

A Língua Portuguesa no Projecto Eurotra

Gabriela Ardisson Mattos
Docente do Departamento de Lingüística
da Faculdade de Letras da Universidade de Lisboa
Investigadora do
Instituto de Lingüística Teórica e Computacional

1. OS OBJETIVOS

O Eurotra é um Projecto de Tradução Automática criado e desenvolvido pela Comunidade Europeia.

O Projecto Eurotra definiu como seus objectivos centrais: (i) a criação de um protótipo pré-industrial de tradução automática para textos escritos de um determinado tipo; (ii) o desenvolvimento de conhecimentos na área da tradução automática ao nível dos países da Comunidade Europeia.

Estes mesmos objectivos funcionaram desde o início como directrizes para os diferentes grupos de língua, entre os quais o núcleo português, o Eurotra-P.

2. A ORGANIZAÇÃO

O *Projeto Eurotra* apresenta uma *estrutura descentralizadora*, em grande parte motivada pela política de alargamento de competência no domínio da tradução automática entre os estados membros.

Assim, enquanto a Direção Geral do Projecto se encontra sediada no Luxemburgo, as equipas de investigadores das diferentes línguas envolvidas encontram-se nos respectivos países, funcionando nas universidades ou em centros de investigação.

Nessa estrutura organizacional destacam-se fundamentalmente os seguintes elementos: o *Director do Projecto*, Sergei Perschke, membro da Direcção Geral da Comissão 13 (DG XIII); os *Grupos de Língua* encarregados da implementação das gramáticas e dicionários monolíngues e da transferência das outras línguas para a sua; o *Liaison Group*, no qual estão representantes dos diferentes grupos de língua, e cuja função é assegurar a coordenação e orientação do trabalho.

O Projecto confere aos grupos de língua um papel eminentemente lingüístico, mais precisamente, de linguística descritiva.

Com efeito, a especificação do suporte lógico (software) e o seu refinamento estiveram a cargo de um *grupo central*, o mesmo acontecendo, aliás, com a definição do quadro de referência lingüístico.

A Decisão do Conselho de 26 de Novembro de 1986 estabeleceu a entrada de Portugal e de Espanha no Projecto Eurotra, na sequência da sua adesão à CEE. A criação oficial do *Grupo Português (Eurotra-P)*, porém, só viria a concretizar-se em Maio de 1987, data em que Portugal assinou o contrato de associação com a CEE para a execução do Projecto.

A ausência de tradição em linguística computacional em Portugal, e a própria arquitectura do Sistema, que prevê a independência do trabalho de codificação linguística relativamente ao suporte lógico (software), estiveram na base da *composição do grupo português*. Assim, o Eurotra-P foi inicialmente constituído por uma equipa de linguistas, docentes na Universidade, com formação em Sintaxe, Semântica e Morfologia, de recém licenciados em linguística, de um engenheiro informático, e de um técnico de programação, perfazendo, no total, 19 pessoas/mês.

A composição do grupo tem variado ao longo do tempo, especialmente no tocante ao número de linguistas seniores, procurando sempre, contudo, adequar-se às necessidades que vão sendo sentidas no seio desta unidade de investigação.

Nas suas linhas gerais, todavia, a estrutura hierárquica inicialmente estabelecida permanece: o Eurotra-P é constituído por um *Director do Projecto* (Maria Helena Mira Mateus), por um *Coordenador Científico*, e por vários *Coordenadores de área*, sendo as áreas consideradas usualmente as seguintes: gramática de análise, gramática de síntese, transferência, dicionários e terminologia.

3. O ESCOPO

As vastas necessidades de tradução da CEE, dado o número de línguas oficiais af presentes, foram um dos motores da criação do Projecto Eurotra.

Assim, o objectivo inicial do Sistema era produzir traduções de boa-qualidade de textos escritos – os documentos oficiais da CEE –, nas nove línguas oficiais da Comunidade.

Para atingir este objectivo a metodologia seguida foi a de estabelecer um *corpus* (Esprit – uma decisão do Conselho das Comunidades), que funcionasse como uma amostra representativa das estruturas gramaticais e do léxico presente nesse tipo de textos.

A descrição linguística no Projecto é assim orientada por um corpus, por vezes remodelado por razões de implementação, para se adequar à progressão da cobertura gramatical atingida.

Para restringir o domínio de aplicação do Sistema foram definidas, ao longo do Projecto, duas *sublinguagens* – a das Tecnologias da Informação, e a das Telecomunicações, mais especificamente, da Comunicação por Satélite.

A *cobertura gramatical*, porém, é vasta e coloca problemas de análise linguística monolíngue e contrastiva para os quais ainda não se conhecem soluções satisfatórias mesmo ao nível da investigação fundamental.

De facto, a cobertura gramatical inicialmente prevista englobava a par de *estruturas de subordinação* largamente estudadas e descritas, *estruturas elípticas* em orações *adverbiais*, *comparativas* e *coordenadas* bem menos trabalhadas pelos linguistas.

Do mesmo modo, fenómenos que se encontram na fronteira entre a Sintaxe e a Semântica colocam problemas só parcialmente resolvidos ao nível da descrição linguística: é o caso, por exemplo, da *determinação*, da *negação*, da *modalidade*, do *modo*, *tempo* e *aspecto*.

O mesmo acontece, aliás, com os fenómenos de *referência pronominal*, que extravazam da gramática da frase para a do discurso.

Vasta é igualmente a *cobertura lexical* prevista: o programa de trabalho inicialmente proposto estipula que os dicionários monolíngues comportem 20.000 entradas lexicais, no final da terceira fase do Projecto, i.e., em Dezembro de 1990.

Em suma, o Projecto Eurotra afigura-se como um dos mais ambiciosos projectos de tradução automática dado o número de línguas envolvidas, a vastidão dos dicionários a manobrar e a complexidade dos problemas de descrição gramatical que a estipulação da existência de um corpus e a delimitação de sublinguagens mal conseguem minimizar.

4. A METODOLOGIA

Admitindo que um sistema de tradução automática, para produzir traduções de boa-qualidade se deve basear em conhecimentos lingüísticos sólidos, o Projecto Eurotra procurou delinear um *suporte lógico* em que os lingüísticos pudessem facilmente codificar as suas gramáticas e dicionários e estabeleceu, como princípio norteador do funcionamento de Sistema, a independência da legislação lingüística relativamente ao suporte lógico.

O formalismo arquitectado actualmente utilizado no Projecto é conhecido sob a designação de E-framework e recorre a dois tipos de mecanismos – geradores e tradutores.

O Projecto Eurotra aborda o processo de tradução utilizando uma estratégia de *transferência*. Assim, a tradução entre línguas comporta como etapas *análise*, *transferência* e *síntese* (ou *geração*).

De facto, embora o número de línguas envolvidas (72 pares de língua) pareça desaconselhar um sistema de tradução baseado na *transferência*, uma abordagem alternativa em termos de uma estratégia de *interlíngua* revelou-se impossível no estado actual dos conhecimentos nomeadamente dos conhecimentos sobre o "significado".

O processo de transferência entre línguas opera sobre representações de interface tanto quanto possível depuradas dos aspectos idiossincráticos dos pares de língua envolvidos. Com efeito, o objectivo visado é a *transferência simples*, idealmente entre unidades lexicais, caracterizadas em função da estrutura argumental em que se inserem.

Assim, a relação entre os textos concretos e os níveis de interface da língua-origem (source language) e da língua-alvo (target language), é mediatizada por estádios intermediários de representação, cada vez mais abstractos. Ou seja, o sistema é *estratificacional*.

Deste modo, gramáticas de análise e de síntese comportam idealmente os seguintes *níveis de representação*: o *ENT* (*Eurotra Normalized Text*), que converte texto em caracteres ASCII e vice-versa; o *EMS* (*Eurotra Morphological Structure*), que decompõe as palavras nas suas unidades constituintes.

(*Configurational Structure*), que dá conta dos constituintes categoriais e da sua ordem superficial na frase; o *ERS* (*Eurotra Relational Structure*), que capta as funções gramaticais exibidas superficialmente pelos constituintes na frase e lhes atribui uma ordem canónica; e, finalmente, o *IS* (*Interface Structure*), que caracteriza os constituintes em termos de funções sintácticas profundas, determinadas pelas relações semânticas que esses constituintes detêm.

Na prática, o Sistema funciona frequentemente sem alguns destes níveis – é o caso do *EMS*, ainda em fase de desenvolvimento na grande maioria dos grupos de língua, ou do *ERS*, que alguns grupos de língua consideram codificar informação repartível pelos níveis contíguos – *ECS* e *IS*.

Na verdade, a *legislação lingüística* do Projecto só é *imperativa* para o nível de *IS*, assumindo os outros níveis, o carácter de recomendações preferenciais.

A legislação lingüística, ainda que inspirada em teorias lingüísticas actuais – Gramática Generativa actual, Gramática Funcional Lexical (ing. Lexical functional Grammar), Gramática da Estrutura Sintagmática Generalizada (ing. Generalized Phrase Structure Grammar) – constitui um quadro de referência próprio, que, tal como os seus modelos inspiradores, não ultrapassa de momento o âmbito da *gramática da frase*.

O corpo central de cada nível de representação é constituído por *regras-b* que permitem codificar a informação gramatical captável por regras de reescrita de uma gramática independente do contexto.

A passagem de um nível para outro, quando envolve alterações na estrutura de constituintes, é operada por regras, denominadas *regras-t*, que não diferem substancialmente das empregadas na transferência de uma língua para a outra.

Conjuntamente com as gramáticas funcionam os *dicionários monolíngües*, para a análise e para a síntese, *bilingües*, para a transferência entre pares de língua.

Na sua globalidade, o Sistema apresenta um cunho marcadamente sintaticista, embora, em última instância, o nível de *IS* procure atin-

gir uma caracterização preponderantemente semântica das frases. De facto, mesmo ao nível da semântica lexical, não é fácil estabelecer uma codificação das unidades lingüísticas sistemática e euroverbalmente adequada.

5. A LÍNGUA PORTUGUESA NO PROJECTO EUROTRA - FRAGMENTOS DE DESCRIÇÃO

No sistema especificado, o *peso da descrição gramatical* recai sobre as *componentes monolingües*, i.e., sobre as gramáticas de análise e de síntese.

É no seio destes módulos que têm de ser encontradas soluções que permitam descrever as particularidades das línguas, e as relações com estruturas de interface tanto quanto possível simples e isomorfas.

No âmbito do Eurotra-P algumas dessas soluções encontram-se descritas em documentos de trabalho do grupo português (cf. *Eurotra-P Working Papers*), ou nos Relatórios de Implementação (cf., em especial, *Portuguese Implementation Report - February 1990*).

A título de *exemplo* destacarei o tratamento proposto em DUARTE, I., G. MATOS, A. NUNES e A. ALVES (1990) para dar conta da *colocação dos pronomes clíticos em Português Europeu*.

Recorde-se que os *pronomes pessoais complemento em Português Europeu* são na sua maioria críticos verbais, ou seja, elementos que dependem acentual e sintacticamente do núcleo verbal da frase em que se encontram inseridos.

Em termos de um *sistema de tradução multilingüe* como o Eurotra, construções exibindo clíticos colocam três problemas *fundamentais*: (i) o da tradução de e para línguas que não possuem pronomes clíticos; (ii) o da tradução entre pares de língua que, exibindo ambas pronomes clíticos, fazem uma diferente repartição do sistema de cliticização; e (iii) o da tradução entre línguas em que os clíticos manifestam padrões de ordem superficial diversos.

As frases em (1) exemplificam o primeiro problema:

- (1) a. A indústria tem-lhes dado apoio financeiro.
b. Industry has given *them* financial support.

O segundo problema é ilustrado por frases como (2) e (3):

- (2) a. O déficite não cessa de aumentar
b. Le déficit ne cesse pas de s'accroître
- (3) a. A Comissão proibiu-os acusativo de divulgar os resultados
b. La Commission *leur*dativo a défendu de divulguer les resultats

Em (2), a divergência entre as frases (a) e (b) é motivada pela inexistência de correspondência perfeita entre os verbos pronominais de uma língua e os de outra.

Em (3) as disparidades entre as frases da língua-origem e da língua-alvo decorrem das estruturas argumentais dos regentes verbais, as quais se reflectem nas marcas de caso morfológico que os clíticos, exibem.

Finalmente, as frases em (4) mostram que as línguas podem diferir nos padrões de ordem dos clíticos:

- (4) a. A Comissão aprovou-o
(cf. a má-formação em Português Europeu de:
"A Comissão o aprovou")
b. La Commission l'a approuvé

Num sistema estratificacional como o do Eurotra, os dois primeiros problemas colocam-se directamente ao nível da componente de *transferência*. Tudo o que os módulos de análise e de síntese podem fazer é minimizar a complexidade da transferência eliminando a informação irrelevante dos níveis de interface.

A *colocação dos clíticos na frase*, porém, é um problema que afecta particularmente as *gramáticas monolingües*:

Na *análise*, há que permitir que os clíticos ocorram em qualquer posição admissível, pois a presença da frase concreta determinará a escolha certa;

Na *síntese*, contudo, há que impedir que os clíticos ocorram em posições possíveis mas não adequadas ao contexto em que se encontram. Assim, *o posicionamento dos clíticos requer na síntese um tratamento específico*.

O carácter monolingüe da tarefa é tanto mais evidente quanto os pronomes clíticos em Português Europeu exibem uma ordenação superficial que não coincide com nenhuma das manifestadas pelas restantes línguas do Projecto que possuem clíticos.

Recorde-se que, o padrão básico de ordem dos clíticos em Francês é a próclise (veja-se o contraste de gramaticalidade entre os exemplos (5a) e (5b), e que no Espanhol, como no Italiano e no Grego, há próclise nas frases finitas (cf. (6a)) e ênclise nas infinitivas (cf. (6b)):

- (5) a. La Commission l'a rejetté
 La Commission a décidé de le rejeter
 b. * La Commission a le rejetté
 * La Commission a décidé de rejeter le
- (6) a. La llevaran de urgencia al médico
 b. Tuvieron que llevarla de urgencia al médico

Em traços gerais, os clíticos em Português apresentam a seguinte distribuição:

(i) Ocorrem em posição proclítica quando precedidos (e comandados) por *Complementadores*, por *sintagmas relativos ou interrogativos* (cf. (7a), (7b) e (7c) ou por *elementos com um estatuto próximo de operadores* (sintagmas quantificados (cf. (7d)), negação (cf. (7e) e (7f)), adverbiais aspectuais, restritivos e de foco (cf. (7g), (7h) e (7i))).

- (7) a. A Comissão disse *que* o aprovava
 b. A Comissão analisou o relatório *que* os representantes *lhe* tinham submetido
 c. *Quem* o aprovou?
 d. *Todos* os representantes *o* aprovaram
 e. *Nem* todos os representantes *o* aprovaram
 f. A Comissão *não* o aprovou
 g. A Comissão *já* o aprovou
 h. Os representantes *também* o aprovaram
 i. *Até* a Comissão *o* aprovou

(ii) Ocorrem em *posição enclítica* em frases finitas ou infinitivas, desde que nenhum dos elementos desencadeadores de próclise esteja presente:

- (8) a. Os representantes aprovaram-*no* na última reunião
 b. Os representantes disseram *tê-lo* aprovado na última reunião

(iii) Ocorrem em *posição mesoclítica* em contextos finitos idênticos aos de ênclise quando a forma verbal se encontra no futuro ou no condicional:

- (9) a. A Comissão aprová-*lo-á* na próxima reunião
 b. A Comissão aprová-*lo-ia* hoje na reunião, se pudesse.

Centremo-nos nos contextos de próclise e de ênclise:

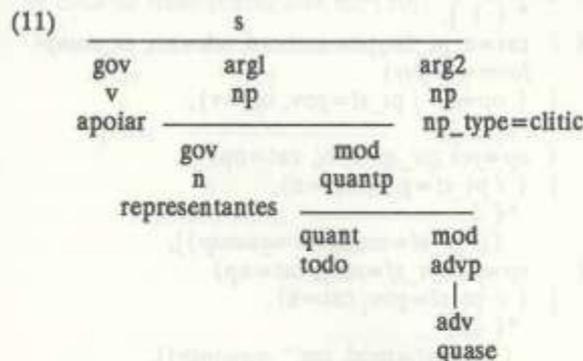
Para a implementação destes padrões de ordem na geração, a estrutura estratificacional do Sistema revela-se particularmente vantajosa.

A título ilustrativo, considere-se a tradução para o Português da seguinte frase, em Inglês:

- (10) Almost all representatives have supported *him*
 => Quase todos os representantes *o* têm apoiado

Na gramática de síntese do Português, em IS, como esperado, dado o carácter de generalidade deste nível de representação, a informação codificada restringe-se ao mínimo – o pronome complemento do verbo é um clítico, ou seja, um elemento caracterizável com o traço, “*np-type = clitic*”.

Deste modo, a frase Portuguesa em (10) é representada em IS pelo seguinte esquema:



Em ERS, a geração dos clíticos na sua posição superficial é preparada pela atribuição de traços específicos – *neg=yes, no* e *op=yes, no*, em que “*neg*” e “*op*” são as etiquetas para “negação” e “operador” – aos nós relevantes, quer através dos construtores (regras-b), quer através de regras de atribuição de traços (regras-f).

As seguintes regras simplificadas, em que os traços pertinentes aparecem a negrito, ilustram a estratégia adoptada:

(12) Regra-b:

```
cS2act = {cat=s, pt_voice, neg=E}
  [ { pt_sframe=123, sf=gov, cat=v, pt_voice=act,
      neg=E, op=Op},
    { pt_sf=subj, cat=np, op=Op},
    ( { pt_sf=obj, cat=Ca,},
      { pt_sf=obj, cat=np, np_type-clitic}),
    ( { pt_sf=iobj, cat=pp};
      { pt_sf=iobj, cat=np, np_type=clitic}),
    { pt_sf=...}]
```

No construtor apresentado, os atributos "neg" e "op" vêm o seu valor instanciado pelas variáveis "E" e "Op". Estas serão substituídas pelas constantes "yes" ou "no", consoante os contextos.

Os contextos que determinam os valores concretamente assumidos pelos atributos "neg" e "op", são captados por regras-f como as seguidamente exemplificadas:

(13) Regras-f:

```
fS_Op1 = { / cat=s, pt_compform ~ =empty}
  [ { op=yes / pt_sf=gov, cat=v},
    * { } ].
fS_Op3 = { / cat=s, pt_fstype=ssubord, wh=no, pt_compform=empty}
  [ { op=no / pt_sf=gov, cat=v},
    * { } ].
fNP_Opyes = { op=yes / pt_sf=subj, cat=np}
  [ { / pt_sf=gov, cat=n},
    * { },
    { / pt_sf=mod, cat=quantp}],
fNP_Opno = { op=no / pt_sf=subj, cat=np}
  [ { / pt_sf=gov, cat=n},
    * { },
    { / pt_sf=mod, cat ~ =quantp}].
```

fS_Op1 e fS_Op2 atribuem ao verbo o traço "op=yes" e "op=no" respectivamente nos contextos de presença e ausência de um complementador lexicalmente realizado.

fNP_Opyes e fNP_Opno atribuem ao Sujeito da frase os traços "op=yes" e "op=no" consoante este comporta ou não um sintagma quantificado.

Em ECS próclise e ênclise serão descritas em termos do traço "hyphen=yes,no". Este, porém, é introduzido em ERS por regras-f sensíveis à presença ou ausência dos traços desencadeadores de próclise, como as seguintes regras ilustram:

(14)

```
f_hyphen_yes = { / cat=s, wh=no},
  { { / role=gov},
    { / role=argl}
    [ { / role=gov},
      ^ { / role=mod, cat ~ =quantp},
      * { / cat ~ =quantp} ],
    {hyphen=yes/pt_sf=?,cat=np,np_type=clitic},
    * { } ]
f_hyphen_no2 = { / cat=s, wh=no},
  [ { { / pt_sf=gov}, cat=v, neg=yes};
    { / pt_sf=gov, cat=v, op=yes}),
  * { } ],
  {hyphen=no / pt_sf=?, cat=np, np_type=clitic},
  * { } ]
```

A aplicação de regras como as mencionadas permite obter objectos representacionais como o seguinte, que especifica a configuração de ERS da frase portuguesa em (10):

(15)

s		
gov	subj	obj
v	np	np
apoiar	_____	np-type=clitic
	gov	mod
	n	quantp
	representantes	_____
	quant	mod
	todo	advp
		adv
		quase

Na passagem de ERS para ECS, a principal alteração estrutural que a gramática tem de captar nas estruturas de cliticização é a seguinte: independentemente de os clíticos representarem ou não argumentos, eles surgem imediatamente adjacentes ao verbo e não na posição tipicamente ocupada por esses argumentos. Assim, na gramática do Português, de ERS para ECS, os tradutores colocam não-ordenada-

mente os constituintes com o traço "np_type=clitic" em *adjacência ao verbo*.

O seguinte tradutor simplificado exemplifica este procedimento:

(16)
 tS = S: {cat=s, pt_style=no;rais, pt_fstype=main},
 [GOV: {cat=v},
 SUBJ: ^ {pt_sf=subj},
 OBJ_NP: ^ {pt_sf=obj, cat=np,np_type~ =clitic},
 OBJ_NPC: ^ {pt_sf=obj, cat=np,np_type=clitic},
 CLIT_DAT: ^ {pt_sf=iobj, cat=np,np_type=clitic},
 IOBJ_NP: ^ {pt_sf=iobj, cat=pp, p_form=a},
 OBL_BYPH: (* { pt_sf=obj};
 { pt_sf=byphrase}),
 MOD: * {pt_sf=mod},
 TRANC_NEG: ^ {pt_sf=tranc, cat=negp}]
 S < MOD,SUBJ,(TRANC_NEG,GOV,CLIT_DAT,OBJ_NPC),MOD
 OBJ_NP, IOBJ, OBL_BYPH, MOD >

Na seqüência linear subsequente à seta, "=>", os elementos envolvidos pelos parentesis curvos encontram-se, por convenção, não ordenados entre si.

Finalmente, em ECS, regras-b, como os construtores simplificados abaixo apresentados, encarregam-se da construção das *configurações de próclise e ênclise* correctas, de acordo com o valor negativo ou positivo assumido pelo atributo "hyphen".

(16)
 cAUXPERFACT = { cat=vp},
 [^ {cat=negp},
 ^ {cat=clitic, hyphen=no, mscase=dat},
 ^ {cat=clitic, hyphen=no, mscase=acc},
 ! {cat=v, pt_vtype=aux, pt_voice=act,
 pt_lu=ter},
 ^ {cat=clitic, hyphen=yes, mscase=dat},
 ^ {cat=clitic, hyphen=yes, mscase=acc},
 {cat=v, pt_vform=pastpart, pt_voice=actc},
 * {cat=advp},
 ^ {cat=ap},
 ^ {cat=np},
 * {cat=pp},
 ^ {cat=sbar, pt_fstype=ssubord}]

Por último, a regra seguinte consolida as estruturas de hifenação patententes nos casos de *ênclise*:

(17)
 cCLLIT = { cat=clitic, hyphen=yes, mscase=M}
 [{cat=diac, pt_lu=hyphen},
 {cat=clitic, hyphen=yes, mscase=M}]

Assim, o objecto representacional de ECS para a frase "Quase todos os representantes o têm apoiado" será o seguinte:

(18)

s						
np			vp			
quantp	art	n	clitic	aux	v	
	os	representantes	o	têm	apoiado	
adv	quant					
		todos				
advp						
quase						

Em suma, o Sistema Eurotra, com a sua arquitectura estratificacional, revela-se adequado para codificar satisfatoriamente pelo menos algumas estruturas das línguas naturais.

Uma condição prévia para o sucesso do empreendimento é obviamente ele repousar em descrições lingüísticas adequadas. Dado o diferente desenvolvimento dos conhecimentos nas diferentes áreas da lingüística fundamental (Sintaxe, Semântica,...) é de esperar que tal adequação atinja níveis diversos, consoante a natureza predominante dos fenómenos envolvidos. Este é, aliás, um dos factores que determina a índole sintacticizante do Projecto no seu estágio actual, bem como a sua limitação à gramática da frase.

6. AS ETAPAS PERCORRIDAS

O desenvolvimento do Projecto Eurotra obedeceu a um plano que previa o seguinte faseamento do trabalho:

1ª fase – 1982-1984: Estabelecimento de infra-estruturas de organização e definição do suporte lógico (software) e das especificações lingüísticas.

2ª fase – 1985-1987: Implementação experimental pelos grupos de língua das estruturas básicas das línguas e de dicionários de dimensão reduzida (2.500 entradas lexicais) – cobertura do 1º ciclo.

3ª fase – 1988-1990: Implementação pelos grupos de língua das estruturas gramaticais complexas representativas do corpus inicial, e alargamento dos dicionários monolíngües até às 20.000 entradas lexicais – cobertura do 2º ciclo.

No fim da terceira fase esperava-se ter obtido como produto um protótipo pré-industrial de tradução automática.

Os resultados progressivamente obtidos foram sujeitos a verificação e avaliação:

No tocante aos grupos de língua, esse controle foi feito indiretamente, através de relatórios detalhados sobre a implementação levada a cabo, relatórios esse que, na última fase do Projecto, passaram necessariamente a acompanhar algumas das entregas periódicas de trabalho.

A avaliação global do Projecto, por sua vez, baseou-se no trabalho de investigação desenvolvido e em demonstrações dos resultados obtidos e esteve a cargo de comissões internacionais de especialistas. O 1º painel de avaliação, presidido pelo Dr. Pannenburg, pronunciou-se em 1987; o 2º, presidido pelo Sr. Danzin, reuniu-se entre o final de 1989 e o início de 1990.

Os resultados alcançados na realidade ficaram, porém, aquém do esperado, mais devido à complexidade da tarefa do que devido ao tempo requerido pela codificação lingüística. O grau de desenvolvimento das gramáticas das línguas que entraram mais tarde no Projecto são disso testemunho.

Com efeito, a entrada tardia de Portugal (e de Espanha) no Projecto poderia deixar antever atrasos na implementação da gramática e dos dicionários no final da 3ª fase. Todavia, no estágio actual a cobertura gramatical e lexical do Eurotra-P, não difere substancialmente da dos outros grupos de língua.

Assim, como o "EUROTRA Programme Assessment Report" enfatiza (cf. DANZIN et al. 1990), o desenvolvimento de um protótipo pré-industrial capaz de produzir traduções de boa qualidade de

um tipo de textos específico não foi ainda alcançado e o sistema produzido pode apenas ser caracterizado como um *protótipo científico*.

Deste modo, dos dois objectivos iniciais do Projecto, apenas o segundo – i.e., a criação de competência específica – foi inteiramente atingido. De facto, o Projecto contribuiu decisivamente para a partilha e desenvolvimento dos conhecimentos nos países da Comunidade Europeia, nas áreas da lingüística contrastiva, da lingüística computacional, da tradução automática e do processamento da linguagem natural.

Assim, com o desenrolar do Projecto e a avaliação progressiva dos resultados obtidos, um novo objectivo foi consensualizado – o da *preservação, através do estudo e informatização das línguas nacionais, dos valores culturais diversificados da Comunidade Europeia*. De facto, como "Eurotra Assessment Report" salienta, numa época de crescente informatização, as línguas dos mercados de maior número de consumidores têm a tendência a ser privilegiadas, e conseqüentemente a impor os seus valores culturais. O Eurotra contrariou esta tendência, ao dar igual atenção a todas as línguas oficiais da Comunidade Europeia.

No que diz respeito a Portugal, estes últimos objectivos foram efectivamente alcançados.

A pesquisa sobre as estruturas do Português, empreendida pelos lingüistas desde há várias décadas, encontrou no Projecto um suporte para a sua divulgação em maior escala, quer entre os especialistas nacionais, que foram no âmbito do Eurotra chamados a partilharem os seus conhecimentos, quer junto dos especialistas de outros países, com os quais levaram a cabo trabalho *lingüístico contrastivo*.

Finalmente, o Projecto Eurotra teve o mérito de ajudar a alicerçar a *lingüística computacional em Portugal*, em particular, a *tradução automática*, domínios de pesquisa até então sem tradições nas Universidades e nos centros de investigação lingüística do País.

Os resultados positivos do Projecto foram suficientes para a Comissão da Comunidade Europeia aceitar a sua continuação e prever o seu *desenvolvimento*, ainda que em novos moldes – o Projecto Eurotra encontra-se incluído no 3º Programa-Quadro para 1993.

Desta pesquisa, parcialmente feita de parceria com a *indústria*, resultará um aperfeiçoamento do Sistema que embora não conduzindo a um protótipo de tradução automática de arquitectura avançada num futuro próximo, encontrará aplicação em produtos menos ambi-

ciosos como sistemas para tradução automática assistida por computador, bases terminológicas ou dicionários computadorizados.

Os anos de 1991 e 1992 são, pois, considerados como uma *fase de transição*, na qual o Projecto se reestrutura para atingir os novos objectivos que se propõe – desenvolvimento do protótipo, reutilização e rentabilização dos produtos decorrentes da pesquisa levada a efeito no âmbito do Projecto.

Assim, dentro do Projecto, reorganizam-se os *grupos centrais* para a investigação de problemas cruciais ao desenvolvimento do protótipo ou para a aplicação dos produtos dele decorrentes, e os *grupos nacionais* empreendem a reformulação dos seus módulos monolíngües com vista a torná-los mais coerentes; constituem subgrupos de duas e três línguas para testar estratégias de transferência; e integram grupos de investigação sobre tópicos lingüísticos de interesse geral.

Nesta nova fase do Projecto, o Eurotra-P encontra-se também presente.

BIBLIOGRAFIA

- ARNOLD, D. (1986) Eurotra: A European Perspective on MT. Proceedings of the IEEE, vol. 74, No 7, Julho de 1986.
- ARNOLD, D. e L. des TOMBE (1987) Basic theory and methodology in EUROTRA, in NIRENBURG, S. ed. (1987).
- BECH, A. e A. NYGAARD (1988) The E-framework: A formalism for natural language processing. Coling Budapest, vol. 1., MTEZ Printing House, Hungria.
- DANZIN, A. (chairman) et alii (1990) Eurotra Programme Assessment Report, March 1990 (Assessment Panel).
- DUARTE, I. e G. MATOS (1988) "Pronominal Constructions in Portuguese". *Eurotra-P Working Papers*, ILTEC, Lisboa.
- DUARTE, I., A. ELISEU e G. MATOS (1989) "Pronominal Constructions – Comments on Steiner's Remarks", *Eurotra - Working Papers*, ILTEC, Lisboa.
- DUARTE, I., G. MATOS, A. NUNES e A. ALVES (1990) "A Strategy for dealing with a puzzling word order problem within Eurotra". *Eurotra-P Working Papers*, ILTEC, Lisboa.
- DURAND, J. e L. SADLER (1990) "Linguistic Research in 1991-1992", Eurotra.
- EUROTRA-P (1990) *Portuguese Implementation Report*, ILTEC, Lisboa.
- KRAWER, S. B. MAEGAARD, S. PERSCHKE e J. PUGH (1990), "Proposal for an overall plan for the content of the Transition Programme". The Eurotra Liaison Group, 13 de Junho de 1990.

- MATEUS, M. H. (1987) "Eurotra, o Projecto Comunitário de Tradução Automática", *Revista da Universidade de Lisboa*, Lisboa, 1987.
- NIRENBURG, S. ed. (1987) *Machine translation – Theoretical and Methodological Issues*, CUP, Cambridge.
- PERSCHKE, S. (1989) "Eurotra", MT Summit II. 16-18 Agosto, Munique, 1989.
- RAW, A., F. van EINDE, E. HACKEN H. HOEKSTRA e B. VANDECAPELLE (1988), "An Introduction to the Eurotra Machine Translation System", van EINDE, F. e P. HAKEN eds. (1988) *Working Papers on Natural Language Processing 1*.
- SLOCUM, J. (1988) "A Survey of Machine Translation: its History, Current Status, and Future Prospects". *Machine Translation Systems*, CUP, Cambridge.
- TUCKER, a. (1987) "Current Strategies in machine translation research and development", in NIRENBURG, S. ed. (1987).

Caminhos do Romance Brasileiro

João Hernesto Weber

156 p. - Série Revisão

Trazendo como subtítulo "De *A Moreninha* a *Os Guaiandás*", *Caminhos do Romance Brasileiro* traça um painel amplo da evolução da ficção brasileira desde meados do século XIX até o presente, buscando na história do país a base de interpretação dos romances analisados. Colocando de lado tendências teóricas que surgem e desaparecem ao sabor das circunstâncias, o autor liga-se àquela tradição humanística que marca a grande crítica ocidental de linha histórico-sociológica, na senda de Lukács, Auerbach, Hauser e outros, na Europa, e que no Brasil se revela em nomes como Otto Maria Carpeaux, Roberto Schwartz, etc. Sem sombra de dúvida, a obra de João Hernesto Weber representa um marco indiscutível e inegável na bibliografia brasileira sobre o assunto, principalmente se levar em conta que são raras as obras que tratam do tema de maneira profunda.

João Hernesto Weber é Mestre em Literatura Brasileira pela Universidade Federal do Rio Grande do Sul.