

JOÃO JOSÉ NETO

"ASPECTOS DO PROJETO DE SOFTWARE DE UM MINICOMPUTADOR"

"Dissertação de Mestrado" apresentada
à Escola Politécnica da Universidade
de São Paulo, para obtenção do título
de Mestre em Engenharia.

Área de Concentração - Engenharia de
Eletrociadade.

Orientador: Prof. Dr. Antonio Marcos de Aguirra Massola

São Paulo
-1975-

*foi feita uma
fazendo um trabalho de
orientação, com meu orientador
científico, o Dr. J. P. G.*

JOÃO JOSÉ NETO

"ASPECTOS DO PROJETO DE SOFTWARE DE UM MINICOMPUTADOR"

"Dissertação de Mestrado" apresentada
à Escola Politécnica da Universidade
de São Paulo, para obtenção do título
de Mestre em Engenharia.

Área de Concentração - Engenharia de
Eletrociadade.

Orientador: Prof. Dr. Antonio Marcos de Aguirra Massola

São Paulo
-1975-

A meus pais e irmã.

AGRADECIMENTOS

Ao Prof. Dr. Antonio Marcos de Aguirra Massola, pela dedicação com que acompanhou, orientou, apoiou e incentivou este trabalho.

Aos Professores Doutores James Gregory Rudolph, Antonio Hélio Guerra Vieira e Tamio Shimizu pelas idéias apresentadas.

Aos Enqs. Benício José de Souza, Ting Kong Sen e Wanner Monteiro Pinheiro, e engenheirandos Luiz Sanches Filho, Mário Tachibana e Charlie Lin, pela participação efetiva no projeto e no aprimoramento dos programas desenvolvidos.

Aos Enqs. Laércio Antonio Marzagão, Antonio Marcos de Aguirra Massola e Benício José de Souza, à Enqa. Selma Shin Shimizu Melnikoff, à Profa. Selenê Cavalcanti Ferrari e aos meus familiares, pelos inúmeros diálogos e constante apoio e estímulo, decisivos para a concretização deste trabalho.

À arta. Sonia Regina Izarelli pelos serviços de datilografia e desenho.

A Valdecir Finco pelo serviços de impressão.

A todos, enfim, que de uma ou outra forma contribuíram para a concepção ou desenvolvimento deste trabalho, particularmente à equipe de estagiários do Laboratório de Sistemas Digitais, sem cuja valiosa participação não teria sido possível a realização deste trabalho.

R E S U M O

O presente trabalho descreve os métodos utilizados no desenvolvimento de alguns dos módulos do "software" básico do Patinho Feio, o primeiro minicomputador desenvolvido no Laboratório de Sistemas Digitais do Departamento de Engenharia de Eletricidade da Escola Politécnica da Universidade de São Paulo.

Para cada módulo apresentado, são discutidos os seus objetivos, sendo, quando conveniente, apresentados alguns dos problemas enfrentados durante seu desenvolvimento específico ou então problemas mais gerais, referentes ao projeto de programas semelhantes ao mesmo. Sempre que possível, são apresentadas alternativas de solução dos referidos problemas, descrevendo-se finalmente alguns detalhes de um exemplo de implementação.

No Capítulo 2 são discutidos mais profundamente os problemas enfrentados durante o projeto e a implementação de um montador, desde a sua concepção até a sua implantação definitiva, enfatizando-se a metodologia empregada na implementação e os critérios adotados nas decisões mais importantes.

No Capítulo 3 é descrito um simulador-interpretador, implementado em outro computador com a intenção de auxiliar o desenvolvimento do "software" básico do Patinho Feio. Algumas técnicas de simulação são discutidas neste Capítulo, ao lado dos algoritmos utilizados pelo interpretador.

Nos demais Capítulos, são descritos mais alguns programas do "software" básico desenvolvidos para o Patinho Feio, tais como um desmontador, um programa para auxiliar a depuração de outros programas e um editor simbólico.

Nos apêndices, são apresentados tópicos julgados convenientes para a complementação de algumas idéias, bem como alguns exemplos de utilização de programas apresentados neste trabalho.

A B S T R A C T

The present work describes the methods that had been employed during the development of some of the basic software modules of the "Patinho Feio", the first minicomputer of the "Laboratório de Sistemas Digitais da Escola Politécnica da Universidade de São Paulo".

For each module, after a brief discussion of their objectives, some of the main problems which had to be solved during their development are presented, as well as more general ones, which arrive when designing programs of the same class. Whenever possible, alternative solutions of these problems are presented, and, at last, an example of implementation is described.

Chapter 2 discusses in detail the problems arrived during the design and implementation of an assembler, from the phase of its conception until that of its final installation, emphasizing the methods employed in the implementation and the criteria which were used when the most important decisions had to be taken.

Chapter 3 describes a simulator-interpreter, which had been implemented in an auxiliary computer, and was intended to help the development of the basic software of the "Patinho Feio". This chapter discusses also some simulation techniques and the algorithms employed in the interpreter.

The remaining chapters describe some additional programs of the basic software developed for the "Patinho Feio", like a disassembler, a debugging routine and a symbolic editor.

The appendices present some subjects considered helpful for completing some ideas, and some execution examples of the programs presented in this work.

I N D I C E

1. INTRODUÇÃO

- 1.1 - Objetivos
- 1.2 - Generalidades
- 1.3 - Observações

2. O MONTADOR

- 2.1 - O conjunto de instruções e a linguagem do montador
 - 2.1.1 - O conjunto de instruções
 - 2.1.2 - Programação em linguagem de máquina
- 2.2 - A conveniência de um montador
 - 2.2.1 - Rotinas auxiliares para elaboração do montador
- 2.3 - Definição das características gerais do montador
- 2.4 - Definição das características externas do montador
 - 2.4.1 - Características do Montador Absoluto
 - 2.4.2 - Características do Montador Relocável
 - 2.4.3 - A sintaxe da Linguagem de Entrada
 - 2.4.4 - Características do Código Objeto gerado pelo Montador Absoluto
 - 2.4.5 - Características do Código Objeto Gerado pelo Montador Relocável
- 2.5 - Definição das Características Internas do Montador
 - 2.5.1 - A representação interna dos rótulos
 - 2.5.2 - A organização da tabela de símbolos
 - 2.5.3 - A manipulação da tabela de símbolos
 - 2.5.4 - A organização e a manipulação de tabela dos mnemônicos
 - 2.5.5 - A representação interna das Constantes
 - 2.5.6 - As pseudos instruções
 - 2.5.6.1 - A pseudo ORG
 - 2.5.6.2 - As pseudos NOME, SUBR, SEGM
 - 2.5.6.3 - A pseudo DEFC
 - 2.5.6.4 - A pseudo COM

- 2.5.6.5 - A pseudo BLOC
- 2.5.6.6 - A pseudo EQU
- 2.5.6.7 - O Controle BLT e D
- 2.5.6.8 - As pseudos DEFE e DEPI
- 2.5.6.9 - A pseudo FIM
- 2.6 - O programa principal
 - 2.6.1 - O primeiro passo
 - 2.6.2 - O segundo passo
 - 2.6.3 - Observações sobre a Ampliação dos Recursos do Montador
 - 2.6.4 - Críticas

3. UM SIMULADOR-INTERPRETADOR PARA A LINGUAGEM DE MÁQUINA DO PATINHO FEIO

- 3.1 - Definição das Especificações do programa de simulação
- 3.2 - O interpretador das instruções
- 3.3 - A simulação do sistema de entrada e saída
 - 3.3.1 - O modelo do Sistema de Entrada e Saída
 - 3.3.2 - A interação entre o interpretador de instruções e o simulador de entrada e saída
- 3.4 - O programa controlador
 - 3.4.1 - A fase de Console
 - 3.4.2 - A fase de Execução
- 3.5 - Comentários, Críticas e Sugestões

4. UM DESMONTADOR PARA A LINGUAGEM DE MÁQUINA DO PATINHO FEIO

- 4.1 - Especificações do Desmontador
- 4.2 - A Lógica do Desmontador
 - 4.2.1 - O primeiro passo do desmontador
 - 4.2.2 - O segundo passo do desmontador
- 4.3 - Alguns detalhes de Implementação
 - 4.3.1 - Desmontador Absoluto
 - 4.3.2 - Desmontador Relocável
- 4.4 - Conclusões

5. A ROTINA DE DEPURAÇÃO DE PROGRAMAS

- 5.1 - A fase de depuração de um programa
- 5.2 - A filosofia de rotina de depuração
- 5.3 - Os problemas enfrentados
- 5.4 - Exemplo de Implementação
 - 5.4.1 - Rotina de Entrada de Dados
 - 5.4.2 - Rotina de Interpretação e Execução de Linguagem de Máquina
 - 5.4.3 - Rotina de Relatório
 - 5.4.4 - Comentários e Observações

6. O EDITOR SIMBÓLICO

- 6.1 - Introdução
- 6.2 - A filosofia do Editor
 - 6.2.1 - Definição do Conjunto de instruções do editor
- 6.3 - A lógica do Editor
- 6.4 - Um exemplo de Implementação
- 6.5 - Comentários sobre alguns detalhes de Implementação

APÊNDICE 1 - O CONJUNTO DAS INSTRUÇÕES DE MÁQUINA DO PATINHO FEIO

- A1.1 - Grupo 1 - Instruções de Referência à Memória
- A1.2 - Grupo 2 - Instruções de Endereçamento Imediato
- A1.3 - Grupo 3 - Instruções de Deslocamentos e Giros
- A1.4 - Grupo 4 - Instruções de Entrada e Saída
- A1.5 - Grupo 5 - Instruções Curtas com operando
- A1.6 - Grupo 6 - Instruções Curtas sem operando

APÊNDICE 2 - ROTINAS AUXILIARES UTILIZADAS NA CONSTRUÇÃO DO "SOFTWARE" BÁSICO

- A2.1 - A rotina de geração de fita perfurada carregável ("dumper")
- A2.2 - O carregador de Fitas Absolutas

- A2.3 - A Rotina de Listagem do Conteúdo da Memória
- A2.4 - A Rotina de Carregamento de Memória a partir de dados em hexadecimal
- A2.5 - Comentários

APÊNDICE 3 - FORMATOS DE FITA OBJETO

- A3.1 - Formato de fita Objeto Absoluta
- A3.2 - Formato de fita Objeto Relocável

APÊNDICE 4 - ALGUNS EXEMPLOS DE UTILIZAÇÃO DOS PROGRAMAS DESENVOLVIDOS

1. INTRODUÇÃO

1.1 - Objetivos

No presente trabalho expõe-se os métodos utilizados no projeto e construção de algumas das peças do "software" básico do Patinho Feio, o primeiro minicomputador desenvolvido no Laboratório de Sistemas Digitais (LSD) do Departamento de Engenharia de Eletricidade da Escola Politécnica da Universidade de São Paulo. Descreve-se as diversas fases do projeto, analisando detalhadamente os principais problemas que cada uma delas apresenta. Esta análise é feita em nível geral, quando se tratar de problemas que devem ser enfrentados independentemente da máquina, ou então em nível particular, em caso contrário. Esta descrição destina-se principalmente à orientação de futuros projetos semelhantes.

A análise dos problemas dependentes da máquina, bem como a dos detalhes de implementação dos programas descritos, restrin ge-se ao caso particular dos exemplos implementados, devendo-se levar em conta a configuração do sistema para o qual foram desenvolvidos. Assim, algumas das soluções apresentadas em tais exemplos podem não ser vantajosas em sistemas que não tenham as mesmas características.

São apresentadas alternativas para as soluções de alguns dos problemas enfrentados, fazendo-se uma análise comparativa das mesmas. Procurou-se, para isto, ressaltar vantagens e desvantagens de cada uma das alternativas mencionadas, terminado-se por optar por uma delas e, sempre que possível, justificando-se a opção adotada na implementação do exemplo, com uma argumentação baseada nas características e limitações do computador utilizado.

Além disto, é descrito aqui o funcionamento dos programas implementados, especialmente com a finalidade de documentar as linhas de raciocínio e a filosofia de projeto e de implementação adotadas para facilitar futuras alterações de tais programas.

1.2 - Generalidades (ref. 11, 14)

Nos próximos Capítulos estão desenvolvidos, dentro das linhas apresentadas em 1.1, os seguintes módulos do "software":

- um montador ("assembler"), cuja finalidade é a de dispensar o programador da tarefa de programação em linguagem de máquina, permitindo que os programas sejam escritos em linguagem simbólica mnemônica. Uma das versões implementadas dâ, além disto, a possibilidade de modularização dos programas, introduzindo a facilidade de se utilizar programas escritos e traduzidos independentemente sem a necessidade de nova tradução a cada utilização dos programas parciais;
- uma rotina de depuração, cuja finalidade é a de facilitar a pesquisa e correção de erros existentes em programas ainda não depurados;
- um desmontador ("disassembler"), cuja finalidade é a de gerar programas escritos na linguagem do montador, a partir de uma fita binária contendo o programa correspondente em linguagem de máquina;
- uma rotina de edição de arquivos ASCII, cuja finalidade é a de mecanizar a correção, a reprodução, a modificação e a listagem de arquivos ASCII, como por exemplo, programas perfurados, em linguagem simbólica, em fita de papel ("programas fonte").

Os quatro programas citados acima foram desenvolvidos para o Patinho Feio estando atualmente à disposição para execução no mesmo.

Alguns programas de utilidade foram implementados no computador Hewlett-Packard 2116-B, com a finalidade de facilitar e de tornar mais versátil a utilização do "software" do Patinho Feio.

Escolheu-se o computador HP-2116-B principalmente pela sua compatibilidade de entrada e saída com a do Patinho Feio (fita de papel). Os programas desenvolvidos para o HP-2116-B não foram implementados diretamente no Patinho Feio devido a algumas restrições atualmente existentes no mesmo, como por exemplo sua pequena capacidade de memória, a inexistência de memória de massa e a baixa velocidade dos periféricos disponíveis. Entretanto, poderão alguns desses programas vir a ser implantados no Patinho Feio assim que estiverem disponíveis no mesmo os periféricos necessários. Os programas aqui apresentados são os seguintes:

- um simulador em nível de registradores, que funciona como um interpretador da linguagem de máquina do Patinho Feio e que também permite a utilização dos periféricos rápidos do HP-2116-B em substituição aos disponíveis no Patinho Feio, bem como o acompanhamento do programa em execução por meio de rastreamento ("traces"), descarregamentos ("dumps") de memória, monitoramento dos processos de entrada e saída, etc. Naturalmente a utilização deste simulador-interpretador é às vezes mais conveniente que a do próprio Patinho Feio, como por exemplo, quando o programa for limitado em velocidade pela entrada e saída (caso de rastreamento e de listagens extensas com pouco processamento);

- um montador ("cross-assembler") equivalente ao desenvolvido para o Patinho Feio, que incorpora uma opção de geração da tabela de referências cruzadas ("cross-reference symbol table") e ordenação alfabética da tabela de símbolos, permitindo utilização total dos recursos do sistema operacional DOS ("disc operating system") do computador HP-2116-B. A tabela de referências cruzadas tem como finalidade facilitar a documentação e mesmo a depuração de programas extensos, gerando, a partir da linguagem simbólica, uma tabela em ordem alfabética, de todos os rótulos ("labels", isto é, nomes simbólicos dados aos endereços das instruções ou dos dados do programa), ao lado de cada qual são listados o número da linha onde o mesmo foi definido e o conjunto de todas as linhas onde foi referenciado. Indica também os rótulos indefinidos, os não utilizados, e o endereço de memória associado a cada rótulo.

1.3 - Observações

- Os dois computadores do LSD em que se efetuou a elaboração dos programas descritos neste trabalho, foram utilizados - com a seguinte configuração:

PATINHO FEIO:

1 Unidade Central de Processamento, com memória de núcleos de ferrite com 4K palavras de 8 bits (1K = 1024).

1 Leitora Ótica de Pita de Papel HP-2737-A, 3000 caracteres por

segundo (máximo).

- 2 Terminais Teletype ASR33 com leitora e perfuradora de fita de papel, entrada por teclado e saída impressa, 19 caracteres por segundo (perfuração opcional).

HEWLETT-PACKARD 2116-B:

- 1 Unidade Central de Processamento, com memória de núcleos de ferrite com 16K palavras de 16 bits.
- 1 Unidade de Disco Magnético HP-2770-A, com capacidade de 3 Megabits, organizados em 64 trilhas, de 92 setores cada, tendo cada setor 64 palavras de 17 bits.
- 1 Impressora HP-2767-A de 89 colunas por linha, 1119 linhas por minuto (máximo).
- 3 Unidades de Fita Magnética HP-7970-A, velocidade máxima 37.5 ips, densidade 899 bpi.
- 1 Console Teletype ASR35, de 199 caracteres por minuto.
- 1 Leitora Ótica de Fita de Papel HP-2748-B, de 599 caracteres por segundo (máximo).
- 1 Perfuradora de Fita de Papel HP-2895-B, de 75 caracteres por segundo.
- 1 Leitora de Cartões de marcas óticas HP-2761-A, de 299 cartões por minuto.

- Os programas desenvolvidos para o Patinho Feio foram inicialmente projetados para a configuração mínima acima descrita. Entretanto, com a inclusão de novos periféricos, é conveniente que se configure os programas existentes, adaptando-os aos novos equipamentos disponíveis, o que pode tornar mais rápida e eficiente a execução de tais programas. Assim, todas as vezes que um

novo periférico é incorporado ao sistema, faz-se a adaptação dos programas às novas condições, permitindo que sejam utilizados os recursos implantados.

Atualmente, dispõe-se dos seguintes periféricos adicionais, já incorporados ao sistema pela equipe do LSD:

- 1 Impressora de linha de 80 colunas HP-2767-A (1110 linhas/minuto).
- 1 Terminal DECWRITER DEC LA 398, de 80 colunas, 39 caracteres por segundo (máximo).
- 1 Perfuradora Rápida de Fita de Papel HP-2735-A, de 115 caracteres por segundo.

Os programas menos extensos permitem a inclusão de uma "seção de configuração", parte do programa que o adapta para a utilização dos periféricos desejados. Os mais extensos, no entanto, não dispõem de memória livre suficiente que possa alojar a seção de configuração. Para estes programas, a configuração pode ser feita manualmente modificando-se certas posições de memória, ou, então montando-se novamente o programa com as devidas alterações.

- Em todos os programas aqui descritos, procurou-se orientar o projeto de tal modo que o programa resultante apresentasse o maior número possível de recursos por unidade de área de memória ocupada. É evidente que, na maioria das vezes, uma tentativa de adicionar mais recursos ao programa nem aumentar a área ocupada pelo mesmo pode trazer problemas no desempenho do programa, tais como queda de velocidade, quebra de estrutura, etc. Em alguns destes casos, como será visto, optou-se por soluções de compromisso, e, em outros, por soluções que dãoem ao programa maiores recursos de utilização.

No presente caso, tem-se os seguintes fatores a considerar :

- a) baixa capacidade de memória (4K palavras de 8 bits)

- b) baixa velocidade de saída dos dados (teletype: 10 palavras por segundo)
- c) deficiência do conjunto de instruções no que se refere a:
- manipulação de dados com mais de uma palavra (p/ex, dados em precisão dupla).
 - comparação de dados
 - extração e inserção de campos dentro de uma palavra (Inclusive mascaramentos)
 - operações aritméticas

Analizando-se os fatores a) e b), levando-se em conta observações realizadas nos programas aqui descritos, chegou-se às seguintes conclusões:

- O tempo de processamento é desprezível em relação ao de entrada e saída. Isto permite concluir que mesmo se o programa for algumas vezes mais lento, o tempo total de execução será praticamente o mesmo.
- Rotinas otimizadas em relação ao tempo de processamento ocupariam, em certos casos, até cerca de 40% a mais de memória em relação a rotinas equivalentes, otimizadas em relação ao número de palavras ocupadas pela mesma na memória. Levando-se em conta o fator a), percebe-se que, quando a dimensão do programa for comparável à dimensão da memória disponível, pode-se chegar até a uma condição de inviabilidade de implementação do mesmo se se insistir em utilizar rotinas otimizadas em velocidade. Deve-se observar que tais rotinas sejam, em média, poucas vezes mais rápidas, o que já foi dito ser praticamente irrelevante no caso.
- Muitas vezes, na tentativa de se reduzir o espaço

ocupado na memória para executar uma certa tarefa, pode ser encontrada uma organização eficiente dos dados que permita a utilização de rotinas rápidas com baixo gasto de memória. Para isso, deve ser tirado o máximo proveito de características particulares da arquitetura da máquina.

- Deve-se levar sempre em conta que este trabalho é dirigido para um computador pequeno e restrito. Assim sendo, algumas das argumentações aqui apresentadas podem não ser válidas para máquinas maiores, e algumas considerações, decisivas neste contexto, podem tornar-se totalmente irrelevantes se no lugar do Patinho Feio for colocado um computador de parte maior.

2. O MONTADOR

Neste capítulo é descrito o programa montador desenvolvido para o Patinho Feio, e cuja finalidade é a de permitir que se programe em uma linguagem simbólica próxima de linguagem de máquina deste minicomputador. Parte-se de uma discussão sobre problemas gerais de programação em linguagem de baixo nível, estudando-se a seguir as características desejáveis para o programa montador, ao lado dos problemas enfrentados na implementação das mesmas num computador do porte do Patinho Feio. A seguir, descreve-se alguns detalhes importantes de projeto e de implementação, finalizando com a descrição de alguns recursos de controle do programa e da lógica do montador implementado.

2.1. O conjunto de instruções e a linguagem do montador

2.1.1. O conjunto de instruções: O Patinho Feio possui um conjunto de 56 instruções, descritas quanto ao funcionamento nas referências (1) e (4). Estas instruções podem ser divididas, para efeito dos algoritmos de montagem que são utilizados em seis grupos, detalhados no apêndice 1:

- instruções de referência à memória;
- instruções de endereçamento imediato;
- instruções de deslocamentos e giros;
- instruções de entrada e saída;
- instruções curtas com operando;
- instruções curtas sem operando.

Todas as instruções pertencentes a um dado grupo, utilizam a mesma rotina de montagem, como será visto em 2.5.3.

2.1.2. Programação em linguagem de máquina: Dado um computador com seu conjunto de instruções, a única maneira de fazê-lo funcionar, se não se dispuser de nenhum outro recurso auxiliar como, por exemplo, um montador em outro computador ("Cross-Assembler"), será programá-lo em linguagem de máquina, isto é, construir uma sequência lógica de zeros e uns, tais que, carregados convenientemente na memória do computador, e a seguir executados, perfaçam a tarefa prevista.

Quando se procede desta maneira, o trabalho de escrever um programa, carregá-lo na memória e em seguida executá-lo é árduo, pois requer muitos cuidados para que o problema que se está atacando seja resolvido corretamente. Assim, dado um problema para ser resolvido em um computador no qual só se dispõe de linguagem de máquina e de um painel, deve-se primeiramente estabelecer um diagrama de blocos da solução do problema, e, em seguida, traduzi-lo para a linguagem de máquina do computador. Como esta tarefa é muito desagradável e trabalhosa para ser efetuada diretamente, convém que o programa não seja imediatamente codificado em linguagem de máquina a partir do diagrama de blocos, mas que se escreva tal programa em uma linguagem intermediária mais acessível, onde cada instrução, que na linguagem de máquina é representada por uma sequência de zeros e uns (que em nada lembram a função que ela deverá executar no programa), será representada por um símbolo, formado de um ou mais caracteres, que de alguma maneira informe ao programador o papel de tal instrução no programa.

Na linguagem simbólica do Patinho Feio, uma instrução qualquer terá no caso mais geral o seguinte atributos (fig. 2.1.2.1):

A	B	C	D
INIC	CART ARM ARM	29 CONTA CONTB	(INICIO DO PROGRAMA)

Fig. 2.1.2.1

Exemplo de instruções em linguagem simbólica do montador

- A - um endereço, que poderá ser simbólico (optativo);
- B - um mnemônico, que indica qual instrução a ser executada;
- C - um operando, que completará a função representada pelo mnemônico;
- D - um comentário, para efeito de documentação do programa;

Por exemplo, um pequeno programa, escrito numa destas linguagens, poderia ter o seguinte aspecto:

1. carregue a variável X no acumulador (inicialmente X = 16);
2. some B8 ao acumulador;
3. guarde o resultado em X;
4. pare;
5. desvie incondicionalmente para 1.

Recodificado na linguagem simbólica do Patinho Feio, o programa acima poderá tomar o seguinte aspecto:

```

UM   CAR   X   CARREGA X NO ACUMULADOR
      SOMI  B8  SOMA B8 AO ACUMULADOR
      ARM   X   ARMAZENA O RESULTADO EM X
      PARE
      PLA   UM  DESVIA INCONDICIONALMENTE PARA UM
      X     DEPC 16  VALOR INICIAL DE X = 16
  
```

Consultando uma tabela de correspondência entre as instruções simbólicas e os respectivos códigos de máquina, e atribuindo arbitrariamente o endereço 180₁₆ à primeira instrução do programa, poder-se-á recodificar o programa, agora em linguagem de máquina (os códigos abaixo estão em rotação hexadecimal):

ENDERECO	CÓDIGO	RÔTULO	MAMÔNTICO	OPERANDO	COMENTÁRIO
180	4189	UM	CAR	X	AC: = X
182	D858		SOMI	88	AC: = AC + B8
184	2179		ARM	X	X: = AC
186	90		PARE		PARE
187	B1F9		PLA	UM	VOLTE P/UM
189	18	X	DEPC	16	X (INICIAL) = 16

Este programa, no seu formato definitivo para utilização do montador, poderia ficar com o seguinte aspecto:

	ORG	/100	ESTABELECE ENDEREÇO DE ORIGEM
UM	CAR	X	AC: = X
	SOMI	89	AC: = AC + 89
	ARM	X	X: = AC
	PARE		PARE
	PLA	UM	VOLTE P/UM
X	DEF.C	16	X (INICIAL) = 16
	FIM	UM	ESTABELECE FIM DO PROGRAMA

As colunas de endereço e de código não aparecem nesta versão.

Observe-se que o programa assim escrito é auto-explicativo, o que não ocorre com o mesmo programa escrito em linguagem de máquina, como se pode notar observando apenas as duas primeiras colunas da penúltima versão.

Como foi visto no exemplo acima, o programa escrito nessa linguagem intermediária pode ter um formato completamente livre, a critério do programador. Se forem estabelecidas algumas regras de sintaxe e padronizações para esta linguagem intermediária, de tal modo que todos os programadores representem com os mesmos símbolos as mesmas operações, terá sido estabelecida uma linguagem de programação para este minicomputador. A esta linguagem dá-se o nome de linguagem do montador ("assembly language").

Desde que a sintaxe da linguagem do montador obedeça fielmente a regras bem definidas, haverá uma relação unívoca entre um programa escrito na linguagem do montador e o programa obtido a partir dele, escrito em linguagem de máquina. Sendo esta montagem algorítmica, poder-se-á escrever um programa relativamente simples que permita a automação desta tarefa de tradução. A este programa dá-se o nome de montador ("assembler").

2.2. A conveniência de um montador

Como foi dito em 2.1., a programação em linguagem de máquina requer inúmeros cuidados para que o programa obtido seja

executado com sucesso. Uma das tarefas mais laboriosas é a da atribuição correta de endereços numéricos aos endereços simbólicos definidos no programa. Isto decorre principalmente pelo fato de que as instruções do computador não ocupam todas o mesmo número de posições de memória, fato que se verifica em todos os computadores cujas instruções não tenham o comprimento uniforme.

Calculados corretamente todos os endereços e montado o programa, este problema surge cada vez que se desejar corrigir o mesmo, nele inserindo, ou dele retirando algumas instruções (por exemplo, quando da sua depuração ou otimização). Nestes casos será sempre necessário recalcular todos os endereços e corrigir todas as referências à memória que houverem sido alteradas pela correção. Além disto, como o comprimento do programa terá mudado, bem como a posição de memória ocupada pelas instruções deverá o programa ser, mais uma vez, carregado na memória integralmente.

Levando em conta que se dispõe apenas de um painel de chaves para a carga de programas, pode-se facilmente verificar que esta tarefa deve ser substituída por outra mais suave e confiável. Se for viável gerar uma fita de papel que possa ser lida por uma leitora de fitas, e que contenha como informação o código de máquina correspondente ao programa que se deseja carregar e executar, o problema estará resolvido. Para isto, será necessário escrever um programa montador, ou em linguagem de máquina, ou então em outro computador.

Optou-se pela alternativa de escrever o montador em linguagem de máquina.

2.2.1 Rotinas auxiliares para a elaboração do montador

Para que o processo da implementação do montador, em linguagem de máquina, se tornasse menos ineficiente, foram escritos, também em linguagem de máquina, algumas rotinas auxiliares, com a função de atenuar as dificuldades encontradas na operação do sistema:

- um carregador de fitas binárias absolutas, cuja função é a de carregar na memória um programa apresentado em fita

de papel, em formato binário;

- um descarregador de código de máquina para fita de papel, em formato aceito pelo carregador de fitas, cuja finalidade é a de salvar o conteúdo da memória em fita de papel, para posterior utilização;

- um descarregador de memória para o terminal, cuja finalidade é a de informar o estado da memória num certo instante, fornecendo no terminal um mapa com o conteúdo da memória em formato hexadecimal;

- um carregador de memória a partir do teclado do terminal, cuja finalidade é a de evitar o uso do painel de chaves, fornecendo um modo mais cômodo e confiável de carregar programas ou dados na memória, a partir de dados escritos em formato hexadecimal do teclado do terminal.

Com estes novos recursos à disposição (apêndice 2), passou-se a elaborar o montador.

2.3. Definição das Características Gerais do Montador

O primeiro passo para a elaboração do montador foi o de estabelecer algumas regras gramaticais para sua linguagem, bem como escolher os mnemônicos que representariam doravante as instruções de máquina. Além disso, decidiu-se também neste estágio sobre a existência ou não de determinados recursos na linguagem. O resultado desta fase foi o seguinte:

A) o formato de entrada é livre. Um comando genérico consta de campos, os quais são separados obrigatoriamente por brancos, e é terminado pela sequência de caracteres especiais "return", "linedeed";

B) um caráter especial "rubout" em qualquer posição da linha, anula inteiramente esta linha, até que seja encontrada a sequência "return", "linefeed";

C) um asterisco na 1a. coluna significa que a linha é de comentário;

D) todos os comandos terão basicamente 4 campos:

- campo de rótulos - que começa na primeira coluna, e que é opcional. Isto quer dizer que, se a primeira coluna estiver em branco, o campo dos rótulos está vazio.

Se existir, deve constar de um ponto ou então de uma sequência de letras, de comprimento qualquer, das quais são consideradas, se existirem, apenas as duas primeiras e a última, para efeito de endereçamento simbólico (esta restrição não é uma limitação séria quanto ao número de rótulos, uma vez que se consegue formar, com as 26 letras do alfabeto, segundo a regra descrita, 18278 rótulos diferentes, embora o número máximo de símbolos permitidos em um programa seja de 256).

- campo dos mnemônicos - que se inicia no primeiro caráter não branco encontrado na linha após o primeiro branco e termina no caráter que precede o próximo branco ou "return" da linha. Este campo é obrigatório. São válidos neste campo os mnemônicos definidos no apêndice 1 e que foram inspirados nos da ref. 1. Também neste campo são utilizados para identificação apenas os dois primeiros e o último caráter do mnemônico. Os demais são ignorados.

É este campo que irá definir o comportamento do montador em relação à linha que o contém, estabelecendo se a linha deve gerar duas, uma ou nenhuma palavra de código, se é apenas controle do montador, etc, ou então, no caso de geração de código, definindo qual o algoritmo de montagem a ser utilizado.

- campo dos operandos - este campo está subordinado ao campo dos mnemônicos. Conforme a classe do mnemônico aí encontrado, o campo dos operandos deve ser composto conforme as regras seguintes:

o instruções de referência à memória: exigem operando, que poderá ser:

- 1) Simbólico puro. Ex.: CAR YA (qualquer símbolo definido no programa)
- 2) Simbólico relativo. Ex.: CAR YA+3 (idem, seguido de deslocamento absoluto)
- 3) Relativo puro. Ex.: CAR *+5 (*seguido ou não de deslocamento absoluto)

- 4) Absoluto. Ex.: CAR /310 (qualquer constante sem sinal ou ***)
- 5) Local puro. Ex.: CAR .-2 (.- ou .+ seguido ou não de um dígito hexadecimal não nulo)
- 6) Local relativo. Ex.: CAR .+ -5 (.- ou .+ seguido ou não de um dígito hexadecimal não nulo, seguido de deslocamento - absoluto)

• instruções de endereçamento imediato: exigem operando que deverá ser uma constante, com ou sem sinal.

Ex.: CARI	25
SOMI	+3P
NAND	-1B2
XOR	-/B1

• deslocamento e giros: exigem operando constante, entre 0 e 4 inclusive.

Ex.: DD	1
DE	/B1
GEV	+4

• entrada e saída: exigem operando constante. Depois de convertido para binário, o operando tem o significado seguinte: seus 4 primeiros bits indicam o canal utilizado e os 4 últimos o tipo de ação de entrada e saída a ser executada.

Ex.: FNC	/D2
SAL	/B2
ENTR	48 ($48_{10} = /30$)
SAI	-48 ($-48_{10} = /D0$)

• curtas com operando: exigem operando constante, a saber:

1 - Testes dos flipflops Transbordo (T) e Vai Um (V):
 o operando deve ser interpretado como booleano (\emptyset ou $\neq \emptyset$) :

Ex.:	SVM	\emptyset	=	SVM \emptyset
	ST	/3	=	ST 1
	SV	-8	=	SV 1
	STM	-GP	=	STM 1

2 - Painel: operando deve ser uma constante entre \emptyset e 7:

Ex.:	PNL	\emptyset
	PNL	/5

• curtas sem operando: estas não exigem operando. Se algo for colocado no campo dos operandos, será ignorado.

Ex.:	TRI	*	=	TRI
	INC	ABCDE	=	INC
	PARE		=	PARE

• pseudo-instruções ("pseudo"): cada uma das pseudo exige operando adequado:

NOME			
SEGM	- Exigem operando simbólico puro	NOME	X
SUBR		EXT	DRIVER
EXT			
ENT	- Exige operando simbólico puro ou relativo	ENT	ABC
		ENT	A+5
ORG	- <u>No montador absoluto</u> , (2.4.1), esta pseudo exige operando absoluto se for o 1º ORG do programa. Caso contrário, o operando poderá ser absoluto, simbólico puro, simbólico <u>re</u> lativo, relativo puro, local puro ou local relativo. <u>Exceção</u> . Se for local, não poderá ser .+ n.		

No montador relocável: (2.4.2), valem as mesmas considerações, exceto que o operando não pode ser absoluto. São permitidos apenas os tipos simbólicos puro ou relativo , relativo puro, e local puro ou relativo desde que não seja .+n .

DEFE } - Exigem operandos simbólicos,absolutos,relativos ou locais.
DEFI } -

BLOC } -
DEFC } - Exigem operando constante decimal, ASCII, ou hexadecimal,
COM } com ou sem sinal.

EQU - Análogo ao ORG absoluto.

FIM - No montador absoluto (2.4.1), exige operando qualquer.
 No montador relocável, (2.4.2), o operando indica endereço de execução se programa principal, ou deve ser zero se for uma subrotina.

- campo dos comentários - no campo dos comentários (opcional) pode figurar qualquer sequência de caracteres que não inclua os caracteres especiais "rubout" "return" ou "linefeed". O campo dos comentários é terminado com a sequência "return" "linefeed", a qual também serve para encerrar a linha do comando.

E - qualquer programa deve ser iniciado por um comando de definição de origem. Há quatro destes comandos: ORG,NOME,SEGM e SUBR. Como será visto em detalhes em 2.5.6, o primeiro é utilizado para definir a origem de programas absolutos, ao passo que os demais servem para associar o endereço relativo "zero" à primeira instrução de um programa principal, segmento ou subrotina, respectivamente;

F - qualquer programa deve ser finalizado por um comando de final de programa. Esta é a pseudo FIM, que, nos programas absolutos e subrotinas relocáveis indica apenas o final físico do programa, e, nos relocáveis, associa além disto no caso de progra

mas principais, o endereço de execução do código gerado;

G - os mnemônicos que representam instruções da máquina serão, quando possível, os mesmos utilizados na documentação dos circuitos correspondentes. Se não for possível o uso dos mesmos mnemônicos, os adotados serão tão próximos quanto possível dos originais;

H - levando em conta o pouco espaço (quantidade de memória) disponível para o programa e para os dados, deve-se escolher uma forma de melhorar o aproveitamento da memória para a construção da tabela de símbolos (2.5). O ideal é maximizar o número permitido de símbolos, e, ao mesmo tempo, minimizar o espaço ocupado por eles na tabela.

Para maximizar o número permitido de símbolos em um espaço fixo de tabela, pode-se por exemplo minimizar o espaço de tabela que cada uma delas ocupa. Assim, por exemplo, seria bom que, na tabela, cada símbolo pudesse ser representado por uma única palavra de 8 bits. Uma possibilidade de realizar isto seria a de utilizar símbolos de dois caracteres, sendo o primeiro alfabético e o segundo numérico. O alfabeto, tendo 26 letras, necessita na sua representação interna, de 5 bits, e as seis combinações restantes poderiam ser preenchidas com caracteres especiais. Os 3 bits que restam na palavra seriam preenchidos pela representação binária de algarismos entre 0 e 7.

O número máximo de símbolos é neste caso 256. Uma vantagem desta solução é a seguinte: se forem reservadas 256 posições na tabela, o símbolo não necessitará estar presente na mesma permanecendo na tabela apenas as informações relativas a ele: o símbolo estaria implícito na posição da tabela em que as informações a ele relativas estariam presentes. Além disto, há uma facilidade adicional pela simplicidade da busca das informações: basta que o registrador de índice seja carregado com o símbolo compactado para que se tenha acesso à informação na tabela com uma única instrução indexada.

Uma desvantagem deste método é clara: os símbolos não são

mneômicos, tornando portanto mais difícil o acompanhamento lógico de um programa escrito na linguagem do montador usando símbolos assim definidos.

Sempre que para um problema apareçam duas soluções diferentes, uma de fácil implementação mas que representa para o usuário maior dificuldade de programação, e outra de implementação mais elaborada, mas que poupa esforço ao usuário, é recomendável que se opte pela segunda desde que isto não venha a causar um problema de insuficiência de área de memória para o restante do programa em questão. Caso isto ocorra, uma eventual solução de compromisso pode ser tentada.

Com base nestas considerações, a idéia de utilização de símbolos de dois caracteres exposta acima foi abandonada, tentando-se um resultado melhor com símbolos de três caracteres alfabéticos, compactados em 15 bits, o que ocupa duas palavras de oito bits, restando ainda um bit para usos eventuais em futuras modificações do programa. Na versão definitiva do montador, os símbolos permitidos são de comprimento qualquer. Internamente, no entanto, são considerados apenas os dois primeiros caracteres e o último. Naturalmente esta solução não é a mais econômica para o sistema, nem a mais cômoda para o usuário, mas foi a que melhor se adaptou às exigências do usuário e do sistema, entre todas as que foram experimentadas;

I - quanto aos tipos de referência à memória permitidos decidiu-se nesta fase permitir, na primeira versão do montador ("bootstrap", isto é, versão inicial, com a ajuda da qual o programa definitivo é construído), apenas referências absolutas e simbólicas puras.

As referências locais, bem como as relativas ao Contador de Instruções e a rótulos foram introduzidas mais tarde, no montador definitivo;

J - quanto às constantes, permitiu-se declarar, no "bootstrap", constantes decimais com ou sem sinal, hexadecimal sem sinal e ASCII sem sinal. Na versão atual são permitidos, além destes tipos, constantes hexadecimais e ASCII com sinal. As constantes apresentam-se em seis formas diferentes. Em princípio não há restrição quanto a sua amplitude. Entretanto, como a conversão é fei-

ta para formato binário de oito bits, o valor final é sempre o número binário formado pelos oito bits menos significativos do número.

FORMATOS DAS CONSTANTES

Sem Sinal:

- a) decimais: As constantes decimais sem sinal constam de uma sequência de dígitos decimais (0 a 9) precedidos e sucedidos por um delimitador.

Ex.: 001; 152; 9999

- b) hexadecimais: As constantes hexadecimais sem sinal constam de uma sequência de dígitos hexadecimais (0 a 9 e A a F), precedidos por um caractere "/" e sucedidos por um delimitador.

Ex.: /001; /AFF; /13E

- c) ASCII: As constantes ASCII sem sinal constam de um caractere especial "@" sucedido por um caractere ASCII qualquer.

Ex.: @1; @P; @ @; @%

Com Sinal:

- d) decimais: As constantes decimais com sinal constam de um sinal (+ ou -) sucedido de uma sequência de dígitos decimais:

Ex.: +53; -18; +0001

- e) hexadecimais: As constantes hexadecimais com sinal constam de um sinal seguido de uma constante hexadecimal sem sinal:

Ex.: +/135; -/12 : -/A8E

f) ASCII: As constantes ASCII com sinal constam, analogamente, de um sinal seguido de uma constante ASCII sem sinal.

Ex.: - @-; + @ @; - @#; + @1

K - decidiu-se também quais as pseudo instruções que deveriam existir.

Implementou-se, no "bootstrap", as pseudos ORG, BLOC, DEFC e FIM.

Na versão definitiva acrescentou-se (cap. 2.5) NOME, ENT, EXT, EQU, COM, SEGM, SUBR, além de dar maior versatilidade às outras. Assim, permitiu-se, na versão definitiva, ORG com operando simbólico e/ou relativo, DEFC com operandos com ou sem sinal, BLOC com operandos constantes quaisquer (no "bootstrap" o operando de BLOC tinha que ser decimal e sem sinal). FIM significava, no "bootstrap", apenas o final do programa. Na versão final, o seu operando pode significar o endereço de execução, dependendo do tipo de programa.

L - decidiu-se quanto ao número de passos e quanto à ocasião em que as deteções de erro iriam ocorrer.

Diz-se que um programa tradutor tem n passos quando o programa fonte e/ou programas dele derivados são lidos, analisados e modificados n vezes pelo programa tradutor, antes que o programa objeto (traduzido) esteja pronto. É muito comum a confusão entre número de passos e número de fases de um programa tradutor. O número de passos é decorrente da divisão lógica de um programa em partes estanques, que são executadas geralmente na mesma sequência. Eventualmente um passo pode ser tão complexo ou extenso que exija uma subdivisão em partes menores chamadas fases, módulos ou ainda segmentos. Estas fases interagem normalmente umas com as outras, e sua sequência de execução é geralmente uma função do particular programa que está sendo traduzido. Nota-se portanto que um programa tradutor é dividido em fases puramente por razões físicas e não lógicas. Durante o projeto de um programa tradutor, o número de passos pode ser definido "a priori", o que não ocorre com o número de fases, que depende fortemente de decisões tomadas durante a etapa de implementação.

Havendo opção entre um ou mais passos, resolveu-se fazer uma análise de vantagens e desvantagens de cada uma delas. A primeira vista, a opção de um passo único é mais atraente, pois neste caso o programa fonte deve ser lido apenas uma vez, e à medida que isto acontece, o código objeto vai sendo gerado e a listagem sendo feita. Esta opção realmente é interessante para máquinas onde haja possibilidade de se armazenar na memória o programa gerado para que, montado o programa todo, seja executado o código logo a seguir ("assemble and go"). No caso isto é impossível, uma vez que a memória está praticamente lotada pelo montador, não havendo espaço onde o programa montado possa ser guardado. Assim, é necessário gerar uma fita de papel com o programa objeto, em formato carregável. Outro problema surge além disto, em montadores de um passo: é o das referências para diante: cada vez que se referenciar um símbolo ainda não definido, é necessário guardar a posição e a instrução onde foi feita a referência, bem como o deslocamento em caso de referência relativa, e o símbolo referenciado. Isto pode ser feito com relativa facilidade usando-se a técnica de guardar estas informações em listas ligadas. Estas listas seriam consultadas na ocasião da definição do símbolo, quando seria gerada uma fita contendo as informações convenientes para que o carregador ("loader") pudesse preencher corretamente as posições indicadas. Uma vez feito isto, pode-se eliminar a lista, guardando-se apenas o endereço correspondente ao símbolo, para uso em futuras referências. Mais uma vez, a escassez de memória se faz sentir, pois listas ligadas são ávidas consumidoras de memória: cada elemento desta lista deveria conter no mínimo as seguintes informações, no presente caso:

- posição onde foi feita a referência - 12 bits
 - qual a instrução - 4 bits
 - informação de deslocamento - 12 bits
 - apontador para próximo elemento da lista - 12 bits
- 48 bits = 5 palavras

No caso de montador "assemble and go" esta lista poderia ser, ao

menos em parte, implementada na própria posição de memória onde o programa está sendo guardado: os apontadores e a informação sobre a instrução poderiam estar nesta área, enquanto os deslocamentos estariam em tabela a parte, e as posições onde foram feitas as referências estariam implícitos na posição dos apontadores.

O esquema de um único passo apresenta também outras desvantagens: se houver um erro em algum ponto do programa, todo o tempo gasto em montar e listar o programa até este ponto, bem como a fita objeto gerada, seriam perdidos. Poder-se-ia contornar o problema montando-se o programa uma vez sem as opções de listagem e geração de fita objeto, com a finalidade de detetar erros. Constatada a inexistência de erros, far-se-ia nova montagem, dessa vez permitindo-se as opções de listagem e geração. Como este procedimento seria rotineiro, haveria necessidade de duas leituras do programa fonte, sempre que se quisesse montá-lo sem correr o risco de perder o tempo de geração de uma listagem e da fita perfurada com o código objeto.

Levando-se em conta estes fatos, bem como a maior facilidade de programação do montador em dois passos, optou-se por este último esquema. Assim, no primeiro passo devem ser detetados todos os erros de sintaxe, gerada a tabela de símbolos, e decidido se a memória pode ou não conter todo o programa. Todas as mensagens de erro são geradas no primeiro passo, de modo que, se o programa conseguir passar por ele sem nenhum erro e sem nenhum símbolo indefinido, poderá ser lido pelo segundo passo sem o problema de perda de tempo devido a erros de codificação. No segundo do passo foram, entretanto, conservadas as detecções de erro e impressão de mensagens para prevenir quanto a possíveis erros de leitura da fita fonte ou manuseio inadequado dos periféricos pelo operador.

Mensagens de erro obtidas no segundo passo são, entretanto, apenas avisos ao operador de que o equipamento ou a operação não desempenham bem seu papel. No caso de ocorrência deste tipo de mensagem, basta ler novamente a fita de modo correto para que o programa seja montado adequadamente.

2.4. Definição das Características Externas do Montador

Definidas as principais características do programa montador, i.é., a sintaxe da linguagem de entrada, o número de passos, e os recursos de que se deseja dotá-lo, passou-se ao seu detalhamento.

2.4.1. Características do Montador Absoluto

No montador absoluto, não são permitidas referências a símbolos globais, isto é, externos ao programa. Assim, um programa absoluto deve ser auto-suficiente, não podendo, portanto, depender de rotinas de biblioteca referenciáveis por nome. Isto não é uma limitação séria se se dispuser de tais rotinas já montadas em posições conhecidas de memória, caso em que as referências a elas poderão ser do tipo absoluto (por endereço). Outro modo de contornar a situação é obter as rotinas desejadas em formato fonte, e anexá-las, no instante da montagem, ao restante do programa, caso em que se deve tomar o cuidado de evitar a duplicação de símbolos.

Num programa absoluto não devem comparecer também as pseudo instruções características de programas relocáveis, a saber: NOME, ENT, EXT, COM, SECM, SUBR.

É no entanto obrigatória a presença de uma pseudo ORG com operando absoluto como primeiro comando do programa, definindo a posição de memória correspondente à origem do código. No decorrer do programa pode-se mudar a origem do código. Neste caso, o operando poderá ser absoluto ou então relativo a um símbolo anteriormente definido.

2.4.2. Características do Montador Relocável

No montador relocável, não são permitidas as pseudos ORG com operando absoluto, pois a alocação da memória deve ser totalmente feita pelo ligador-relocador (ref. 2): um ORG com operando absoluto provocaria a geração do código subsequente em posições absolutas de memória, o que iria tirar do ligador-relocador o controle da alocação de memória. É entretanto obrigatória

a presença, como primeiro comando do programa, de uma pseudo NO-ME, SIGN, ou SUBR, cuja função é dar à origem do programa um endereço relocável zero e um nome para efeito de geração dos mapas de memória na ocasião da ligação (fase em que são atribuídos endereços absolutos às instruções dos diversos trechos relocáveis de um programa, e construídos, a partir deles, módulos executáveis de código objeto absoluto).

Nun programa relocável é permitida a presença de pseudos ENT, EXT, COM, cuja função será de definir pontos de entrada como rótulos globais (isto é, referenciáveis, em outros programas, pelo nome), nomes de rótulos globais definidos em outro programa, e áreas de memória comuns a diversos programas ou subrotinas. Tais pseudos irão gerar informações para o ligador-relocador, permitindo que este execute a relocação conveniente dos programas referenciados, e aloque convenientemente a memória para os diversos subprogramas e dados.

Como foi dito, a pseudo FIM num programa relocável indica, se se tratar de programa principal ou segmento, que o endereço da execução é o indicado pelo seu operando.

2.4.3 A Sintaxe da Linguagem de Entrada

Com base nas decisões e definições anteriormente estudadas, chegou-se às características sintáticas da linguagem simbólica de entrada cujo programa tradutor é o objeto deste capítulo. Descrição pormenorizada das funções das diversas instruções e pseudo-instruções implementadas, bem como das regras de sintaxe da linguagem encontram-se na ref. 7.

2.4.4 Características do código objeto gerado pelo montador absoluto

A geração do código objeto é executada no segundo passo do montador absoluto, e obedece a algumas regras, que são detalhadas a seguir.

Não havendo espaço na memória para a geração do código "in loco" para um procedimento do tipo "assemble and go", tal código deve ser gerado em algum dispositivo de saída, em um formato carregável. Fixado este formato pelas especificações do carregador absoluto (ref. 2) chegou-se ao formato de fita objeto que consta de blocos de dados e as informações adicionais seguintes: nº de "bytes" de que o bloco se compõe, endereço do 1º "byte" de dados na memória, os dados propriamente ditos à imagem da memória e um "byte" de teste de soma longitudinal.

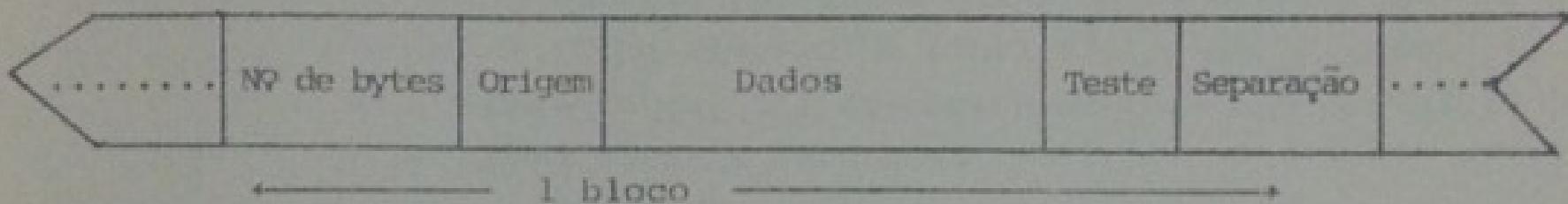


Fig. 2.4.4.1 - Formato lógico de cada bloco de dados na fita objeto absoluta.

Os blocos de que consta a fita absoluta são construídos pelo 2º passo do montador em uma área de memória, a qual deverá ser descarregada para a fita de papel cada vez que uma das seguintes condições ocorrer:

- a) preenchimento completo da área de memória: a área deve ser esvaziada para permitir que novos dados aí sejam armazenados;
- b) o número de linhas de programa fonte atingiu um múltiplo de 64. Neste caso, se houver um único dispositivo de saída para código objeto e listagem, o número múltiplo de 64 indica final de página, impondo que nesta ocasião seja descarregado o programa objeto, para que a listagem não seja prejudicada;
- c) ocorrência de alguma pseudo instrução de mudança de origem (ORG, BLOC). Como o bloco de dados é sempre

preenchido com dados que ocuparão posições contíguas, a ocorrência de uma pseudo instrução de mudança de origem acarreta a geração de dados que geralmente não irão ocupar posições contíguas às anteriores. Pode-se resolver este problema forçando-se a finalização e descarregamento do bloco que estava sendo preenchido e iniciando-se a construção de um novo bloco;

- d) ocorrência de uma pseudo instrução de FIM. Esta pseudo força o descarregamento do último bloco, formado pelos dados que ainda não haviam sido descarregados devido à não ocorrência de nenhum dos eventos citados em a) b) ou c). Como FIM indica final físico do programa, ela deverá forçar tal descarregamento, caso contrário o programa ficaria truncado.

Resumindo: o código construído pelo montador absoluto é gerado à imagem da memória e está pronto para a execução, bastando que seja carregado nas posições convenientes de memória para que possa ser corretamente interpretado e executado pelo computador. Assim, montado um programa numa certa área de memória, este programa, por referenciar endereços absolutos, poderá não funcionar se for carregado em outra região de memória. Por esta razão é necessário que se informe ao programa carregador o endereço em que o bloco de dados deve ser carregado. Além disto, é necessário especificar para o programa carregador o comprimento do bloco, pois como foi visto, os blocos têm comprimento variável. Tem-se, portanto, na fita objeto, 4 informações básicas: o número de palavras de que se compõe o bloco, o endereço em que o bloco deve ser carregado e o bloco de dados propriamente dito. Devido à baixa confiabilidade das operações de entrada e saída em qualquer periférico, é conveniente acrescentar às três informações anteriores mais uma, cuja finalidade é a de, se não eliminar, ao menos reduzir as consequências devidas à incidência de erros de leitura por ocasião da carga do programa. Assim, introduziu-se uma palavra de teste longitudinal de erros de leitura, a qual

é construída tomando-se o complemento de dois da somatória dos números binários correspondentes às instruções que compõem o bloco de dados, às duas palavras de endereço e à palavra que indica o número de dados do bloco. Se, durante a carga do programa, a somatória de todas as palavras lidas, desde o contador de dados até a palavra de teste, não for zero, terá sido detetada a ocorrência de um ou mais erros de leitura, o que deverá ser comunicado pelo programa carregador para que o operador possa tentar nova leitura ou investigar a causa do problema. Finalmente, entre um bloco e outro existe uma sequência de quatro zeros, que servem como separadores de blocos, e que são ignorados pelo carregador. Sua finalidade é exclusivamente a de facilitar ao operador a operação de tentar novamente a leitura de um bloco por ocasião de ocorrência de erro de leitura.

2.4.5. Características do Código Objeto Gerado pelo Montador Relocável

No código objeto absoluto, sendo ele gerado à imagem da memória, está implícito que os dados contidos no bloco de código são todos do mesmo tipo, isto é, números binários que deverão ser guardados na memória talis como elas são, não exigindo portanto nenhuma manipulação durante a operação de carga. No caso do código relocável, alguns dados, tais como as constantes e as instruções que não referenciam a memória, terão esta característica. Com as demais instruções nem sempre acontece o mesmo. Em alguns casos a referência é a uma posição fixa de memória, logo o código correspondente será absoluto, e portanto deverá ser gerado como se fosse uma constante. Mas, na maioria dos casos, uma instrução deste tipo contida em um programa relocável pode referenciar posições externas (chamadas de rotinas do sistema, por exemplo), endereços internos simbólicos relativos à origem do programa (por exemplo, desvios dentro do programa) ou então endereços internos simbólicos - relativos a uma área de dados comum a diversos programas montados separadamente.

Como se pode notar, há 4 tipos diferentes de dados em um programa relocável:

- dados não relocáveis;
- dados relocáveis em relação à origem do programa;
- dados relocáveis em relação à origem de uma área comum de dados;
- dados que indicam referências a posições externas (variáveis globais).

Tem-se portanto que fornecer ao programa ligador-relocador tais informações adicionais quanto à natureza dos dados que este deverá manipular. O programa não pode ser carregado diretamente por não estar em formato imagem da memória, devendo portanto, ser manipulado e relocado adequadamente (ref. 2) antes que possa ser executado. Apesar da desvantagem de um procedimento adicional intermediário entre montagem e execução, ganha-se muito em flexibilidade, pois isto permite que a manipulação de programas em formato relocável seja vantajosa na grande maioria dos casos em relação ao procedimento de montar novamente o programa fonte utilizando o montador absoluto.

O conteúdo de uma fita objeto relocável (apêndice 3) estar portanto condicionado à maneira pela qual ele será tratado pelo programa carregador. Deve-se ter, de acordo com as especificações do ligador-relocador (ref. 2) as seguintes informações numa fita objeto relocável:

- um bloco de nome do programa (bloco NOME) onde figura o nome, o comprimento e o tipo (programa ou subrotina);
- blocos que indicam o nome de todas as posições externas referenciadas no programa (blocos EXT), que informam ao ligador-relocador quais rotinas de biblioteca devem ser carregadas junto com o programa em questão;
- blocos que indicam os nomes pelos quais posições internas ao programa serão chamados externamente (bloco ENR) e seus endereços internos;

- blocos de dados, cuja estrutura é análoga à do bloco de dados do montador absoluto, sendo que neste caso cada dado é acompanhado por uma informação de relocação (informação que associa a cada dado, um dos quatro tipos descritos anteriormente);
- bloco de FIM, onde é guardado o endereço relocável de execução do programa (se fôr o caso de um programa principal ou segmento).

Todos os blocos têm, como primeira informação, o número de palavras, apresentando em seguida, o tipo de bloco, e como última informação, a palavra de teste de erros de leitura, sendo separados entre si por uma sequência de quatro zeros, sem valor lógico.

2.5. Definição das Características Internas do Montador

Estabelecidos os aspectos gerais de funcionamento e os detalhes de algumas alternativas possíveis de solução de problemas relativos ao tratamento do programa fonte, passou-se a considerar os aspectos internos da manipulação das informações nele contidas, visando à obtenção do código objeto a partir de tais informações.

2.5.1. A Representação Interna dos Rótulos

Como foi visto em 2.3., um rótulo, apesar de poder possuir mais de três caracteres no programa fonte, será sempre reduzido ao formato interno padrão de três caracteres, sendo válidos para tal fim os dois primeiros e o último caracteres da sequência de entrada. Caso tenha menos de três caracteres, o símbolo será completado com caracteres especiais "0" à direita até que tenha preenchido o número de três caracteres (fig. 2.5.1.1.).

FORMATO EXTERNO	FORMATO INTERNO
A	A @ @
AB	AB @
ABC	ABC
ABCD	ABD

Fig. 2.5.1.1

Exemplo da correspondência entre os formatos externo e interno dos rótulos

O algoritmo de transformação dos rótulos para o formato binário interno é o seguinte:

Sabendo-se que as letras do alfabeto são representadas, no código ASCII sem paridade, por uma sequência de números começando em /41 e terminando em /5A, e que o símbolo se compõe apenas de caracteres alfabéticos, bastarão 5 bits para representá-lo. Convencionando-se representar a letra "A" pelo número 00000_2 , "B" por 00010_2 , e assim subtrair /40 do código ASCII correspondente à letra para obter o correspondente compactado nos 5 bits menos significativos da palavra. O carácter "@" será representado por 00000_2 , e será injetado artificialmente no caso da ocorrência de um número de caracteres menor que 3 no símbolo original.

Chamando de N_i o número compactado definido acima correspondente ao carácter de índice i ($i = 1$ correspondente ao primeiro carácter do símbolo, etc), tem-se a seguinte definição do símbolo compactado:

$$C = \sum_{i=1}^L N_i \times 2^{5*(i-1)} \quad \text{em 16 bits}$$

Desmembrando-se em duas palavras de 8 bits tem-se

$$C_1 = \text{int } (C/2^8) \quad (\text{8 primeiros bits de } C)$$

$$C_2 = C - 2^8 \times C_1 \quad (\text{8 últimos bits de } C)$$

que definem a representação interna dos rótulos.

EXEMPLO: Para o Rótulo "AB"

A 00001

B 00010

@ 00000

Logo C =

0	00000	00010	00001
---	-------	-------	-------

ou seja: $C_1 = \boxed{00000} \boxed{00010}$

$C_2 = \boxed{00001} \boxed{00000}$

2.5.2. A Organização da Tabela de Símbolos

Em um montador, seja ele de um ou dois passos é necessário gerar-se uma tabela de correspondência entre os nomes dos rótulos e os respectivos endereços. Em montadores de um passo, para computadores pequenos, esta tabela será, ao menos em parte, residente na memória no instante da execução do programa que está sendo montado, pois em tais casos, recorre-se normalmente ao uso de posições de ligação ("links"), posições de memória que servem como apontadores para o endereço de memória correspondente ao símbolo que elas representam. Assim, em um montador de um passo para computadores com endereçamento indireto (ref. 8), utiliza-se a técnica de reservar, numa tabela, uma posição de ligação, toda vez que ocorrer uma referência a um símbolo ainda não definido, sendo então substituída a instrução que a ele se refere por uma outra que referencia indiretamente este apontador gerado. Como o endereço destes apontadores é conhecido e como estes vêm acompanhados, na tabela de símbolos por uma marca ("tag") que os identificam como apontadores, a geração

do código poderá ser feita, com relativa facilidade, por este processo. Assim que o símbolo for definido, a marca de apontador é retirada, substituindo-se o endereço do mesmo pelo endereço real do símbolo. Para computadores com maior capacidade de memória, pode-se escolher outras estratégias de tratamento de símbolos ainda indefinidos, uma das quais seria a de, detetada uma referência a um símbolo ainda não definido, guardar numa lista ligada os endereços onde ocorreram as referências ao mesmo, juntamente com a informação de qual a instrução que o referencia em cada caso. Definido o símbolo, gera-se código objeto nos endereços indicados na lista (técnica de "backtracking"), com as instruções também indicadas, utilizando o endereço real do símbolo para a montagem do código, sem utilizar o recurso do endereçamento indireto. Este processo é indicado quando não se dispõe de endereçamento indireto ou quando se deseja minimizar a área ocupada por informações que não pertencem explicitamente ao programa.

Tira-se da lista, em seguida, todo o conjunto de informações relativo ao símbolo, guardando-se apenas o seu endereço na tabela de símbolos convencional. Em montadores que geram código na própria memória, pronto para a execução ("assemble and go"), a geração da lista ligada pode ser feita "in loco", isto é, a posição da memória onde ocorreu uma referência a símbolo não definido é preenchida com informações sobre a instrução que ali deve ser montada, e com um apontador para a próxima referência ao mesmo símbolo. O apontador para o primeiro elemento de tais listas figura na tabela de símbolos, e o último elemento deverá conter uma marca identificadora de "fim de lista". (Fig. 2.5.2.1.)

TABELA DE SÍMBOLOS

NOME	ENDEREÇO	OUTRAS INF.
X		INDEFINIDO

(ANTES DA DEFINIÇÃO)

ÁREA DO PROGRAMA OBJETO

(programa montador)

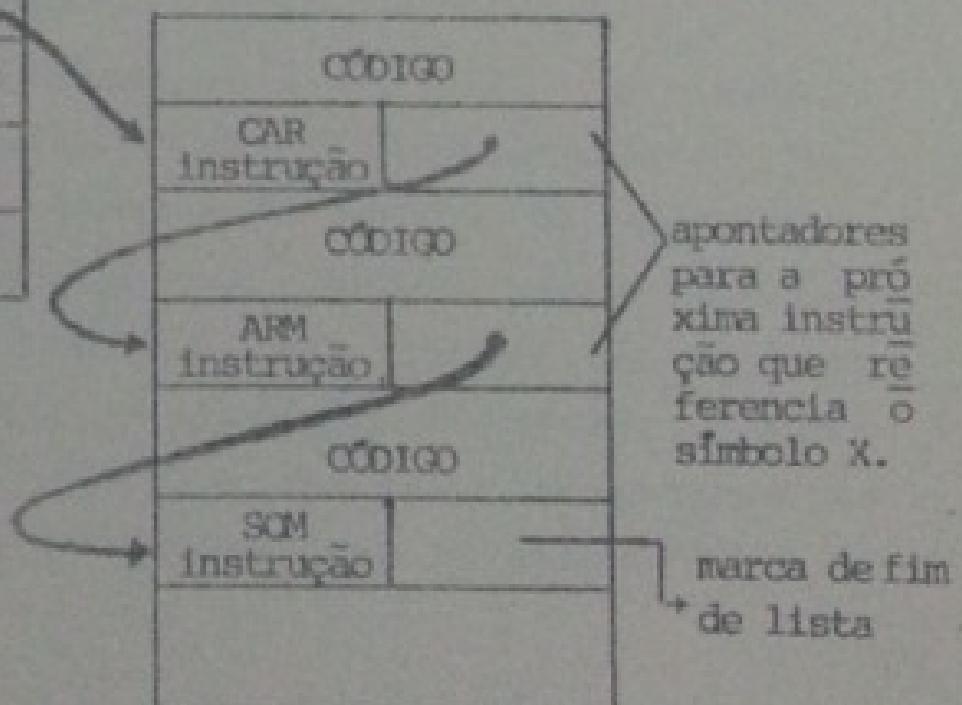
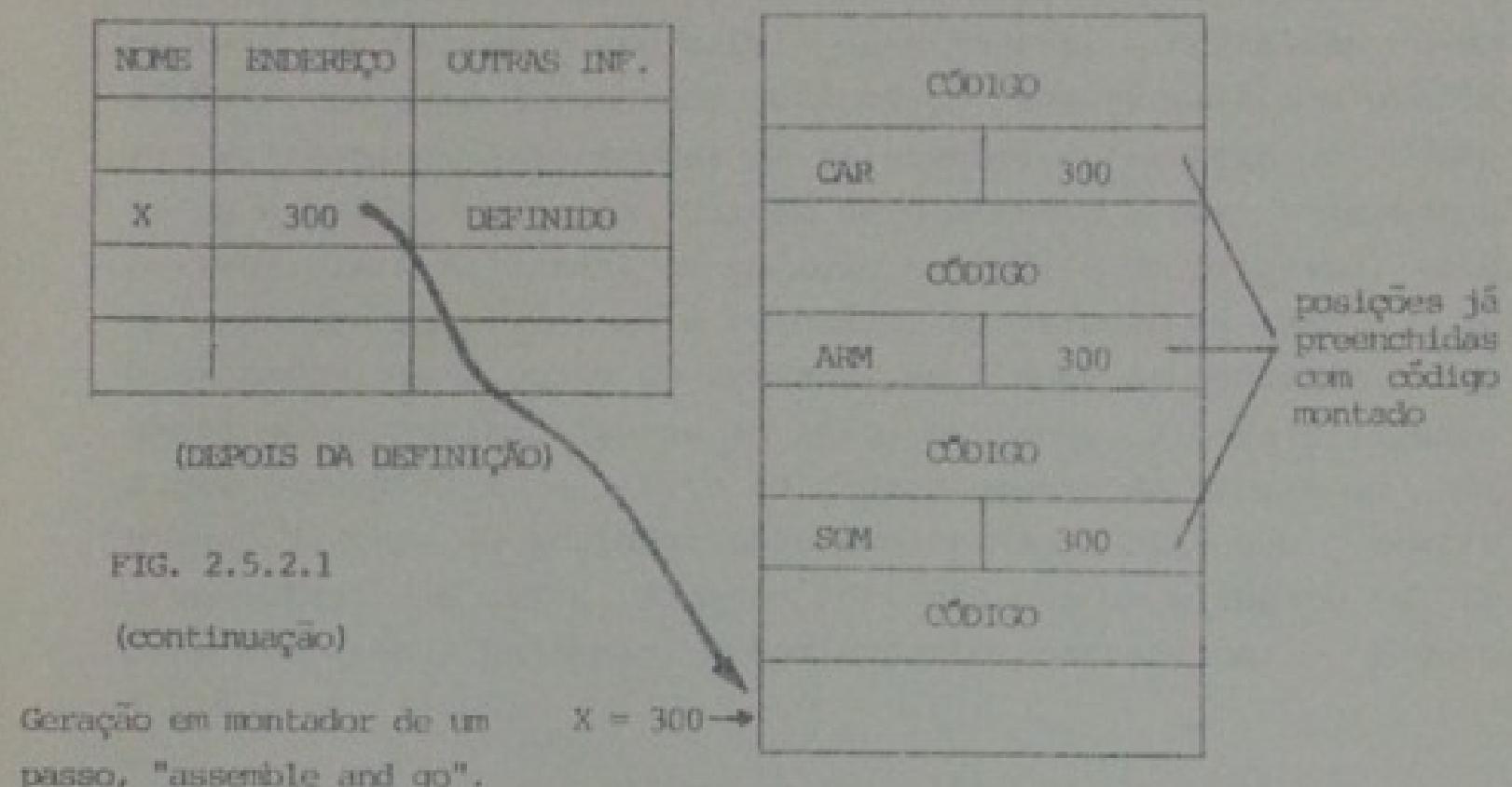


FIG. 2.5.2.1

(continua)



Em um montador de dois passos este problema não existe, pois a tabela de símbolos é produto da leitura completa do programa fonte, o qual é relido no segundo passo, e portanto a geração do código objeto utilizará informações provenientes do programa fonte e da tabela de símbolos gerada no primeiro passo. Após a montagem do programa objeto, a tabela de símbolos poderá ser destruída, pois não haverá necessidade de sua presença na memória na ocasião da execução do programa montado. Na fita objeto relocável, entretanto, deverão ser guardadas algumas informações provenientes da tabela de símbolos para posterior utilização, por ocasião da ligação e relocação do programa. Tais informações são guardadas em tabelas à parte: nomes de referências externas, nome do programa, áreas comuns a diversos programas, comprimento do programa, informações sobre o tipo do programa e outras (ref. 2).

Com base nas decisões tomadas em 2.3 e 2.4, e levando em conta a escolha do formato interno dos símbolos, pode-se passar em seguida à elaboração do formato da tabela de símbolos.

As informações que um elemento da tabela deverá conter devem ser as necessárias e suficientes para uma decisão simples e rápida sobre a natureza e sobre os atributos do mesmo. Como foi visto em 2.5.1., o nome do símbolo, uma das informações que o elemento da tabela deve conter, ocupará duas palavras. Como o

endereço a ele correspondente deverá ter 12 bits, haverá necessidade de mais duas palavras, sendo que uma delas será preenchida parcialmente com a parte do endereço correspondente ao símbolo. Um bit será necessário para indicar a definição do símbolo. No caso do montador relocável, são necessários outros dados, chamados informações de relocação, referentes ao símbolo em questão, os quais indicam se o símbolo é uma referência externa, se é relocável em relação ao inicio do programa, se é um endereço absoluto, se é relocável em relação ao inicio da área comum, se é ponto de entrada, se é nome de programa, segmento ou subrotina. Tem-se ao todo, portanto, 8 possibilidades o que consumirá mais 3 bits em cada elemento da tabela de símbolos. No caso do montador absoluto, estes 3 bits são sempre preenchidos com zeros. Fica-se portanto com um elemento da tabela ocupando 4 palavras, e com aspecto mostrado na figura 2.5.2.2.

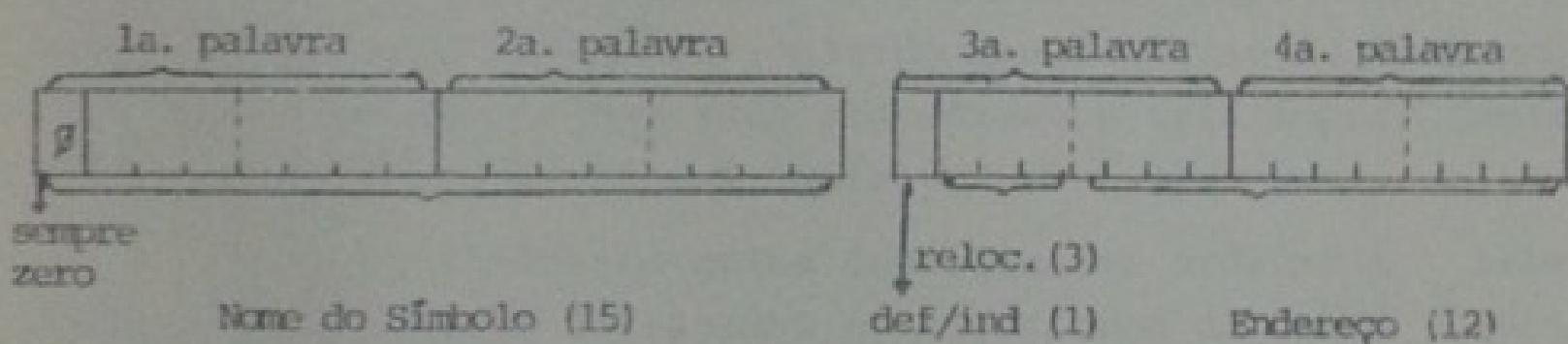


Fig. 2.5.2.2 - Estrutura de um elemento da tabela de símbolos

A ocorrência de um símbolo mais de uma vez no campo de rótulos (múltipla definição) não é considerada na tabela de símbolos, produzindo apenas uma mensagem de erro. O símbolo permanecerá com a primeira definição ocorrida no programa.

Dispondo-se de mais memória e de armazenamento auxiliar em disco ou fita magnética, poder-se-ia incorporar ao elemento da tabela de símbolos mais informações, como número da linha em que ocorreu a definição, e apontador para lista cujos elementos conteriam os números das linhas em que o símbolo foi referenciado no programa. Isto seria opcional, e execu-

cem (fig. 2.5.2.3)

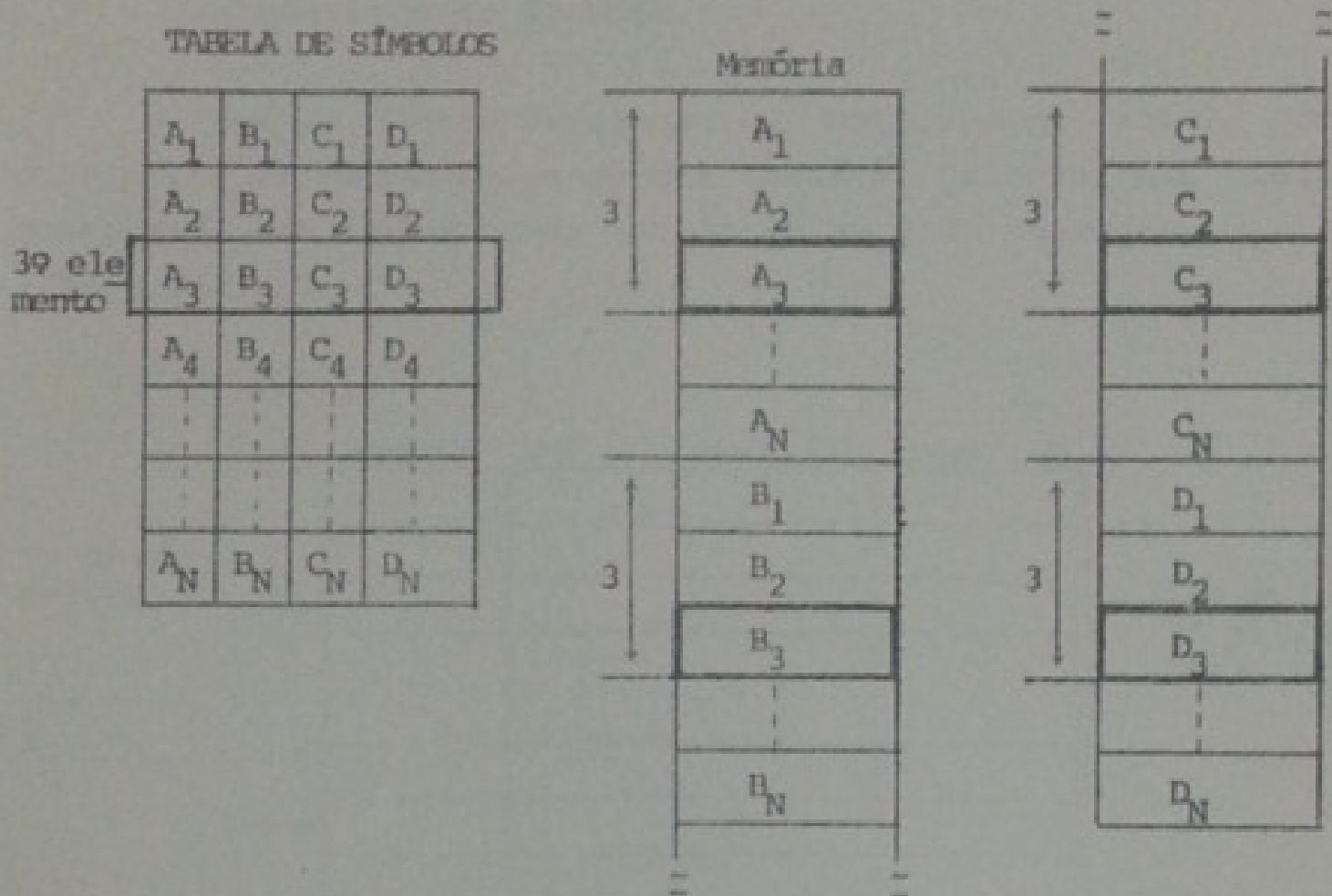


Fig. 2.5.2.3 - Organização na memória da tabela de Símbolos do Montador. Está realçado o terceiro elemento da tabela.

Assim, utilizando-se o mesmo indexador, pode-se ter acesso às quatro palavras, bastando para isto que sejam utilizadas instruções indexadas (ref. 8) referenciando os inícios das quatro subtabelas resultantes da divisão efetuada.

Com isto são economizados os cálculos de índices e possíveis cálculos de endereços efetivos caso haja necessidade de dar ao indexador um valor maior que 256.

2.5.3. A Manipulação da Tabela de Símbolos

Passa-se a seguir, com base no que foi concluído em 2.5.2, à definição das rotinas básicas para a criação, manipulação e pesquisa da tabela de símbolos. Como se sabe, a tabela em si contém as seguintes informações lógicas:

- nome do símbolo
- indicador de relocação (tipo de símbolo)
- indicador de definição
- endereço do símbolo

Como o número de elementos contidos na tabela não é constante, há necessidade de uma variável externa que indica quantos são os elementos da tabela em um dado instante.
(fig. 2.5.3.1)

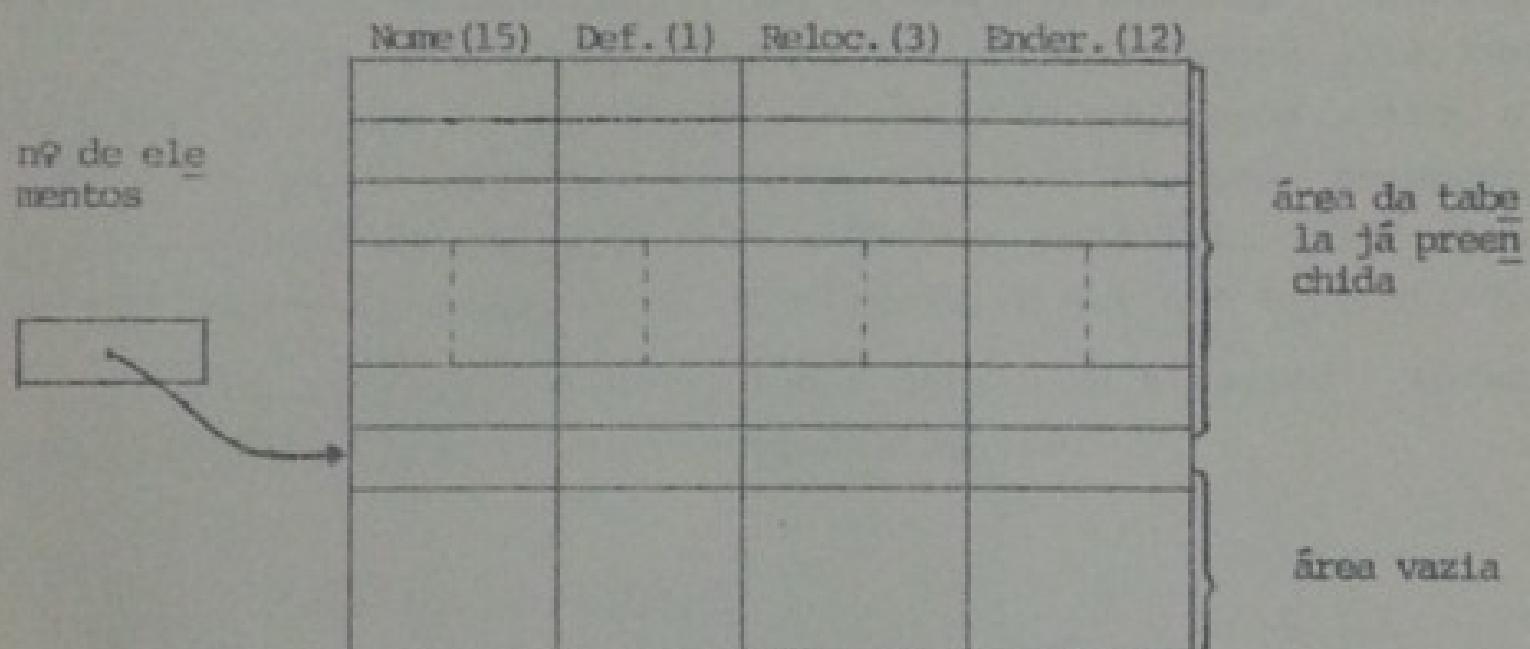


Fig. 2.5.3.1 - Esquema Lógico a Tabela de Símbolos

Três problemas a resolver se apresentam ao se escolher este esquema para a tabela de símbolos:

- a) Tem-se um símbolo, obtido a partir do programa fonte, e deseja-se saber se ele está na tabela, bem como a sua posição em caso afirmativo;
- b) Sabe-se que o símbolo não está na tabela, e deseja-se acrescentá-lo aos já existentes;
- c) Sabe-se que o símbolo está na tabela, e deseja-se conhecer ou modificar seus atributos.

O problema a) ocorre sempre que um símbolo for detetado no campo dos operandos de instruções de referência à memória, ou então no campo dos rótulos.

O problema b) ocorre quando, no caso anterior, o símbolo não for encontrado na tabela. Neste caso é necessário que se inclua o novo símbolo na mesma. O campo de definição deverá ser feito "definido" caso o símbolo compareça no campo de rótulos, e "indefinido" no caso contrário.

O problema c) ocorre sempre que um símbolo apareça no campo dos rótulos, tendo aparecido anteriormente no programa como operando. Neste caso, tal símbolo já está na tabela, mas indefinido, desejando-se torná-lo definido, sendo para isso modificados o campo de definição e o campo de endereços. Caso o símbolo já tenha aparecido no campo dos rótulos, uma mensagem de erro deverá ser enviada pelo montador, pois isto significa uma tentativa de redefinição do símbolo, o que não é permitido.

As rotinas que manipulam a tabela de símbolos foram implementadas conforme a descrição a seguir.

1. Rotina de Pesquisa

A rotina TAB, de pesquisa na tabela de símbolos, tem como entrada o símbolo a pesquisar, compactado em duas palavras de oito bits. Sua função é a de varrer a tabela, comparando cada símbolo nela encontrado com o símbolo fornecido, e retornando com duas informações:

- a) se encontrou ou não o símbolo na tabela;
- b) apontador para a posição da tabela onde o símbolo foi encontrado, ou para a primeira posição livre da tabela, se o símbolo não foi encontrado.

2. Rotina de Inclusão

A rotina COLOC, de inclusão de um novo elemento na tabela de símbolos, limita-se a utilizar os resultados da roti-

na de pesquisa, que deve ser chamada previamente colocando o novo elemento na primeira posição livre de tabela de símbolos, e ligando o bit de indefinição do símbolo. Esta rotina fornece também uma mensagem de erro no caso de a tabela estar totalmente preenchida.

Em uma versão do montador com mais memória disponível, poder-se-ia incluir nesta rotina uma ordenação alfabética dos símbolos, para efeito de maior facilidade de utilização da listagem da tabela pelo usuário.

3. Rotina de Tratamento geral dos Símbolos

A rotina LABEL, de tratamento dos símbolos, tem por finalidade construir a tabela de símbolos no primeiro passo do montador. Para isso, analisa cada linha do programa fonte, extraíndo de mesmo todos os símbolos, classificando-os, calculando quando necessário seus endereços e montando com estas informações a tabela de símbolos. O procedimento é o seguinte:

- a) Varre-se a linha, extraíndo-se e compactando-se o rótulo, se este existir. A rotina de compactação se encarrega de detetar invalidade de rótulos;
- b) Chama-se a rotina TAB (de pesquisa) se o rótulo existir;
- c) Se o rótulo for encontrado na tabela, verifica-se se já estava definido, caso em que uma mensagem de erro é fornecida. Se não estava definido, atribui-se ao campo de endereços o valor do contador de instruções corrente e define-se o símbolo;
- d) Se o rótulo não foi encontrado, chama-se a rotina COLOC, definindo-se a seguir o símbolo;
- e) Tratado o campo dos rótulos, verifica-se se a instrução apresenta operando simbólico. Em caso

afirmativo, pesquisa-se este símbolo na tabela, e se não for encontrado, chama-se a rotina COLOC, que o acrescentará aos demais, com o bit de indefinição ligado.

4. Rotina de Listagem e Teste de Consistência da Tabela de Símbolos

A rotina KONSIST, que testa a consistência de tabela de símbolos, deve ser chamada quando fôr detetado o final físico do programa, isto é, o aparecimento do pseudo FIM. Sua função é a de analisar um a um os elementos de tabela de símbolos, fornecendo mensagem de erro sempre que algum símbolo da tabela aparecer com o bit de indefinição ligado. Além disso, a rotina testa se o programador deseja uma listagem da tabela de símbolos, fornecendo, em caso afirmativo, para cada elemento da tabela, uma saída impressa do seu nome, agora descompactado e convertido para o formato padrão de três caracteres, ao lado do endereço que consta da tabela, e das informações de relocação (estas apenas no caso de programa relocável). No final, é impresso o número de símbolos indefinidos - encontrados na tabela.

Todas estas rotinas são utilizadas no primeiro passo do montador. O segundo passo utiliza apenas a rotina TAB de pesquisa, e não altera o conteúdo da tabela de símbolos, obtendo dela apenas as informações colhidas durante o primeiro passo.

2.5.4. A Organização e a Manipulação da Tabela dos Mnemônicos

Definidos em 2.3 como símbolos de três caracteres, os mnemônicos ocupam, a exemplos dos rótulos, depois de compactados, duas palavras cada um.

Representando operações, comandos ou instruções, os mnemônicos podem ser classificados de acordo com a sua função. Assim pode-se agrupar os mnemônicos em conjuntos menores cujos elementos têm todos as mesmas características sintáticas e semânticas. Com isso, constatada a validade do mnemônico e des

coberto o grupo a que ele pertence, pode-se, por inspeção das características de seu grupo, decidir qual será a ação a tomar em relação ao respectivo campo de operandos ou na geração do código objeto.

Os grupos de mnemônicos, em número de sete, são os seguintes:

- referências à memória (grupo 1)
- imediatas (grupo 2)
- curtas sem operando (grupo 3)
- deslocamento e giros (grupo 4)
- entrada e saída (grupo 5)
- curtas com operando (grupo 6)
- pseudos (grupo 7)

Desta maneira, instruções do mesmo grupo terão tratamentos análogos, sendo tal tratamento diferente do que receberá uma instrução de outro grupo. Logo, a informação do grupo ao qual a instrução pertence é muito importante para as rotinas de tratamento de rótulos e para a montagem do código objeto.

Há diversas alternativas de construção das tabelas de mnemônicos, todas dirigidas para uma pesquisa rápida, e contendo maior ou menor número de informações sobre a ação do montador em relação ao mnemônico. Um método eficiente para uma busca rápida é organizar tabelas para busca logarítmica (refs. 7,9). Encontrado o mnemônico, pode-se ler na posição correspondente de uma outra tabela, as informações a ele relativas e que poderia incluir a informação sobre o número do grupo a que a instrução pertence, além de outros detalhes sobre a instrução específica, como por exemplo o seu código de operação, se a instrução precisa ou não de operando, se é longa ou curta, etc. Quanto mais informações esta segunda tabela contiver tanto maior será o espaço de memória por ela ocupado,

porém será mais rápido e mais simples obter todas as informações de que se necessita para o tratamento do programa fonte. Outros métodos mais complicados para pesquisa de mnemônicos existem , porém, sua grande maioria exige memória adicional para tabelas - auxiliares, para apontadores e mesmo para o algoritmo de busca, o que não os torna aconselháveis no presente caso.

A tabela de mnemônicos foi organizada para busca linear por motivos de simplicidade do algoritmo de busca. Assim, adotou-se a opção de obter as informações sobre a instrução na posição em que o mnemônico foi encontrado. Isso leva a uma economia de memória pois dispensa tabelas adicionais e algoritmos complicados.

Em vista disso e levando-se em conta que esta tabela não varia no decorrer do processamento, decidiu-se ordená-la de tal maneira que os grupos mais usados fossem encontrados mais rapidamente na busca linear. Assim, o primeiro grupo é das instruções de referência à memória, e o último, o das pseudos. Com isto diminui-se o tempo de busca de um mnemônico na tabela. Uma segunda tabela, a de apontadores do início das subtabelas de instruções do mesmo grupo, tem como finalidade estabelecer os limites de tais subtabelas, permitindo um cálculo simples do número da tabela a que o mnemônico pesquisado pertence. Obtido este número, indexa-se com ele uma terceira tabela, a dos atributos do grupo . Esta tabela contém as seguintes informações: PSEUDO ou não, LONGA ou curta, OPERANDO ou não, NUMÉRICO ou não, e NÚMERO da subtabela a que o mnemônico pertence. Estas informações foram organizadas em formato decodificado, isto é, cada bit representa direta e independentemente a variável a ele correspondente para maior facilidade de teste, e agrupadas em uma palavra para cada grupo de mnemônicos. Na fig. 2.5.4.1 está , esquematicamente,a organização de tabela de mnemônicos adotada.

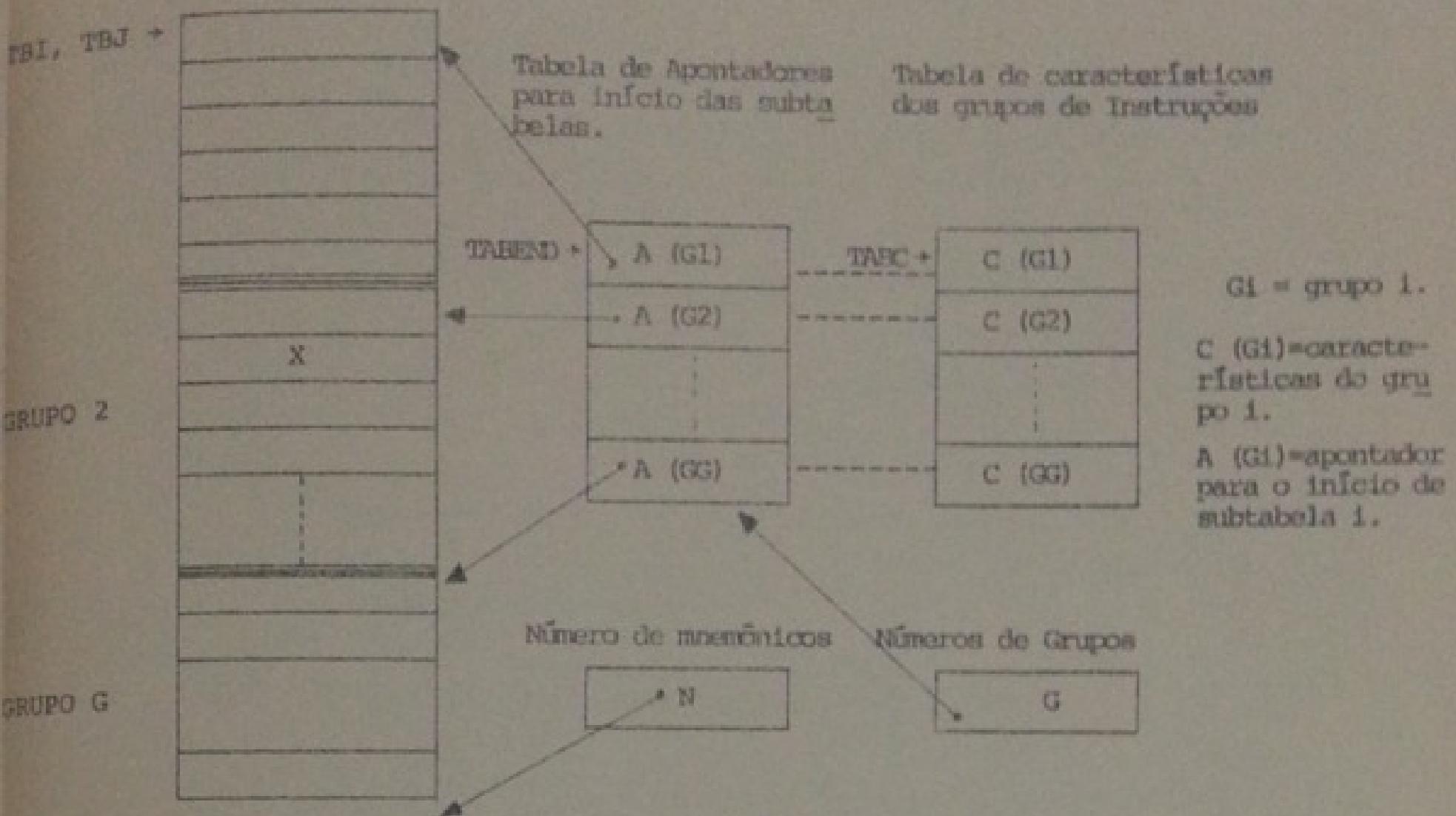


Fig. 2.5.4.1 - Organização da Tabela de mnemônicos

A rotina MNEM é a responsável pela pesquisa e identificação dos mnemônicos, constituindo a parte central do montador. No primeiro passo, sua função é basicamente de pesquisa e de detecção de mnemônicos inválidos e de auxiliar na construção de tabela de símbolos, mas no segundo passo sua participação se faz presente na geração do código objeto. Simplificadamente, o algoritmo utilizado para implementar a rotina MNEM é o seguinte (fig. 2.5.4.1):

- compara-se o mnemônico a pesquisar com cada um dos elementos da tabela de mnemônicos, até que ele seja encontrado na posição i da tabela ou então que a tabela seja esgotada (neste último caso, uma mensagem de erro é fornecida e o controle é devolvido ao programa que chamou a rotina MNEM);
- verifica-se, por comparação com os elementos da tabela de apontadores TABEND, a que intervalo, ou seja, a qual subtabela, o mnemônico pertence. Isto é feito pro-

curando-se TABEND (j) e TABEND (j+1) tais que TABEND (j) \leq i $<$ TABEND (j+1), caso que o grupo (subtabela) ao qual o mnemônico pertence será j;

- c) As informações a respeito do grupo j de instruções, ao qual o mnemônico em questão pertence, são obtidas imediatamente na posição j da tabela TABC de características dos diversos grupos de instruções.

Observe-se que a tabela de mnemônicos é subdividida em duas tabelas, TBI e TBJ porque cada elemento da mesma ocupa 16 bits (a exemplo dos rótulos).

2.5.5. A Representação Interna das Constantes

Como foi visto em 2.3, as constantes podem assumir diversas representações externas:

- com ou sem sinal
- hexadecimal, decimal ou ASCII

Para transformar constantes, escritas em uma das possíveis seis combinações acima, para o formato binário interno em complemento de dois, foi elaborada uma rotina de conversão geral, (CONVERT) que identifica o tipo da constante e em seguida chama rotinas particulares de conversão, uma para cada tipo de dado. As rotinas DECBIN e HEXBIN, que convertem constantes sem sinal, escritas em decimal e hexadecimal, respectivamente, para binário, utilizam-se da definição da representação de um número em uma base dada:

Sendo A_n, A_{n-1}, \dots, A_0 os dígitos que representam um número N (sem sinal) na base B, a representação de N na base B será $A_n A_{n-1} \dots A_1 A_0$ e o valor de N será dado por

$$N = \sum_{i=0}^{B-1} B^i A_i$$

Desta fórmula decorre imediatamente o algoritmo de conversão implementado:

- a) Posicionar N em zero; Apontar para A_n
- b) Se o dígito mais à direita (A_0) já foi manipulado, terminou a conversão (isto é verificado analisando-se o carácter que segue o último dígito tratado, verificando-se se o mesmo é ainda um dígito ou não)
- c) Caso contrário, $N \leftarrow N + B \times N + \text{dígito apontado}$
- d) Voltar para b) apontando o próximo dígito

Como não se tem instrução de multiplicação, e como os dados a converter normalmente são endereços, ocupando portanto duas palavras, foi necessário implementar pequenas rotinas específicas para efetuar multiplicação por 16 e por 16, e soma em dupla precisão. Foram usados tais algoritmos específicos no lugar de uma rotina geral de multiplicação por serem muito mais econômicas e eficientes para esta aplicação particular (único ponto do montador onde se faz cálculos aritméticos). Multiplicar por 16 equivale, em aritmética binária de complemento de 2, a deslocar o número para a esquerda de 4 bits. Multiplicar por 16 equivale a multiplicar por 8 e somar o resultado ao dobro do número. Como multiplicar por 2 é o mesmo que deslocar para a esquerda de 1 bit, e multiplicar por 8, deslocar de 3 bits, então uma rotina DED de deslocamento à esquerda em dupla precisão, chamada várias vezes na sequência conveniente, e associada à rotina SOMD de soma em dupla precisão é suficiente para que se possa implementar a subrotina MDD de multiplicação por 16 em dupla precisão. O uso das rotinas MDD, SOMD, e DED permitem a construção das rotinas de conversão DECBIN e HEXBIN, que implementam o algoritmo descrito. Executada a conversão, retorna-se para a rotina CONVERT, a qual se encarrega de acertar o sinal do número convertido.

Deve-se observar que a constante decimal ou hexadecimai vem escrita em caracteres ASCII, sendo necessário portanto uma pequena conversão prévia para BCD - (binary-coded decimal) dígito por dígito. Mensagens de erro são fornecidas sempre que for detetado algum dígito ilegal.

No caso de a constante a converter ser ASCII, não haverá necessidade de manipulação das informações da linha, exceto quanto ao sinal. Por isso não foi escrita uma subrotina para o tratamento de tais constantes, sendo este tratamento executado na própria rotina CONVERT.

As informações em formato interno, apesar de convenientes para o computador, não o são para o usuário. Por isso, ao lado das rotinas de conversão do formato externo para o interno, deve-se ter rotinas que fazem a conversão oposta. No caso, foi implementada uma rotina de conversão binário-hexadecimal CONVHEX, a qual converte um número binário de 8 bits para caracteres hexadecimais, imprimindo-os no dispositivo de saída. Esta rotina é utilizada nas listagens do código objeto, da tabela de símbolos e dos endereços associados às instruções, na listagem do programa fonte.

Os algoritmos de conversão binário-hexa e binário-decimal podem ser resumidos na seguinte descrição :

Seja B a base em questão, e N o número binário a converter. Suponha-se conhecido "a priori" que o número N vai ter no máximo $n+1$ dígitos na base B. Nestas condições:

- a) Posicionar i + n
- b) Fazer $A_i = \text{int}(N/B^i)$ obtendo A_i em BCD
- c) Fazer $N = N - A_i \times B^i$
- d) Fazer $i = i - 1$. Se i for negativo, passar para o passo e; caso contrário, voltar para o item b.
- e) Converter os A_i ($i = 0, 1, 2, \dots, n$) para ASCII e imprimir.

Levando em conta que não se dispõe de instruções de multiplicação nem de divisão, vê-se que não é possível a implantação imediata do algoritmo. Por isso, foi usado o recurso de guardar B^1 em tabelas, e de efetuar a divisão por subtrações sucessivas de B^1 . Para subtrair em dupla precisão, utilizou-se a rotina SOMD de soma em dupla precisão, associada com uma rotina de complementação de dois em dupla precisão, CMPD. Esse processo, apesar de demorado e ineficiente, permite a implementação do algoritmo em um espaço de memória bastante reduzido, o que não ocorreria no caso de se programar um algoritmo geral de divisão, que seria utilizado apenas neste ponto do montador.

2.5.6. As Pseudo Instruções

O conjunto das pseudo instruções disponíveis num montador define uma grande parte da versatilidade do mesmo. Pseudo-instruções têm, via de regra, tratamento individual, o que exige uma disponibilidade razoável de memória no caso de se desejar dotar o montador de muitos recursos adicionais. Por essa razão, decidiu-se implementar, na primeira versão do montador, apenas as indispensáveis, acrescentando-se novas pseudo instruções na medida do possível, tendo sempre em vista a quantidade de memória disponível e a vantagem da adição da nova pseudo instrução. Dessa maneira, a primeira versão do montador apresentou apenas as pseudo instruções ORG (para definir a origem do programa), DEFC (para definir uma constante), BLOC (para definir área de dados) e FIM (para definir o final físico do programa fonte). Além disso, tais pseudo instruções tinham formato inflexível para facilidade de implantação rápida. Assim, ORG admitia apenas operandos constantes sem sinal, BLOC admitia apenas operando decimal e FIM servia apenas para finalizar o programa. Implantada a primeira versão do montador, obteve-se com ele um recurso extra para a programação, que fôra até este ponto feita totalmente em linguagem de máquina. Passou-se a escrever o restante do montador na sua própria linguagem, conseguindo-se com isto reescrever e melhorar uma grande parte do mesmo, dando-lhe característica de montador relocável , e também acrescentando-lhe novos recursos, entre os quais as novas pseudo instruções EQU, NOME, ENT, EXT, DEFE, DEFT, COM, SEG, SUBR.

Detectada a presença de uma pseudo instrução, desvia-se para a respectiva rotina de tratamento. A seguir, descreve-se o funcionamento e o tratamento de cada pseudo instrução implementada.

2.5.6.1. A Pseudo ORG

Em um programa escrito em linguagem de baixo nível, isto é, linguagem na qual o programador precisa preocupar-se com os detalhes da máquina com a qual está trabalhando, há necessidade de se informar ao montador qual a posição de memória a partir da qual se deseja que o programa seja montado. Evidentemente é disso que depende a informação necessária à montagem das instruções de referência à memória, uma vez que o endereçamento é simbólico e que a cada símbolo é associada um posição de memória.

Portanto, é necessário que se estabeleça em qual posição de memória se deve iniciar a montagem. Isto pode ser implementado facilmente com a escolha "a priori" de uma posição fixa de memória, e criando-se um apontador para esta posição no início da execução do montador. Esta solução é, entretanto, inflexível, exigindo para a modificação de uma posição de memória qualquer, que se monte novamente todo o programa.

Contornar-se esta situação com facilidade criando-se uma pseudo instrução que modifica o apontador para a posição de memória onde se deseja montar o código. Esta pseudo operação, a de ORG do código, poderia ter outras utilidades que não a de apenas definir a posição da primeira palavra do código. Por exemplo, reservar posições de memória para área de trabalho, sobrepor um novo código a outro já montado, etc. Para isto, é desejável que o operando desta pseudo instrução seja o mais flexível possível. Procurou-se por isso generalizá-lo como uma referência qualquer à memória, evidentemente com algumas restrições, que serão vistas adiante.

Deverá ser o primeiro comando no caso de um programa absoluto. Isto ocorre porque é necessário que se estabeleça de inicio se o programa vai ser absoluto ou relocável e, caso seja absoluto, deve-se indicar a partir de qual posição de memória se deseja montá-lo. Observe-se que nestas circunstâncias é necessário que o operando indique uma posição absoluta de memória, conhecida, não podendo ser, portanto, referência simbólica

pura ou relativa, nem uma referência local, uma vez que, sendo este o primeiro comando do programa, tais referências estariam indefinidas.

Uma variante para a opção de usar ORG como 1º comando seria estabelecer um valor padrão "default", a ser utilizado no caso da não especificação explícita da origem. Neste caso, haveria a necessidade de criar uma nova pseudo para indicar apenas se o programa é absoluto ou relocável. Novamente, ter-se-ia a opção de adotar um "default" que admitisse ser o programa relocável (ou absoluto com uma origem fixa), no caso da omissão de tal pseudo.

Ainda no caso de um programa absoluto, é interessante que se possa definir, a qualquer momento no programa, que o trecho a seguir deve ser montado a partir de uma nova origem, mediante a utilização da pseudo ORG. Neste caso, pode-se admitir referências absolutas, relativas, simbólicas e locais, uma vez que é possível neste caso, que tais referências tenham sido definidas anteriormente. Deve-se notar que, caso a referência não seja absoluta, deverá ela relacionar-se obrigatoriamente com posições já definidas de memória, pois se isto não ocorrer, torna-se, embora possível, bastante complicada a manipulação desta pseudo, da tabela de símbolos e da geração do código, o que torna inviável a elaboração de tal tratamento nas atuais condições. Sendo totalmente dispensável na grande maioria dos casos, este recurso foi abandonado.

Num programa relocável, é conveniente que se possa definir uma nova origem para o código a qualquer altura do programa. Analogamente ao caso do programa absoluto, tal origem definida pela pseudo instrução ORG deve ser, neste caso, uma posição conhecida de memória, relativa ao início do programa.

Sendo sempre relativa a origem permitida em programas relocáveis, a pseudo ORG não poderá ser o primeiro comando de tais programas, pois sendo obrigatoriamente relocável (simbólico, relativo ou local), seu operando não teria sentido, pela não ocorrência prévia da definição de tais referências.

2.5.6.2. As pseudos NOME, SUBR, SEGM

Quando se trabalha com programas relocáveis, é comum que um mesmo programa ou subrotina tenha mais de um ponto de acesso global. Assim, se o programa fôr parte de um arquivo de programas relocáveis, como é o caso de uma biblioteca de subrotinas, é conveniente que ele tenha um nome pelo qual possa ser identificado. Tal nome teria, além disso, a função de representar simbolicamente o endereço da primeira palavra do código objeto do programa, endereço este em relação ao qual as referências relocáveis seriam relativas.

Para definir o nome de um programa relocável, criou-se as pseudos SUBR, NOME e SEGM cujas características são as seguintes:

1. São mutuamente exclusivas, pois, definido o tipo de programa, e sendo este único, a ocorrência de mais de uma destas pseudos no mesmo programa será uma tentativa de redefinição do tipo do mesmo;
2. Não deverão estar presentes em programas absolutos, pois sua ocorrência, além de atribuir um nome à origem relocável do programa, serve também para indicar ao montador que o programa em questão é relocável;
3. Deverão ser o primeiro comando de um programa relocável, para forçar que o programa tenha uma identificação.

(Uma alternativa para esta imposição seria a de permitir que um nome "default" pudesse ser atribuído ao programa. Esta possibilidade pode criar problemas, no caso de existir um sistema operacional no qual o programa objeto faça parte na qualidade de arquivo. A tentativa de dois usuários diferentes guardarem dois arquivos relocáveis com o mesmo nome poderia criar confusões para os próprios usuários ou para o operador. Assim, seria conveniente que o sistema operacional pudesse atribuir, ao nome do

arquivo, não apenas o nome fornecido pelo usuário, mas também alguma identificação que ligue o arquivo ao usuário que o gerou).

4. Seu operando, deverá ser somente do tipo simbólico puro, pois não teria sentido como nome de arquivo se não o fosse.

A) - A Pseudo "NOME"

O funcionamento da pseudo NOME é o seguinte:

1. No passo 1, o resultado do tratamento recebido por esta pseudo será o de criar um elemento da tabela de símbolos, correspondente ao nome do programa. Além disso, sua presença caracterizará o programa como relocável.
2. No passo 2, seu tratamento será o de gerar na fita objeto, relocável, um bloco de NOME, no qual estão resumidas as características do programa, tais como comprimento, tamanho da área de variáveis comuns a diversos programas e informação sobre o fato de o código ser um programa principal. Estas informações não são todas provenientes do operando da pseudo NOME, mas sim coletadas durante a execução de todo o primeiro passo do montador, para utilização nesta ocasião.

Poder-se-ia juntar algumas destas informações no operando da pseudo NOME. Por exemplo, além do nome do programa, poder-se-ia declarar ali algumas outras informações como por exemplo o tamanho da área de variáveis comuns, o fato de usar ou não rotinas do sistema para o atendimento de interrupção, o tipo do programa (subrotina, programa principal, segmento) etc.

Por motivos de simplicidade de tratamento, resolveu-se manter a política de coletar tais dados durante o primeiro passo, e criar outras pseudos para identificar subrotinas e segmentos.