RESUMO

Ao longo dos Últimos vinte e cinco anos da história da Inteligência Artificial - estudo da inteligência como computacao – constituíram-se corpos de conhecimento em redor do desenvolvimento e aplicação de várias metodologias e técnicas, relacionadas com a resolução de Problemas. De entre. entre esses corpos. a Engenharia da Linguagem sobressaiu na. ultima década.

Entendemos por Engenharia da Linguagem, a disciplina cujo objecto é o estudo.. projecto e construção de sistemas para o Processamento da comunicação e da significação. Estes sistemas destinam-se à resolução de Problemas através da interacção coloquial (em língua natural) entre um homem e o computador.

o. Introdução

Apresentam-se alguns aspectos que descrevem o significado de uma nova disciplina, a Engenharia da Linguagem (Coelho. 1980). a qual é definida através de uma síntese de diversas pesquisas. assentes fundamentalmente na construção de três sistemas computacionais {1} e elaboradas desde 1974 em redor de vários Projectos complementares.

O objectivo da Presente comunicação é dar unidade e corpo aos estudos até agora realizados, Perspectivando a sua motivação
e iluminando o interesse nacional pelo desenvolvimento de uma tecnologia intimamente ligada ao futuro da Informática. Umatal investigação Permitirá, Por um lado. evitar a importação de tecnologia e mesmo fornecer tecnologia não importável. E, por outro lado, proporcionará a construção de tecnologia exportável para. entre outros. os países de Língua Portuguesa. Defende-se a continuação de uma tal investigação tendo em conta as condições Já criadas. quer através do desenvolvimento do saber. quer pela exploração de extensões Julgadas imprescindíveis rara viabilizar a utilidade Pratica de tais sistemas. inclusive no LNEC.

(1) O Programa TUG;AO, construído para o domínio das informações sobre Pessoas. serviu. Por um lado. de suporte experimental rara o teste de vários mecanismos necessários à interacção coloquial. E, por outro lado. Permitiu o desenvolvimento dos Programas CALCC (Cotta & silva, 1978). para a consulta da legislação sobre a construção civil. e do TUGA; (Coelho 1979),Para a implementação de um serviço de biblioteca.

1. Definição

A Engenharia da Linguagem é uma disciplina em formação, congregando conhecimentos desenvolvidos principalmente no âmbito das Ciências da Computação, da Inteligência Artificial e da Linguística, e cujo suporte de experimentação e um computador digital. A sua ênfase recai na construção de sistemas que sejam práticos. eficazes, utilizáveis, transportáveis. modificáveis, e com comunicação versátil.

A síntese teórica de varias famílias de ideias é obtida, sobretudo através da combinação de várias técnicas das Ciências da Computação com ideias de organização da Engenharia, por intermédio dos métodos da inteligência Artificial.

A Engenharia da Linguagem é distinta da Linguistica. O linguista vê na linguagem um conjunto de mecanismos completos que necessitam de uma descrição;. enquanto sistema; a sua descrição recai sobre os aspectos fonético, morfológico, informativo, sintáctico, lexical. semântico. lógico.

Pragmático; os seus componentes (unidades estruturas e significados) completam-se, definindo-se uns pelos outros. e o conjunto forma um todo coerente e único: a linguagem. que o linguista estuda enquanto tal. O engenheiro da linguagem opta por um quadro conceptual mais restrito, Pois está interessado arenas na componente linguistica da cognição isto é na atribuição de significado às entidades linguisticas Assim, defini de uma só vez um certo mundo, capaz de descrever o domínio de aplicação e uma certa Procura nesse mundo. Isto quer dizer para um conjunto de aplicações se selecciona um subconjunto da língua natural, e que o domínio do discurso e a base de dados são restritos. Desde logo. embora as frases Possam ser comparáveis às que. o linguista estuda como material experimental. a sua descrição visa uma representação que assegure uma interpretação utilizável nos termos e para os objectivos definidos Pelo sistema: exploração de bases de dados e Programas, elaboração de respostas calculadas, simulação de processos de compreensão. Deste modo. a diferença essencial entre as duas investigações ressalta dos objectivos da Engenharia da Linguagem, que. são a construção de um sistema/conjunto de Programas capazes de compreender uma língua natural e de responder a perguntas de um modo suficientemente eficaz para tornar o computador mais acessível e utilizável em aplicações práticas. A construção deste tipo de Programas estimula as pesquisas sobre a inteligência e a linguagem. do mesmo modo que a física Experimental e a Engenharia interactuam com a Física Teórica.

Sobre o plano geral do estudo da linguagem. as duas investigações são perfeitamente complementares na medida em que:

1)Por um lado. é necessário estudar os fenómenos que constituem a linguagem, e

2)por outro lado. o estudo deve apresentar um carácter tal que os resultados a que ele conduz sejam obtidos de modo incontestável; a utilização de um computador é. por certo, a segurança da reprodutividade do método.

A Engenharia da Linguagem é também distinta das Ciências da Computação. na medida em que estas se interessam principalmente pelos processos formais (algoritmos) e Pelas estruturas de dados e mecanismos de controlo associados.

A Engenharia da Linguagem e uma disciplina cuja génese se compara, por exemplo. à Engenharia do Software {2} e engenharia da Programação (3). A sua autonomização visa fazer convergir um conjunto de conhecimentos dispersos (4). mas eivados de um eixo e propósito comuns: colocar o computador ao alcance de cada vez mais utilizadores. que pensam e comunicam em língua natural.

2. OBJECTO E CONTEXTO

O objecto da Engenharia da Linguagem e o estudo. projecto e construção de sistemas para o processamento da comunicação e da significação.
Estes sistemas denominados coloquiais.são, fundamentalmente. programas de computador. escritos em linguagens de alto nível. destinados a manipular significados. através da combinação de algoritmos e de estrutura; de dados. Destinam-se à resolução geral de problemas através da interacção coloquial (em língua natural) entre um homem e um computador.

O objecto de estudo da problemática que envolve a construção destes sistemas é a compreensão da comunicação em língua natural. através da consideração dos processos usados para a interpretação e geração de frases. trocadas durante um dialogo.

O contexto de engenharia desta disciplina reside na aplicação de conhecimentos das ciências da Computação (Por exemplo. matemáticos) e da Linguística à construção de técnicas e mecanismos que satisfaçam um dado desiderato. e através dos métodos da Inteligência Artificial.

Este contexto inscreve-se no quadro geral de qualquer engenharia {5}. o qual se apresenta em seguida. Um problema de engenharia típico pode ser idealizado como se segue: dado um conjunto inicial de dispositivos de comportamento conhecido e um conjunto de regras através dos quais os possamos combinar para produzir entidades mais complexas. construir um mecanismo composto cujo comportamento satisfaça certas propriedades especificadas.

No caso particular da comunicação em língua natural. o objectivo é a construção de um sistema capaz de evidenciar inteligência. baseada sobretudo na compreensão daquela língua. O projecto de engenharia é o conjunto dos métodos empregues para abordar tais problemas. De facto. para cada problema de projecto, um engenheiro deve estabelecer a forma da resposta. No caso particular do objecto do nosso estudou a forma da resposta é uma frase de uma língua natural, que seja coerente em relação à frase de entrada, Se o Problema Pertence a um tipo conhecido, escolhem-se várias formas possíveis. Na maioria dos domínios da engenharia, a forma da resposta é a descrição do mecanismo desejado como um conjunto de componentes e das suas ligações. Em geral, esta descrição tem muitos parâmetros indeterminados. O Problema do engenheiro consiste em determinar se e possível instanciar uma destas formas gerais da resposta, de acordo com as imposições do Problema particular. Se o problema do projecto não é típico, opta-se quer pela sua reformulação num problema típico, quer Pela sua decomposição na combinação de vários problemas típicos (6). A composição de soluções dos sub-problemas pode, no entanto, conduzir a interacções que provoquem a necessidade de resolver outros problemas (Por exemplo, avarias).

3. motivação

A Engenharia da Linguagem serve para tornar o computador acessível a um maior numero de utilizadores, nomeadamente os utilizadores casuais e inexperimentados.

(6) Esta forma de olhar para um Problema não é mais do que uma das metodologias correntes da resolução de problemas em Inteligência Artificial.

O computador é cada vez mais um instrumento indispensável porque:

1) cada vez mais é necessário guando se trata de manipular informação Para calcular, organizar, planear;

2) a tendência de colocar o computador a executar tarefas habitualmente reservadas ao homem não Pára de se desenvolver; e

3)a revolução da micro-informática vai acentuar ainda mais a ligação da Informática às nossas actividades humanas e a um maior Público.

Daqui que o computador terá de falar, ouvir e compreender a nossa própria língua. para atingir este objectivo existem fundamentalmente duas vias complementares de investigação:

(1)via tecnológica, isto é a construção de dispositivos análogos ao olho e ao ouvido, e capazes de reconhecerem um sinal numa curva gráfica, ou numa sequência de níveis acústicos.

(2)via formal, isto é a representação e formalização do conhecimento humano, e de modo a modelar o fenómeno da compreensão da língua e do objecto do discurso.

Esta segunda via é a da Engenharia da Linguagem.

4.Método

O método da Engenharia da Linguagem assenta na experimentação e/ou elaboração de teorias de Processamento da informação (Por exemplo, formalismos para a representação e uso do conhecimento sobre a língua natural e do domínio do Problema a resolver), linguísticas (por exemplo, sobre a referência, imprescindível no desenrolar de um diálogo) e psicológicas (Por exemplo, compreensão da

língua), presentes na maquinaria dedutiva, apta a suportar as diferentes etapas de um processo de interacção coloquial homem-computador (Por exemplo, a identificação automática do discurso, e a produção do significado de uma frase de uma língua natural)

Este método de trabalho adopta a via da engenharia para a resolução dos problemas da interacção coloquial propriamente dita. O uso da noção de plano como abstracção é central para esta via.

O Plano descreve o mecanismo projectado em vários níveis de pormenorização. Note-se que a essência da compreensão de um mecanismo consiste no conhecimento dos propósitos de cada um dos seus componentes. Isto envolve a construção da descrição do mecanismo (ror exemplo, a conversa entre um homem e o computador) a qual combina cada um dos componente; com os seus papeis nos planos apropriados.

5.LIMITES

O modelo é usado comoaproximação do verdadeiro comportamento de um componente, processo e/ou mecanismo (Por exemplo, modelos computacionais para o processamento da linguagem, para os processos humanos de decisão ou para o uso da linguagem).

Uma das caracterís ticas do ser humano e a sua capacidade em produzir uma imagem (modelo) cognitiva do ambiente que o rodeia. Esta imagem é sempre uma abstracção da realidade escolhida em relação a um determinado objectivo. Na compreensão da linguagem, as frases sue tratam da realidade estão relacionadas com um modelo cognitivo (atribuição de significado), e causam reacções tais como respostas, modificações ou avaliações dos modelos

A adopção de modelos é uma das limitações típicas de vários domínios da Engenharia. Isto é óbvio na Engenharia Civil, onde as Propriedades actuais dos materiais, tais como solo e cimento, não são completamente especificáveis e com rigor. Na Engenharia Mecânica existe um melhor controlo das propriedades dos materiais. mas processos importantes como o desgaste, lubrificação e vibração são ainda modelados de forma incompleta. A Engenharia Electrotécnica tem modelos muito precisos para alguns dos seus componentes básicos. como por exemplo o modelo Ebers -Moll de um transístor. No entanto, modelos muito Precisos introduzem um outro tipo de dificuldade: as equações resultantes do uso de tais modelos são demasiado complexas para serem resolvidas. As Ciências da Computação tem a mesma dificuldade com modelos Precisos. mas não utilizáveis. Por exemplo a maior Parte dos computadores tem instruções de adição com vírgula flutuante, cujas propriedades são somente definidas de modo preciso no seu manual de hardware, através da linguagem maquina. Na Engenharia da Linguagem, as bases de dados abstraem apenas fatias da realidade, e por isso cobrem uma faixa limitada de situações.

6.CAMPO DE ACÇÃO

A Engenharia da Linguagem aborda uma forma de comunicação em sue a linguagem desempenha um papel primordial.

A interacção coloquial com o computador envolve pelo menos quatro linguagens. Uma, a linguagem de utilização. por exemplo uma língua natural como o português, suporta a comunicação simples com um Programa A outra, a linguagem de controlo, governa as formas estruturadas, de interacção (por exemplo, par Pergunta-resposta, dialogos imbricados). A outra, a linguagem de Programação. comunica ao computador o texto do Programa. E, a outra. a linguagem interna do computador, na qual o Programa é traduzido, suporta o Processamento da informação.

Ointeresse Pelo estudo da linguagem tem sido manifestado, ao longo da história das ciências da Computação e da Inteligência Artificial, através; da Produção ininterrupta de linguagens artificiais, cada vez mais sofisticadas e evoluídas (ou de alto nível), e mais ou menos orientadas para problemas Particulares.. O recente interesse Pela língua natural (a linguagem de mais alto nível) decorre pois naturalmente desse estudo e ambição. Embora a comunicação em língua natural com o computador tenha uma longa tradição em informática, somente no fim da década de 60 foram criadas condições

que Possibilitaram um grande surto de desenvolvimento e aplicações na década de 70. Essas condições dizem respeito à evolução tecnológica do computador e à Produção de novos conhecimentos linguísticos. Mas e apesar do recente Progresso, a língua natural é restrita uma dimensão, razoável e necessária para o uso dos sistemas coloquiais. Sob este Condicionalismo. a língua natural é considerada também como uma linguagem formal

7.Instrumentos

Os instrumentos essenciais da Engenharia da Linguagem são o computador digital, a linguagem de Programação de alto nível e a gramática.O computador encarrega-se do Processamento. A linguagem de Programação fornece o suporte para a representação do conhecimento. A gramática (por exemplo, as gramáticas semânticas asseguram a tradução de frases de uma língua natural numa representação semântica) Permite a descrição das linguagens. A sua maquinaria dota os sistemas coloquiais com o Poder de comando da compreensão da linguagem.

A ideia de uma gramática é um conceito natural, formalizado originalmente pelos linguistas, na sua Procura de uma descrição estrutural adequada às frases de uma língua natural. Mais recentemente, esta ideia foi usada com maior êxito Pelos cientistas da computação na descrição das linguagens de Programação e das estruturas de Programa;,

Se bem que o Ponto de Partida desta ideia se deva a Tue. que em 1914 investigava regras de reescrita ou Produções. foi Post que por volta de 1936 a traduziu no enunciado seguinte: às construções legais ou teoremas de uma linguagem formal devem ser concebidas como cadeias de signos, e existem Produções que actuam como regras de inferência para Permitir a geração de novas cadeias, a Partir das velhas'.

Post considerava que a matemática podia ser vista como um processo de manipulação de signos; um teorema independentemente da interpretação que lhe possamos dar, é em ultima analise justamente uma cadeia de signos. De acordo com este pensamento Post desenvolveu sistemas canónicos os quais permitiam a substituição concatenação e combinação de cadeias de signos.

Por volta de 1956 Chomsky aplicou a ideia de Post à língua natural com o objectivo de dar pela primeira vez uma descrição mais precisa e matemática de pelo menos uma parte da Linguística. Introduziu a ideia de gramática e de linguagem, e discutiu os Processos de geração e análise de frases.

A via formal de Chomsky excitou o interesse dos cientistas da computaçao. e em especial dos escritores de compiladores. Estes estavam. também. interessados em expressar precisamente o processo de compilação. Este aspecto particular da interacção Linguística com as Ciências da Computação forneceu um dos suportes para a revolução informática desde então em curso (7).

(7) O surto de desenvolvimento da informática nas duas ultimas décadas deve-se também a disciplinas como a Inteligência Artificial ou a Electrónica, No caso da Inteligência Artificial citam-se. a titulo de exemplo, as epistemologias dos processos. dos programas e das interacções entre objectos. das quais surgiram conceitos como 'time-sharing', linguagem de alto nível. e arquitecturas de maquinas e linguagens inovadoras. respectivamente.
A gramática é um instrumento geral, servindo não só para o reconhecimento de frases de uma língua natural ou de uma linguagem de programação, mas também para o reconhecimento de formas de conversa ou de situações em. que elas estejam inseridas. De facto, as linguagens são similares na medida em que os objectos a serem reconhecidos se podem reduzir aos símbolos básicos ou primitivos e às classes sintácticas. O método sintáctico. traduzido através da escrita de gramáticas, permite o uso de primitivas para descrever detalhes locais e regras de Produção para descrever a estrutura global.

8. APLICAÇÕES

A Engenharia da Linguagem visa, alem do mais, tornar acessível os sistemas informáticos {8}, armazenados num
computador, a todo o utilizador neófito, isto é, visa dotá-los de capacidade de comunicação. E. por isso, adopta a língua natural como a linguagem de utilização.

As principais aplicações da Engenheira da Linguagem são:

1)sistemas para a interacção coloquial com grandes bases de dados,
2)'front-ends para sistemas informáticos (Por exemplo, sistemas de gestão de bases de dados. terminais inteligentes ou sistemas com varias capacidades, de informação, de supervisão, de investigação. etc.),
3)interfaces para utilização fácil de programas (Por exemplo, programa; especialistas num dado domínio do conhecimento como a analise estrutural), 4) assistentes do computador (por exemplo para fornecer informações sobre as características dos ficheiros nele armazenados), e 5) sistemas para a instrução assistida por computador (por exemplo. o ensino da Geometria Euclidiana Plana nas escolas secundárias). As três primeiras aplicações são particularmente interessantes para o LNEC pois abrem facilidades de acesso imprescindíveis para uma maior exploração de programas e dados. Em relação as duas primeiras existem já estudos e trabalho experimental (por exemplo, a instalação do Programa CALCC para consulta da legislação da construção civil, na biblioteca do LNEC).

A difusão da teleinformatica, prevista para um futuro bem próximo caracteriza-se pelo Pequeno numero de utilizadores especializados e pelo grande numero de utilizadores casuais (isto é aqueles que estão interessados em interrogar aquelas bases). Estes embora o façam por vezes, não estão por isso motivados para aprenderem linguagens de interrogação especializadas (artificiais), cada vez mais complexas e de apreensão morosa. A língua natural aparece assim como esperança e solução para o alargamento da informática. pois os utilizadores são dispensados de aprenderem uma nova linguagem, e terão apenas de observar restrições numa linguagem já deles conhecida. O acesso às bases de dados em língua natural torna-se assim uma vantagem em áreas como a Engenharia, a Administração Publica, a Gestão Industrial. a medicina, a Farmácia, etc. De facto, o acesso passa a ser encarado como Parte de um Processo de resolução de problemas -- o utilizador poderá empregar a linguagem na qual geralmente formula e soluciona os seus problemas.

As linguagens artificiais de interrogação apresentam um caracter de complexidade Porque impõem. além do mais, só' fazerem referencia aos dados efectivamente presentes na base. Isto implica que o utilizador seja obrigado a conhecer de modo preciso a Própria organização da base e a natureza exacta de cada um dos dados recenseados. Esta obrigação provoca que haja necessidade em usar apenas perguntas não ambíguas, para permitir uma maior concentração dos utilizadores na estrutura e tratamento da base. Por outro lado, a formulação de frases numa linguagem artificial não é livre e obedece ao uso de um numero restrito de palavras, ditas naturais.

9. EXPERIMENTAÇÃO

A definição do objecto e campo da Engenharia da Linguagem aprofundar-se-á na medida em que os sistemas protótipos construídos, conduzam a uma experimentação que vise descobrir as suas insuficiências, portabilidade e aceitabilidade por parte dos seus utilizadores.

A experimentação dos sistemas consiste na sua exploração por grupos diversificados de utilizadores (Cotta, 1980). Essa exploração traduz-se em protocolos, os quais são registados para posterior análise e avaliação. Três pontos de vista estão presentes nessa análise e avaliação: o da cobertura linguística da gramática (análise das construções sintácticas e do dicionário), o do desvio entre o processamento linguístico e o dependente do domínio do problema (por exemplo. histogramas do comprimento da frase em palavras e numero de frases, testes das variações sintácticas à volta de um conceito simples) e o da exploração da utilização (facilidade de cooperação com o programa, nível

de entendimento e velocidade de resposta).

10.INVESTIGAÇÃO

A investigação prosseguida dentro da Engenharia da Linguagem, abarca a comunicação em língua natural. a qual requer uma combinação de capacidades especificamente linguísticas, e baseadas na dedução. O seu objectivo é alargar e melhorar as características dos sistemas para a interacção coloquial entre o homem e o computador (Pereira, 1977)

A interacção será tanto mais ampla e livre na medida em que os sistemas sejam capazes de conversar do mesmo modo ou. os seres humanos. deste modo, abrem -se quatro vias principais de trabalho:

1)o desenvolvimento de modelos teóricos sobre o uso da linguagem,

2)a construção de sistemas escritos numa única linguagem de programação. Se. como o Prolog para, por exemplo. o acesso e/ou criação de bases de dados (por exemplo. o Programa TUGA (Coelho, 1979) que opera um serviço de biblioteca em inteligência Artificial), 3) interfaces que processem língua natural (Por exemplo. para programas de grande utilidade prática e exploração; um caso Particular é o 'front-end' para sistemas de gestão de bases de dados. como o DBMS-10). e 4) o desenvolvimento de sistemas integrados, autónomos e inteligentes em microcomputador {9}, comunicando em língua natural (por exemplo, especialistas de calculo. gestores de bases de dados. supervisores de experiências). Em qualquer uma destas vias, a investigação sobre os mecanismos ligados à comunicação entre múltiplos agentes assume especial importância, e retoma a investigação realizada sobre a res olução de problemas e a dedução focada exclusivamente em problemas que um simples agente podia resolver. Assim, o conhecimento dos mecanismos da conversa (sistemas, modelos e formas de cooperacão) e da referencia (Por exemplo. a anáfora, referencia Pronominal e uso de contextos) aparece naturalmente, em nossa opinião, como um centro de preocupações,

Destas quatro grandes vias de trabalho. somente as ultimas três são interessantes para o LNEC, pois aproveitam investigações que vem sendo realizadas nos Grupos de Lógica Computacional (GLC), de Bases de Dados (GBD), e de desenvolvimento de Periféricos (GDP) do Centro de Informática. Nomeadamente, a via 2) apoia-se na construção de 2 Programas (Cotta & Silva, 1978 Coelho, 1979), a qual sugere novas facilidades do Prolog (Por exemplo, a gestão de ficheiros em disco) e o desenvolvimento da gramática de diálogos

(9) Por exemplo. um microcomputador de 16 bits como o INTEL 8086, contendo um sistema Prolog para a escrita dos Programas e dados, memória, para armazenar os Programes e as bases de dados, uma fonte de alimentação. e uma impressora para a entrada e saída de informação (e de situações), com a inclusão de mecanismos para a manipulação da referencia e do contexto. A via 3) assenta em experimentação, traduzida Pela construção de alguns módulos de uma interface para o DBMS-10 (Coelho, 1976), e na colaboração do GBD para a completa especificação da interface. A via 4) é sugerida pelo trabalho de (Colmeraver et al 1979), e tem sido objecto de intensa discussão com L. Moniz Pereira, F. C, Pereira, J. C Cotta e A. Porto, por forma e constituir um projecto de investigação a médio Prazo (3-4 anos) e de colaboração com a UNL.

Esta investigação não ocorre isolada, e não culmina com o domínio da interacção coloquial em Português com o computador. Os trabalhos até aqui desenvolvidos resultam de uma frutuosa cooperação com o Departamento de Inteligência Artificial da Universidade de Edimburgo, e com o Grupo de Inteligência Artificial da Universidade de Aix-Marselha. Estes trabalhos apontam, a nível externo para um projecto europeu de cooperação sobre o uso das línguas latinas na comunicação com o computador (em colab oração com a Universidade de Aix-Marselha), e a nível interno, no LNEC, para. o desenvolvimento de uma Engenharia do Conhecimento, especialmente dedicada aos modos

de expressar, reconhecer e usar formas diversas e Particulares do conhecimento.

Referencias

COELHO, H. (1976) On a conversational interface between users and a data base

DI (LNEC) Working report

COELHO,H. (1979) A Procram conversing in Portuguese Providing a library service

University of Edinburgh, Ph.D. Thesis

COELHO, H. (1980) Elementos Para uma Engenharia da Linguagem

LNEC. Tese Para Especialista (em Preparação)

COLMERAUER,A.;KANOUI,H. ;VAN CANEGHEM,M.(1979)

Etude et realization d'um sisteme Prolog

Université d' Aix-Marseille, GIA

COTTA,j.C.; SILVA, A.P. (1978)

Interacção com bases de dados

LNEC

COTTA, J C. (1980)

Experiência de utilização da língua natural no acesso a bases de dados

Comunicação ao CPI80

PEREIRA,L.M. (1977)

A compreensão da linguagem natural em Inteligência Artificial

LNEC