigos do processamento de linguagem natural (ainda da altura em que o seu nome predilecto era Compreensão de Linguagem Natural). Veja-se a análise de histórias de acidentes (Lewis, 1991), de transacções comerciais e financeiras (McDonald, 1997), ou de nomeações oficiais (Chandrasekar e Srinivas, 1997).

Uma outra aplicação, bem implantada já a nível industrial, é a leitura óptica – e correcção respectiva – de textos (veja-se o International Workshop on Performance Evaluation Issues in Multilingual OCR (IWPEIMO), enquanto que a leitura de textos manuscritos (e em particular assinaturas) é também uma área participada (IWFHR).

A utilização do reconhecimento de voz para certo tipo de entrada de dados (por exemplo actualização de bases) parece ser uma realidade em domínios restritos, como já foi referido na secção anterior a propósito da catalogação de imagens. (Hunt, 1997) relata a avaliação de um sistema para tradutores profissionais de entrada de texto através da fala. Neste contexto, convém referir que garantir a mobilidade do utilizador permitindo ao mesmo tempo a comunicação falada traz problemas específicos que é também preciso investigar (Flanagan et al., 1991).

4.10 Segurança e identificação

Sistemas baseados em características únicas do falante podem ser usados para permitir que um sistema apenas reaja à "voz do dono". Da mesma forma, é possível conceber reconhecedores de assinaturas manuscritas (OSCAR).

Por outro lado, sistemas de atendimento e socorro público têm vantagens em usar um identificador automático de língua falada (Caseiro, 1998) de forma a redirigir a chamada para um operador que possa comunicar com a pessoa que precisa de ajuda.

De forma menos premente, mas cada vez mais exigível, a identificação da língua do cliente em qualquer outra aplicação permite uma interacção muito mais cordial (O'Hagan, 1996).

A forma de escrever também pode ser uma pista para a identificação de um autor, o que pode ter, além de interesse literário (Kenny, 1982), valor nos tribunais.

4.11 Outras

Existem outras aplicações do processamento computacional da língua, que, à falta de um rótulo abrangente, incluo aqui.

O planeamento político e linguístico pode partir de um inquérito (ou corpus) previamente estabelecido (Neustupný, 1978), assim como se pode, por exemplo, avaliar os resultados de uma reforma, ortográfica ou outra. Também da mesma forma se pode medir as principais dificuldades na apreensão e uso de uma língua e, a partir daí, estabelecer programas pedagógicos ou recomendações neste sentido (Sim-sim e Ramalho, 1993).

Através da manipulação estatística do conteúdo dos textos, investigadores há que produzem estudos sociológicos, históricos e mesmo psicológicos de uma dada população ou época. Ou, através dos textos publicados numa dada área (ou dos seus resumos), fazem a análise de uma disciplina científica (Moscarola e Bolden, 1998).

É possível também investigar o próprio processo de desenho de documentos ("document design") através de técnicas de PLN, como é exemplificado numa recente conferência sobre "Document Design" (DD98). Utilizando conceitos linguísticos tentou-se igualmente analisar a própria formação de conceitos em ciência (Cornuejols et al., 1999).

E pode-se medir o peso de uma cultura a partir de identificadores e expressões associadas a essa cultura em línguas estrangeiras, requerendo tal processo, contudo, consideração linguística não trivial (PLCLI).

Finalmente, alguns investigadores dedicam-se à identificação de formas de comunicação humana através da gravação e análise de diálogos, analisando fenómenos como confronto, mal-entendidos, humor e ironia (Linell, 1995); enquanto outros consideram a especificidade cultural nas novas comunidades electrónicas (Ess, 1998).

5 Algumas recomendações

Após mencionar tantos domínios de investigação diferentes, envolvendo em maior e menor grau conhecimentos (também diferentes) sobre a língua, existirá alguma coisa que os permita unificar? E, por sua vez, analisar sob outros ângulos? Discutirei aqui estas questões aqui apenas na perspectiva do desenvolvimento de sistemas, apresentando algumas reflexões para quem cria aplicações:

1. É preferível ter uma aplicação menos ambiciosa mas com um ambiente utilizável do que uma aplicação muito poderosa desligada do resto do ambiente de trabalho.
2. É preferível resolver o problema num número suficiente de casos e deixar o ser humano tomar o controlo em casos mais complicados do que criar um programa que, por vezes, produz resultados completamente inaceitáveis ou se descontrola.
3. A aplicação deve ser pensada na perspectiva do utilizador, não na do cientista nem na da empresa que pretende vender. A maior parte das pessoas (incluindo os decisores nas empresas que financiariam a introdução de PLN em produtos ou serviços) tem expectativas irrealistas quanto ao que se pode esperar do processamento de linguagem natural.

4. Nem todas as soluções concordam com a teoria linguística da moda, ou mesmo com **qualquer** teoria linguística.
5. A elegância de uma solução é sempre contrariada pela necessidade de optimização: ao optimizar, melhorando o desempenho, consegue-se sistemas mais eficientes e colaborantes, mas diminui-se em geral drasticamente a facilidade de alterar o sistema ou de o propor como modelo em conferências.
6. Há um cotovelo no diagrama complexidade/desempenho: a partir de um certo ponto a melhoria de qualidade implica um enorme dispêndio de recursos.

Visto que uma língua é um sistema aberto, é impossível garantir um certo nível de desempenho qualquer que seja o sistema, o falante, o registo, o domínio. Não é possível sequer testar antecipadamente um número representativo de exemplos (excepto em domínios muito limitados). Isto torna a engenharia da linguagem mais complexa do que outro tipo de processamento informático, mas também mais interessante.

De facto, estamos "condenados" a aprender muito sobre a língua quando desenvolvemos qualquer aplicação. Não há melhor recompensa, a nível profissional, do que ter um sistema que funciona e desempenha de forma satisfatória uma dada função tendo a ver com a nossa própria língua.

Resta-me, pois, desejar que muitos leitores do presente artigo venham a ter a satisfação de contribuírem para a existência de várias aplicações em português e para portugueses.

Agradecimento

Agradeço a Jan Engh, Signe Oksefjell e Elisabete Ranchhod os comentários pertinentes sobre a organização e conteúdo do presente artigo.

Referências

- AARTS, Bas; Nelson, Gerald; Buckley, Justin (1998), "The Internet Grammar of English: New horizons in grammar pedagogy", Renouf, Antoinette (ed.), *Explorations in Corpus Linguistics*, Rodopi: Amsterdam/Atlanta (pp.251-257).
- AISA. ESCA ETRW workshop Accessing information in spoken audio, <http://svr-www.eng.cam.ac.uk/~ajr/esca99.html>
- ALTA VISTA. <http://www.altavista.net/>
- AMITAY, E. (1998), "Using common hypertext links to identify the best phrasal description of target web documents", *Proceedings of the SIGIR'98 Post-Conference Workshop on Hypertext Information Retrieval for the Web (Melbourne, Australia, August 28, 1998)*, disponível em http://www.mri.mq.edu.au/~einat/publications/sigir_98.ps.
- ANDRE, E.; Herzog, G.; Rist, Th. (1988), "On the Simultaneous Interpretation of Real World Image Sequences and their Natural Language Description: the System SOCCER", *Proceedings of ECAI 88* (pp. 449-).
- ANSARI, Daniel; Hirst, Graeme (1998), "Generating warning instructions by planning accidents and injuries", *Proceedings of the 9th International Workshop on Natural Language Generation (Niagara-on-the-Lake, 5-7 August 1998)*, acessível de <ftp://ftp.cs.toronto.edu/pub/gh/Ansari+Hirst-Warnings-98.ps>.

- ASTON, Guy; Burnard, Lou (1996), *The BNC Handbook: Exploring the British National Corpus with SARA*, Edinburgh: Edinburgh University Press.
- ATWELL, Eric (1999), *The Language Machine: The impact of Speech and Language technologies on English Language Teaching*, The British Council, July 1999.
- AVACCM. Advanced Voice-enabled Applications for the Car Concertation Meeting (Brussels, 26 October 1998), <http://www.linglink.lu/le/events/speech-in-car/resume.html>
- BALL, J. Eugene; Ling, Daniel T. (1993), "Natural Language Processing for a Conversational Assistant", Microsoft Research Report MSR-TR-93-15, October 1993.
- BALL, Gene (1997), "Dialogue Initiative in a Web Assistant", Susan Haller & Susan McRoy (eds.), *Computational Models for Mixed Initiative Interaction: Papers from the 1997 AAI Spring Symposium*.
- BALL, Gene; Ling, Dan; Kurlander, David; Miller, John; Pugh, David; Skelly, Tim; Stankosky, Andy; Thiel, David; Van Dantzich, Maarten; Wax, Trace (1997), "Lifelike Computer Characters: The Persona Project at Microsoft", J. Bradshaw (ed.), *Software Agents*, Menlo Park, Ca.: AAI/MIT Press.
- BANCO DE QUESTÕES (LÍNGUA PORTUGUESA 9º ANO). Porto Editora, duas disquetes.
- BATES, Madeleine; Ellard, Dan; Peterson, Pat; Shaked, Varda (1991), "Using Spoken Language to Facilitate Military Transportation Planning", *Proceedings of the DARPA Speech and natural Language Workshop* (Pacific Grove, California, February 19-22, 1991), San Mateo, Calif.: DARPA (pp.217-220).
- BICK, Eckhard (1997), "Internet Based Grammar Teaching", Ellen Christoffersen & Bradley Music (eds.), *Datalogvistisk Forenings årsmøde 1997 - DALF '97* (Kolding, 1997) (pp. 86-106).
- BOITET, Christian (1990), "Towards Personal MT: general design, dialogue structure, potential role of speech", Hans Karlgren (ed.), *Proceedings of COLING'90* (Helsinki, 1990), Vol. 2 (pp.30-35).
- BOITET, Christian; Tomokiyo, Mutsuko (1996), "Theory and Practice of Ambiguity labelling with a View to interactive Disambiguation in Text and Speech MT ", *Proceedings of COLING'96* (Copenhagen, 5-9 August 1996), Copenhagen: Center for Sprogteknologi (pp.119-124).
- BOURBEAU, L.; Carcagno, D.; Goldberg, E.; Kittredge, R.; Polguère, A. (1990). "Bilingual generation of Weather Forecasts in an Operations Environment", Hans Karlgren (ed.), *Proceedings of COLING'90* (Helsinki, 1990), Vol. 1 (pp.90-92).
- BOSC, Patrick; Liétard, Ludovic; Pivert, Olivier (1998), "Extended Functional Dependencies as a Basis for Linguistic Summaries", Jan M. Zytkow & Mohamed Quafafou (eds.), *Principles of Data Mining and Knowledge Discovery: Proceedings of the Second European Symposium, PKDD'98* (Nantes, September 1998) [Lecture Notes in Artificial Intelligence 1510], Springer Verlag (pp.255-263).
- BRITANNICA. <http://corporate.britannica.co.uk/>
- BRONISZ, Didier; Grossi, Thomas; Jean-Marie, François (1990), "Advice-Giving Dialogue", Hans Karlgren (ed.), *Proceedings of COLING'90* (Helsinki, 1990), Vol. 1 (pp.41-43).

- BROUSSEAU, J.; Drouin, C.; Foster, G.; Isabelle, P.; Kuhn, R.; Normandin, Y.; Plamondon, P. (1995), "French Speech Recognition in an Automatic Dictation System for Translators: the TransTalk Project", *Proceedings of Eurospeech 95*, Madrid, Spain (pp.193-196).
- CALEJO, Miguel; Pereira, Luis Moniz; Porto, António (1986) "Linguagem Natural por Menus", *Actas do Segundo Encontro Português e Inteligência Artificial, EPIA-86* (Lisboa, 8-10 de Outubro, 1996), Lisboa (pp.161-174).
- CARBONELL, J.R. (1970), "AI in CAI: An Artificial Intelligence approach to Computer-Aided Instruction", *IEEE Transactions on Man-Machine Systems*, Vol. 11, pp.190-202.
- CASEIRO, Diamantino António (1998), "Identificação Automática da Língua em Fala Contínua", Tese de Mestrado, Instituto Superior Técnico, Março de 1998.
- CHANDRASEKAR, R.; Srinivas, B. (1997), "Gleaning information from the Web: Using Syntax to Filter out Irrelevant Information", *Natural Language Processing for the World Wide Web, Papers from the 1997 AAAI Symposium* (Stanford, March 24-26, 1997), Menlo Park, California: AAAI Press (pp.27-34).
- CHEN, Hsin-Hsi; Lee, Jen-Chang (1996), "Identification and Classification of Proper Nouns in Chinese Texts", *Proceedings of COLING'96* (Copenhagen, 5-9 August 1996), Copenhagen: Center for Sprogteknologi, Vol 1. (pp.222-229).
- COCH, José (1996), "Evaluating and comparing three text-production techniques", *Proceedings of COLING'96* (Copenhagen, 5-9 August 1996), Copenhagen: Center for Sprogteknologi, Vol. 1 (pp.249-54).
- COLAÇO, António Manuel Maduro (1994), "O vocabulário do discurso científico de vulgarização, Agosto de 1988 - Outubro de 1990", Tese de Mestrado, Universidade Nova de Lisboa, 1994.
- COLLIER, Alex (1998), "Identifying diachronic change in semantic relations", Renouf, Antoinette (ed.) *Explorations in Corpus Linguistics*, Rodopi: Amsterdam/Atlanta (pp.259-268).
- CORNUEJOLS, A., Tiberghien, A. & Collet, G. (1999), "Decomposing the Scientific Discovery Process Using Multiple Interpretations of "Notions"", L. Magnani, N.J. Nersessian & P. Thagard (eds.), *Model-based Reasoning in Scientific Discovery*, New York: Kluwer Academic/Plenum Publishers, 1999, resumo em <http://philos.unipv.it/courses/progra1.html>
- COSTA, Ernesto; Duchénoy, Sylvie; Kodratoff, Yves (1986), "A resolution based method for discovering students' misconceptions", *Actas do Segundo Encontro Português de Inteligência Artificial, EPIA-86* (Lisboa, 8-10 de Outubro, 1996), Lisboa: EPIA (pp.261-270).
- CURSO DE INGLÊS - WHO IS OSCAR LAKE. <http://www.textoeditora.pt/DD98>. <http://cwis.kub.nl/~fdl/research/tw/docdes98/>
- DÉJA VU. <http://www.atril.com/whatsdv.html>
- DIMARCO, Chrysanne; Foster, Mary Ellen (1997), "The automated generation of Web documents that are tailored to the individual reader", *Natural Language Processing for the World Wide Web, Papers from the 1997 AAAI Symposium* (Stanford, March 24-26, 1997), Menlo Park, California: AAAI Press (pp.44-53).
- DRAGON NATURALLYORGANIZED.
<http://www.dragonsys.com/products/naturallyorganized.html>

- DREAMWEAVER. <http://www.macromedia.com/software/dreamweaver/index.shtml>
- EU APRENDO PORTUGUÊS, 9.º ANO. Porto Editora Multimedia, 1997.
- ESS, Charles (1998), "First Looks: CATaC'98", C.Ess & F. Sudweeks (eds.), *Proceedings of Cultural Attitudes Towards Communication and Technology '98* (University of Sydney, Australia) (pp.1-17), acessível de http://www.it.murdoch.edu.au/~sudweeks/catac98/01_ess.html
- FELDMAN, Ronen; Fresko, Moshe; Kinar, Yakkov; Lindell, Yehuda; Liphstat, Orly; Rajman, Martin; Schler, Yonatan; Zamir, Oren (1998), "Text Mining at the Term Level", in Jan M. Zytkow & Mohamed Quafafou (eds.), *Principles of Data Mining and Knowledge Discovery: Proceedings of the Second European Symposium, PKDD'98* (Nantes, September 1998) [Lecture Notes in Artificial Intelligence 1510], Springer Verlag (pp.65-73).
- FLANAGAN, J.L.; Mammone, R.; Elko, G.W. (1991), "Autodirective Microphone Systems for natural communication with speech recognizers", *Proceedings of the DARPA Speech and natural Language Workshop* (Pacific Grove, California, February 19-22, 1991), San Mateo, Calif.: DARPA (pp. 160-163).
- FLiP. <http://www.priberam.pt/Flip/>
- FONOGRÁFE - (Win) - CD-Rom, Aprendizagem da Língua Portuguesa, Porto Editora, 1997.
- FONSECA, Ana Cristina de Sena Raposo Paiva (1993), "Comunicação em Linguagem Natural para um Tutor Inteligente", Tese de Mestrado, Instituto Superior Técnico, Junho de 1993.
- FOSTER, G.; Isabelle, P.; Plamondon, P. (1997), "Target-Text Mediated Interactive Machine Translation", *Machine Translation*, **12**:1-2, pp.175-194.
- FRACZAK, Lidia; Lapalme, Guy; Zock, Michael (1998), "Automatic generation of subway directions: Saliency gradation as a factor for determining message and form", *Proceedings of the Ninth International Workshop on Natural Language Generation* (Niagara-on-the-Lake, 5-7 August 1998), disponível em <http://www.iro.umontreal.ca/~scriptum/FraczakINLG98.ps.gz>
- FRANKENBERG-GARCIA, Ana (1999), "Crosslinguistic influence as a key to extracting second language teaching materials for monolingual classes from translation corpora", Granger, Sylviane (ed.), *Proceedings of the Workshop Contrastive Linguistics and Translation Studies: Empirical Approaches* (Louvain-la-Neuve, 5-6 February 1999).
- GAGNOULET, Christian; Sorin, Christel (1991), "Field test evaluations and Optimization of Speaker independent speech recognition for telephone applications", *Proceedings of the DARPA Speech and natural Language Workshop* (Pacific Grove, California, February 19-22, 1991), San Mateo, Calif.: DARPA (pp.160-163).
- GARCIA, B. Borges; Lopes, Gabriel P. (1999), "Medidas de plausibilidade para determinar a aceitação de informação em diálogos", *Actas do Workshop sobre Linguística Computacional da APL* (Lisboa, Maio 1998), Lisboa: APL.
- GoalGetter. <http://iris19.ip0.tue.nl:9000/english.html>
- Grammar Safari. <http://deil.lang.uiuc.edu/web/pages/grammarsafari.html>
- GRANGER, Sylviane (1998), "The Computer Learner Corpus: A Testbed for Electronic EFL Tools", John Nerbonne (ed.), *Linguistic Databases*, CSLI Publications (pp.175-188).

- GUERREIRO, Pedro (1983), "Sílabas, o formatador-maravilha: Manual de utilização", CIL-1/83, Centro de informática do LNETI, Março 1983.
- GUINAN, Catherine; Smeaton, Alan F. (1992), "Information Retrieval from Hypertext Using Dynamically Planned Guided Tours", D. Lucarella, J. Nanard, M. Nanard & P. Paolini, *Proceedings of the ACM Conference on Hypertext* (Milão, 30 Novembro – 4 de Dezembro, 1992), New York: ACM (pp.122-130).
- HEYER, Gerhard; Kese, Ralf; Oemig, Frank; Dudda, Friedrich (1990), "Knowledge Representation and Semantics in a Complex Domain: The UNIX Natural Language Help System GOETHE", Hans Karlgren (ed.), *Proceedings of COLING'90* (Helsinki, 1990), Vol. 3 (pp. 361-363).
- HIRSCHMAN, Lynette; Seneff, Stephanie; Goodine, David; Phillips, Michael (1991), "Integrating Syntax and Semantics into Spoken Language Understanding", *Proceedings of the DARPA Speech and natural Language Workshop* (Pacific Grove, California, February 19-22, 1991), San Mateo, Calif.: DARPA (pp.366-371).
- HOVY, Eduard; Ide, Nancy; Frederking, Robert; Mariani, Joseph; Zampolli, Antonio (eds.) (1998) "Multilingual Information Management: Current Levels and Future Abilities", <http://www.cs.cmu.edu/~ref/mlim/>, July 1998.
- HOME PAGE READER. IBM Home Page Reader for Windows. <http://www.austin.ibm.com/sns/hpr.html>
- HUNT, Melvyn J. (1997), "Practical Large-Vocabulary Speech Recognition in a Multilingual Environment", *Speech Communication* Vol. 23, No. 4, December 1997, pp.297-305.
- IBM TRANSLATION MANAGER. <http://www.qsoft.de/ibmtrans/tm2.htm>
- ISABELLE, Pierre (1987), "Machine Translation at the TAUM Group", Margaret King (ed.), *Machine Translation: The State of the Art*, Edinburgh: Edinburgh University Press (pp. 247-318).
- ISABELLE, Pierre; Dymetman, Marc; Foster, George; Jutras, Jean-Marc; Macklovitch, Elliot; Perrault, François; Ren, Xiaobo; Simard, Michel (1993), "Translation Analysis and Translation Automation", *Proceedings of the Fifth International Conference on Theoretical and Methodological Issues in Machine Translation, TMI'93* (Kyoto, July 14-16, 1993) (pp.201-217).
- ISAÍAS, Pedro; Carvalho, Tiago; Assis, Ana Cristina (1995), "Bibliotecas Digitais na World Wide Web - Uma Proposta para a Universidade Nova de Lisboa (UNL)", *Actas da Conferência Nacional WWW, Informação Multimédia na Internet* (Braga, 6-8 de Julho de 1995), <http://www.di.uminho.pt/IMI/Proceedings/index.html>.
- IWFHR. The 6th International Workshop on Frontiers in Handwriting Recognition (Taejon, Aug 12-14,1998), <http://ai.kaist.ac.kr/iwfhr98/>
- IWPEIMO. International Workshop on Performance Evaluation Issues in Multilingual OCR (Bangalore, September 19, 1999) <http://www.cfar.umd.edu/~kanungo/workshop/mlocr.html>
- JONES, Karen Sparck (1999), "What is the role of NLP in text retrieval?", Tomek Strzalkowski (ed.), *Natural Language Information Retrieval*, Dordrecht/Boston/London: Kluwer Academic Publishers (pp.1-24).
- JULIA, Luc; Neumeyer, Leonardo; Charafeddine, Mehdi; Cheyer, Adam; Dowding, John (1997), "HTTP://WWW.SPEECH.SRI.COM/DEMOS/ATIS.HTML", *Natural Language Processing for the World Wide Web*,

- Papers from the 1997 AAAI Symposium* (Stanford, March 24-26, 1997), Menlo Park, California: AAAI Press (pp.72-76)
- KAY, Martin; Gawron, Jean Mark; Norvig, Peter (1994), *Verbmobil: A Translation System for Face-to-Face Dialog*, CSLI Lecture Notes, No. 33, Stanford, Calif.: Center for the Study of Language and Information, 1994.
- KENNY, Anthony (1982), *The computation of style: an introduction to statistics for students of literature and humanities*, Oxford: Pergamon, 1982.
- KLAVANS, Judith; Jacquemin, Christian; Tzoukermann, Evelyne (1997), "A natural language approach to multi-word term conflation", *Proceedings of the DELOS Workshop on Cross-Language Information Retrieval*, ETHZ, Zurich, ERCIM: European Consortium for Informatics and Mathematics, 1997.
- LANDAU, D.; Feldman, R.; Zamir, O.; Aumann, Y.; Fresco, M.; Lindell, Y.; Lipshtat, O. (1998), "TextVis: An Integrated Visual Environment for Text Mining", Jan M. Zytkow & Mohamed Quafafou (eds.), *Principles of Data Mining and Knowledge Discovery: Proceedings of the Second European Symposium, PKDD'98* (Nantes, September 1998) [Lecture Notes in Artificial Intelligence 1510], Springer Verlag (pp.56-64).
- LARSSON, Arne; Merkel, Magnus (1993), "Semiotics at Work: Technical Communication and Translation in a Multilingual Corporate Environment", *Proceedings of the Ninth Nordic Conference on Computational Linguistics, NODALIDA'93* (Stockholm, 1993) (pp.155-163).
- LAVIE, Alon; Gates, Donna; Gavalda, Marsal; Mayfield, Laura; Waibel, Alex; Levin, Lori (1996), "Multi-lingual Translation of Spontaneously Spoken Language in a Limited Domain", *Proceedings of COLING'96* (Copenhagen, 5-9 August 1996), Copenhagen: Center for Sprogteknologi, Vol. 1 (pp.442-447).
- LESTER, James; Bares, William; Callaway, Charles; Towns, Stuart (1998), "Natural Language Generation Journeys to Interactive 3D Worlds", *Proceedings of the Ninth International Workshop on Natural Language Generation* (Niagara-on-the-Lake, 5-7 August 1998) (pp.2-7), disponível em <http://www.csc.ncsu.edu/eos/users/l/lester/www/imedia/papers.html#inlg-98>.
- LEWIS, David D. (1991), "Evaluating text categorization", *Proceedings of the DARPA Speech and natural Language Workshop* (Pacific Grove, California, February 19-22, 1991), San Mateo, Calif.: DARPA (pp.312-318).
- LEXIQUEST. <http://www.erli.com/products/guidedtourt.htm>
<http://demo.erli.com/LQC/explain.htm>
- LEXONLINE. <http://www.jurinform.pt/online/>
- Lingua. http://www.clul.ful.pt/sectores/projecto_portuguesfalado.html
- LINELL, Per (1995), "Troubles with mutuality: Towards a dialogical theory of miscommunication and misunderstanding", I. Markova, C.F. Graumann, K. Foppa (eds.), *Mutualities in dialogue*, Cambridge: Cambridge University Press.
- LOEHR, D. (1998), "Can Simultaneous Interpretation Help Machine Translation", David Farwell, Laurie Gerber, Eduard Hovy (eds.), *Machine Translation and the Information Soup (Proceedings of the Third Conference of the Association for Machine Translation in the Americas, AMTA'98)* (Langhorne, October 28-31, 1998), Springer Verlag (pp.213-224).
- Ludilangue. Projet LUDILANGUE.
<http://lilla2.unice.fr/demos/FLE/EAO/ludilang/pg0.htm>

- MACHADO, Altamiro; Sá, Vítor; Sun, Tao (1996), "Multimedia Encyclopaedia for the teaching Portuguese/Chinese/Portuguese", *Actas da 8ª Conferência de Reconhecimento de Padrões, RECPAD'96* (Braga, 1996) (pp.487-90).
- MACKLOVITCH, Elliott (1995), "TransCheck - or the Automatic Validation of Human Translations", *Proceedings of the MT Summit V* (Luxembourg, 1995).
- MACKLOVITCH, Elliott (1996), "Peut-on vérifier automatiquement la cohérence terminologique?", *META*, Vol. 41, no. 3.
- MARIA, Nuno; Gaspar, Pedro; Ferreira, António; Silva, Mário J. (1998), "Information Preservation in ARIADNE", *Proceedings of the 6th DELOS Workshop* (Tomar, 17-19 Junho de 1998).
- MARTINS, R.T.; Hasegawa, R.; Nunes, M.G.V.; Montilha, G.; Oliveira Jr., O.N. (1998), "Linguistic issues in the development of ReGra: a Grammar Checker for Brazilian Portuguese", *Natural Language Engineering*, Vol. 4 (Part 4, December 1998), pp. 287-307, acessível de <http://nilc.icmc.sc.usp.br/download/arq.zip>
- MCCOY, Kathleen; Demasco, Patrick; Jones, Mark; Pennington, Christopher; Rowe, Charles (1990), "Applying Natural Language Processing Techniques to Augmentative Communication Systems", Hans Karlgren (ed.), *Proceedings of COLING'90* (Helsinki, 1990), Vol. 3 (pp.413-415).
- MCDONALD, David D. (1997), "Markup is Markup", *Natural Language Processing for the World Wide Web, Papers from the 1997 AAAI Symposium* (Stanford, March 24-26, 1997), Menlo Park, California: AAAI Press (pp.104-111).
- MEDEIROS, José Carlos (1992), "Ferramentas de processamento de corpora usando o PALAVROSO", Santos, Diana (ed.), "Processamento de corpora no INESC", Vol. 1, Relatório INESC RT-65/92.
- MELBY, Alan (1987), "On human-machine interaction in translation", Sergei Niremburg (ed.), *Machine Translation: Theoretical and Methodological Issues*, Cambridge: Cambridge University Press.
- MERKEL, Magnus; Nilsson, Bernt; Ahrenberg, Lars (1994), "A phrase-retrieval system based on recurrence", *Proceedings of the Second Annual Workshop on Very Large Corpora (WVLC-2)* (Kyoto, August 1994), Kyoto, 1994 (pp.99-108).
- MICROSOFT ENGLISH QUERY. <http://www.microsoft.com/sql/70/gen/eqmain.htm>
- MILLER, George A.; Teibel, Daniel A. (1991), "A proposal for lexical disambiguation", *Proceedings of the DARPA Speech and natural Language Workshop (Pacific Grove, California, February 19-22, 1991)*, San Mateo, Calif.: DARPA (pp.395-399).
- MOOsaico. <http://moo.di.uminho.pt/~pmoo/>
- MOSCAROLA, J.; Bolden, R. (1998), "From the Data Mine to the Knowledge Mill: Applying the Principles of Lexical Analysis to the Data Mining and Knowledge Discovery Process", Jan M. Zytkow & Mohamed Quafafou (eds.), *Principles of Data Mining and Knowledge Discovery: Proceedings of the Second European Symposium, PKDD'98* (Nantes, September 1998) [Lecture Notes in Artificial Intelligence 1510], Springer Verlag (pp.405-413).
- NEALE, Helen; Stanton, Danaë (1999), "KidStory: Evaluating Children's Collaborative Story Creation", Mimo Caenepeel, David Benuon & Duncan Smith (eds.), *Community of the Future: i3 Annual Conference* (Sienna, 20-

- 22 October 1999), Edinburgh: The Human Communication Research Centre, the University of Edinburgh (pp.30-35).
- NETO, João Paulo da Silva (1998), "Reconhecimento da Fala Contínua com aplicação de técnicas de Adaptação ao Orador", Tese de Doutoramento, Instituto Superior Técnico, Fevereiro de 1998.
- NEUSTUPNÝ, J.V. (1978), *Post-structural approaches to language: Language theory in a Japanese Context*, Tokyo: University of Tokyo Press.
- NOVAIS, A.Q.; Vilela, Mário (1997), "Prontuário Multimédia: Diciomédia", *Revista da Faculdade de Letras "Língua e Literatura" XIV*, pp.529-535.
- NW. <http://www.rdues.liv.ac.uk/newwds.html>
- O'HAGAN, Minako (1996), *The coming industry of teletranslation*, Clevelon / Philadelphia / Adelaide: Multilingual Matters Ltd., 1996.
- OLIVEIRA, Luís Miguel Veiga Vaz Caldas (1996), "Síntese de Fala a Partir de Texto", Tese de Doutoramento, Instituto Superior Técnico, Outubro de 1996.
- OSCAR. An Offline Script and Character Recognition Toolset, <http://vasawww.essex.ac.uk/newoscar/>
- PARIS, Cécile; Vander Linden, Keith (1996), "Building Knowledge Bases for the Generation of Software Documentation", *Proceedings of COLING'96* (Copenhagen, 5-9 August 1996), Copenhagen: Center for Sprogteknologi, Vol. 2 (pp.734-739).
- PEREIRA, Rui Miguel; Calado, Pável Pereira; Oliveira, Arlindo Limede (1997), "GAMA: um sistema para a filtragem de informação baseada em perfis de interesse", Relatório Técnico INESC, Dezembro 1997.
- PLCLI. "A presença das línguas e das culturas latinas na Internet", União Latina, 28 de Setembro de 1998, acessível em http://www.unilat.org/dtil/lenguainternet/pt/lingua/lingua_indice.htm.
- PICCHI, Eugenio; Peters, Carol; Marina, Elisabetta (1992), "A translator's workstation", *Proceedings of COLING'92* (Nantes, 23-28 July 1992) (pp.972-976).
- PONTE, M. Nunes da; Azevedo, L. (1993), "Using Microcomputers for Alternative and Augmentative Communication. A Project with Very Young Cerebral Palsied Children", *Closing the Gap Conference* (Minnesota, USA, 1993).
- POWERPOINT. Microsoft PowerPoint 97 SR-1.
- PÚBLICONLINE. <http://pesquisas.publico.pt/>
- PUSTEJOWSKY, J; Boguraev, B.; Verhagen, M.; Buitelaar, P.; Johnston, M. (1997), "Semantic Indexing and Typed Hyperlinking", *Natural Language Processing for the World Wide Web, Papers from the 1997 AAI Symposium* (Stanford, March 24-26, 1997), Menlo Park, California: AAI Press (pp.120-128).
- QUARESMA, Paulo; Lopes, José Gabriel (1993), "Reconhecimento de Intenções para uma Interação Robusta com Bases de Conhecimento Médicas", *Actas do 1º Encontro de Processamento de Língua Portuguesa (Escrita e Falada) - EPLP'93* (Lisboa, 25-26 de Fevereiro de 1993) (pp.27-31).
- QUARESMA, Paulo; Rodrigues, Irene Pimenta; Lopes, Gabriel; Almeida, Teresa; Garcia, Elsa; Lima, Ana (1999), "Um sistema de pesquisa de informação para bases de textos em Português", *Actas do IV Encontro para o Processamento Computacional da língua portuguesa escrita e falada - PROPOR'99* (Évora, 21-22 Setembro de 1999) (pp.209-218).
- QUIQUOIÒÙ. <http://quiquoiou.wanadou.fr/quiquoiou/html/plus4.html>

- RAMALHO, J.C.; Almeida, J.J.; Henriques, P.R. (1996), "Document Semantics: two approaches", *Proceedings of Celebrating a Decade of SGML*, 1996.
- RANCHHOD, Elisabete Marques; Mota, Cristina (1999), "Dicionários electrónicos de léxicos terminológicos. "Seguros"", *Actas do Workshop sobre Linguística Computacional da APL* (Lisboa, Maio 1998), Lisboa: APL.
- RAVIN, Yael (1993), "Grammar Errors and Style Weaknesses in a Text-Critiquing System", K. Jensen, G. Heidorn & S. Richardson (eds.), *Natural Language Processing: The PLNLP Approach*, Boston: Kluwer Academic Press (pp.65-76).
- RESNIK, Philip (1997), "Evaluating Multilingual Gisting of Web Pages", *Natural Language Processing for the World Wide Web, Papers from the 1997 AAAI Symposium* (Stanford, March 24-26, 1997), Menlo Park, California: AAAI Press (pp.129-135).
- RICHARDSON, Stephen; Braden-Harder, Lisa (1993), "The Experience of Developing a Large-Scale Natural Language Processing System: Critique", K. Jensen, G. Heidorn & S. Richardson (eds.), *Natural Language Processing: The PLNLP Approach*, Boston: Kluwer Academic Press (pp.77-89).
- PARDO, Thiago Alexandre Salgueiro; Rino, Lucia Helena Machado (1999), "O Processo Computacional e Explicativo de uma Interface de Redução entre Formalismos Gramaticais", *Actas do IV Encontro para o Processamento Computacional da língua portuguesa escrita e falada - PROPOR'99* (Évora, 21-22 Setembro de 1999) (pp.193-197).
- SAKAI, Keiichi; Yagisawa, Tsuyoshi; Fujita, Minoru (1996), "A CD-ROM Retrieval System with Multiple Dialogue Agents", *Proceedings of COLING'96* (Copenhagen, 5-9 August 1996), Copenhagen: Center for Sprogteknologi, Vol. 2 (pp.400-405).
- SANTOS, Diana (no prelo), "Toward language-dependent applications", *Machine Translation*.
- SELIGMAN, Mark (1997), "Evaluating Multilingual Gisting of Web Pages", *Natural Language Processing for the World Wide Web, Papers from the 1997 AAAI Symposium* (Stanford, March 24-26, 1997), Menlo Park, California: AAAI Press (pp.142-148).
- SIM-SIM, I.; Ramalho, G. (1993), *Como lêem as nossas crianças? Caracterização do nível de literacia da população escolar portuguesa*, Lisboa: GEP, Ministério da Educação.
- SILBERZTEIN, Max (1993), *Dictionnaires électroniques et analyse automatique de textes : le système INTEX*, Paris: Masson Ed.
- SOMERS, H.; Black, B.; Ellman, J.; Gilardoni, L.; Lager, T.; Multari, A.; Nivre, J.; Rogers, A. (1997), "Multilingual Generation and Summarization of Job Adverts: The TREE Project", *Proceedings of the Fifth Applied Natural Language Processing Conference* (Washington D. C., April 1997).
- SRIHARI, Rohini K.; Zhang, Zhongfei; Chopra, Rajiv (1997), "Show&Tell: Using Speech Input for Image Interpretation and Annotation", *Intelligent Integration & Use of Text, Image, Video, and Audio Corpora: Papers from the 1997 AAAI Spring Symposium*, Menlo Park: AAAI Press (pp.17-24).
- SYSTRAN. <http://www.systransoft.com/products.htm>
- TAKAHASHI, Masahito; Shinchu, Tsuyoshi; Yoshimura, Kenji; Shudo, Kosho (1996), "Processing Homonyms in the Kana-to-Kanji Conversion",

- Proceedings of COLING'96* (Copenhagen, 5-9 August 1996), Copenhagen: Center for Sprogteknologi, Vol. 2 (pp.1135-1138).
- TONG, Loong Cheong (1990), "An Explanation Facility for a Grammar Writing System", Hans Karlgren (ed.), *Proceedings of COLING'90* (Helsinki, 1990), Vol. 2 (pp.359-364).
- TOOLE, J.; Turcato, D.; Popowich, F.; Fass, D.; McFetridge, P. (1998), "Time-Constrained Machine Translation", David Farwell, Laurie Gerber, Eduard Hovy (eds.), *Machine Translation and the Information Soup (Proceedings of the Third Conference of the Association for Machine Translation in the Americas, AMTA'98)* (Langhorne, October 28-31, 1998), Springer Verlag (pp.103-112).
- TRADOS WORKBENCH. <http://www.trados.com/workbench/index.html>
- TRANSLLEXIS. <http://www.qsoft.de/ibmtrans/translex.htm>
- TRIBBLE, C.; Jones, G. (1990), *Concordances in the Classroom: A resource book for teachers*. Essex: Longman.
- Verboteca. <http://www.verboteca.com/>
- VIA VOICE. <http://www.software.ibm.com/speech/>
- WATANABE, Hideo (1996), "A Method for Abstracting Newspaper Articles by Using Surface Clues", *Proceedings of COLING'96* (Copenhagen, 5-9 August 1996), Copenhagen: Center for Sprogteknologi (pp.974-999).
- WEBBER, Bonnie Lynn; Di Eugenio, Barbara (1990), "Free Adjuncts in Natural Language Instructions", Hans Karlgren (ed.), *Proceedings of COLING'90* (Helsinki, 1990), Vol. 2 (pp.395-400).
- WEIZENBAUM, J. (1966), "ELIZA – A Computer Program for the Study of Natural Language Communication between Man and Machine", *Communications of the ACM*, Vol. 9, No. 1, pp.36-44.
- WOJCIK, R.H.; Holmback, H. (1996), "Getting a Controlled Language Off the Ground at Boeing", *Proceedings of the First International Workshop on Controlled Language Applications*, Katholieke Universiteit Leuven, Belgium, 1996 (pp. 22-31).
- WORD. Microsoft Word 97 SR-1.
- ZARRI, Gian Piero (1997), "Natural Language Indexing of Multimedia Objects in the Context of a WWW Distance Learning Environment", *Natural Language Processing for the World Wide Web, Papers from the 1997 AAAI Symposium* (Stanford, March 24-26, 1997), Menlo Park, California: AAAI Press (pp. 155-158).

¹ Cf. as chamadas para as grandes conferências COLING (International Conference on Computational Linguistics), os Annual Meetings of the Association for Computational Linguistics (ACL), as International Conferences on Spoken Language Processing (ICSLP), por sua vez descendentes das conferências ainda mais amplas AAAI, ICASSPs, etc, com sub-áreas como NLP ou Speech.

² Como a ANLP (Conference on Applied Natural Language Processing) ou a IANLP (International conference on Industrial Applications of Natural Language Processing).

³ Como, por exemplo, as International Conferences on Theoretical and Methodological Issues in Machine Translation (TMI), Conferences on Computational Lexicography and Text Research (COMPLEX), International Conferences on Language Resources and Evaluation (LREC), International Workshops on Speech Synthesis, Workshops on Lexicon Driven Information

Extraction, International Congresses on Expert Systems & their Applications, Symposia on Natural Language Processing for the World Wide Web, International Workshops on Natural Language Generation (INLG), International Congresses on Terminology and Knowledge Engineering (TKE), etc., etc.

⁴ Não considerarei aqui a definição alternativa, ou pelo menos complementar, em que o processamento computacional da língua é um meio para investigar a própria língua. Se aceitarmos também essa definição, nessa acepção o PLN é uma sub-área da psicologia, da filologia, e de uma forma geral das ciências cognitivas.

⁵ A palavra "aplicações" é, na minha opinião, infeliz (aplicações de quê?), mas já está suficientemente disseminada para impedir a sua substituição.

⁶ Refira-se que estas duas propriedades implicariam um trabalho gigantesco de pesquisa bibliográfica e um conhecimento profundo de cada área muito além da minha competência actual.

⁷ Veja-se (Colaço, 1994) para uma introdução ao discurso de divulgação.

⁸ Será, pois, de deixar bem claro que a menção ou omissão de produtos no presente artigo não obedece a quaisquer intuítos comerciais nem resulta de qualquer avaliação do mercado.

⁹ Veja-se <http://www.priberam.pt/FLiP/hifenizador.htm>.

¹⁰ Por exemplo de <http://www.alta-vista.net/>.

¹¹ Para o português, consulte-se o catálogo em <http://www.portugues.mct.pt> (Recursos: Material didáctico).

¹² Acessível para a língua portuguesa em <http://visl.hum.ou.dk/Linguistics.html>.