

CINTHIA O. DE A. FREITAS



**USO DE MODELOS ESCONDIDOS DE MARKOV PARA
RECONHECIMENTO DE PALAVRAS MANUSCRITAS**

Tese apresentada à Pontifícia Universidade
Católica do Paraná para a obtenção do título
de Doutor em Ciências.

Curitiba
2001

CINTHIA O. DE A. FREITAS



USO DE MODELOS ESCONDIDOS DE MARKOV PARA RECONHECIMENTO DE PALAVRAS MANUSCRITAS

Tese apresentada à Pontifícia Universidade
Católica do Paraná para a obtenção do título
de Doutor em Ciências.

Área de Concentração: Processamento e
Análise de Documentos

Orientador: Prof. Dr. Flávio Bortolozzi
Co-orientador: Prof. Dr. Robert Sabourin

Curitiba
2001

Freitas, Cinthia Obladen de Almendra

Uso de Modelos Escondidos de Markov para Reconhecimento de Palavras Manuscritas. Curitiba, 2001.

188p.

Tese(Doutorado) – Pontifícia Universidade Católica do Paraná. Programa de Pós-Graduação em Informática Aplicada.

1. Documentos manuscritos 2.Reconhecimento de palavras manuscritas 3. Modelos Escondidos de Markov (MEM) 4. Cheques bancários. I.Pontifícia Universidade Católica do Paraná. Centro de Ciências Exatas e de Tecnologia. Programa de Pós-Graduação em Informática Aplicada II-t



ATA DA SESSÃO PÚBLICA DE DEFESA DE TESE DE DOUTORADO
 DO PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA APLICADA
 DA PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ

DEFESA DE TESE Nº 002

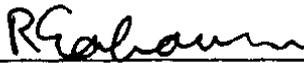
Aos 05 dias do mês de junho de 2001 realizou-se a sessão pública de defesa da tese "Uso de Modelos Escondidos de Markov para Reconhecimento de Palavras Manuscritas", apresentada por Cinthia Obladen de Almendra Freitas como requisito parcial para a obtenção do título de **Doutor em Ciências**, perante uma Banca Examinadora composta pelos seguintes membros:

Prof. Dr. Flávio Bortolozzi
 PUCPR (Presidente)


 assinatura

APROVADO
 parecer

Prof. Dr. Robert Sabourin
 ÉTS Montréal


 assinatura

Aprovado
 parecer

Prof. Dr. Jacques Facon
 PUCPR


 assinatura

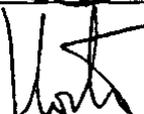
Aprovado
 parecer

Prof. Dr. Antonio N. Santa Rosa
 UnB


 assinatura

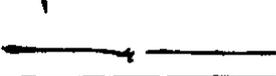
Aprovado
 parecer

Prof. Dr. Luciano da F. Costa
 USP - São Carlos


 assinatura

APROVADO
 parecer

Prof. Dr. Dibio Leandro Borges
 PUCPR (suplente)


 assinatura

APROVADO
 parecer

Prof. Dr. Edson Scalabrin
 PUCPR (suplente)


 assinatura

Aprovado
 parecer

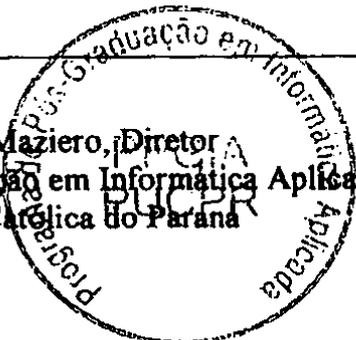
Conforme as normas regimentais do PPGIA e da PUCPR, o trabalho apresentado foi considerado APROVADO SEM RESTRIÇÕES (aprovado sem restrições, aprovado com exigências ou reprovado), segundo a avaliação da maioria dos membros da Banca Examinadora acima indicada.

Observações, exigências e/ou restrições da Banca Examinadora, quando houver:

NÃO HÁ

continuar no verso, se necessário

Prof. Dr. Carlos Alberto Maziero, Diretor
 Programa de Pós-Graduação em Informática Aplicada
 Pontifícia Universidade Católica do Paraná





Ao meu marido José Junior e
aos meus filhos Jessica e José Filho.

Agradecimentos

Ao amigo e orientador Prof. Dr. Flávio Bortolozzi pela orientação segura e constante incentivo.

Ao Prof. Dr. Robert Sabourin pelas importantes contribuições dadas à realização deste trabalho.

Ao Prof. Dr. Abdenaïm El Yacoubi pelos esclarecimentos sobre Modelos Escondidos de Markov.

À Pontifícia Universidade Católica do Paraná – PUCPR pelo apoio recebido durante a realização deste trabalho.

Aos meus pais, Ziloá e Nicolau, pelo carinho e apoio nos momentos difíceis.

Ao meu querido marido José Junior pelo amor e incansável compreensão.

Aos meus queridos filhos, Jessica e José Filho, pelos risos e brincadeiras.

Aos amigos de jornada, Justino e Alceu, pela amizade sincera e constantes incentivos.

Aos amigos de jornada Marisa, Luiz e José Eduardo, pelo sentimento de juventude.

Aos ex-alunos, Glauco e Anderson, pelo respeito e por seguirem alguns passos já trilhados por mim.

A todos que direta ou indiretamente colaboraram na execução deste trabalho.

Sumário

Agradecimentos	ii
Sumário	iii
Lista de Figuras	vii
Lista de Tabelas	xi
Lista de Abreviaturas e Símbolos	xiii
Resumo	xv
Abstract	xvii
Capítulo 1	
Introdução	1
1.1. Desafio	2
1.2. Motivação.....	3
1.3. Proposta.....	3
1.4. Contribuição.....	6
1.5. Organização.....	6
Capítulo 2	
Palavras Manuscritas	8
2.1. O Problema: Palavras Manuscritas em Cheques Bancários Brasileiros	9
2.2. Escrita, Conhecimento Prévio e Leitura.....	14
2.3. Leitura e Reconhecimento.....	22
2.4. Reconhecimento Estático de Palavras Manuscritas	27
2.4.1. Restrições dos Métodos de Reconhecimento.....	27
2.4.1.1. Tipo de Aquisição da Imagem	28
2.4.1.2. Tipos de Escrita.....	29
2.4.1.3. Números de Escritores	29
2.4.1.4. Tamanho do Léxico.....	30

2.4.1.5. Qualidade da Imagem	31
2.4.1.6. Fatores Humanos.....	33
2.4.1.7. Outros Fatores	34
2.4.1.8. Características da Aplicação	35
2.4.2. Etapas de um Método de Reconhecimento	36
2.4.2.1. Aquisição Imagem	36
2.4.2.2. Pré-Processamento	37
2.4.2.3. Estratégias de Segmentação	44
2.4.2.4. Extração de Primitivas	46
2.4.2.5. Metodologias para Reconhecimento	55
2.4.3. MEM para Reconhecimento de Palavras	60
2.4.4. Discussão	65

Capítulo 3

Base de Dados

72

3.1. Bases de Dados	72
3.2. Criação da Base de Dados de Laboratório de Cheques Bancários Brasileiros	75
3.2.1. Critérios de Estabelecimento da Base de Dados	76
3.2.2. Vocabulário de Cheques Bancários	77
3.2.3. Geração dos Valores dos Cheques Bancários	78
3.2.4. Estatística Prévia da Base de Dados	79
3.2.5. Processo de Coleta dos Cheques de Laboratório	81
3.2.6. Armazenamento e Recuperação das Informações	83
3.2.7. Características da Base de Dados de Treinamento	86
3.2.7.1. Estilo da Escrita.....	86
3.2.7.2. Símbolos Iniciais e Terminais do Extenso dos Cheques	87
3.2.8. Considerações Finais.....	88

Capítulo 4

Um Método para Reconhecimento de Palavras Manuscritas do Extenso de Cheques Bancários Brasileiros

89

4.1. Pré-Processamento das Palavras Manuscritas.....	90
4.1.1. Procedimentos Preliminares.....	90
4.1.2. Segmentação do Extenso Manuscrito em Palavras.....	91
4.1.3. Correção da Inclinação Vertical da Escrita.....	95
4.1.4. Detecção do Corpo da Palavra	97
4.2. Primitivas Perceptivas	99
4.2.1. Conjunto 01 - Primitivas Perceptivas.....	100
4.2.2. Conjunto 02 - Primitivas Perceptivas, Concavidades e Convexidades.....	105

4.2.3. Conjunto 03 - Primitivas Perceptivas, Concavidades e Convexidades Rotuladas.....	107
4.2.4. Escolha de Alfabetos de Grafemas	109
4.2.4.1. O que é um Grafema?.....	110
4.2.4.2. Escolha de uma Medida de Semelhança	111
4.2.4.3. Entropia de Shannon e Informação Mútua.....	111
4.2.4.4. Alfabetos Hierárquicos.....	114
4.2.4.5. Validação dos Alfabetos Hierárquicos através da Informação Mútua.....	116
4.2.5. Reconhecimento de Palavras Manuscritas do Extenso de Cheques Brasileiros.....	118
4.2.5.1. Fundamentação Teórica dos Modelos Escondidos de Markov.	119
4.2.5.2. Condições Iniciais	123
4.2.5.3. O Treinamento: Algoritmo <i>Baum-Welch</i>	125
4.2.5.4. A Validação do Treinamento: Algoritmo de Validação Cruzada.	132
4.2.5.5. O Reconhecimento: Algoritmo <i>Forward</i>	133
4.2.5.6. Vantagens e Desvantagens dos MEM.....	135
4.2.5.7. Experimentos Realizados	137
4.2.5.8. Análise de Erros	140
4.3. Primitivas Geométricas	150
4.3.1. Vetor de Características	151
4.3.2. Método de Classificação: Quantização Vetorial.....	153
4.3.3. Reconhecimento das Palavras Manuscritas com MEM.....	156
Capítulo 5	
Conclusão	159
Referências Bibliográficas	162
Apêndice A	
Estrutura dos Cheques Bancários Brasileiros	170
A.1. Modelos para Cheques: Distribuição e Preenchimento dos Campos.....	170
A.2. Dimensões do Cheque.....	170
A.3. Elementos Componentes	171
A.4. Formulários dos Cheques.....	173
A.5. Características do Valor Numérico por Extenso.....	174
Apêndice B	
Comparativo de Métodos de Reconhecimento de Palavras Manuscritas	175
B.1. Bases de Dados.....	176

B.2. Pré-Processamento das Palavras.....	179
B.3. Extração de Primitivas.....	182
B.4. Metodologias de Reconhecimento	182
B.5. Resultados Obtidos.....	183
B.6. Análise dos Resultados.....	187

Lista de Figuras

Figura 1.1	Exemplos de estilos de letras.....	2
Figura 1.2	Esquema do reconhecimento de palavras manuscritas.....	5
Figura 2.1	Valor numérico e palavras-chaves.....	9
Figura 2.2	Classes de palavras do léxico em português.....	10
Figura 2.3	Exemplos de escrita: a) cantos Ojibwa, b) catecismo asteca, c) inscrições antigas [CAGLIARI, 1991].....	15
Figura 2.4	Exemplos de escrita ideográfica: escrita egípcia [CAGLIARI, 1991].....	15
Figura 2.5	Estilo mistral: enlace entre as diversas letras [SOLO, 1989].....	16
Figura 2.6	Exemplos da letra “a” em diferentes alfabetos [CAGLIARI, 1991].....	17
Figura 2.7	Variações com base no traçado [CAVANAGH, 1974].	18
Figura 2.8	Estilos de escrita cursiva [CAVANAGH, 1974] e [SOLO, 1989]	19
Figura 2.9	Alfabeto e estilo de escrita para alfabetização no Brasil.....	19
Figura 2.10	Verificação do princípio de fechamento [MACHADO & TEIXEIRA, 1996]	21
Figura 2.11	Tipos de escrita [TAPPERT, 1990]	29
Figura 2.12	Exemplos de ruídos gerados no processo de aquisição da imagem: a) ruído “salpicado”, b) ruído gerado pelo alimentador, c) incorreta operação do <i>scanner</i> [KIM, 1996].....	31
Figura 2.13	Exemplos da qualidade de imagem associado ao meio utilizado: a) papel vegetal e b) papel sulfite branco.....	32
Figura 2.14	Exemplos de processo de binarização: a) imagem original em níveis de cinza, b) <i>threshold</i> = 217 (método anisotropia) c) <i>threshold</i> = 110 (método OTSU).....	32
Figura 2.15	Inclinação: a) da linha de base (<i>skew</i>) b) vertical dos caracteres (<i>slant</i>)	33
Figura 2.16	Exemplos de: a) traços contíguos, b) traços desconectados, c) posição errônea.....	33
Figura 2.17	Erros de ortografia: a) AE0017a e b) AE0233a	34
Figura 2.18	Ortografia diferenciada para uma mesma palavra.....	34
Figura 2.19	Correção da inclinação vertical dos caracteres: a) imagem original, b) remoção das linhas horizontais de percurso longo, c) remoção das faixas de pequeno comprimento, d) janelas isoladas e faixas conectando o centro de gravidade e ângulos de inclinação mostrados a cada janela, e) imagem corrigida [BOZINOVIC & SRIHARI, 1989]	39
Figura 2.20	Correção através da aplicação do histograma de projeção vertical: a) imagem limiarizada a partir da imagem original, b)	

	histograma de projeção vertical a partir da imagem contida em (a), c) imagem corrigida, d) histograma de projeção vertical da imagem contida em (c) [BUSE et al., 1997].....	40
Figura 2.21	Transformação de uma imagem com inclinação de 60 graus [GUILLEVIC, 1995].....	40
Figura 2.22	Exemplo de falha na correção da inclinação vertical dos caracteres [GUILLEVIC, 1995].....	40
Figura 2.23	Correção da inclinação dos caracteres: a) imagem original, b) pixels de bordas não horizontal, c) imagem corrigida [KIMURA et al., 1993].....	40
Figura 2.24	Correção vertical da inclinação dos caracteres: a) imagem original, b) contorno, c) e d) linhas verticais inferiores e superiores, e) imagem corrigida com base no ângulo estimado a partir das linhas detectadas em (c) e (d) [KIM & GOVINDARAJU, 1997].....	41
Figura 2.25	Correção da inclinação vertical: a) histograma de projeção vertical das duas retas de ângulos 30 graus e 0 graus, b) correção da inclinação [EL YACOUBI, 1996].....	41
Figura 2.26	Determinação das linhas de referência do corpo das minúsculas [CÔTE, 1997].....	43
Figura 2.27	Estimação das linhas de referência do corpo das minúsculas [BERCU & LORETTE, 1993].....	43
Figura 2.28	Normalização do corpo das minúsculas: a) imagem original, b) detecção dos máximos locais do contorno superior, c) filtragem dos máximos indesejáveis, d) imagem normalizada utilizando uma transformação não linear [EL YACOUBI, 1996].....	44
Figura 2.29	Versões transformadas do dígito “5”: a) imagem original, b) rotação, c) escala, d) estiramento, e) inclinação, f) correção da inclinação, g) espelhamento [TRIER et al., 1996].....	48
Figura 2.30	Estágios da extração de primitivas: a) esqueleto com <i>grid</i> , b) representação dos segmentos de linha parametrizados, c) primitivas extraídas [SENIOR & ROBINSON, 1998].....	49
Figura 2.31	Regularidades e singularidades de uma palavra [CÔTE, 1997]....	50
Figura 2.32	Alfabeto de 12 traços [AVILA, 1996].....	51
Figura 2.33	Exemplos de descrição de palavras através do alfabeto de 12 traços [AVILA, 1996].....	51
Figura 2.34	Vales e perfil superior [LECOLINET, 1990].....	52
Figura 2.35	Primitivas com base em 8 traços [DZUBA et al., 1998].....	53
Figura 2.36	Diagrama de um método de reconhecimento de formas.....	56
Figura 2.37	MEM para o caracter [EL YACOUBI, 1996].....	61
Figura 2.38	MEM para o espaçamento inter-palavras [EL YACOUBI, 1996].....	62
Figura 2.39	Modelo global para o prefixo “rue” [EL YACOUBI et al., 1994]	62
Figura 2.40	Métodos de análise de uma palavra manuscrita: a) caracter-caracter, b) caracter-ligação-caracter [CHO et al., 1995].....	63
Figura 2.41	Topologias dos modelos: a) caracter-caracter, b) caracter-ligação-caracter [CHO et al., 1995].....	64

Figura 2.42	MEM para palavras de cheques franceses [KNERR et al., 1996].....	64
Figura 2.43	MEM para palavras [GILLOUX, 1994].....	64
Figura 2.44	Modelo para as letras cursivas [GILLIES, 1992].....	65
Figura 2.45	Variáveis e metodologias no reconhecimento de palavras manuscritas.....	70
Figura 3.1	Exemplo de cheque de laboratório preenchido.....	83
Figura 3.2	Exemplares de estilos de valores por extenso.....	84
Figura 3.3	Esquema lógico da codificação dos cheques de laboratório.....	85
Figura 3.4	Exemplos de símbolos: a) inicial e b) terminal.....	88
Figura 4.1	Exemplos de filtragem <i>soft</i> : a,c,e) imagens originais, b,d,f) imagens filtradas.....	91
Figura 4.2	Exemplos de espaçamentos intra e entre palavras.....	92
Figura 4.3	Exemplos de correção da inclinação vertical dos caracteres: a) correção de 40 graus, b) correção de 23 graus, c) correção de 58 graus, d) imagem extraída de [GUILLEVIC, 1995] com 7 graus de inclinação (ver resultado obtido em [GUILLEVIC, 1995] na Figura 2.22), e) correção de 29 graus.....	96
Figura 4.4	Pré-processamento das palavras: a)imagens originais, b)correção da inclinação vertical dos caracteres e suavização do contorno.....	97
Figura 4.5	Zonas de uma palavra.....	97
Figura 4.6	Determinação das zonas de palavras "quatrocentos".....	98
Figura 4.7	Extração de primitivas: conjunto 01 - PP - primitivas perceptivas.....	104
Figura 4.8	Pseudo-segmentação das palavras.....	104
Figura 4.9	Extração de primitivas: conjunto 02 - PPCC.....	106
Figura 4.10	Extração de primitivas: conjunto 03 - PPCCR - primitivas perceptivas, concavidades e convexidades rotuladas.....	108
Figura 4.11	Tipos de estrutura dos modelos de Markov: a) modelo sem restrições, b) modelo sequencial, c) modelo paralelo [EL YACOUBI, 1996]	121
Figura 4.12	Modelo de estados e modelo de arcos: a) modelo de estados e b) modelo de arcos [EL YACOUBI, 1996].....	122
Figura 4.13	Topologia MEM para cada uma das palavras.....	130
Figura 4.14	Reconhecimento de palavras isoladas por MEM [RABINER & JUANG, 1983].....	135
Figura 4.15	Número de estados dos modelos das palavras.....	137
Figura 4.16	Exemplos de confusão no reconhecimento de palavras - conjunto PPCCR.....	149
Figura 4.17	Exemplo da rotulação do corpo da palavra para primitivas geométricas.....	152
Figura 4.18	Distribuição de frequência dos rótulos.....	153
Figura 4.19	Fronteiras para dois símbolos compostos cada qual por duas amostras.....	155
Figura 4.20	Esquema de reconhecimento: primitivas perceptivas e geométricas.....	157
Figura A.1	Modelo para distribuição das informações [KOERICH, 1997].....	171

Figura A.2	Modelo padrão de preenchimento dos campos e áreas de um cheque [OLIVEIRA, 1998].....	171
Figura A.3	Dimensões mínima e máxima de cheques bancários.....	172
Figura A.4	Faixas de um cheque bancário brasileiro [OLIVEIRA, 1998].....	173

Lista de Tabelas

Tabela 2.1	Métodos de Correção da Inclinação Vertical.....	39
Tabela 2.2	Métodos de Correção da Inclinação Vertical.....	42
Tabela 2.3	Extração de Primitivas.....	47
Tabela 3.1	Características e aplicações de trabalhos com bases de dados existentes.....	74
Tabela 3.2	Vocabulário de cheques bancários para o domínio estudado.....	77
Tabela 3.3	Geração aleatória dos valores numéricos para a base de treinamento.....	79
Tabela 3.4	Total de dígitos aleatórios gerados.....	79
Tabela 3.5	Exemplos de valores gerados.....	80
Tabela 3.6	Exemplos de valores numéricos por extensos.....	80
Tabela 3.7	Contagem dos caracteres.....	80
Tabela 3.8	Contagem das palavras.....	81
Tabela 3.9	Exemplos da codificação dos cheques de laboratório.....	85
Tabela 3.10	Distribuição dos tipos de escrita na base de treinamento.....	87
Tabela 4.1	Conjunto de Primitivas 01 – PP – Primitivas Perceptivas.....	100
Tabela 4.2	Conjunto de Primitivas 02 – PPCC – Primitivas Perceptivas, Concavidades e Convexidades.....	106
Tabela 4.3	Conjunto de Primitivas 03 – PPCCR – Primitivas Perceptivas, Concavidades e Convexidades Rotuladas.....	107
Tabela 4.4	Alfabeto Hierárquico – PP – Primitivas Perceptivas.....	115
Tabela 4.5	Alfabeto Hierárquico PPCC - Primitivas Perceptivas + Côncavo e Convexo.....	115
Tabela 4.6	Alfabeto Hierárquico PPCCR - Primitivas Perceptivas + Concavidades e Convexidades Rotuladas.....	116
Tabela 4.7	Validação das Fusões para o Conjunto de Primitivas PPCCR.....	118
Tabela 4.8	Conjuntos de Seqüências de Observações (¹ PC = palavras-chaves)	124
Tabela 4.9	Taxas de Reconhecimento das Palavras para os Conjuntos de Primitivas.....	138
Tabela 4.10	Taxas de Reconhecimento das Palavras para os Conjuntos de Primitivas Perceptivas.....	139
Tabela 4.11	Taxas de Reconhecimento das Palavras para os Conjuntos de Primitivas Perceptivas para Equação (4.12).....	140
Tabela 4.12	Matriz de Confusão para o Conjunto de Primitivas PP.....	141
Tabela 4.13	Matriz de Confusão para o Conjunto de Primitivas PPCC.....	142
Tabela 4.14	Matriz de Confusão para o Conjunto de Primitivas PPCCR.....	143
Tabela 4.15	Análise de Erros – PP.....	145
Tabela 4.16	Análise de Erros – PPCC.....	146
Tabela 4.17	Análise de Erros – PPCCR.....	147
Tabela 4.18	Rótulos do Vetor de Características.....	152

Tabela B.1a	Características das Bases de Dados – Palavras do Idioma Inglês.....	177
Tabela B.1b	Características das Bases de Dados – Palavras do Idioma Francês..	177
Tabela B.1c	Características das Bases de Dados – Palavras do Idioma Português.....	177
Tabela B.2	Léxico em Português, Inglês e Francês.....	178
Tabela B.3	Distribuição das Palavras nas Bases de dados de Treinamento.....	180
Tabela B.4	Pré-Processamento de Palavras Manuscritas.....	181
Tabela B.5	Extração de Primitivas.....	183
Tabela B.6	Metodologias de Reconhecimento.....	183
Tabela B.7	Resultados Obtidos – Idioma Francês.....	184
Tabela B.8	Resultados Obtidos – Idioma Inglês.....	186
Tabela B.9	Resumo dos Resultados Obtidos.....	188
Tabela B.10	Resultados de Sistemas de Reconhecimento.....	188

Lista de Abreviaturas e Símbolos

<i>bmp</i>	<i>bit map</i>
CDVDHMM	<i>Continuous Density Variable Duration Hidden Markov Models</i>
CEDAR	<i>Center of Excellence in Document Analysis Recognition</i>
CENPARMI	<i>Center of Pattern Analysis and Machine Intelligence</i>
<i>dpi</i>	<i>dots per inch</i>
ETS/Canada	<i>École de Technologie Supérieure – Montréal - Canada</i>
<i>jpeg</i>	<i>Joint Photographers Expert Group</i>
HMM	<i>Hidden Markov Models</i>
HWR	<i>Handwritten Word Recognition</i>
LARDOC	Laboratório de Análise e Reconhecimento de Imagens de Documentos
LRPC	Laboratório de Reconhecimento de Padrões e Redes de Computadores
MEM	Modelos Escondidos de Markov
MICR	<i>Magnetic Ink Character Recognition</i>
NIST	<i>National Institute Standards and Technology</i>
PP	Primitivas Perceptivas
PPCC	Primitivas Perceptivas, Concavidades e Convexidades
PPCCR	Primitivas Perceptivas, Concavidades e Convexidades Rotuladas
PUC-PR	Pontifícia Universidade Católica do Paraná
RPM	Reconhecimento de Palavras Manuscritas
<i>tiff</i>	<i>Tagged Image Format File</i>
k	Número de pixels vizinhos mais próximos
N	Número máximo de estados do modelo de <i>Markov</i>
X, Y	Variáveis aleatórias discretas
$H(P)$	Entropia de Shannon
$H(. .)$	Entropia condicional
$I_M(X, Y)$	Informação Mútua entre X e Y
$I_M(M, G_K)$	Informação Mútua para as palavras M e os grafemas G_K
$max(.)$	Função que retorna o máximo valor entre uma lista de valores
P_i	Probabilidade de ocorrência
$P(. .)$	Probabilidade condicional
$\lambda=(Q, V, A, B, \pi, T)$	Conjunto necessário na definição do modelo de <i>Markov</i>
Q	Conjunto de estados do modelo de <i>Markov</i>
V	Conjunto de símbolos observáveis
A	Matriz de transição entre os estados da cadeia de <i>Markov</i>
B	Matriz de probabilidade de observação de símbolos condicionada pelos estados da cadeia de <i>Markov</i>
π	Conjunto de densidade de probabilidade inicial
T	Comprimento da seqüência de observações

t	Intervalo de tempo
O	Seqüência de observações
$\hat{\lambda}$	Modelo estimado
$\operatorname{argmax}(\cdot)$	Função que retorna o índice relacionado ao argumento que maximiza uma função
α	Probabilidade <i>Forward</i>
β	Probabilidade <i>Backward</i>
i, j, k	Valores inteiros
$\log P(\cdot)$	Logaritmo da probabilidade
w	Número de símbolos ou <i>codebooks</i>
D_i	Distância Euclidiana mínima

Resumo

O presente trabalho apresenta o “Uso dos Modelos Escondidos de Markov para Reconhecimento de Palavras Manuscritas”.

Os processos que visam o tratamento automático de palavras manuscritas devem explorar os aspectos relevantes dos processos de escrita e leitura, associados com os mecanismos de conhecimento prévio desenvolvidos e realizados pelos seres humanos. Desta forma, apresenta-se uma revisão associativa destes processos e mecanismos com o objetivo principal deste trabalho, ou seja, responder as seguintes perguntas visando o reconhecimento de palavras manuscritas:

- Quais são as primitivas perceptivas relevantes no processo de reconhecimento de palavras manuscritas?
- Como representar palavras cursivas sem a presença de primitivas perceptivas?

Apresenta-se uma revisão bibliográfica do Reconhecimento de Palavras Manuscritas – RPM enfocando todas as etapas necessárias ao reconhecimento de palavras manuscritas.

Os Modelos Escondidos de Markov – MEM foram escolhidos como método de reconhecimento para as palavras manuscritas. Portanto, apresenta-se a formulação e os algoritmos aplicados nas etapas de treinamento, validação e testes dos modelos. Apresenta-se ainda, a definição, geração, coleta e armazenamento das bases de dados de laboratório de cheques bancários brasileiros.

Com base nos materiais e métodos estudados apresenta-se a metodologia desenvolvida responsável pelo reconhecimento das palavras isoladas do valor expresso por extenso em cheques bancários brasileiros. São descritas todas as etapas do reconhecimento, desde o pré-processamento das imagens até o reconhecimento propriamente dito das palavras.

A abordagem adotada para o reconhecimento é Global, na qual não existe a necessidade de segmentação explícita das palavras em letras ou pseudo-letras. Cada

palavra do léxico composto por 39 palavras é tratada como uma única unidade a ser reconhecida. O método caracteriza-se por ser omni-escriptor, ou seja, não existem restrições ao estilo da escrita.

As primitivas extraídas das palavras permitem uma representação global através de primitivas perceptivas e primitivas geométricas das formas a reconhecer. Para os experimentos realizados com as primitivas perceptivas obtém-se uma taxa de reconhecimento de 70,6%.

O método desenvolvido está voltado para resolver problemas específicos do léxico de palavras utilizadas nos cheques bancários brasileiros. O léxico pode ser decomposto em 4 conjuntos de palavras: palavras terminadas por “entos”, por “enta”, palavras que compõem os conjuntos da dezena e da unidade, e ainda, o conjunto de palavras-chaves (mil, reais ou real, centavos ou centavo).

Finalmente, apresenta-se a conclusão e os trabalhos futuros que podem ser realizados como continuidade do presente trabalho.

Palavras-Chaves: documentos manuscritos, reconhecimento de palavras manuscritas, HMM, cheques bancários.

Abstract

This work presents a study about the application of Hidden Markov Models in handwritten words recognition. The methodology involved with the handwritten words automatic processing explores the relevant aspects from the writing and reading human process associated with the priori knowledge mechanisms. Then, a review associating these processes with the main goal of this work is presented. A bibliography review of the research area called Handwritten Word Recognition - HWR is also presented. The principal goal of this work is to answer two following questions:

- What are the relevant perceptual features in the handwritten recognition process?
- How to represent cursive words with an absence of perceptual features?

The Hidden Markov Models - HMM was chosen as the recognition method applied to handwritten words. Therefore, the theoretical formulation and algorithms applied to the training, validation and testing stages are presented. The works show the definitions, collection and storage of the laboratory database including Brazilian bankchecks.

Based on the materials and methods studied the work presents the developed methodology responsible for the isolated word recognition in the legal amount lexicon. All the stages in the recognition are described, from image preprocessing to word recognition.

The recognition is based on a Global Approach, meaning that it permits the word segmentation stage to be avoided by extracting global features from the words, therefore avoiding explicit segmentation. Each word in the lexicon corresponds an unique model. The lexicon are composed by 39 words and the system is an omni-scriptor, there are not restrictions to the writing style.

The feature extraction stage permits a global representation of the words to be recognized using perceptual features and a geometric feature. The recognition rate obtained with perceptual features is 70.6%. A significant increase in the recognition rate

is observed with the use of concave and convex features, showing a better word representation especially for the words with an absence of perceptual features.

The method is faced to solve specific problems with the Brazilian bankchecks vocabulary. The lexicon was divided in the following four sub-sets: words finished by "entos" and "enta", the words that represents the "dezena" (tens) and "unidade" (unity), and the sub-set of key-words: "mil", "reais/real", "centavos/centavo".

Finally, the work presents the conclusion and the future works that can be done in continuity to this one.

Keywords: Handwritten Word Recognition, Hidden Markov Models, Bankcheck Applications.

Capítulo 1

Introdução

A análise e reconhecimento de documentos, especialmente os documentos manuscritos, desperta grande interesse de um modo geral, uma vez que esses documentos apresentam componentes de grande complexidade.

Assim, o Reconhecimento de Palavras Manuscritas – RPM (*Handwritten Word Recognition* – HWR) caracteriza-se por ser uma sub-área de pesquisa em Análise e Reconhecimento de Documentos. Entre os documentos manuscritos destaca-se o preenchimento de cheques bancários.

Milhões de cheques bancários circulam como documentos financeiros em todas as partes do mundo. Esses cheques, preenchidos à mão ou por meio mecânico/eletrônico, tem sua importância centrada em vários campos, tais como, o número da conta corrente a quem pertence o cheque, o valor numérico do cheque, a expressão do valor numérico por extenso a data e a assinatura do emitente. O detalhamento da estrutura dos cheques bancários brasileiros está descrita no Apêndice A deste trabalho.

Com tantas informações importantes contidas em um cheque, ainda hoje, as operações de processamento deste documento não são realizadas de forma automática. Isto deve-se ao fato da grande complexidade de tratamento de documentos, principalmente aqueles que contém informações manuscritas.

Atualmente, no Brasil, a população tem acesso aos cheques e faz uso desse documento diariamente. Na compensação bancária de cheques um operador humano lê a informação contida no campo numérico e digita em um terminal de computador. Os campos relativos à data, à assinatura, ao extenso e aos demais campos são verificados visualmente sem maiores cuidados ou detalhamento. Quando um cheque é apresentado

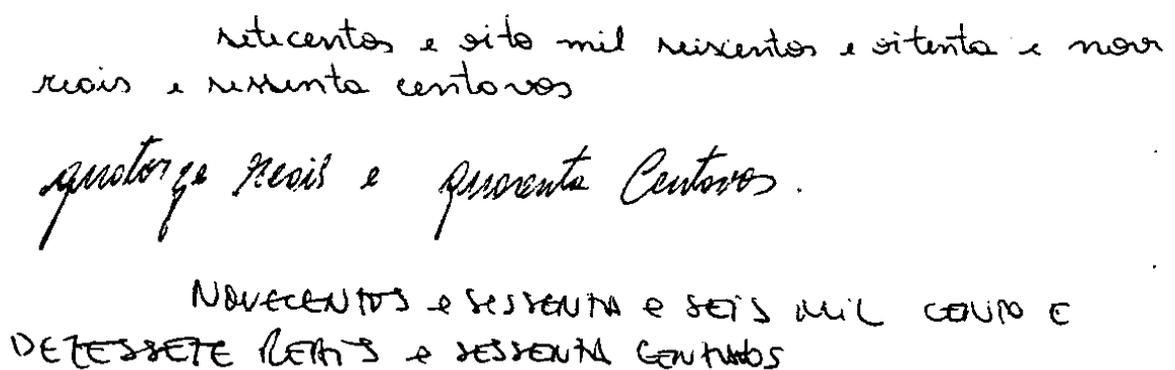
em um caixa de uma agência bancária, ou seja ao portador, o mesmo tem seus campos inspecionados e é realizada a comparação visual da assinatura do cheque com o modelo no cartão de assinaturas existente no banco, dependendo do valor do cheque apresentado. Todo esse processamento manual demanda tempo e custo elevado, além é claro, de não ser eficiente em algumas situações.

1.1. Desafio

Pesquisas estão sendo realizadas em diversas partes do mundo buscando o processamento automático de cheques bancários, podendo-se citar os trabalhos de: [HEUTTE, 1994], [GUILLEVIC, 1995], [AGARWAL et al., 1995], [CONGEDO et al., 1995], [FILATOV et al., 1995], [LEPLUMEY et al., 1995], [MONTOLIU, 1995], [AVILA, 1996], [KNERR et al., 1996], [CÔTE, 1997], [OLLIVIER, 1999].

Entende-se por processamento automático, a busca de um método automático para processamento de imagens de cheques, processo *off-line* ou estático, envolvendo, principalmente, a análise e o reconhecimento de informações manuscritas.

Documentos manuscritos apresentam componentes de grande complexidade, tais como: diferentes estilos de escrita, diferentes tipos de números manuscritos, contexto da escrita, contexto da aquisição do documento (*on-line* - dinâmico ou *off-line* - estático) e ainda, o tamanho do léxico a ser reconhecido e o número de escritores. A Figura 1.1 mostra alguns exemplos da diversidade do estilo de escrita retirados da base de dados de treinamento de cheques bancários brasileiros.



Novecentos e setenta e seis mil seiscentos e setenta e nove
 reais e noventa centavos
 quatorze reais e noventa centavos.
 NOVECENTOS e SETENTA e SEIS MIL COITO e
 DEZESSETE REAIS e DESSENTA CENTAVOS

Figura 1.1:Exemplos de estilos de escrita

Deste modo, o desafio do presente trabalho é utilizar os processos de escrita e leitura realizados pelo ser humano como subsídios ao reconhecimento de palavras

manuscritas.

Todas estas razões fazem com que as inúmeras pesquisas que estão sendo desenvolvidas tenham como objetivo encontrar soluções “ótimas” para documentos com informações manuscritas de diferentes formatos, origens e finalidades.

Mesmo considerando os progressos já realizados, principalmente nas áreas de reconhecimento de manuscritos, nenhum método, abordagem ou método resolveu o problema do processamento e análise automática de cheques bancários brasileiros.

1.2. Motivação

O presente estudo tem por motivação a busca dos aspectos relevantes dos processos de escrita e leitura de maneira a contribuir com a resolução de problemas específicos referentes ao reconhecimento de palavras manuscritas no escopo do extenso de cheques bancários brasileiros.

Com base nos processos de escrita e leitura, adota-se uma abordagem Global ou *Global Approach*, sem segmentação explícita das palavras em letras ou pseudo-letras. Cada palavra do léxico é tratada como uma única unidade a ser reconhecida e, portanto, a seleção de primitivas a serem extraídas das palavras trata de responder duas importantes questões:

- **Quais são as primitivas perceptivas (*perceptual features*) relevantes no processo de reconhecimento de palavras manuscritas?**
- **Como representar palavras cursivas sem a presença de primitivas perceptivas?**

Para o reconhecimento de palavras manuscritas existe uma abundância de técnicas que são capazes de descrever a similaridade das formas que pertencem a uma mesma classe e que permitem ao mesmo tempo distinguir as representações de formas separadamente entre classes diferentes. Portanto, a motivação está em conseguir um método de reconhecimento que se aproxime ao máximo do ser humano em sua capacidade de ler uma palavra manuscrita qualquer, sem restrição.

1.3. Proposta

O presente trabalho tem por objetivo modelar e avaliar computacionalmente o reconhecimento de palavras manuscritas no contexto de cheques bancários brasileiros. Para isso, propõe-se a consideração de características perceptivas relacionadas à escrita

e a leitura, e ao uso de Modelos Escondidos de Markov – MEM para processamento da informação [MADHVANATH & GOVINDARAJU, 1998] [SCHOMAKER & SEGERS, 1998].

O extenso de um cheque corresponde a um valor numérico, ao qual se aplica uma gramática conhecida no momento da grafia deste valor por extenso. Portanto, o léxico é composto por 39 palavras, a saber: {novecentos, oitocentos, setecentos, seiscentos, quinhentos, quatrocentos, trezentos, duzentos, cem / cento, noventa, oitenta, setenta, sessenta, cinquenta / cincoenta, quarenta, trinta e vinte, dezenove, dezoito, dezessete, dezesseis, quinze, quatorze, treze, doze e onze, nove, oito, sete, seis, cinco, quatro, três, dois, um / hum, mil, reais / real, centavos / centavo}.

Dois fatos aumentam a complexidade do problema de reconhecimento no contexto tratado. Primeiramente, a grande similaridade de formas entre as palavras do léxico. E, segundo, o fato da gramática portuguesa aceitar diferentes grafias para uma mesma palavra, por exemplo: “um” e “hum”, “três” e “treis”, “quatorze” e “catorze”, “cinquenta”, “cinqüenta” e “cincoenta”.

Geralmente o reconhecimento de palavras manuscritas baseia-se em duas abordagens do problema: Local ou *analytical approach* realizada em nível dos caracteres [LECOLINET, 1990], [FAVATA & SRIHARI, 1992], [GILLOUX et al., 1995], [KIM, 1996] e Global ou *global approach* em nível das palavras [GUILLEVIC, 1995] e [CÔTE, 1997]. Considerando-se as possíveis abordagens, adota-se uma abordagem Global, sem segmentação explícita das palavras em letras ou pseudo-letras.

Trabalha-se com a extração de primitivas globais do tipo perceptivas e geométricas buscando-se diminuir a variabilidade de formas intra-classes e aumentar a discriminação inter-classes.

Para o reconhecimento existem, basicamente, duas classes de métodos: os estruturais e os estatísticos. O método estatístico escolhido foram os Modelos Escondidos de Markov – MEM. Estes modelos oferecem um enfoque probabilístico ao enfoque estrutural das primitivas e permitem uma treinamento automático dos parâmetros a estimar. O interesse dos MEM reside na sua possibilidade de modelar eficazmente diferentes fontes de conhecimento. Seu ponto forte está na integração coerente dos diferentes níveis de modelagem (morfológico, léxico, sintático) [AVILA,

1996], [EL YACOUBI, 1996], [CHO et al., 1995], [CHEN et al., 1995a], [AGAZZI & KUO, 1993].

Assim, a Figura 1.2 resume esquematicamente o método proposto, mostrando cada uma das etapas necessárias ao reconhecimento das palavras isoladas do extenso manuscrito de cheques bancários brasileiros através dos MEM.

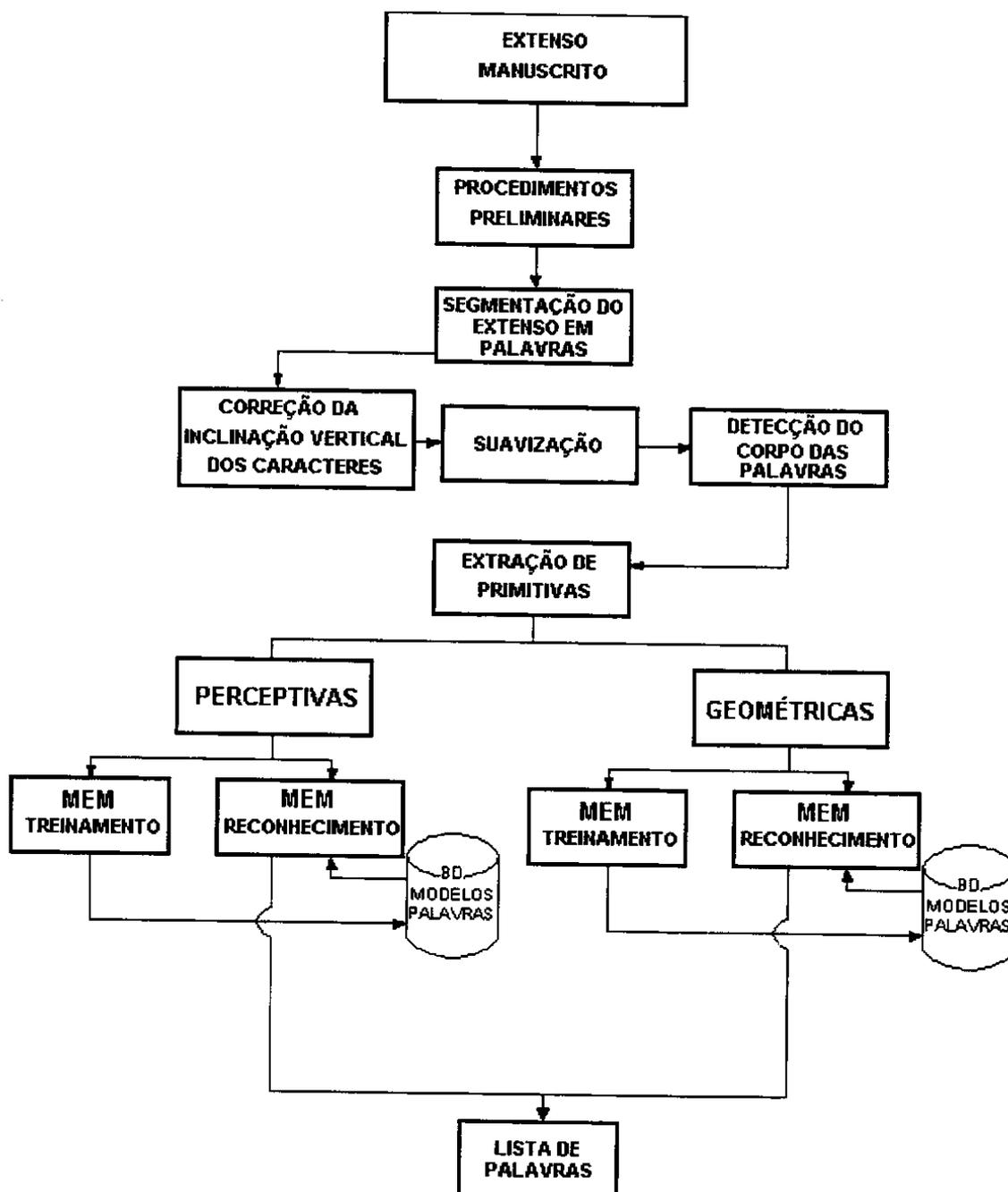


Figura 1.2: Esquema do método de reconhecimento de palavras manuscritas

1.4. Contribuição

O presente trabalho envolve uma área de pesquisa inovadora no Brasil, uma vez que trata do reconhecimento de palavras manuscritas no contexto de cheques bancários brasileiros.

O método proposto parte da percepção humana das palavras manuscritas para definir primitivas perceptivas a serem extraídas das formas a reconhecer. O presente trabalho contribui para o entendimento dos processos de escrita e leitura realizados pelo ser humano com o objetivo de relacionar o homem e os métodos automáticos de tratamento de palavras manuscritas.

Outra contribuição refere-se a aplicação dos Modelos Escondidos de Markov no reconhecimento de palavras manuscritas provenientes de cheques bancários. Além, da contribuição de sua aplicação em problemas de reconhecimento de palavras do idioma português. Ressalta-se, ainda, a contribuição do presente trabalho na criação das bases de dados de cheques de laboratório.

O presente trabalho coloca-se, ainda, como mais uma contribuição na área de Análise e Reconhecimento de Documentos, podendo vir auxiliar outras aplicações, tais como: reconhecimento de palavras manuscritas em formulários, projetos e anotações de campo, identificação dos procedimentos de recebimento e despacho de documentos para recuperação e consultas, conversão de desenhos e anotações manuscritas para recuperação ou união com Projetos Auxiliados por Computador – PAC.

1.5. Organização

Nesse capítulo foram apresentados os desafios, a motivação, os objetivos e as contribuições desse trabalho. O Capítulo 2 apresenta o problema do reconhecimento das palavras manuscritas, bem como, os aspectos relevantes dos processos de escrita e leitura. A revisão bibliográfica mostra o estado da arte do reconhecimento de palavras manuscritas e apresenta os fundamentos teóricos dos Modelos Escondidos de Markov.

O Capítulo 3 apresenta as diversas etapas da formação da base de dados de cheques bancários brasileiros com o objetivo de atender os diversos trabalhos de pesquisa em realização na PUC-PR, na área de análise e reconhecimento de documentos manuscritos. Descreve-se a metodologia empregada no estabelecimento da base, coleta, aquisição digital e armazenamento das imagens. Destacam-se as principais

características dos cheques coletados quanto ao valor numérico por extenso preenchido à mão em um cheque. Apresenta-se a situação atual da base de dados.

O Capítulo 4 refere-se a todas as etapas necessárias ao reconhecimento das palavras isoladas do extenso, destacando-se: pré-processamento (procedimentos preliminares, segmentação do extenso em palavras, correção da inclinação vertical dos caracteres, detecção do corpo das palavras), extração de primitivas perceptivas e geométricas, treinamento e validação dos modelos MEM das palavras e reconhecimento através dos modelos MEM. Apresenta-se os experimentos realizados e os resultados obtidos realizando-se uma detalhada análise de erros.

O Capítulo 5 apresenta a conclusão e as possibilidades de trabalhos futuros visando a continuidade das pesquisas na área de reconhecimento de palavras manuscritas na PUCPR.

Capítulo 2

Palavras Manuscritas

Este capítulo permite uma visão geral da área de Reconhecimento de Palavras Manuscritas – RPM (HWR – *Handwritten Word Recognition*). Inicialmente, apresenta-se uma descrição do problema de reconhecimento das palavras manuscritas do extenso de cheques bancários brasileiros. Para tanto, busca-se entender os aspectos relevantes dos processos de escrita e leitura realizados pelo ser humano.

A pesquisa realizada faz uma associação dos métodos utilizados no aprender a escrever e no aprender a ler pelo ser humano com os métodos utilizados pelos pesquisadores em seus métodos de reconhecimento automático de palavras manuscritas. Ressalta-se que o ser humano acumula conhecimentos prévios ao longo de sua vida de maneira a utilizá-los futuramente nos processos de leitura e reconhecimento de palavras já conhecidas anteriormente.

São apresentadas também as propriedades de palavras manuscritas, os métodos e técnicas empregados no problema do reconhecimento de manuscritos e pesquisas realizadas na área. Descreve-se a conceituação e formulação teórica dos Modelos Escondidos de Markov – MEM escolhidos como método de reconhecimento das palavras manuscritas.

O objetivo do capítulo é resgatar os aspectos relevantes dos processos de escrita e leitura humana, associados ao conhecimento prévio do ser humano. Desta forma, é possível fornecer subsídios ao estudo das relações existentes entre os processos realizados pelo ser humano e um método de reconhecimento automático de palavras manuscritas.

2.1. O Problema: Palavras Manuscritas em Cheques Bancários Brasileiros

O extenso de um cheque corresponde a um valor numérico, ao qual se aplica uma gramática conhecida no momento da grafia deste valor por extenso. Portanto, a partir do valor numérico é possível definir duas características do problema: as palavras-chaves e os blocos internos às palavras-chaves formados por palavras que representam os valores numéricos, conforme mostrado na Figura 2.1.

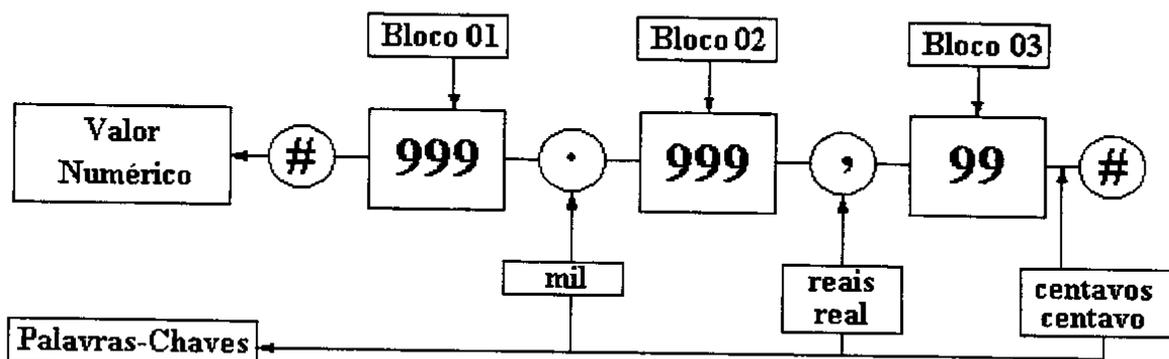


Figura 2.1: Valor numérico e palavras-chaves

Entende-se por palavras-chaves as 3 palavras que identificam 3 grandes blocos no extenso de um cheque. Considerando que as bases de dados foram formadas para atender valores entre R\$ 0,01 (um centavo) e R\$ 999.999,99 (novecentos e noventa e nove mil, novecentos e noventa e nove reais e noventa e nove centavos), as palavras-chaves são justamente as palavras correspondentes no léxico ao ponto da indicação do valor da **milhar**, a vírgula da indicação do valor da parte inteira em **reais/real** e ao terminal da indicação do valor da parte decimal em **centavos/centavo**.

Entende-se por bloco o conjunto de palavras grafadas por extenso usadas para representar a quantia expressa em uma das partes do valor numérico do cheque. Deste modo, a análise dos blocos permite identificar a formação interna de cada um através de palavras do léxico. Os blocos 1 e 2, conforme Figura 2.2, são idênticos entre si. Esses blocos têm como terminais na gramática o mesmo conjunto de palavras. Por outro lado, o bloco 3, que representa o valor numérico referente a parte dos “centavos/centavo” tem a característica de não apresentar o conjunto de palavras terminadas por “entos”.

O léxico de 39 palavras pode, então, ser classificado de acordo com a ordem correspondente ao algarismo no valor numérico resultando em 5 classes de palavras, a saber:

- classe “entos”: {novecentos, oitocentos, setecentos, seiscentos, quinhentos, quatrocentos, trezentos, duzentos, cem / cento},
- classe “enta”: {noventa, oitenta, setenta, sessenta, cinquenta / cincoenta, quarenta, trinta e vinte},
- classe dezena: {dezenove, dezoito, dezessete, dezesseis, quinze, quatorze, treze, doze e onze},
- classe unidade: {nove, oito, sete, seis, cinco, quatro, três, dois, um / hum },
- classe palavras-chaves: {mil, reais / real, centavos / centavo}.

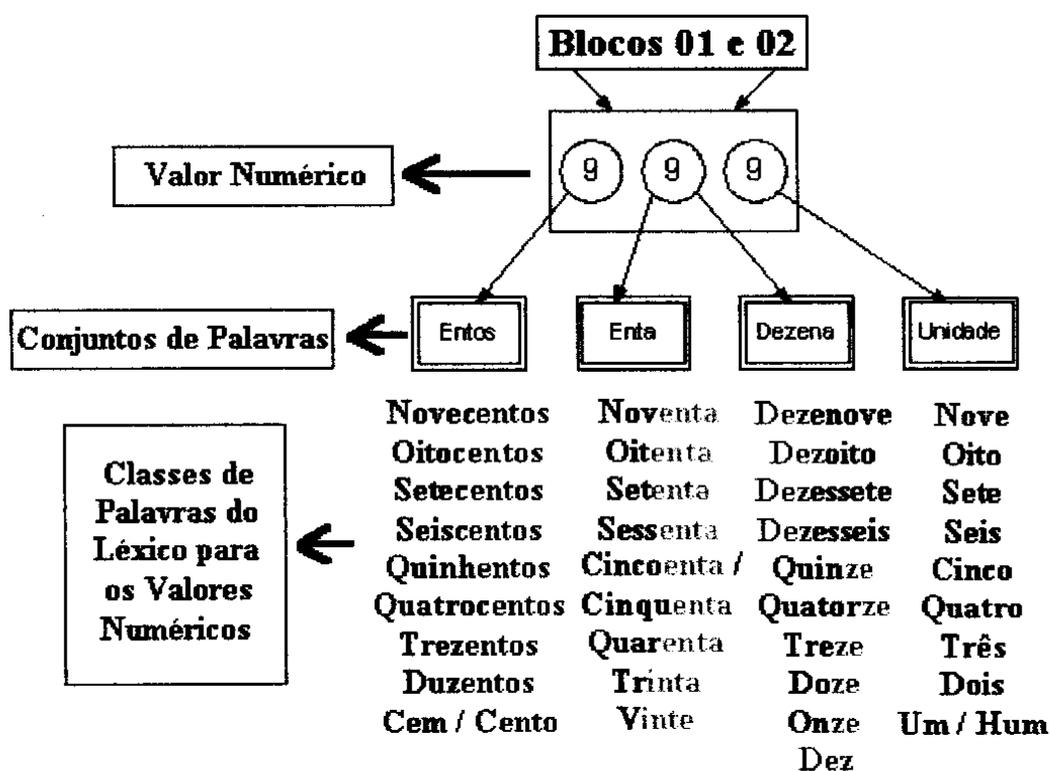


Figura 2.2: Classes de palavras do léxico em português

Dois fatos aumentam a complexidade do problema de reconhecimento no contexto tratado. Primeiramente, a grande similaridade de formas entre as palavras do léxico. E, segundo, o fato da gramática portuguesa aceitar diferentes grafias para uma mesma palavra, por exemplo: “um” e “hum”, “três” e “treis”, “quatorze” e “catorze”, “cinquenta”, “cinqüenta” e “cincoenta”.

Para o reconhecimento de palavras manuscritas existe uma abundância de técnicas que são capazes de descrever a similaridade das formas que pertencem a uma mesma classe e que permitem ao mesmo tempo distinguir as representações de formas separadamente entre classes diferentes. Portanto, o objetivo do reconhecimento automático é o de desenvolver um método que se aproxime ao máximo do ser humano em sua capacidade de ler uma palavra manuscrita qualquer (sem restrição). Deve-se lembrar que o ser humano é capaz de ler palavras mal escritas graças à exploração das informações de contexto.

Geralmente o reconhecimento de palavras manuscritas baseia-se em duas abordagens do problema: Local (*analytical approach*) realizada em nível dos caracteres [LECOLINET, 1990], [FAVATA & SRIHARI, 1992], [GILLOUX et al., 1995], [KIM, 1996] ou Global (*global approach*) realizado em nível das palavras [GUILLEVIC, 1995] e [CÔTE, 1997].

Na primeira abordagem, o método se depara com a necessidade de segmentação das palavras em caracteres/letras ou pseudo-letras. Esta abordagem caracteriza-se pela dificuldade em se definir a fronteira entre os caracteres. E, portanto, o método de reconhecimento dependerá também do sucesso do processo de segmentação. Assim, estes métodos são constituídos por etapas distintas para segmentação e reconhecimento, ou associando segmentação e reconhecimento em uma mesma etapa.

Por outro lado, a abordagem global permite evitar a etapa de segmentação das palavras, extraíndo-se primitivas globais das palavras sem necessidade de segmentação explícita das mesmas. Esta abordagem procura explorar as informações do contexto das palavras, sendo que aspectos baseados em modelos psicológicos podem ser inseridos [GUILLEVIC, 1995] e [CÔTE, 1997]. Porém, é uma abordagem restrita às aplicações de léxico com pequenas dimensões.

As duas abordagens, local e global, não são estanques podendo-se trabalhar com uma abordagem híbrida [CHEN & KUNDU, 1993]. Deste modo, é possível combinar as vantagens e diminuir as desvantagens.

Considerando as possíveis abordagens do problema, a etapa de extração de primitivas caracteriza-se por uma das tarefas importantes no sucesso dos métodos de reconhecimento de palavras manuscritas. A função de um conjunto selecionado de primitivas é diminuir a variabilidade de formas intra-classes e aumentar a discriminação inter-classes. Diversos tipos de primitivas podem ser combinadas visando obter métodos

mais eficazes através de um conjunto pequeno de primitivas, porém, com alta discriminação das classes consideradas [TRIER et al., 1996] e [EL YACOUBI, 1996].

Considerando-se o exposto, adota-se uma abordagem Global, sem segmentação explícita das palavras em letras ou pseudo-letras. Cada palavra do léxico é tratada como uma única unidade a ser reconhecida e, portanto, a seleção de primitivas a serem extraídas das palavras trata de responder duas importantes questões:

- **Quais são as primitivas perceptivas (*perceptual features*) relevantes no processo de reconhecimento de palavras manuscritas?**
- **Como representar palavras cursivas sem a presença de primitivas perceptivas?**

A resposta está em reunir os aspectos relevantes dos processos de escrita, leitura e conhecimento prévio realizados pelos seres humanos com os trabalhos de [MADHVANATH & GOVINDARAJU, 1998] e [SCHOMAKER & SEGERS, 1998]. No primeiro trabalho encontra-se a definição de primitivas perceptivas ou *perceptual features*, que são as principais características utilizadas na representação da forma de palavras, especialmente nos casos de aplicações com léxico de tamanho pequeno, tal como o extenso em cheques bancários. Então, o primeiro conjunto de primitivas tem por base a extração das seguintes primitivas perceptivas – PP: ascendentes, descendentes e laços, representados por símbolo e posição.

As primitivas perceptivas estão presentes na grande maioria das classes de palavras cursivas do léxico em questão. Entretanto, ao analisar-se cuidadosamente o vocabulário do extenso em cheques bancários brasileiros verifica-se a presença de um sub-conjunto de palavras que não possui este tipo de característica, tais como: “um”, “cinco”, “seis”, “nove”, “cem”, “reais”. Ou ainda, caracteriza-se por uma palavra curta: “sete”, “oito”, “dois”, “três”, “mil”. Assim, trabalha-se com dois outros conjuntos de primitivas globais agregando-se primitivas baseadas em concavidades e convexidades existentes no corpo das palavras. Estes conjuntos são denominados de primitivas perceptivas com concavidades e convexidades – PPCC e primitivas perceptivas com rotulação das concavidades e convexidades – PFCCR, respectivamente.

Propõe-se ainda, um conjunto de primitivas geométricas com base na rotulação do fundo da imagem para os pixels do corpo das palavras. Este conjunto auxilia na exploração das regularidades de uma palavra, tal qual [CÔTE, 1997].

Assim, partindo-se da extração de primitivas pode-se trabalhar com reconhecimento das palavras isoladas. Para o reconhecimento existem, basicamente, duas classes de métodos: os estruturais e os estatísticos. Um modelo híbrido formado pela utilização conjunta de métodos estruturais e estatísticos é uma solução ótima para o problema de reconhecimento de formas. Pois, essa estratégia permite, ao mesmo tempo, beneficiar-se das vantagens de cada um dos dois enfoques.

Os métodos estruturais são utilizados para descrever as características intrínsecas de uma forma, efetuando-se uma modelagem em termos de componentes estruturais, chamados de primitivas, ligados a natureza de cada forma.

Contrariamente as aproximações estatísticas, as aproximações estruturais introduzem a noção de ordem na descrição de uma forma. Assim, os métodos estruturais são particularmente adaptáveis às aplicações onde a ordem é justamente primordial. Este é o caso do reconhecimento de palavras manuscritas onde o aspecto sequencial das letras de esquerda para direita deve ser corretamente modelado.

A força do enfoque estatístico reside na sua capacidade de modelar a variabilidade das formas sem considerar as limitações inerentes ao problema.

Considerando-se o exposto os Modelos Escondidos de Markov – MEM oferecem um modelo misto formado pela utilização conjunta de métodos estruturais de representação das formas a reconhecer e de um método estatístico para treinamento dos modelos e reconhecimento das palavras propriamente ditas. Portanto, este é o método escolhido para o reconhecimento das palavras manuscritas no presente trabalho. Outros autores aplicam os MEM para palavras manuscritas: [AGAZZI & KUO, 1993], [CHO et al., 1995], [CHEN et al., 1995a], [AVILA, 1996] e [EL YACOUBI, 1996].

A diferença entre a abordagem adotada e outros trabalhos realizados por diferentes autores está no fato da modelagem proposta trabalhar com primitivas perceptivas baseadas na percepção humana das palavras manuscritas, cabendo ao processo de extração de primitivas a classificação das mesmas e a geração das seqüências de observações.

O interesse dos MEM reside na sua possibilidade de modelar eficazmente diferentes fontes de conhecimento. Seu ponto forte está primeiro no fato, da integração coerente dos diferentes níveis de modelagem: morfológico ou traçado da palavra, léxico ou vocabulário de palavras estudadas e sintático ou o modelo da palavra propriamente dita. E segundo, na existência de algoritmos eficazes para determinar o valor ótimo dos

parâmetros fornecendo a melhor adequação entre o modelo e a base de dados (conhecimento), chamada de base de treinamento.

Deste modo, busca-se nos processos de escrita e leitura realizados pelo ser humano os aspectos relevantes de ambos para que se possa entender e obter subsídios ao estabelecimento de uma método de reconhecimento para palavras manuscritas no escopo do extenso de cheques bancários brasileiros.

2.2. Escrita, Conhecimento Prévio e Leitura

O que é escrever? Como concluir que algo é escrita? Qual a evolução da escrita? Como a leitura se processa? O que é conhecimento prévio? Como este conhecimento pode ser utilizado? O que é ler? Ler e escrever são processos análogos? [SUEN & LI, 1997].

Cada uma destas perguntas possui resposta e auxiliará na compreensão do problema do reconhecimento da escrita humana, mais especificamente, a escrita cursiva.

O que é escrever? A etimologia da palavra (em grego, *gráfein*, em latim, *scribere*) atesta a origem física da ação de escrever. No Aurélio encontra-se que escrever é representar por letras, redigir, compor, fixar, gravar, dirigir (cartas) e dirigir carta (a alguém). De acordo com [CAGLIARI, 1991], escrever é também uma forma de expressão artística e até um passatempo. Ou ainda, citando [TEBEROSKY & TOLCHINSKY, 1996], escrever é uma atividade intelectual em busca de uma certa eficácia e perfeição, que se realiza por meio de um artefato gráfico-manual, impresso ou eletrônico, para registrar, comunicar, controlar ou influir sobre a conduta dos outros, que possibilita a produção e não só a reprodução, e que supõe tanto um efeito de distanciamento como uma intenção estética.

A escrita, por sua vez, seja ela qual for, conforme [CAGLIARI, 1991], tem como objetivo primeiro a leitura. A leitura é uma interpretação da escrita que consiste em traduzir os símbolos escritos em fala. Portanto, como concluir que algo é escrita? [TEBEROSKY & TOLCHINSKY, 1996] responde esta questão explicando que a escrita, ou sejam marcas gráficas produzidas “no lugar de” algo, é formada, no caso da escrita alfabética, por unidades de notação e pela combinação dessas unidades de forma a constituir um sistema. As unidades de notação, por sua vez, denotam a linguagem. Então, são usadas como marcas “no lugar de” alguma unidade de linguagem. Desta

forma, conclui-se que algo é escrito por estar representado por unidades de linguagem aceitas e compreendidas pelo sistema em questão.

A história da escrita delinea sua evolução a partir das escritas pictóricas, das escritas ideográficas e das escritas alfabéticas. A escrita pictórica caracteriza-se por uma escrita através de desenhos ou pictogramas. Este tipo de escrita aparece em inscrições antigas, sendo exemplos os cantos Ojibwa da América do Norte, o catecismo asteca e mais recentemente a estória em quadrinhos, conforme Figura 2.3 extraída de [CAGLIARI, 1991].

A escrita ideográfica baseia-se no uso de desenhos especiais chamados ideogramas, tem por princípio representar só o significado, referindo-se a um pensamento não necessariamente de natureza linguística, ver exemplos da Figura 2.4.

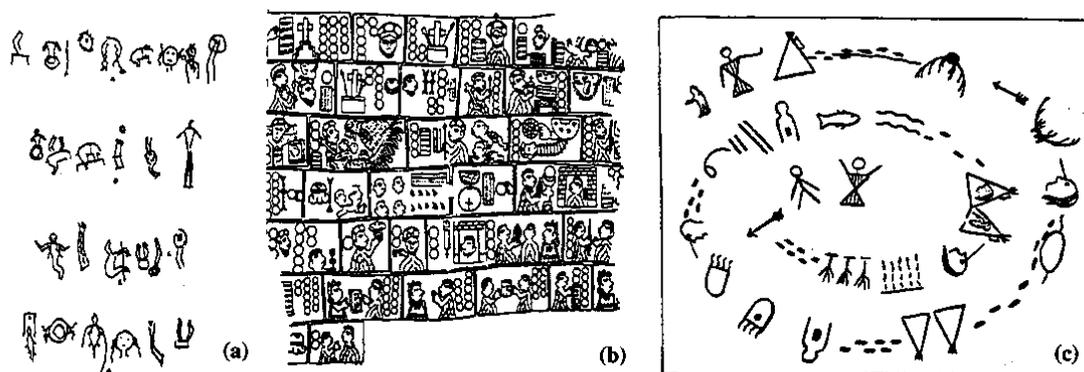


Figura 2.3: Exemplos de escrita pictórica: a) cantos Ojibwa; b) catecismo asteca; c) inscrição antiga [CAGLIARI, 1991]

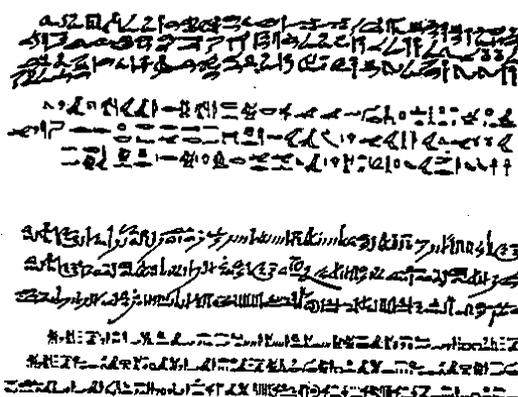


Figura 2.4: Exemplos de escrita ideográfica: escrita egípcia [CAGLIARI, 1991]

Ao longo dos tempos os ideogramas foram perdendo alguns dos traços mais representativos das figuras retratadas e tornaram-se uma simples convenção de escrita. A origem das letras está nos ideogramas, mas estes perderam o valor ideográfico, assumindo uma nova função de escrita: a representação puramente fonográfica, ou seja,

a escrita baseada no significante e não mais no significado. O ideograma perdeu seu valor pictórico e passou a ser simplesmente uma representação fonética [CAGLIARI, 1991] e [TEBEROSKY & TOLCHINSKY, 1996]. As letras do alfabeto atualmente utilizado derivam dessa evolução. A evolução da escrita mostra que os gregos, escrevendo consoantes e vogais, criaram a escrita alfabética. E, portanto, a escrita alfabética se caracteriza pelo uso de letras.

O uso da letra manuscrita iniciou-se na França, conforme [PARRAMÓN, 1991], durante o reinado de Luis XIV, época em que a França dominava quase toda a Europa. A letra manuscrita francesa chamava-se *rondé*, era vertical e semelhante a letra gótica no seu início. Havia ainda outros estilos: a letra italiana, denominada *italienne*, e a tipicamente *inglesa*, desenvolvida a partir da italiana com um desenho que perdura até os dias de hoje por seu estilo clássico. O estilo Mistral, membro da família dos tipos manuscritos ou caligráficos, é o estilo que consegue um enlace entre todas as letras [PARRAMÓN, 1991], conforme Figura 2.5.



Figura 2.5: Estilo mistral: enlace entre as diversas letras [SOLO, 1989]

De maneira geral, atualmente faz-se uso de um sistema alfabético, porém na verdade, de acordo com [CAGLIARI, 1991] esse sistema não possui uma única forma e nem é completamente alfabético, pois utilizam-se outros caracteres de natureza ideográfica, tais como os sinais de pontuação e os números. Segundo o autor, pode parecer paradoxal, mas pouco percebe-se ou explica-se que se faz uso de várias formas de representação gráfica, ou seja, vários tipos de alfabeto em uso, e em geral, de maneira misturada. A escrita cursiva é diferente da escrita de forma (letra tipo bastão) e, como mencionado, entre um extremo e outro a escrita cursiva pode ser classificada em diferentes categorias distintas. A classificação de [TAPPERT et al., 1990] está apresentada neste capítulo e rege as observações sobre o estilo da escrita cursiva neste trabalho.

Quando um sistema alfabético é utilizado por um número grande de pessoas em diferentes lugares, para usos diversos, a forma das letras do alfabeto, que inicialmente

era única, passa a admitir variantes [CAGLIARI, 1991]. No mundo antigo, as variantes se restringiam a uns poucos casos. O latim, por exemplo, não tinha as letras maiúsculas. Hoje, encontramos uma enorme variedade de tipos de alfabetos, um exemplo é o Livro: *Lettering and Alphabets*, de [CAVANAGH, 1974], que apresenta 85 tipos de alfabetos completos, e ainda muitos outros [SOLO, 1989], [PARRAMÓN, 1991] e [ROSA, 1991].

A primeira letra do alfabeto, a letra *a*, pode aparecer escrita de diferentes maneiras conforme Figura 2.6, sendo cada um dos exemplos pertencente a um tipo de alfabeto diferente. Segundo [CAGLIARI, 1991] a letra *A* é tão diferente de *a*, quanto *p* é de *m*, por exemplo: *p*, *b*, *d* e *q* são muito mais semelhantes entre si do que *b* e *B*, *g* e *G*. Realmente, não se percebe a cada instante da escrita ou da leitura estas diferenças, para uma pessoa adulta um *A* é um *A*, seja ele escrito como for [CAGLIARI, 1991].

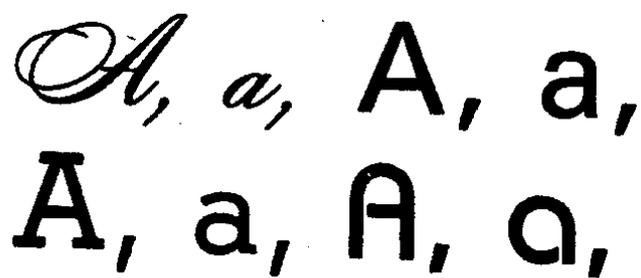


Figura 2.6: Exemplos da letra “a” em diferentes alfabetos [CAGLIARI, 1991]

A escrita cursiva é a mais complicada das escritas existentes no mundo de acordo com [CAGLIARI, 1991], [PARRAMÓN, 1991] e [ROSA, 1991], porque varia enormemente, seguindo as idiossincrasias de cada usuário. A escrita cursiva só é menos difícil para quem está acostumado com o escrever e com o modo de traçar as letras de quem escreveu, caso contrário, sabe-se muito bem, é difícil ler o que os outros escrevem e, às vezes, até mesmo o que nós próprios escrevemos [CAGLIARI, 1991].

A Figura 2.7 apresenta as variações geradas na escrita cursiva com base em três tipos básicos de traçados apresentadas por [CAVANAGH, 1974]. O primeiro traçado origina uma escrita cursiva mais arredondada e com um laço extra na letra *t*, o segundo traçado é o oposto do primeiro originando uma escrita cursiva mais acentuada no que se refere as letras *h*, *s* e *m*. Por outro lado, o terceiro traçado origina uma escrita cursiva onde as letras minúsculas (corpo da palavra) é mais alto, apresentando um resultado intermediário entre o primeiro e segundo traçados. [CAVANAGH, 1974] apresenta diferentes tipos de escrita cursiva e cita que muitas pessoas geram derivações próprias

(diferentes inclinações, condensamento ou expansão dos caracteres) dos dois tipos de escrita cursiva: Formal e Informal. O objetivo de Cavanagh é ensinar a escrever, de forma manual, diferentes tipos de alfabetos para diferentes ocasiões ou necessidades, ou seja, tem o objetivo de tornar o leitor um calígrafo. Entre os 85 alfabetos apresentados por [CAVANAGH, 1974] e os 100 alfabetos apresentados por [SOLO, 1989] apresentam-se na Figura 2.8 os alfabetos que mais se aproximam da nossa forma de escrever.

mmmm Christmas
uuu Christmas
mmm Christmas

Figura 2.7: Variações com base no traçado [CAVANAGH, 1974]

A B C D E F G a b c d e f g h i
H I J K L M j k l m n o p q r
N O P Q R S T s t u v w x y z
U V W X Y Z 1 2 3 4 5 6 7 8 9

A B C D E F G U V W X Y Z
H I J K L M a b c d e f g h i j
N O P Q R S T k l m n o p q r s
t u v v w w x y z

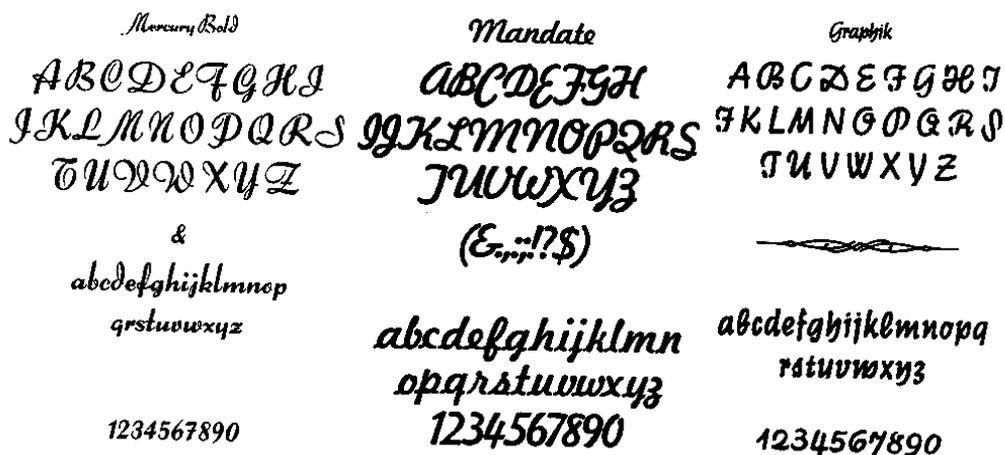


Figura 2.8: Estilos de escrita cursiva [CAVANAGH, 1974] e [SOLO, 1989]

Deve-se ressaltar que ao se alfabetizar uma criança é apresentado à ela um conjunto de letras que formam o alfabeto utilizado no Brasil, conforme Figura 2.9, o qual permitirá a criança concatenar perfeitamente letra a letra, e portanto, produzir palavras. Ao longo de sua vida cada criança irá gerar um estilo próprio de escrita, alternando várias vezes a forma de desenhar as letras até que na vida adulta defina um estilo pessoal de escrever. Porém, muitas das características do traçado, das proporções e das ligações entre os caracteres podem ser mantidas.

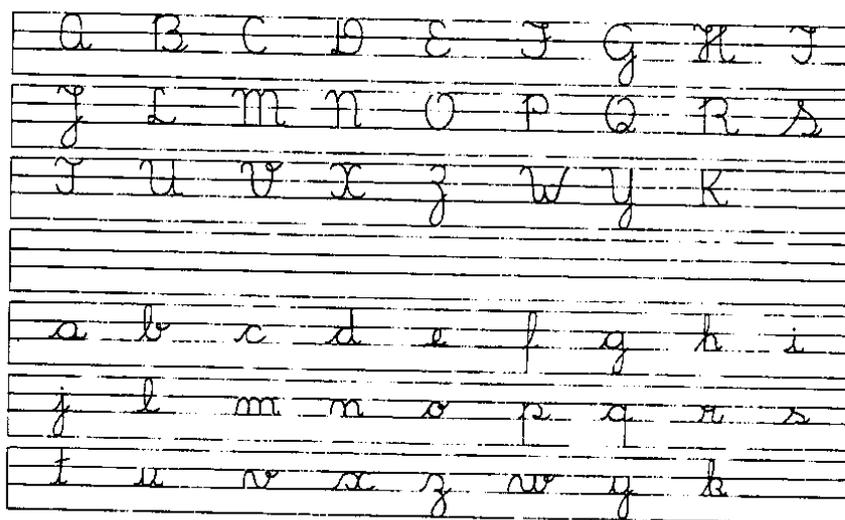


Figura 2.9: Alfabeto e estilo de escrita cursiva para alfabetização no Brasil

Agora, é possível analisar uma nova atividade, ou seja, a leitura. Segundo [KATO, 1990] a maioria das pessoas e escolas supõem que à produção segue-se automaticamente a recepção. Em outras palavras, se for ensinada a escrita, o aluno aprenderá automaticamente a ler. Contudo, cita [KATO, 1990] que a leitura pode ser adquirida independentemente da escrita. O exemplo está em leitores que são proficientes leitores de uma língua estrangeira, porém, nada escrevem nessa língua. O

autor não quer mostrar que a prática da produção não interfere favoravelmente na capacidade de recepção, mas sim, que existe uma interferência recíproca entre os dois processos.

Pergunta-se então: O que é ler? Por leitura se entende toda manifestação linguística que uma pessoa realiza para recuperar um pensamento formulado por outra e colocado em forma de escrita [CAGLIARI, 1991]. A leitura é a realização do objetivo da escrita. No Aurélio encontra-se que ler é percorrer com a vista o que está escrito, proferindo ou não, mas conhecendo as respectivas palavras, conhecer, interpretar por meio da leitura, pronunciar em voz alta, recitar, ver e estudar (coisas escritas), decifrar, interpretar o sentido de, explicar, adivinhar, predizer, prelecionar, ver e interpretar o que está escrito. A maioria dos idiomas supõe uma leitura horizontalizada, da esquerda para a direita.

Para [MACHADO & TEIXEIRA, 1996] ao ler associa-se com as informações lidas à imensa bagagem de conhecimentos que temos armazenados em nosso cérebro e, então, somos capazes de criar, imaginar e sonhar. Citam que ler com velocidade torna a leitura mais proveitosa, estimula a mente e desenvolve o senso de antecipação e expectativa, torna-a mais participativa e não possibilita a fragmentação do assunto. Contudo, solicita do usuário o bom senso de ler com velocidade variada, dependendo da situação.

O processo de alfabetização faz com que as pessoas sejam submetidas a processos com enfoques fônicos, globais ou lingüísticos que, de uma forma geral, estimulam uma área muito pequena da leitura, visando as sílabas ou as palavras. Normalmente, a leitura é realizada letra a letra, sílaba a sílaba ou palavra a palavra, padrão mantido desde o período de alfabetização. A leitura acelerada propõe uma maior captação de elementos - um bloco de palavras que formam uma idéia, dinamizando a leitura.

O olho humano é o órgão do sentido que possibilita a leitura, além dos dedos que, através do tato, permitem aos deficientes visuais decifrarem o texto em *braille*. Os olhos funcionam como uma máquina fotográfica sofisticada, levando o conteúdo captado para o cérebro, onde será interpretado.

Sabe-se então, que a visão, como os outros sentidos, é responsável por informar ao cérebro sobre o estímulo recebido. A percepção traduz as mensagens sensoriais que

ela informou, para que ocorra a compreensão. Dois fatores interferem na percepção: os estímulos (com toda a sua gama de variedades) e os fatores pessoais.

Uma característica dos estímulos interessa ao contexto do presente trabalho: o fechamento. Entende-se por fechamento o preenchimento de lacunas na informação. Através do princípio do fechamento, a Gestalt explica que a mente humana tem a capacidade de perceber figuras ou idéias incompletas, conforme Figura 2.10. Além disso, através da raiz (parte principal) das palavras pode-se perceber seu significado [MACHADO & TEIXEIRA, 1996].

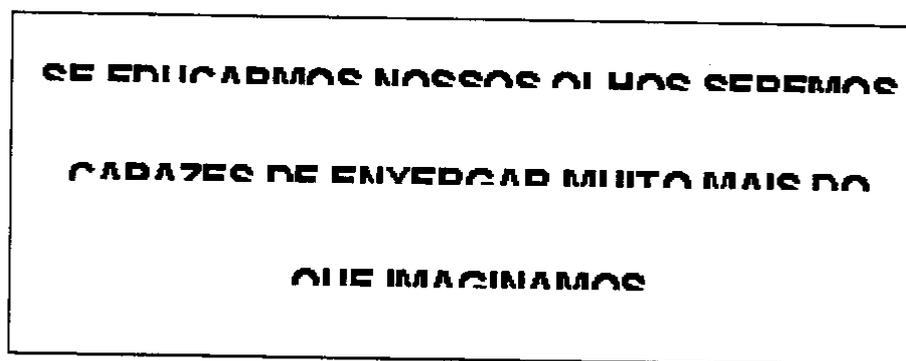


Figura 2.10: Verificação do princípio de fechamento[MACHADO & TEIXEIRA, 1996]

Para o processo de leitura, é relevante o conhecimento prévio que o usuário carrega com si próprio e não pode-se deixar de lado os aspectos do conhecimento linguístico, o conhecimento do mundo e o conhecimento partilhado e a armazenagem e ativação do conhecimento na memória.

O conhecimento linguístico diz respeito aos elementos linguísticos, tais como os itens lexicais e as estruturas sintáticas. Com base neste conhecimento, ativado em memória, o usuário utilizará os elementos linguísticos como “pistas”, afim de estabelecer uma relação entre o linguístico e o conceitual-cognitivo. Deve-se lembrar que o contexto linguístico está associado também a este conhecimento, de modo que a interação entre estes níveis permitirá ao usuário apreender o sentido veiculado pelo texto.

Agora, pode-se refletir sobre o reconhecimento de palavras manuscritas e o desenvolvimento de métodos que aproximem-se o máximo possível da capacidade do ser humano de ler uma palavra qualquer. A revisão bibliográfica apresentada descreve as técnicas e métodos utilizados no desenvolvimento de métodos para reconhecimento de palavras manuscritas considerando-se os trabalhos já realizados pelos pesquisadores ao longo dos últimos 30 anos. O reconhecimento de palavras manuscritas tem sido

tratado como ler o que está escrito, no mesmo sentido da escrita, ou seja, da esquerda para a direita, letra a letra, sílaba a sílaba e algumas vezes palavra a palavra.

NEMIROVSKY em [TEBEROSKY & TOLCHINSKY, 1996] pergunta: Por que algumas crianças podem ler seu nome e de outras pessoas, mas ao escrever seu próprio nome utilizam um conjunto correspondente de letras, mas não o conjunto original? Por que muitas crianças escrevem da esquerda para a direita e lêem da direita para a esquerda? Por que algumas crianças não controlam a quantidade de grafias ao ler sua própria escrita? Se ler fosse simplesmente o inverso de escrever e escrever o inverso de ler, nenhuma dessas situações apareceriam, ou seriam casos excepcionais restritos a algumas crianças que, por razões estranhas, atuariam de maneira insólita. No entanto, cita a autora que estas situações são habitualmente presenciadas por professores, pesquisadores e psicólogos. Na verdade, ler e escrever não são atividades inversas mas sim diferentes, que exigem das pessoas conhecimento e recursos distintos [TEBEROSKY & TOLCHINSKY, 1996].

A seguir apresenta-se uma revisão associando os aspectos de leitura e reconhecimento de palavras com a finalidade de dar sustentação aos modelos MEM propostos no presente trabalho.

2.3. Leitura e Reconhecimento

A revisão realizada sobre as atividades de escrita e leitura, bem como a associação destes processos com o conhecimento prévio do indivíduo em questão, permite que o presente trabalho passe a considerar o indivíduo em questão como sendo o computador. Desta forma, fazendo um paralelo com uma criança, o desenvolvimento de um método de reconhecimento de palavras manuscritas assemelha-se ao desejo de iniciar o processo de leitura. Nesta Seção apresenta-se os diversos métodos de alfabetização e reconhecimento das palavras realizados pelo ser humano, sendo que a estes métodos estão associadas também as formas de segmentação: de seqüência em palavras, de palavras em sílabas, ou mesmo de palavras em letras ou pseudo-letas.

Com base em todos estes métodos, técnicas e conhecimentos apresentados, pergunta-se tal qual [KATO, 1990]: em que medida, porém, a discriminação visual precisa e um processamento analítico em nível de unidades menores que as palavras são necessários para a leitura? Segundo o autor, na medida que o aprendiz desenvolve sua

capacidade de se apoiar em estruturas cada vez maiores, em seu conhecimento do mundo, esse tipo de operação será cada vez menos em nível de unidade grafêmicas e silábicas e cada vez mais a nível de fatias informacionais significativas.

O reconhecimento das formas será em grande parte determinado por processo inferenciais e de predições ditadas pelo conhecimento lingüístico e extralingüístico do leitor. Um leitor proficiente é capaz de compreender um texto escrito em letra cursiva bem pouco legível, exatamente porque ele faz uso dessa habilidade para ler sem decodificar cada grafema [KATO, 1990]. Afinal, sabe-se com base na experiência e nos autores [CAGLIARI, 1991] e [PARRAMÓN, 1991], que aquele que escreve fluentemente, em letra cursiva, também não tem o cuidado de desenhar todos os traços distintivos de cada letra, contando com a capacidade do leitor de ler pelo significado.

Alguns resultados obtidos por [LAVINE, 1972] e ainda, confirmados por [REGO, 1982] ambos apud [TEBEROSKY & TOLCHINSKY, 1996], no contexto brasileiro e apresentados por [KATO, 1990], mostram que as diferentes fases da leitura, em um processo de alfabetização, são importantes para o entendimento das variáveis inerentes ao problema. Estes resultados serão apresentados de maneira resumida fazendo-se uma comparação com os métodos de computação desenvolvidos e em desenvolvimento para o reconhecimento de palavras manuscritas.

Inicialmente, para [TEBEROSKY & TOLCHINSKY, 1996] a criança não diferencia a função do texto e da figura, achando que esta última também é lida. A figura é vista com função pictográfica pela criança. [LAVINE, 1972] apud [TEBEROSKY & TOLCHINSKY, 1996] mostra que aos três anos de idade, as crianças rejeitam figuras e desenhos como escrita, porém segundo os autores, ainda associam um valor icônico à escrita. É interessante ressaltar que os sujeitos desses estudos acreditam que objetos grandes são representados por palavras mais longas. Assim, ao escrever-se *casa* e *casinha*, o entendimento ocorrerá achando que é a primeira que representa o conceito da segunda. Nesta fase, o sujeito ainda não atingiu a fase fonográfica, pois ele não tem consciência de que *casinha* é mais longa porque a palavra tem mais sons do que *casa*. Esta fase é chamada de pré-silábica, [TEBEROSKY & TOLCHINSKY, 1996]. Alguns métodos reconhecedores utilizam como informação o tamanho da palavra, porém este tipo de informação não é muito recomendada pelos pesquisadores, uma vez que recomenda-se o uso de características invariantes. Isto devido ao fato de que, ao

comparar-se diferentes palavras de mesmo tamanho pode-se cometer erros no processo de avaliação [TRIER et al., 1996].

Após a fase pré-silábica [TEBEROSKY & TOLCHINSKY, 1996] distinguem três outras fases: a silábica, a silábico-alfabética e a alfabética, a medida que a criança atribui a cada símbolo o valor de uma sílaba, de uma sílaba ou de um segmento fonético e, finalmente, apenas de um segmento fonético. Neste ponto a comparação com os métodos reconhecedores cabem aos que se baseiam na segmentação das palavras em sílabas, letras ou pseudo-letras, tais como: [BOZINOVIC & SRIHARI, 1989], [LECOLINET, 1990], [MOREAU et al. 1991], [LEROUX et al., 1991], [FAVATA & SRIHARI, 1992], [KIMURA et al., 1993], [GILLLOUX, 1994], [HEUTTE, 1994], [CHEN et al., 1995a], [CHO et al., 1995], [MOHAMED, 1995], [CONNELL, 1996], [KNERR et al., 1996], [EL YACOUBI, 1996].

Para [FROMKIN & RODMAN, 1974] apud [KATO, 1990], a criança só pode aprender o sistema alfabético quando cada segmento sonoro de sua língua possuir para ela uma realidade psicológica. Em outras palavras, a alfabetização se torna possível quando a criança tiver consciência da relação símbolo gráfico e som oral de sua língua. Novamente, fazendo um paralelo com os trabalhos de reconhecimento de manuscritos os autores [GUILÉVIC, 1995], [KIM, 1996] e [CÔTE, 1997] procuram inserir, no desenvolvimento de seus métodos, aspectos baseados em modelos psicológicos.

[KATO, 1990] apresenta as diferentes hipóteses de aprendizagem, levando-se em conta as características do sistema ortográfico da língua portuguesa, a saber:

- hipótese fonética,
- hipótese de que há arbitrariedades no sistema ortográfico,
- hipótese de que há regularidades contextuais entre fala e escrita e
- hipótese de que há regularidades oriundas da morfologia.

A cada hipótese de aprendizagem, temos os métodos correspondentes para a efetivação do processo de aprendizagem:

- **Método Global Puro:** supõe que a apreensão do estímulo visual se dá de forma ideográfica, sem análise das partes que o compõem,
- **Método Global Analítico-Silábico:** supõe que a concepção da criança sobre a palavra escrita é que os estímulos são decomponíveis em unidades menores, silábicas, ou que é possível introduzir a ela esta noção,

- **Método Global Analítico-Fonêmico:** pressupõe que essa decomposição é possível de ser feita a nível fonêmico-grafêmico,
- **Método Silábico-Sintético:** supõe que a criança seja capaz de perceber uma entidade mais abstrata que a palavra, a sílaba, e a partir de sua representação grafêmica chegar às unidades significativas como a palavra e a frase,
- **Método Fônico-Sintético:** supõe que a criança seja capaz de captar unidades sonoras físicas, menores que a sílaba (o fone) para, a partir de sua representação grafêmica, chegar às unidades significativas.

Assim, os métodos globais simulam melhor o que ocorre naturalmente quando uma criança ou um adulto se depara com o universo visual que o cerca. Portanto, em [KATO, 1990] a percepção não parte de segmentos isolados, mas sim da discriminação dos traços distintivos dos estímulos que estão sendo observados. Por exemplo, ao observar-se o anúncio da *Coca-Cola*, uma pessoa poderá vir a segmentar a sílaba *co*, dada a sua dupla ocorrência, e na comparação do *co* com o *ca*, poderá vir a diferenciar o *a* do *o* e isolar o *c*, e assim por diante. Uma vez percebidas essas unidades, que em si não têm significado, as pessoas podem combinar as unidades e chegar a unidades significativas.

Com base nestes itens, [KATO, 1990] apresenta que as pesquisas em leitura, principalmente na área de psicologia e psicolinguística, são unânimes em afirmar que, na leitura proficiente, as palavras são lidas não letra por letra ou sílaba por sílaba, mas sim como um todo analisado, isto é, por reconhecimento instantâneo e não por reconhecimento analítico-sintético. Citando, [SMITH, 1978] apud [KATO, 1990], o reconhecimento de palavras se dá como o reconhecimento de outro objeto qualquer, como por exemplo: árvore, carro, casa e, da mesma forma que identifica-se um objeto através de sua configuração geral, pode-se reconhecer uma palavra através do todo (seu contorno e extensão, entre outros) sem análise de suas partes. Conforme já mencionado por [MACHADO & TEIXEIRA, 1996] e apresentado anteriormente, [KATO, 1990] afirma que pode-se identificar uma árvore enxergando-se apenas uma parte de sua copa, a palavra pode ser reconhecida ou adivinhada sem enxergarmos a sua totalidade (Figura 2.10). A leitura de uma palavra por um leitor competente é feita, pois, de maneira ideográfica.

Neste ponto ressalta-se que o trabalho desenvolvido tem por base explorar esta habilidade humana no desenvolvimento de um método reconhecedor de palavras

manuscritas do extenso em cheques bancários. Para [EL YACOUBI, 1996] os métodos reconhecedores devem ser o mais próximo possível da habilidade humana de ler qualquer palavra escrita. Assim, a maioria dos métodos tenta atingir este objetivo trabalhando com as palavras de modo global ou analítico utilizando como unidade geradora de informação o grafema. Considera-se como um grafema uma entidade que pode corresponder a uma parte de uma letra, uma letra ou de letras conectadas. Este conceito está devidamente explorado no Capítulo 4. Portanto, o método desenvolvido baseia-se na Abordagem Global.

[SMITH, 1969 e 1978] apud [KATO, 1990] reconhecem na literatura que, quanto mais eficiente o leitor, maior o seu vocabulário visual, cerca de 50.000 itens, e que o processo de análise e síntese da palavra em unidades menores só será usado para itens estranhos a esse universo. Os autores acreditam ainda, que raramente o leitor competente chega a analisar tais palavras, pois o contexto normalmente fornece “pistas” suficientes para seu entendimento.

Para um leitor iniciante, porém, cujo vocabulário visual ainda é muito limitado, mesmo para aqueles alfabetizados pelo método global, o processo de leitura envolve muito pouco reconhecimento visual instantâneo, consistindo a leitura, mais freqüentemente, em operações de análise e síntese, sendo apreensão do significado mediada quase sempre pela decodificação em palavras auditivamente familiares [KATO, 1990].

As operações de análise e síntese, depois de certo tempo, podem basear-se em segmentos maiores que a letra: sílabas, morfemas e, acredita [KATO, 1990], também em partes maiores da palavra que não constituem unidades lingüísticas, mas que tem uma incidência muito grande na língua, como é o caso do segmento *-ola* que aparece em palavras como *bola*, *mola*, *cebola* e muitas outras. Trata-se, pois de uma operação de decomposição e composição, que envolve também o reconhecimento instantâneo, porém parcial, da palavra.

No caso do vocabulário utilizado para o preenchimento do valor por extenso de um cheque existem três classes de palavras que compartilham a característica do segmento *-ola*, a classe de palavras terminadas por “-enta” e a classe de palavras terminadas por “-entos” e a classe de palavras iniciadas por “dez-”.

Para [KATO, 1990] se uma palavra for apenas reconhecida parcialmente, a leitura da parte estranha exigirá decomposições sucessivas até a identificação formal

total da palavra. O conhecimento prévio do leitor quanto às restrições fonotático/ortográficas e seu vocabulário auditivo podem atuar para minimizar esse trabalho de análise e síntese.

Resumidamente, é como se cada pessoa trabalhasse ora como um interpretador, ora como um compilador. O método desenvolvido segue este mesmo formato trabalhando, inicialmente, com um conjunto de primitivas globais composto por primitivas perceptivas (*perceptual features*) e posteriormente, com um conjunto de primitivas geométricas extraídas sobre o corpo das palavras.

A seguir apresenta-se uma revisão dos métodos e técnicas utilizados na área de Reconhecimento de Palavras manuscritas – RPM relacionados com aplicações de cheques bancários.

2.4. Reconhecimento Estático de Palavras Manuscritas

O reconhecimento de palavras manuscritas, como já mencionado, possui diversas variáveis e estratégias de abordagens do problema. Apresenta-se um resumo das variáveis e estratégias possíveis na busca de soluções, ficando evidente que as opções existentes dependem muito da aplicação, ou seja, do contexto em que as palavras manuscritas estão inseridas. Evidencia-se também, o caminho geralmente considerado pelos métodos de reconhecimento de palavras manuscritas já desenvolvidos: aquisição, pré-processamento, segmentação, extração de primitivas e reconhecimento.

2.4.1. Restrições dos Métodos de Reconhecimento

Sabe-se que o estudo de palavras manuscritas caracteriza-se como um problema difícil pela própria natureza da escrita cursiva, conforme mencionado anteriormente, uma vez que as mesmas são muito instáveis na sua forma e qualidade do traçado.

Sete fatores principais permitem julgar a complexidade do problema de reconhecimento da escrita cursiva:

- tipo da escrita;
- tipo de aquisição da imagem: estática (*on-line*) ou dinâmica (*off-line*);
- o número de escritores e a aprendizagem ou não do estilo do escritor e
- o tamanho do léxico;

- qualidade da imagem;
- fatores humanos e
- outros fatores, tais como: linhas guia e linhas de separação.

Estes fatores encontram-se apresentados a seguir, visando uma melhor caracterização do problema do reconhecimento de palavras manuscritas, além de determinarem as diferentes operações a serem efetuadas, os métodos de reconhecimento a utilizar, permitindo a organização de um método. Apresenta-se, também as características da aplicação de interesse do presente trabalho.

2.4.1.1. Tipo de Aquisição da Imagem

Os métodos de reconhecimento distinguem-se entre si a partir do tipo de aquisição da imagem. De acordo com [BUNKE et al., 1995], o reconhecimento pode ser dito:

- Reconhecimento dinâmico ou *on-line*: quando a imagem é adquirida dinamicamente, ou seja, no momento da escrita por processo de mesa digitalizadora, papel eletrônico ou caneta ótica;
- Reconhecimento estático ou *off-line*: quando a imagem é adquirida estaticamente, ou seja, após a escrita por processo de *scanner*.

No reconhecimento dinâmico o sinal é constituído de informação temporal, dinâmica e uniforme. O escritor é parte do processo e trabalha como um supervisor da escrita, pois pode identificar e corrigir eventuais erros. Desta forma, a natureza da informação extraída a partir da aquisição facilita o processo de reconhecimento.

Por outro lado, no reconhecimento estático o escritor é ausente. O sinal obtido é bi-dimensional, não contém informação temporal e o traçado possui espessura variável. Este processo de aquisição pode gerar ruídos na imagem, os quais deverão ser tratados na fase de pré-processamento da imagem. Observa-se que dependendo do tipo de ruído gerado ou do processo de pré-processamento empregado nas imagens, a possibilidade de corrigir um erro praticamente não existe, e na maioria das vezes, a imagem poderá ser simplesmente rejeitada.

As imagens adquiridas a partir do documento original podem gerar imagens binárias (branco de preto), em níveis de cinza e coloridas, dependendo do documento original. A resolução da imagem adquirida é outro fator de influência na qualidade da mesma, pois ao se efetuar a aquisição de uma imagem em 100, 200 ou 300 *dpi* o

resultado será diferente. Sabe-se também que uma melhor resolução gera imagens de melhor qualidade, porém necessitam de maior espaço de armazenamento e tempo de processamento.

2.4.1.2. Tipos de Escrita

Diversos autores distinguem duas categorias principais de tipos de escrita: a escrita bastão derivada dos modelos de caracteres impressos, e a escrita cursiva. Para [TAPPERT et al., 1990] existem 5 categorias de escrita, situadas entre as duas principais categorias (caixa alta e cursiva pura), conforme Figura 2.11, classificadas por ordem crescente de complexidade:

- palavras em caracteres disjuntos contidos em retângulos pré-impressos (caixa alta);
- palavras em caracteres disjuntos com espaçamento regular;
- palavras em caracteres disjuntos com a presença de vínculos entre eventuais caracteres;
- palavras em escrita cursiva pura, ou seja, todos os caracteres de uma palavra são conectados dois a dois;
- palavras em escrita mixada, ou seja, misturando os demais tipos de escrita.

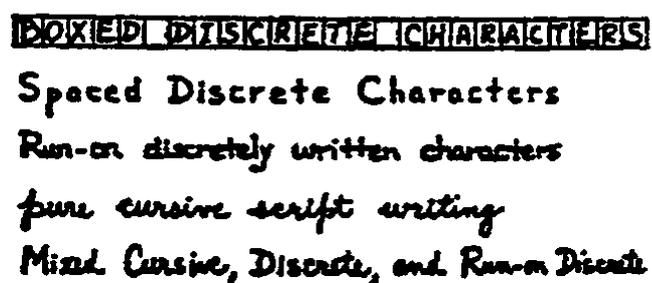


Figura 2.11: Tipos de escrita [TAPPERT et al., 1990]

Observa-se que no primeiro tipo a escrita é forçada, enquanto que nos demais casos nenhuma restrição é imposta ao escritor. A variabilidade de formas é a característica mais importante.

2.4.1.3. Número de Escritores

A variabilidade das formas a serem consideradas pelos métodos depende então do número de escritores. Três categorias de métodos são encontrados em função do número de escritores:

- mono-escritor: um único escritor,
- multi-escritor: número limitado de escritores e
- omni-escritor: um número muito grande de escritores diferentes.

Para as duas primeiras categorias, os métodos podem aprender o estilo do escritor, ao contrário da última categoria, onde a aprendizagem de um estilo de escrita é impossível, pois a variabilidade de estilos é diretamente proporcional ao número de escritores.

2.4.1.4. Tamanho do Léxico

A maioria dos métodos de reconhecimento da escrita cursiva tem como base um dicionário de palavras. Os métodos visam, então, encontrar uma solução entre uma lista de palavras contidas no léxico ou rejeitar a palavra recebida como entrada. Nestas condições, quanto maior o tamanho do léxico, maior a complexidade e maior a lista de soluções possíveis, visto que existe uma probabilidade maior de existirem palavras parecidas em um léxico de grandes dimensões. O inconveniente destes métodos está no insucesso do reconhecimento de novas palavras (não contidas no léxico). Os métodos podem ser melhorados ou aprimorados, mas ainda não existe uma solução geral para o problema. Desde modo, pode-se simplificar a tarefa dividindo o problema em sub-problemas. Então, o mais complexo será o reconhecimento de palavras do tipo mixada (cursiva e discreta), obtida a partir de um grande número de escritores que utilizam um léxico de pequeno tamanho.

Para [CÔTE, 1997] pode-se considerar o léxico como sendo:

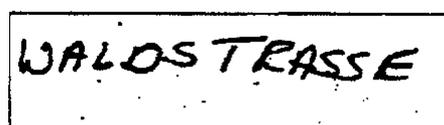
- pequeno: quando contém de 1 a 100 palavras, sendo que as aplicações mais freqüentes de léxico com pequeno tamanho são as que possuem um número muito grande de escritores diferentes [omni-escritores];
- médio ou intermediário: quando contém de 100 a 500 palavras e
- grande: quando contém mais de 500 palavras.

A precisão contida no léxico também exerce influência sobre o reconhecimento. Assim, um léxico composto de palavras curtas onde as formas gerais são similares é de difícil tratamento (por exemplo: “sete” e “oito”, “mil” e “real”) quando comparado com palavras que são longas e possuem estruturas bem diferentes (por exemplo; “quarenta”, “trezentos”, “quinhentos”).

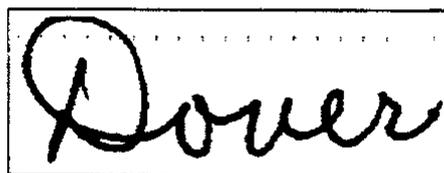
2.4.1.5. Qualidade da Imagem

A qualidade da imagem é determinada por diferentes fatores relacionados entre si, a saber:

- a) **Ruídos gerados pelo processo de aquisição da imagem:** As imagens são coletadas através de *scanner* de forma única, ou seja, com uma única configuração (contraste, brilho, intensidade de pixels, entre outros) independente da imagem a ser coletada. A geração de ruídos esparsos é frequente neste processo. Existe ainda a possibilidade de linhas tracejadas ou pontilhadas geradas pelo mecanismo de alimentação automática de folhas no *scanner*. A incorreta operação do *scanner* pode gerar grandes defeitos na imagem resultante do processo de *scanner*. A Figura 2.12 exemplifica os ruídos gerados pelo processo de aquisição da imagem [KIM, 1996];



(a)



(b)



(c)

Figura 2.12: Exemplos de ruídos gerados no processo de aquisição da imagem: a) ruído “sal e pimenta”, b) ruído gerado pelo alimentador e c) incorreta operação do *scanner* [KIM, 1996]

- b) **Qualidade do meios utilizados na escrita (caneta e papel):** As características da caneta e do papel são fatores importantes na qualidade da imagem resultante, como mostrado na Figura 2.13, onde o exemplo (a) foi obtido sobre papel vegetal e o exemplo (b) sobre papel sulfite branco. As cores e o padrão de fundo do papel causam problemas relacionados as características dos dispositivos envolvidos no processo de *scanner* e algoritmos. Após a obtenção da imagem, texturas tornam-

se padrões de fundo extremamente ruidosos. Isto ocorre em envelopes decorados (quando do processamento da parte manuscrita de endereços postais) e em cheques bancários (quando do processamento da parte manuscrita de cheques). Entretanto, padrões de fundo complexos com muitos níveis de cinza ou cores não conseguem ser eliminados pela maioria dos processos de binarização global. A espessura da caneta e a pressão aplicada na escrita determinam a espessura dos traços e suas variações. Além disso, a textura do papel e qualidade da caneta podem gerar espaços em branco nos traços e bolhas nos contornos;

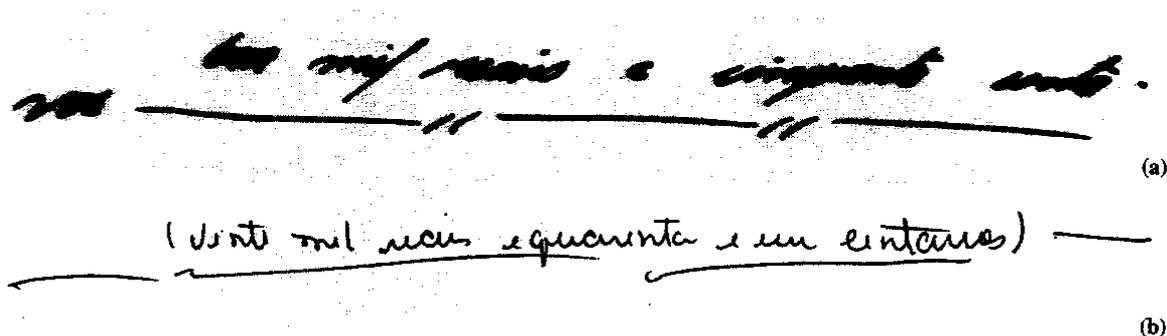


Figura 2.13: Exemplos da qualidade de imagem associado ao meio utilizado:
a) papel vegetal e b) papel sulfite branco

c) **Processo de binarização:** A maioria dos programas que processam imagens de manuscritos são projetados para trabalhar sobre imagens binárias [LIU & SRIHARI, 1994] e [OTSU, 1979]. Processos de binarização inteligentes e dinâmicos são procurados para poder levar em conta não somente a intensidade dos traços mas também a intensidade do fundo. A Figura 2.14 apresenta alguns exemplos de diferentes processos de binarização das imagens, demonstrando ser esta etapa importante para a qualidade das imagens a serem trabalhadas.



Figura 2.14: Exemplos de processo de binarização: a) imagem original em níveis de cinza, b) $threshold = 217$ (método anisotropia) e c) $threshold = 110$ (método OTSU)

2.4.1.6. Fatores Humanos

As características de um manuscrito são determinadas na maioria das vezes pelo escritor. Pode-se citar ainda o estilo e adicionar um número de fatores relevantes tais como:

a) **Inclinação vertical:** escrever inclinado é muito comum, especialmente na escrita cursiva, apesar dos escritores iniciarem a aprendizagem da escrita, na escola, sem inclinação. A inclinação horizontal da escrita (da linha de base) é denominada *skew* e a inclinação vertical é denominada *slant*, ambos os casos estão exemplificados na Figura 2.15. Além do escritor, a inclinação horizontal pode ser inserida a partir das fontes de aquisição da imagem, por exemplo, quando o papel é colocado no *scanner*;

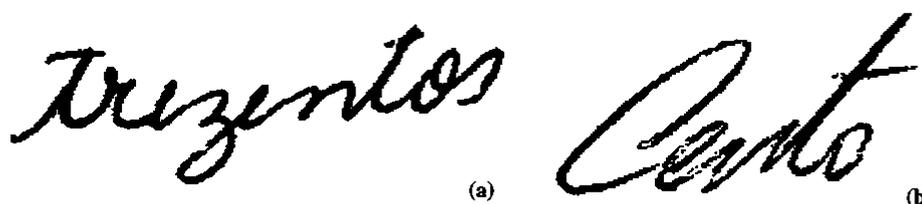


Figura 2.15: Inclinação: a) da linha de base (*skew*) e b) vertical dos caracteres (*slant*)

b) **Traços contíguos/desconectados:** Alguns caracteres escritos cursivamente e a maioria dos caracteres discretos são formados por mais de um traço, daí a possibilidade de traços desconectados. Traço(s) de um caracter podem ser posicionados erroneamente ou inteiramente esquecidos. Os pingos dos caracteres i e j ou a barra de corte do caracter t, são alguns exemplos. Além disso, traços exagerados podem tocar outros caracteres. E ainda, traços podem ser fragmentados durante o processo de binarização resultando em pedaços de traços desconectados. Considerando que o caracter é a unidade básica na maioria dos processos de reconhecimento, agrupamentos de traços e cortes de traços são importantes. A Figura 2.16 mostra alguns desses problemas;

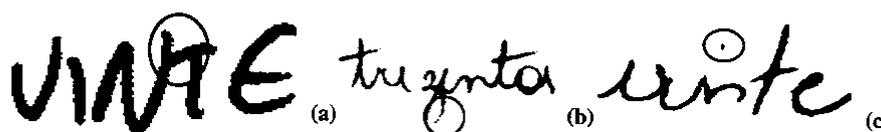


Figura 2.16: Exemplos de: a) traços contíguos, b) traços desconectados e c) posição errônea

c) **Erros de ortografia:** Este é ainda um outro fator que torna o reconhecimento difícil, especialmente, quando os paradigmas do reconhecimento levam vantagem da informação contextual. O uso de abreviações de palavras é outro problema similar. No caso de cheques, as abreviações não são utilizadas, nem tão pouco autorizadas. Porém, os casos com erros de ortografia são comumente encontrados, conforme exemplificado na Figura 2.17. Outros casos são as palavras que possuem diferente ortografia, por exemplo: “cinquenta” e “cincoenta”, conforme Figura 2.18, e ainda, “catorze” e “quatorze”, “três” e “treis”, “um” e “hum”. Para estes casos considera-se um único modelo para as palavras. Porém os demais casos de erros ortográficos, por exemplo: “dose”, “quatrocentos”, “trese”, “cesenta” e muitos outros encontrados na base de dado, não recebem tratamento específico;

(oitocentos e quatro mil , seiscientos e deze-
(a)

Trezentos e trinta e nove mil , quinhentos e sesenta
(b)

Figura 2.17: Erros de ortografia: a) AE0217a e b) AE0233a

cinquenta cincoenta

Figura 2.18: Ortografia diferenciada para uma mesma palavra

2.4.1.7. Outros Fatores

Na maioria das aplicações reais, a extração de unidades manuscritas é uma tarefa não trivial. A sobreposição das linhas guias com as respostas e a interferência entre linhas escritas consecutivamente aumentam a complexidade do problema, como descrito a seguir:

a) **Linhas guia:** Formulários, envelopes pré-impresos e cheques bancários fornecem frequentemente linhas guia para eliminar a possibilidade de inclinação horizontal. Porém, as imagens binárias não fazem a diferenciação entre as linhas guia e o traçado dos caracteres em termos de suas intensidades. Algoritmos devem

ser desenvolvidos para remover as linhas guia [SANTOS, 1997]. O processo de eliminação, na maioria da vezes, introduz artifícios, tais como a eliminação de alguns traços-chaves [KOERICH, 1997].

b) Linhas de separação: A presença de ascendentes e descendentes na escrita cria a possibilidade de sobreposição vertical entre as linhas. Os algoritmos para separação de linhas geram, na maioria dos casos, segundo [KIM, 1996], resultados imperfeitos.

2.4.1.8. Características da Aplicação

A aplicação de interesse do presente estudo, relacionada com o reconhecimento do extenso manuscrito em cheques bancários, tem por características:

- **Tipo de aquisição:** estática, através de *scanner*;
- **Tipo de escrita:** pode-se encontrar nas bases de dados qualquer um dos tipos definidos por Tappert em [TAPPERT et al., 1990];
- **Número de escritores:** omni-escritores. No caso das bases de dados do presente estudo, considerou-se um escritor diferente para cada cheque preenchido;
- **Tamanho do léxico:** pequeno (39 palavras);
- **Qualidade das imagens:** os ruídos provenientes do processo de aquisição das imagens em 300 *dpi* foram minimizados através de uso de *scanner* de alta resolução e processo cuidadoso de escolha dos parâmetros de aquisição. Quanto a qualidade dos meios utilizados na escrita (caneta e papel) não foram feitas restrições ao tipo de caneta, sendo o papel sulfite branco;
- **Processo de binarização:** as imagens foram binarizadas através do método OTSU [OTSU, 1979];
- **Inclinação vertical:** a inclinação vertical existente nas imagens adquiridas sendo somente proveniente do estilo de escrita dos escritores;
- **Traços contíguos/desconectados:** podem ser encontrados nas bases de dados dos cheques de laboratórios traços contíguos ou desconectados, bem como posicionamento errôneo de traços de corte ou pingos;
- **Erros de ortografia:** em primeira instância foram considerados os casos especiais de ortografia já conhecidos e esperados “um” e “hum”, “três” e “treis”, “catorze” e “quatorze”, “cincoenta” e “cinquenta”. Cada modelo

engloba todas as ortografias encontradas na base dados de treinamento, não havendo exclusão de exemplares;

- **Linhas guia e linhas de separação:** estes problemas não foram considerados sendo tratados na fase de extração das informações manuscritas dos cheques.

2.4.2. Etapas de um Método de Reconhecimento

Aborda-se as etapas necessárias ao reconhecimento de palavras manuscritas, a saber: aquisição, pré-processamento, segmentação, extração de primitivas, reconhecimento.

2.4.2.1. Aquisição das Imagens

A etapa de pré-processamento das imagens adquiridas por *scanner* reveste-se de importância, uma vez que prepara as imagens para as demais etapas, ou seja, a correção da linha de base, a correção da inclinação vertical das palavras, a normalização do corpo de letras minúsculas e a segmentação e extração de primitivas.

Inicialmente a fase de pré-processamento relaciona-se com a generalização dos parâmetros para obtenção das imagens através do processo de *scanner*, no que diz respeito ao brilho e contraste. Isto posto, devido a grande quantidade de imagens a serem adquiridas, sendo portanto inviável o ajuste do equipamento de *scanner* a cada imagem ou cheque a ser adquirido. Os autores de um modo geral, pouco se referem a estes parâmetros, porém a boa seleção destes parâmetros visa reduzir os ruídos nas imagens adquiridas, bem como facilitar o processo de limiarização das imagens adquiridas em níveis de cinza.

Um segundo passo, então, é a limiarização, ou seja, para o caso das imagens em estudo, a passagem de uma imagens em 256 níveis de cinza para 2 níveis: branco e preto. Existem muitos algoritmos de morfologia matemática que são empregados neste procedimento: Anisotropia, Entropia, Kittler, Otsu e ainda, escolha manual do limiar. [CHEN et al, 1995a] determinam o valor ótimo de limiar considerando a criação de uma função de regressão sendo conhecido o histograma do mapa de probabilidade da imagem. [MOHAMED, 1995] utiliza o procedimento de Otsu para binarização das imagens da base de dados CEDAR, com nomes de cidades, estados e códigos de endereçamento postal (CEP ou *ZIP code*). A escolha entre os diversos procedimentos

ficará a cargo do pesquisador, levando-se em conta o processo de aquisição da imagens e os ruídos gerados pelo processo.

Será necessário, então, reduzir os ruídos gerados nas imagens. Os trabalhos apresentados por [CHEN et al., 1995a] e [BOZINOVIC & SRIHARI, 1989] utilizam as técnicas de morfologia matemática e procedimentos de suavização das imagens, respectivamente, para redução de ruídos. [KIM, 1996] realiza a redução de ruídos através da identificação de pixels isolados nas imagens e rejeita as imagens com fundo extremamente ruidoso. A rejeição é realizada através da análise do número e do tamanho dos componentes conectados na imagem e do número de componentes no contorno exterior da palavra. [GUILLEVIC, 1995] trabalha com a remoção ou inserção de pixels isolados e de conjuntos formados por mais de um pixel, através de operações de suavização. [BUSE, 1997] utiliza o Filtro de Gabor para trabalhar com as imagens em níveis de cinza, eliminando desde modo as operações de binarização, esqueletização e normalização de tamanhos.

As ferramentas de morfologia matemática são muito utilizadas nesta etapa, tendo sido empregadas também por: [HEUTTE, 1994], [BERTILLE et al., 1994], [CHO et al., 1995], [FILATOV et al., 1995] e [MOHAMED, 1995]. A maior dificuldade encontrada nesta etapa é o número de parâmetros que devem ser adotados e aplicados de maneira geral, ou seja, de forma automatizada e não caso a caso.

2.4.2.2. Pré-Processamento

O pré-processamento das imagens é composto, geralmente, pelas seguintes etapas:

- Correção da linha de base (*skew*),
- Correção da inclinação vertical dos caracteres (*slant*) e
- Detecção e normalização do corpo das palavras.

a) Correção da Linha de Base

A correção da linha de base da escrita visa detectar o ângulo de inclinação com a horizontal e corrigir este ângulo de tal forma que a escrita se torne horizontal. Esta inclinação pode ser gerada pela inserção do documento inclinado no equipamento de aquisição de imagem ou ser uma inclinação natural do estilo de escrita do escritor em questão.

De um modo geral, os métodos aplicados na correção de linha de base podem ser globais ou locais. Os métodos globais realizam uma estimativa da inclinação da palavra considerando que a inclinação é válida para a palavra como um todo. Por outro lado, os métodos locais consideram que a inclinação horizontal das palavras não é constante e igual para toda a palavra, realizando assim pequenas correções localizadas.

A maioria dos métodos de correção da inclinação da linha de base tem por princípio tratar a inclinação como um valor constante ao longo do texto ou mesmo de uma palavra. Os resultados obtidos por diversos autores demonstram que a solução global é suficiente para palavras isoladas na maioria dos casos: [KIMURA, 1993], [GUILLEVIC, 1995], [MOHAMED, 1995], [EL YACOUBI, 1996], [MORITA, 1998].

O presente trabalho não aplica nenhum procedimento de correção da linha de base, uma vez que o preenchimento do extenso nos cheques conta com duas linhas de referências auxiliares.

b) Correção da Inclinação Vertical dos Caracteres

Entende-se por inclinação vertical dos caracteres o ângulo formado entre o eixo correspondente a direção de escrita dos caracteres e o eixo vertical. O objetivo da correção desta inclinação é tornar a palavra o mais vertical possível, visando reduzir a variabilidade da escrita e, posteriormente, simplificar a etapa de reconhecimento, uma vez que as primitivas a serem detectadas, em sua grande maioria, são sensíveis a essa inclinação. As técnicas mais simples supõem que esta inclinação é uniforme para uma mesma palavra, realizando uma correção global da inclinação. Outras técnicas trabalham isolando segmentos de palavra e efetuando a correção individual de cada segmento [BOZINOVIC & SRIHARI, 1989]. A Tabela 2.1 apresenta alguns dos métodos aplicados por diferentes autores e suas características.

A Figura 2.19 apresenta as etapas de correção da inclinação vertical desenvolvida por [BOZINOVIC & SRIHARI, 1989]. A Figura 2.20 mostra os resultados obtidos pelo método de [BUSE et al., 1997].

A técnica desenvolvida por [BUSE et al., 1997] apresenta a desvantagem de que a inclinação de um caracter maiúsculo ou de grandes proporções, no que diz respeito à altura, tende a influenciar o cálculo pois este contém uma grande quantidade de pixels. Porém, o efeito é mínimo uma vez que o autor assume a inclinação como uniforme para toda a palavra.

A técnica de [GUILLEVIC, 1995] mostra que para imagens com pequena inclinação não ocorrem mudanças significativas através da operação de correção. Por outro lado, a qualidade dos resultados diminui à medida que o ângulo de inclinação aumenta. De fato, o que ocorre é que a correção da altura da palavra não é significativa enquanto a largura da imagem provavelmente é alterada. O pior caso observado na base de dados do autor é o caso de um extenso escrito com pouca altura e com ângulo de inclinação de 60 graus, conforme Figura 2.21. Outro problema não resolvido é o caso de traços longos horizontais com inclinação diferente de 0 grau, uma vez que estes traços influenciam a determinação da inclinação vertical da palavra, como mostrado na Figura 2.22.

Tabela 2.1: Métodos de Correção da Inclinação Vertical

Método	Autores	Características
Global	[KIMURA et al., 1993]	Contorno externo, considerando os pixels dos segmentos de borda que não são horizontais
	[GUILLEVIC, 1995]	Histogramas de projeção horizontal
	[MOHAMED, 1995] e [EL YACOUBI, 1996]	Morfologia matemática
	[KNERR, 1996]	Baseado no contorno sugerido por [KIMURA, 1993]
	[KIM & GOVINDARAJU, 1997]	Extração das linhas verticais e oblíquas do contorno
	[BUSE et al., 1997]	Entropia mínima do histograma de projeção vertical e método do domínio da frequência
Local	[BOZINOVIC & SRIHARI, 1989]	Cálculos de percursos com parâmetros heurísticos

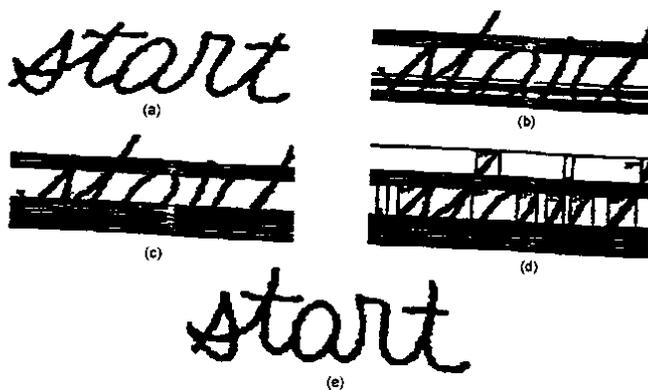


Figura 2.19: Correção da inclinação vertical dos caracteres: a) imagem original; b) remoção das linhas horizontais de percurso longo, c) remoção das faixas de pequeno comprimento, d) janelas isoladas e faixas conectando o centro de gravidade e ângulos da inclinação mostrados a cada janela, e) imagem corrigida [BOZINOVIC & SRIHARI, 1989]



Figura 2.20: Correção através da aplicação do histograma de projeção vertical: a) imagem limiarizada a partir da imagem original, b) histograma de projeção vertical a partir da imagem contida em (a), c) imagem corrigida, d) histograma de projeção vertical da imagem contida em (c) [BUSE et al., 1997]

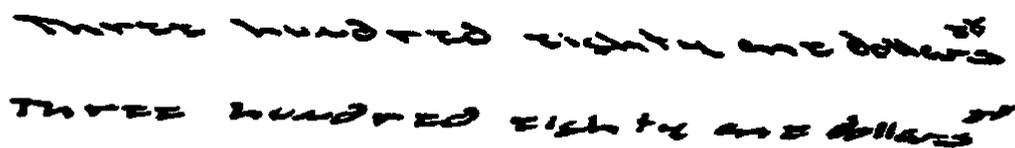


Figura 2.21: Correção de uma imagem com inclinação de 60 graus [GUILLEVIC, 1995]



Figura 2.22: Exemplo de falha na correção da inclinação vertical dos caracteres [GUILLEVIC, 1995]

A Figura 2.23 apresenta a técnica aplicada por [KIMURA et al., 1993]. O algoritmo de [KIM & GOVINDARAJU, 1997] assume que o ângulo da inclinação vertical das palavras está entre -45° e 45° , sendo um exemplo de resultado apresentado na Figura 2.24.

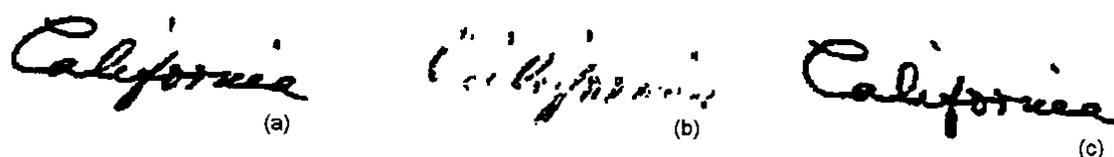


Figura 2.23: Correção da inclinação dos caracteres: a) imagem original, b) pixels de bordas não horizontal, c) imagem corrigida [KIMURA et al., 1993]

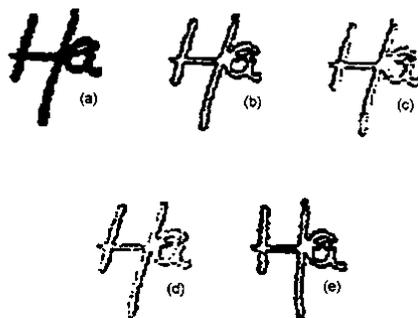


Figura 2.24: Correção vertical da inclinação dos caracteres: a) imagem original, b) contorno, c) e d) linhas verticais inferiores e superiores, e) imagem corrigida com base no ângulo estimado a partir das linhas detectadas em (c) e (d) [KIM & GOVINDARAJU, 1997]

Em [EL YACOUBI, 1996] utiliza-se o histograma de projeção horizontal que por sua vez utiliza a transformada de *Hough*. Gera-se um conjunto de retas de inclinações diferentes, com a mesma origem, com orientações regularmente espaçadas. Este conjunto define o plano delimitado pela primeira bissetriz (45 graus) e pela segunda bissetriz (-45 graus). Cada reta do conjunto possui um ângulo de inclinação com a horizontal, sendo cada ângulo múltiplo de um ângulo elementar escolhido com valor bem pequeno ($\pi/80$). O objetivo é fornecer precisão na representação do espaço das inclinações possíveis em uma palavra. Posteriormente, calcula-se o histograma de projeção dos segmentos dos pixels pretos em cada uma das retas definidas. A Figura 2.25 apresenta um exemplo do procedimento proposto.

A vantagem do procedimento descrito em [EL YACOUBI, 1996] é que ele funciona para todo os tipos de escritas e, além disso, pode estimar globalmente a inclinação média dos caracteres, pois é fácil calcular a inclinação de um caracter isolado. A grande desvantagem é o tempo de processamento.

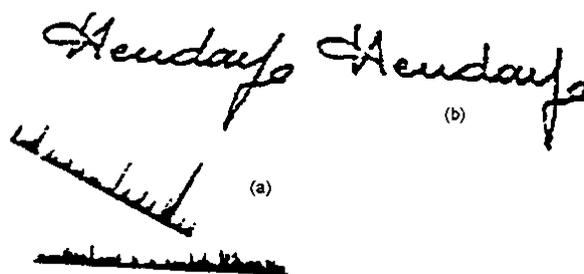


Figura 2.25: Correção da inclinação vertical: a) histograma de projeção vertical de duas retas de ângulos: 30 graus e 0 graus e b) correção da inclinação [EL YACOUBI, 1996]

Para compensar a desvantagem do tempo de processamento [EL YACOUBI, 1996] propôs outra metodologia semelhante a utilizada por [KIM & GOVINDARAJU, 1995]. A diferença está em não realizar uma nova detecção do contorno superior, aproveitando o contorno utilizado na detecção da linha de base. [EL YACOUBI, 1996] comenta que este método é menos preciso que o método dos histogramas, porém apresenta um tempo de processamento reduzido.

Muitas são as metodologias de correção da inclinação vertical dos caracteres que apresentam bons resultados. Ressalta-se, portanto, que a aplicação dessas diferentes metodologias tem como maiores limitações o emprego de limiares empíricos e o tempo de processamento.

c) Detecção e Normalização do Corpo das Palavras

Entende-se por corpo das palavras a parte das palavras cursivas formadas pelas letras minúsculas. Desde modo, a detecção e normalização do corpo das palavras é tarefa importante a ser realizada com o objetivo de auxiliar no reconhecimento das mesmas.

As técnicas diferenciam-se no fato de considerar ou não o corpo das palavras uma faixa necessariamente horizontal. A Tabela 2.2 apresenta alguns métodos utilizados para detecção do corpo das palavras por diferentes autores.

Tabela 2.2: Métodos de Detecção do Corpo das Palavras

Método	Autores	Características
Horizontal	[BOZINOVIC & SRIHARI, 1989]	Histograma de projeção horizontal com busca de valores máximos
Não Horizontal	[BERCU & LORETTE, 1993]	Extremos locais dos traços de direção vertical
Horizontal	[EL YACOUBI, 1996]	Segmentos de reta formados pelos pontos máximos do contorno externo, transformação não linear
Não Horizontal	[CÔTE, 1997]	Entropia

[CÔTE, 1997] propõe um método para detecção do corpo das palavras com base em histogramas do contorno das palavras selecionados por Entropia. A Figura 2.26 mostra o resultado obtido por [CÔTE, 1997]. O procedimento apresenta algumas dificuldades, uma vez que utiliza a heurística para determinar qual área do histograma corresponde ao corpo das palavras.

A técnica de [BERCU & LORETTE, 1993] tem o inconveniente de considerar a linha superior como uma reta não necessariamente horizontal, Figura 2.27. [EL YACOUBI, 1996] comenta que esta hipótese está longe de ser verificada no caso da escrita manuscrita sem restrições.

A técnica desenvolvida por [EL YACOUBI, 1996] é semelhante a técnica que corrige a linha de base. A desvantagem da técnica é o fato de utilizar limiares empíricos, porém a vantagem deste método é não necessitar de uma segmentação prévia das palavras. A Figura 2.28 mostra o resultado obtido com a palavras "Rennes".

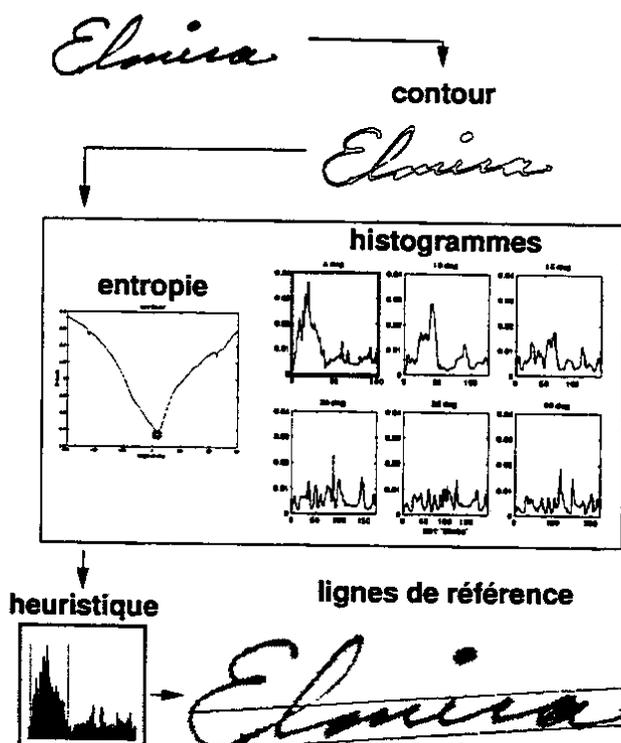


Figura 2.26: Determinação do corpo de palavras [CÔTE, 1997]

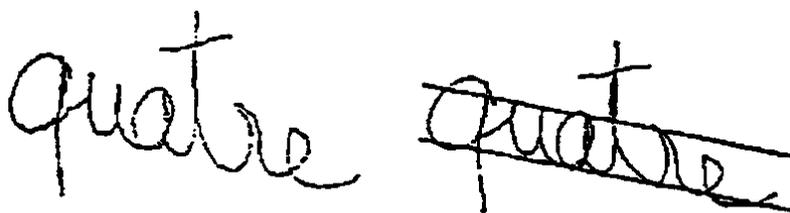


Figura 2.27: Estimação das linhas de referência do corpo das minúsculas [BERCU & LORETTE, 1993]

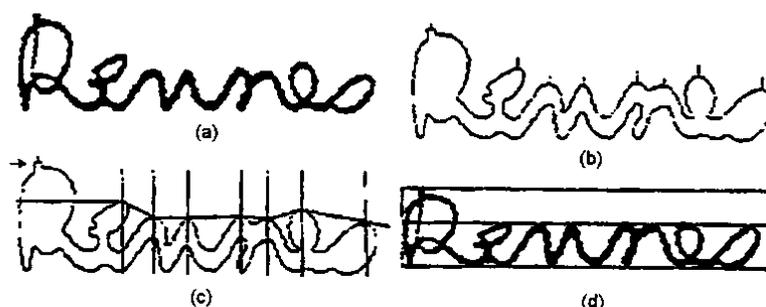


Figura 2.28: Normalização do corpo das minúsculas: a) imagem original, b) detecção dos máximos locais do contorno superior, c) filtragem dos máximos indesejáveis, d) imagem normalizada utilizando uma transformação não linear [EL YACOUBI, 1996]

A detecção das linhas de referência do corpo das palavras não é somente importante para a etapa de normalização do corpo das palavras, mas sim também para a determinação das zonas principais de uma palavra. Estas zonas são normalmente 3 ou mais áreas, denominadas geralmente de: ascendente, corpo e descendente. Nestas áreas são extraídas as primitivas que caracterizam as palavras propriamente ditas, utilizadas posteriormente no processo de reconhecimento.

2.4.2.3. Estratégias de Segmentação

O reconhecimento de palavras manuscritas tem base em 3 estratégias de segmentação:

- estratégia local, analítica ou baseada em caracteres (*Analytical Approach*),
- estratégia global (*Global Approach*) e
- estratégia híbrida.

a) Estratégia Analítica ou Baseada em Caracteres

Esta estratégia envolve a segmentação das palavras em letras ou pseudo-letras e então aplicar um reconhecedor de caracteres diretamente a cada segmento. Entretanto, determinar significativamente os pontos de segmentação sem considerar a entrada de dados a partir do reconhecimento de caracteres é quase impossível. Dependendo do número de estágios na segmentação e da metodologia de reconhecimento a ser utilizada, as estratégias de segmentação baseadas em caracteres podem ser classificadas como a seguir:

- Segmentação Fixa-Uniforme: a palavra é fragmentada em intervalos fixos gerando uma série de pontos de fragmentação. Infelizmente, o

intervalo “ideal” não pode ser determinado facilmente. Portanto, a palavra sofre, normalmente, uma segmentação excessiva e depende de um processo adequado de agrupamento de segmentos durante o reconhecimento [CHEN & KUNDU, 1993] e [KIMURA et al., 1993];

- **Segmentação Heurística:** supera alguns dos inconvenientes da segmentação fixa-uniforme. As ligações e concavidades nas imagens das palavras são usadas como característica básica para determinar os prováveis pontos de segmentação. A espessura média dos traços e o número de caracteres são utilizados como indicações dos pontos de segmentação. Este método minimiza o número de pontos de segmentação e reduz a complexidade da pesquisa. Porém, através deste método é difícil segmentar caracteres contíguos e certos pares de caracteres [FAVATA & SRIHARI, 1992];
- **Segmentação Multi-Estágios:** o módulo de reconhecimento e a informação contextual (tais como a estatística obtida a partir do léxico) são utilizados em uma segmentação prévia. O tamanho do léxico pode ser reduzido durante os primeiros estágios (por exemplo, baseado em primitivas geométricas [KIMURA et al., 1993]). Estágios subsequentes geram os pontos de segmentação mais precisamente [BOZINOVIC & SRIHARI, 1989].

b) Estratégia holística: Este enfoque utiliza algoritmos que tratam a palavra como um todo. Alguns exemplos de primitivas extraídas nesta estratégia são: número de traços verticais, ascendentes, descendentes e laços. Esta estratégia supera a dificuldade de segmentação da palavra. [HARMON, 1962] implementou um leitor de manuscritos holístico. [FARAG, 1979] propôs um reconhecimento de palavras utilizando a seqüência da direção dos traços como primitivas. [HULL, 1988] e [MADHVANATH & GOVINDARAJU, 1992] apresentam resultados experimentais onde a estratégia holística é aplicada. Entretanto, encontram-se dificuldades na seleção de primitivas holísticas apropriadas a diferenciação entre palavras de um léxico de grande tamanho.

c) Estratégia híbrida: Esta estratégia é uma combinação das duas outras estratégias apresentadas. Primitivas holísticas são utilizadas no estágio prévio de redução do léxico. Nos estágios seguintes, reconhecedores baseados na segmentação operam para reduzir o léxico. Esta estratégia apresenta alta *performance* em termos de

velocidade e precisão para aplicações com léxico de grande tamanho [MONTOLIU, 1995].

2.4.2.4. Extração de Primitivas

Esta seção apresenta os conjuntos de primitivas selecionados por diversos autores que influenciaram a escolha das primitivas relativas ao presente trabalho. Inicialmente, apresenta-se os detalhes relevantes destes trabalhos e, posteriormente, realiza-se uma discussão sobre a busca e seleção dos conjuntos de primitivas utilizados: Perceptivas (*Perceptual Features*) e Geométricas.

Todos os estudos de reconhecimento de padrões e mais especificamente de palavras manuscritas tem seu ponto forte na seleção de conjuntos de primitivas capazes de representar e discriminar as diferentes formas encontradas. A tarefa de escolher primitivas ditas “ótimas” não é uma tarefa fácil ou trivial. Muitas técnicas de extração de primitivas tem sido desenvolvidas e aplicadas ao reconhecimento de manuscritos, podendo-se resumir em 3 classes [HEUTTE, 1994] e [EL YACOUBI, 1996]:

a) **Primitivas baseadas em transformadas global e séries de expansão:** Transformadas e séries de expansão, tais como, Fourier, Walsh, Harr e outras, tornam as primitivas invariantes para algumas deformações globais, por exemplo, translação e rotação. Entretanto, tais técnicas apresentam um custo computacional alto no que se refere a tempo;

b) **Primitivas baseadas na distribuição estatística dos pontos:** Estas primitivas incluem *zoning*, momentos, n-tuplas, *crossing* e distâncias. São tolerantes com distorções e levam em conta, para alguns casos, as variações de estilo. Provêm alta velocidade e baixa complexidade de implementação;

c) **Primitivas geométricas e topológicas:** Estas são as primitivas mais empregadas para representar global e localmente as propriedades dos caracteres. Estão incluídas nesta classe os ascendentes, descendentes, laços, traços, barras em diferentes direções, pontos finalizadores, interseções de segmentos de linha, laços, *stroke relations* e propriedades angulares. Estas primitivas tem alta tolerância a distorções, variações de estilo, translação e rotação.

As aplicações na área de RPM combinam os diferentes tipos de primitivas visando obter métodos mais robustos através de um conjunto pequeno de primitivas porém, com alta representatividade do problema. Portanto, diferentes tipos de formas

precisam de diferentes tipos de primitivas, que por sua vez precisam de diferentes tipos de classificadores. A Tabela 2.3 resume alguns conjuntos de primitivas selecionados por diversos autores no contexto de palavras manuscritas.

Tabela 2.3: Extração de Primitivas

Tipo de Primitivas	Autores	Características
Geométrica	[BOZINOVIC & SRIHARI, 1989]	Traços, curvas, laços, pontos isolados, picos e contorno
Geométrica	[LECOLINET, 1990]	Componentes conexos
Global	[GUILLEVIC, 1995]	Posição relativa dos: ascendentes e descendentes, traços verticais e horizontais, diagonais sul-leste e sul-oeste, laços e estimativa do comprimento da palavra
Geral	[TRIER et al., 1996]	Recomenda conjuntos de primitivas em função do tipo de caracter: dígitos e letras
Global	[SENIOR & ROBINSON, 1998]	Primitivas notáveis
Híbrido	[CHEN et al, 1995b]	Momentos, primitivas topológicas, geométricas, distribuição dos <i>pixels</i> na imagem e linhas de referência
Global	[AVILA, 1996]	Alfabeto de 12 traços
Global	[EL YACOUBI, 1996]	Conjuntos diferenciados para letras maiúsculas e minúsculas
Híbrido	[KIM, 1996]	Cadeias de código (<i>chain code</i>) Freeman
Global	[KNERR, 1996]	Três conjuntos complexos de primitivas
Global	[CÔTE, 1997]	Ascendentes, descendentes e laços, concavidades e convexidades
Geométrica e Topológica	[HAN & SETHI, 1997]	Traços horizontais, verticais e laços, <i>end points</i> , <i>branch points</i> , <i>crossing points</i> , pontos côncavos e convexos, curvas simples
Global	[DZUBA et al., 1998]	Alfabeto de 8 traços
Global	[MADHVANATH & GOVINDARAJU, 1998]	Ascendentes, descendentes, comprimento (<i>Perceptual Features</i>)

[TRIER et al., 1996] recomenda que as primitivas a serem extraídas possuam as seguintes características:

- **invariante:** são primitivas que possuem aproximadamente o mesmo valor para exemplos de um mesmo caracter, ou seja, não devem sofrer influência de parâmetros, tais como: translação, rotação, estiramento (*stretched*), inclinação, e espelhamento (*mirrored*), conforme Figura 2.29. Quando

primitivas invariantes não podem ser escolhidas, a alternativa é padronizar a imagem original. Porém, erros de discretização podem ser inseridos com esse procedimento,

- **reconstrução:** são as primitivas que guardam informações sobre os caracteres originais de onde essas primitivas foram extraídas. Ao se reconstruir os caracteres com base nas primitivas extraídas é possível saber se o conjunto de primitivas escolhido é capaz de capturar a estrutura principal dos caracteres.

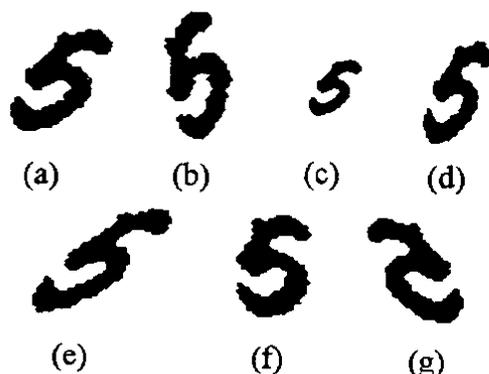


Figura 2.29: Versões transformadas do dígito “5”: a) imagem original, b) rotação, c) escala, d) estiramento, e) inclinação, f) correção da inclinação g) espelhamento [TRIER et al., 1996]

As primitivas indicadas por [TRIER et al., 1996] para imagens binárias de caracteres do tipo letra de imprensa manuscrita (Figura 2.11 – estilo *spaced discrete characters* [TAPPERT, 1990]) são as seguintes:

- *template matching*,
- transformadas (por exemplo, *Karhunen-Loeve transform*);
- histogramas de projeção horizontal e vertical,
- zoneamento (*zoning*),
- momentos geométricos invariantes (média e variância).

Para [BOZINOVIC & SRIHARI, 1989] as primitivas geométricas escolhidas representam propriedades topológicas, espaciais (zonas), de curvatura e pontos extremos (*cusps-shape*) das imagens das palavras, sendo o conjunto formado por:

- traço na zona alta (*upper stroke*),
- traço longos na zona alta (*long upper stroke*),
- traço na zona baixa (*lower stroke*),
- curva na zona alta (*upper curve*),

- curva para a esquerda (*left curve*),
- curva na zona baixa (*lower curve*),
- curva para a direita (*right curve*),
- pico ou vértice isolados (*i-shape*),
- picos ou vértices (*n-shape*),
- contorno longo na zona alta (*long lower contour*),
- laço grande na zona média (*large middle loop*),
- laço pequeno na zona média (*small middle loop*),
- laço na zona alta (*upper loop*),
- laço na zona baixa (*lower loop*),
- interrupção ou corte (*slash*) e
- ponto (*dot*).

[SENIOR & ROBINSON, 1998] acreditam que o desempenho de um reconhecedor pode ser melhorado utilizando-se mais informações sobre as *primitivas notáveis* de uma palavra. As primitivas são extraídas sobre o esqueleto codificado da imagem da palavra, conforme mostrado na Figura 2.30. As primitivas notáveis selecionadas pelos autores foram:

- pontos sobre as letras “i” e “j”, ou ainda traços isolados sobre ou acima da linha média são marcadores potenciais dos pontos nestas letras (*dots*),
- junções entre dois traços ou cruzamentos (*junctions*),
- pontos finalizadores de traços (*end points*),
- pontos que caracterizam a mudança de direção do traçado (*turning points*),
- laços (*loops*).

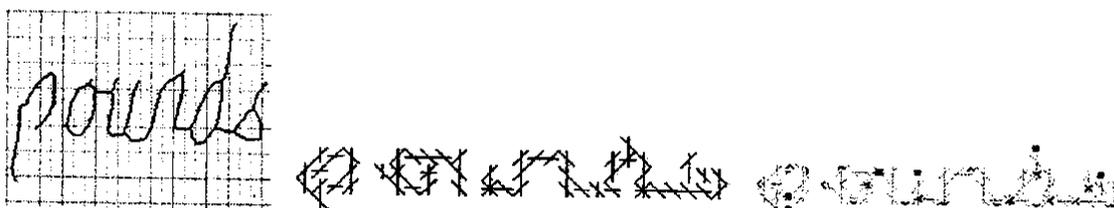


Figura 2.30: Estágios da extração de primitivas: a) esqueleto com *grid*,
 b) representação dos segmentos de linha parametrizados c) primitivas extraídas
 [SENIOR & ROBINSON, 1998]

[CHEN et al, 1995b] apresentam a extração de primitivas para reconhecimento de palavras manuscritas através de CDVDHMM - *Continuous Density Variable*

Duration Hidden Markov Models. Um conjunto de 35 primitivas foi selecionado incluindo: momentos, primitivas topológicas, geométricas, distribuição dos *pixels* na imagem e linhas de referência.

[CHEN et al, 1995b] trabalham com uma classificação das primitivas em *singulares* e *regulares*. Esta mesma classificação é utilizada por [CÔTE, 1997] que define as palavras como o conjunto formado pelas regularidades e singularidades da mesma, conforme Figura 2.31. As regularidades são as primitivas côncavas e convexas do corpo das palavras e as singularidades são os ascendentes e descendentes.

[CÔTE, 1997] apud [SUEN, 1993] trabalha com o que os autores denominam de letras-chaves, ou sejam, os ascendentes (letras: “t”, “l”, “b”), os descendentes (letras: “p”, “g”, “q”), os ascendentes-descendentes (letras: “f” e “gh”) e os laços no corpo das palavras (letras: “o”, “e”). Outras primitivas selecionadas são as denominadas condicionais, compreendendo as primitivas que estão associadas entre si através de uma condição, por exemplo o laço da letra “d” com o ascendente desta mesma letra.

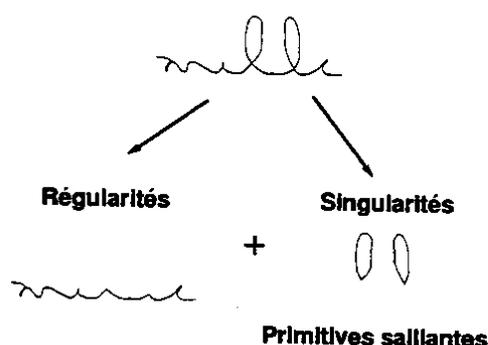


Figura 2.31: Regularidades e singularidades de uma palavra[CÔTE, 1997]

[EL YACOUBI, 1996] apresenta um trabalho detalhado sobre a extração de primitivas para aplicação em reconhecimento de endereços postais. Yacoubi utiliza a segmentação das palavras em letras. Apresenta a extração de primitivas definindo dois conjuntos distintos: letras maiúsculas e minúsculas. No caso das letras maiúsculas, foi utilizada a análise dos histogramas de transição dos contornos de cada padrão segmentado (por letra). Estes histogramas permitem encontrar 2, 4 ou 6 transições em cada direção (horizontal e vertical), estando a atenção voltada para a parte central do histograma, uma vez que esta é a área mais estável para uma letra. Desta forma, foi estabelecido um alfabeto de tamanho 19, sendo 14 símbolos para as formas dos histogramas (projeção horizontal)-letras e 5 para os pontos de segmentação.

Para as letras minúsculas [EL YACOUBI, 1996] estabelece um alfabeto de tamanho 38, sendo 33 símbolos para as formas dos segmentos-letras e 5 símbolos para os pontos de segmentação.

[EL YACOUBI, 1996] trata também os espaços inter-segmentos (entre os segmentos de palavra), de maneira a estabelecer as ligações entre as letras. Assim, o autor cria uma categorização dos pontos de segmentação em 5 classes. Esta preocupação é justificada diante de uma proposta que trabalha com a segmentação explícita das palavras.

[HAN & SETHI, 1997] propõem um método de reconhecimento do extenso manuscrito em cheques bancários utilizando um conjunto de primitivas geométricas e topológicas para descrever cada palavra. Considera-se a distribuição espacial destas primitivas na imagem da palavra. A extração das primitivas é realizada sobre imagens normalizadas (sem efeito de translação, escala e rotação).

[AVILA, 1996] propõe a extração de primitivas para reconhecimento de palavras manuscritas em cheques franceses com base em um alfabeto de 12 traços. A Figura 2.32 apresenta o alfabeto, sendo que as formas elementares são obtidas a partir dos pontos de ancoragem do esqueleto da palavra original sobre a linha da média das palavras. A Figura 2.33 mostra alguns exemplos de descrição de palavras utilizando este alfabeto.

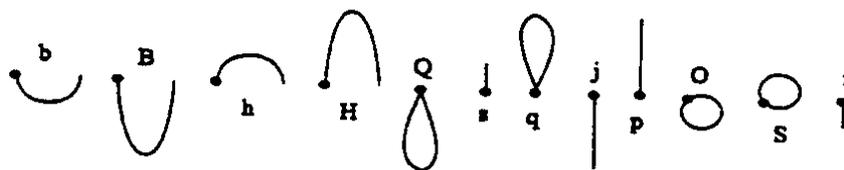


Figura 2.32: Alfabeto de 12 traços [AVILA, 1996]

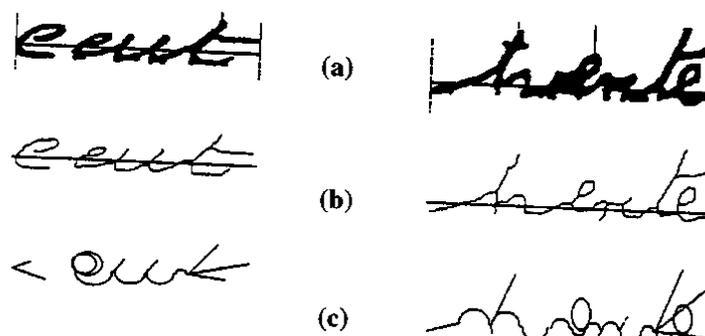


Figura 2.33: Exemplos de descrição de palavras através do alfabeto de 12 traços: imagem original, b) esqueleto, c) alfabeto de 12 traços [AVILA, 1996]

[LECOLINET, 1990] apresenta um método para reconhecimento de palavras em cheques franceses baseado na análise dos componentes conexos, que segundo o autor, constituem a “essência” de uma imagem. O objetivo é estudar a fronteira entre o interior e o exterior de cada forma, tratando de problemas como os vales simples e aos vales laterais escondidos, conforme Figura 2.34.

[FREEMAN, 1961] apud [KIM, 1996] utiliza cadeias de código ou *chain code* de Freeman baseadas em primitivas globais e locais para o reconhecimento de palavras no contexto de nomes de ruas e de valores por extenso em cheques. Duas primitivas globais foram escolhidas: média e traço médio de toda a palavra. Primitivas locais em 8 direções são extraídas em sub-imagens (3x3) da imagem original.

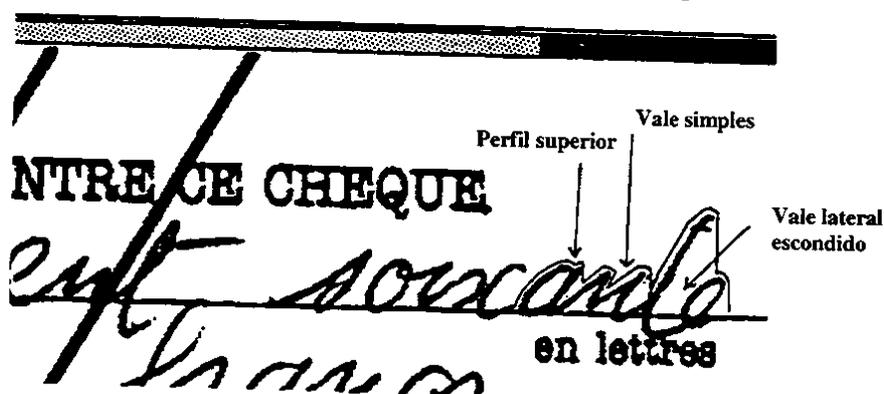


Figura 2.34: Vales e perfil superior [LECOLINET, 1990]

[KNERR et al., 1996] apresenta o sistema A2iA para reconhecimento de cheques manuscritos. Este sistema trabalha com um complexo conjunto de primitivas divididas em 3 tipos:

- Tipo 1: primitivas gerais que descrevem a imagem do caracter (18 primitivas). Exemplos: altura do caracter, tipo do perfil direito/esquerdo/superior/inferior do caracter, número de componentes conectados, entre outros;
- Tipo 2: primitivas que descrevem a posição relativa de partes do caracter em relação a imagem do caracter (9 primitivas). Exemplos: ascendentes, descendentes, traços horizontais/verticais/ +45 graus/-45 graus, laços, entre outros;
- Tipo 3: primitivas que dependem da posição do caracter em relação a linha de texto no cheque (3 primitivas). São elas: posição vertical e horizontal do caracter em relação a linha de texto e número de conexões entre caracteres vizinhos de uma mesma linha de texto.

[DZUBA et al., 1998] apresenta um conjunto de 8 traços a serem extraídos de palavras manuscritas em processo dinâmico, conforme Figura 2.35, as quais podem ser comparadas ao alfabeto de 12 traços proposto por [AVILA, 1996].

A extração das características é feita sobre 3 formas de representação das palavras: imagem binária, seqüência de pontos dos contornos interno e externo da imagem e representação através de grafo. São extraídas as seguintes primitivas:

- máximos e mínimos locais dos pontos pretos e brancos;
- máximos e mínimos brancos e pretos que ultrapassam horizontalmente e estão localizados dentro da espessura do traço formando um arco na região superior ou inferior;
- máximos e mínimos brancos e pretos que ultrapassam horizontalmente e estão localizados dentro da espessura do traço formando pequenos traços horizontais;
- os laços extraídos são classificados segundo a presença de arcos: os laços fechados são aqueles que possuem um arco superior e inferior formando um ciclo, os demais são ditos arcos superior ou inferior somente;
- traços longos (*cusp*);
- arcos com ponto finalizador e traços horizontais de cruzamento (tais como na letra "t").

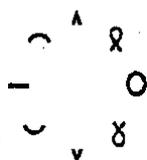


Figura 2.35: Primitivas com base em 8 traços [DZUBA et al., 1998]

Os trabalhos de [DZUBA et al., 1998] e de [MADHVANATH & GOVINDARAJU, 1998], assim como o trabalho de [EL YACOUBI, 1996] realizam a extração de primitivas sobre a seqüência de pontos dos contornos interno e externo da imagem das palavras através da representação em cadeias de código.

Acredita-se que sobre uma cadeia de código pode-se definir um conjunto de primitivas [LI & YEUNG, 1997] ou traços [DZUBA et al., 1998] e [AVILA, 1996] a serem extraídos das palavras. [LU & SHRIDHAR, 1996] é exemplo de outro trabalho que também trata as palavras como um conjunto de regularidades e singularidades, assim como [CÔTE, 1997].

A variedade de primitivas selecionadas nos diferentes trabalhos citados permite compreender a complexidade e riqueza de informações contidas nas palavras manuscritas. Esta riqueza de informações permite, ao ser bem explorada, uma redução da complexidade do problema. Isto devido ao fato de que as primitivas evidenciadas pelos diferentes autores caracterizam-se como representativas do problema: palavras manuscritas.

Cabe, então, a cada pesquisador buscar um conjunto de primitivas representativo, invariante, independente entre si, de tamanho adequado ao tempo de processamento e adequado ao método de reconhecimento a ser utilizado. Fica então a seguinte pergunta: **Quais são as primitivas relevantes (*perceptual features*) no processo de reconhecimento de palavras manuscritas?**

A resposta está em reunir os aspectos relevantes dos processos de escrita e leitura apresentados neste capítulo com os trabalhos de [MADHVANATH & GOVINDARAJU, 1998] e [SCHOMAKER & SEGERS, 1998]. O primeiro fornece uma definição de *perceptual features* que são as principais características utilizadas na representação e reconhecimento de palavras pelo ser humano, especialmente nos casos de aplicações com léxico de tamanho pequeno, tal como em valores por extenso em cheques bancários. São assim reconhecidas como primitivas perceptivas: ascendentes, descendentes e tamanho das palavras, representados por valores e posições.

Estas primitivas vem ao encontro dos aspectos relevantes dos processos de escrita e leitura realizados pelo ser humano, bem como dos métodos de aprendizagem e reconhecimento descritos anteriormente (Capítulo 2, Seções 2.2 e 2.3).

O trabalho de [SCHOMAKER & SEGERS, 1998] descreve um experimento realizado com o objetivo de determinar as primitivas relevantes no processo de leitura de palavras cursivas realizado pelo ser humano. Os autores propõem um método para inserir mais detalhes sobre primitivas geométricas as quais os seres humanos utilizam para ler. O método consiste em computar o número de *clicks* que uma pessoa realiza com o *mouse* sobre a imagem de uma palavra (idioma inglês) em condições de luminosidade precária, apresentada na tela do computador. O trabalho confirma o alto grau de utilização dos ascendentes, descendentes, cruzamentos e pontos de alta curvatura no reconhecimento do padrão manuscrito. As principais conclusões de [SCHOMAKER & SEGERS, 1998] são as seguintes:

- em um primeiro nível as pessoas utilizam os ascendentes (d,k,l,h,t,b) descendentes (q,y,j,g,p), sendo a letra f um caso especial, pois possui ambas as características,
- as consoantes possuem uma maior importância no processo de leitura do que as vogais, sendo possível ler ou reconhecer uma palavra sem a presença dessas letras (*handwriting = hndwrtnng*),
- as vogais (a,e,i,o) geram um mesmo comportamento do leitor, porém a letra `u` necessita de mais informações, isto porque deve ser diferenciada de uma letra `w` ou `m`,
- a primeira e a última letras são muito importante no processo de reconhecimento (estas letras fornecem `pistas` sobre as palavras e permitem uma ligação com o conhecimento prévio do leitor sobre o padrão a ser analisado e reconhecido – Seções 2.2 e 2.3),
- a última letra é muito utilizado no processo de reconhecimento,
- palavras curtas necessitam de um maior número de informações para serem reconhecidas,
- palavras longas necessitam de mais informações no final das mesmas,
- o final das palavras, a barra de corte da letra `t` e o ponto da letra `i` deterioram o processo de reconhecimento quando são mal interpretados.

Finalmente, [SUEN, 1998] apresenta um conjunto de perguntas que visam estabelecer a interdependência do conhecimento humano sobre os processos de escrita e leitura com os métodos de reconhecimento automático de palavras manuscritas. Afirma que para se ter todas as respostas às perguntas apresentadas, devem ser desenvolvidas mais e mais experiências psicológicas, destinadas a evidenciar as partes mais pertinentes dos caracteres manuscritos. Somente assim, comenta o autor, pode-se identificar as primitivas interessantes e discriminantes.

2.4.2.5. Metodologias para Reconhecimento

O reconhecimento consiste em a partir de uma forma (uma palavra) desconhecida estabelecer sobre um conjunto de formas conhecidas, quais entre as formas desse conjunto mais se assemelha a forma desconhecida, e sobre tudo isso tomar uma decisão da melhor hipótese. Essa decisão é efetuada geralmente medindo-se a semelhança da forma desconhecida com um conjunto de referência (ou modelos)

armazenados na memória e descritos em uma representação análoga. As referências ou modelos são obtidos através da etapa denominada de treinamento. A Figura 2.36 apresenta um diagrama de um método de reconhecimento.

Nos métodos de reconhecimento, cada produção manuscrita de uma palavra dada é ligada a uma forma particular. Após a fase de aquisição do sinal, a extração de primitivas permite transformar a forma em uma representação, mais fácil de manipular que a forma original. O reconhecimento consiste em decodificar essa representação atribuindo-se uma das classes conhecidas ou referências.

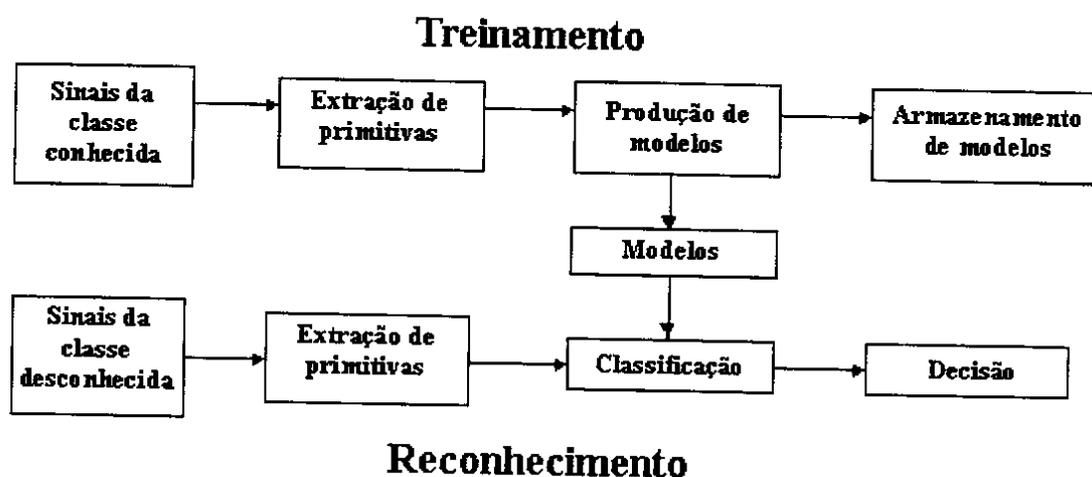


Figura 2.36: Diagrama de um método de reconhecimento de formas

Deste modo, duas são as metodologias de reconhecimento utilizadas, a saber:

- Estatística e
- Estrutural.

a) Métodos Estatísticos

Os métodos estatísticos exploram as propriedades métricas das formas a serem reconhecidas efetuando-se uma modelagem em termos de componentes ligados a natureza de cada forma.

Para [EL YACOUBI, 1996] no domínio dos métodos estatísticos as aproximações beneficiam-se dos métodos de treinamento automático que se apoiam sobre bases teóricas, tais como, a teoria da decisão bayesiana, os métodos de classificação não supervisionados, a análises em componentes principal e agrupamentos dinâmicos. Duas aproximações principais são geralmente utilizadas: os métodos paramétricos e os não-paramétricos.

a.1) Métodos Paramétricos (bayesianos)

Estes métodos operam sempre sobre a hipótese que as classes estudadas seguem uma distribuição de probabilidades de uma certa forma. Supõem-se que são conhecidas as leis de probabilidade das classes, ou seja, os métodos não podem ser aplicados de maneira realística a não ser sobre hipóteses gaussianas com estimativa de parâmetros correspondentes. A decisão consiste em determinar a classe para a qual a forma desconhecida apresenta a probabilidade de pertencer considerando-se um máximo de pertinência [EL YACOUBI, 1996].

Estes métodos são muito rápidos nos cálculos, entretanto exigem muito das bases de treinamento bastante para uma estimação correta dos parâmetros da distribuição de classes suposta. A técnica denominada Distância de Mahalanobis é muito utilizada em reconhecimento de formas. Esta técnica vem diretamente da aplicação da teoria bayesiana e mede a distância entre o ponto representando a forma desconhecida e os pontos representando as classes conhecidas. [HEUTTE, 1994] utiliza a Distância de Mahalanobis no reconhecimento de envelopes postais e valores numéricos em cheques.

a.2) Métodos Não-Paramétricos

Nestes métodos, as leis de probabilidade em cada classe são desconhecidas. O problema está na pesquisa de fronteiras de decisão entre as classes. Três são as técnicas utilizadas no reconhecimento de formas considerando este tipo de método: o método do k vizinhos mais próximos, o método das Janelas de Parzen e o método de Correspondência por Máscaras (*matching*).

- **Métodos dos k Vizinhos mais Próximos:** A forma desconhecida é afetada pela classe dos seus k vizinhos mais próximos no conjunto do treinamento. A regra de implicação está fundamentada, em geral, sobre o cálculo da Distância Euclidiana Clássica entre a forma e as classes pré-determinadas, mas pode se entender a aplicação em todo o espaço munido de uma distância qualquer. Este método é simples de ser executado e fornece bons resultados. Entretanto ele não é bastante rápido devido ao fato do número elevado de distâncias que se deve calcular. [OLLIVIER, 1999] utiliza este método para o reconhecimento de palavras no contexto de cheques bancários. [NESCHEN, 1995] aplica este método no reconhecimento de dígitos e caracteres maiúsculos manuscritos.

- **Janelas de Parzen:** O objetivo das janelas de Parzen é de fornecer uma estimação de probabilidade da forma desconhecida para uma classe conhecida. Em cada ponto do espaço, conta-se o número de pontos das classes conhecidas próximas da forma desconhecida e atribui-se a estes pontos uma ponderação, tanto mais forte quanto o ponto é mais próximo da forma desconhecida. As janelas de Parzen fornecem excelentes resultados quando não se dispõe de informações *a priori* sobre a natureza da distribuição das formas. Mas o método, comparado ao dos k Vizinhos mais Próximos, consome muito tempo de cálculo. Por isso ele é menos utilizado apesar do seu interesse teórico.
- **Métodos de Correspondência por Máscaras:** Esta técnica chamada de *Template Matching* consiste em definir cada classe por um conjunto padrão de protótipos armazenados na memória. A comparação se faz diretamente e a classe reconhecida é aquela que contém o protótipo o mais correlacionado à forma desconhecida. Os métodos de comparação podem ser igualmente simples como um método de comparação pixel a pixel, ou também complexas como uma análise de árvore de decisão nas quais apenas alguns pixels selecionados são testados. Esse tipo de técnica sofre com a sensibilidade aos ruídos e não se adapta as diferenças encontradas nos estilos de escrita como citado por [EL YACOUBI, 1996].

b) Métodos Estruturais

Os métodos estruturais são utilizados para descrever as características intrínsecas de uma forma, efetuando-se uma modelagem em termos de componentes estruturais, chamados de primitivas, ligados a natureza de cada forma.

Contrariamente as aproximações estatísticas, as aproximações estruturais introduzem a noção de ordem na descrição de uma forma. Deste fato, os métodos estruturais são particularmente adaptáveis às aplicações onde a ordem é justamente primordial. Este é o caso naturalmente do reconhecimento de palavras manuscritas onde o aspecto seqüencial das letras de esquerda para direita deve ser corretamente modelado.

Dado a manipulação de um espaço de representação estrutural, as fases do treinamento e de decisão são tratadas sobre um ângulo diferente. Estes métodos tratam de comparar duas formas pela intermediária de suas descrições respectivas levando em conta a ordem das primitivas presentes.

A técnica mais conhecida entre os métodos estruturais é a Programação Dinâmica. A Programação Dinâmica é a área do conhecimento utilizada para resolver problemas de decisão. O problema consiste em determinar o mínimo custo conforme apresentado em [RABINER & JUANG, 1993]. Esta técnica permite detectar o caminho ótimo que corresponde à distância acumulada mínima pela correspondência entre duas cadeias ou formas. Os métodos heurísticos são também utilizados para determinar o caminho ótimo, mas estes requerem mais da memória e do tempo de cálculo relativamente ao da programação dinâmica [KIM, 1996].

A técnica da Programação Dinâmica foi introduzida entre os anos 70 e 80 em estudos de padrões da fala a partir de estudos de [RABINER & JUANG, 1993]. Esta técnica foi utilizada para reconhecimento de caracteres isolados por [BURR, 1983] e, mais recentemente, para reconhecimento de manuscritos por [KIMURA et al., 1993].

Apesar de sua natureza diferente, os enfoques estatísticos e estruturais podem ser utilizados nos mesmos domínios de aplicação. A escolha de um enfoque pode estar simplesmente ligado às restrições materiais, tais como, o tamanho da base de treinamento disponível, o tempo de cálculo requisitado e o tamanho da memória necessária.

A diferença entre os dois enfoques é clara de um ponto de vista teórico, porém, na prática numerosos métodos são mistos. Por exemplo, os métodos estruturais utilizam os cálculos estatísticos para a estimação da frequência das primitivas em estudo. Inversamente encontramos métodos estatísticos que manipulam elementos estruturais, tais como os Modelos Escondidos de Markov - MEM.

Deste modo, um modelo misto formado pela utilização conjunta de dois métodos é uma solução para o problema de reconhecimento de formas. Essa estratégia permite ao método beneficiar-se das vantagens de cada um dos dois enfoques e evitar alguns inconvenientes.

Os MEM permitem efetuar uma modelagem em termos de componentes estruturais, chamados de primitivas, ligados a natureza de cada forma. Utiliza-se dos métodos estruturais a possibilidade de inserção nos modelos da noção de ordem na descrição de uma forma. Assim, os métodos estruturais são particularmente adaptáveis às aplicações onde a ordem é justamente primordial. Este é o caso do reconhecimento de palavras manuscritas onde o aspecto seqüencial das letras de esquerda para direita deve ser corretamente modelado.

Por serem modelos estatísticos, no que se refere ao método de reconhecimento, força dos MEM está na sua capacidade de modelar a variabilidade das formas sem considerar as limitações inerentes ao problema.

O interesse dos MEM reside na sua possibilidade de modelar eficazmente diferentes fontes de conhecimento. Seu ponto forte está primeiro no fato, da integração coerente dos diferentes níveis de modelagem: morfológico ou traçado da palavra, léxico ou vocabulário de palavras estudadas e sintático ou o modelo da palavra propriamente dito. E segundo, na existência de algoritmos eficazes para determinar o valor ótimo dos parâmetros fornecendo a melhor adequação entre o modelo e a base de dados (conhecimento), chamada de base de treinamento [EL YACOUBI, 1996].

Considerando-se o exposto os Modelos Escondidos de Markov – MEM foram escolhidos como o método para o reconhecimento das palavras manuscritas no presente trabalho. Outros autores aplicam os MEM para palavras manuscritas: [AVILA, 1996], [EL YACOUBI, 1996], [CHO et al., 1995], [CHEN et al., 1995a], [AGAZZI & KUO, 1993].

A seguir apresenta-se a uma revisão bibliográfica de trabalhos realizados no contexto de palavras manuscritas com aplicação desses modelos, visto ser esta a metodologia escolhida pelo presente trabalho. A fundamentação teórica dos MEM está descrita no Capítulo 4 juntamente com o método desenvolvido para o reconhecimento de palavras manuscritas.

2.4.3. MEM para Reconhecimento de Palavras

Muitos trabalhos podem ser encontrados no escopo de palavras manuscritas e MEM, tais como: [GILLIES, 1992], [AGAZZI & KUO, 1993], [BERCU & LORETTE, 1993], [BERTILLE, et al., 1994], [BUNKE et al., 1995], [CHO et al. 1995], [GILLOUX, et al., 1995], [MOHAMED, 1995], [MONTOLIU, 1995], [AVILA, 1996], [CONNELL, 1996], [KNERR et al., 1996], [EL YACOUBI, 1996], [KUNDU, 1997], [SAION & BELAÏD, 1997], [MARTI & BUNKE, 1998], [SENIOR & ROBINSON, 1998].

[BERTILLE et al, 1994], [GILLOUX et al, 1995] e [EL YACOUBI, 1996] apresentam e descrevem os caminhos para implementar e, posteriormente, melhorar um método para reconhecimento de palavras manuscritas usadas no contexto de nomes de cidades em correspondências via correio e cheques bancários. Os trabalhos envolvem o

pré-processamento das imagens de palavras, a segmentação da palavra em letras ou pseudo-letras, extração de primitivas dos segmentos encontrados e, finalmente, o reconhecimento das palavras baseado em MEM. Estes trabalhos são sem dúvida um marco nas atividades de pesquisa de métodos reconhedores de palavras manuscritas.

Os autores definem MEM do tipo esquerda-direita, sendo a etapa de treinamento realizada com base no algoritmo de Baum-Welch e o reconhecimento no algoritmo de Viterbi.

Por se tratar de uma aplicação com grande variabilidade nas palavras manuscritas, o MEM definido por [EL YACOUBI, 1996] permite representar a substituição, a omissão e a inserção de símbolos. A Figura 2.37 mostra estas características do MEM escolhido. A transição $0 \rightarrow 4$ modela a omissão de símbolos. A transição $0 \rightarrow 3$ seguida de $3 \rightarrow 4$ modela o caso do símbolo (O_i) correspondente a letra em curso seguida de um ponto de segmentação (S_i). E por fim, o caminho dos estados $0 \rightarrow 1 \rightarrow 2 \rightarrow 3 \rightarrow 4$ modela a segmentação excessiva dos caracteres em dois fragmentos.

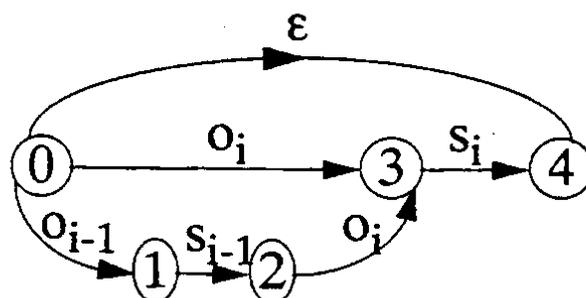


Figura 2.37: MEM para o caracter [EL YACOUBI, 1996]

[EL YACOUBI, 1996] realizou um estudo detalhado dos espaçamentos inter-segmentos. O autor criou uma categorização dos pontos de segmentação em 5 classes. Esta preocupação é justificada diante de uma proposta que trabalha com a segmentação explícita das palavras em letras ou pseudo-letras. A Figura 2.38 apresenta o MEM definido para modelagem dos espaçamentos intra-palavras, ou seja, o espaçamento entre duas letras adjacentes fisicamente conectadas ou ligadas virtualmente (caso onde as letras pertencem a uma mesma palavra, porém encontram-se desconectadas).

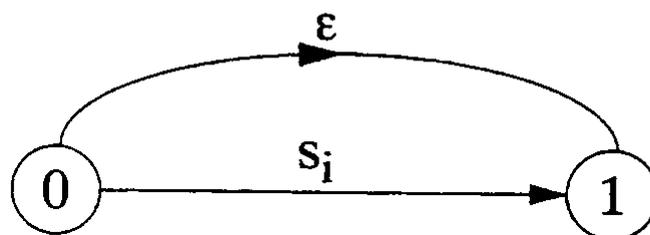


Figura 2.38: MEM para o espaçamento inter-palavras [EL YACOUBI, 1996]

Para modelar as diferentes palavras, [EL YACOUBI, 1996] define dois modelos diferentes para cada um dos tipos de letras tratadas: maiúsculas e minúsculas, conforme já mencionado anteriormente. Portanto, o modelo global é construído através da justaposição dos dois modelos, compondo assim as palavras. A Figura 2.39 apresenta o MEM para o prefixo “rue”. Desta forma, o autor apresenta a possibilidade dos modelos serem compostos para cada palavra em função dos tipos de escrita que se encontram numa mesma palavra: cursiva, bastão e mixada.

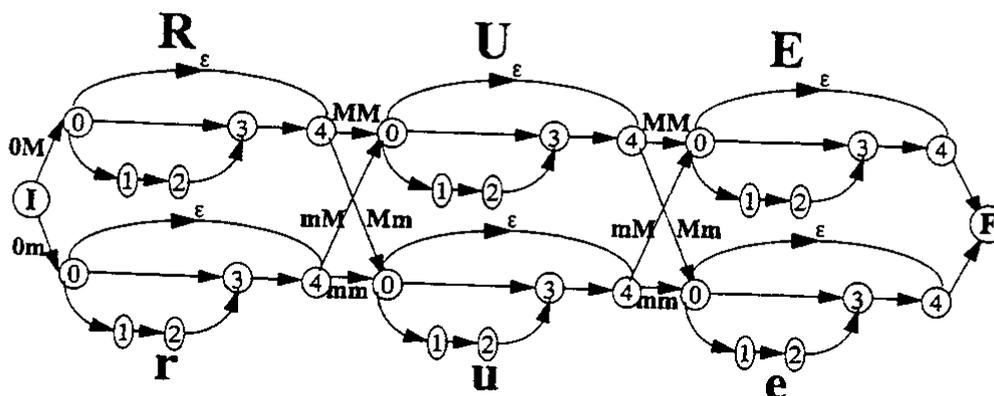


Figura 2.39: Modelo global para o prefixo “rue” [EL YACOUBI, 1996]

[CHO et al., 1995] apresenta duas maneiras de modelar palavras manuscritas. A primeira considera a palavras em estilo cursivo puro como uma seqüência de caracteres e a segunda como uma seqüência de caracter-ligação-caracter. A Figura 2.40 apresenta estes dois pontos de vista.

Para [CHO et al., 1995] os caracteres são, normalmente, mais complexos que as ligações e, portanto, os MEM para caracteres necessitam de mais estados e transições do que os MEM para ligações. Cada modelo de caracter contém um diferente número de estados, os quais são determinados com base no comprimento médio das seqüências dos

exemplares de treinamento. As topologias dos dois diferentes modelos estão apresentadas na Figura 2.41.

O sistema A2iA para reconhecimento de cheques manuscritos descrito por [KNERR et al., 1996] apresenta uma estrutura baseada em Redes Neurais e MEM. As redes neurais são utilizadas para classificação das 74 primitivas extraídas e determinação do número de estados para cada MEM de cada palavra. Os MEM são então, utilizados no treinamento das probabilidades de transição (algoritmos de Viterbi e Baum-Welch) e, posteriormente no reconhecimento (algoritmo *Forward*), associados às probabilidades das redes neurais e regras de Bayes. A Figura 2.42 exemplifica a topologia dos MEM definidos para diversas palavras do contexto de cheques franceses.

[GILLOUX, 1994] apresenta a topologia do MEM adotado para o reconhecimento de palavras manuscritas em cheques franceses. Cada classe de palavras é representada por um MEM esquerda-direita, conforme Figura 2.43. O modelo escolhido leva em conta deleções, inserções ou substituições de símbolos. Para cada classe de palavras o modelo correspondente é definido com um número de estados dependendo do comprimento mais provável dos exemplos desta classe.

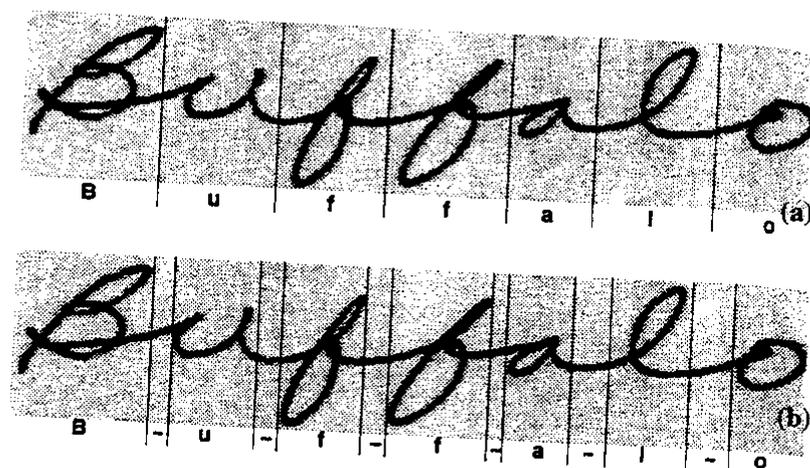


Figura 2.40: Métodos de análise de uma palavra manuscrita:
a) caracter-caracter, b) caracter-ligação-caracter [CHO et al., 1995]

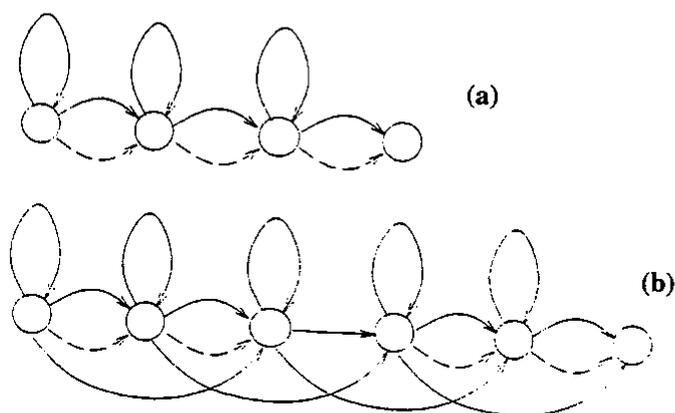


Figura 2.41: Topologias dos modelos:
a) caracter-caracter, b) caracter-ligação-caracter [CHO et al., 1995]

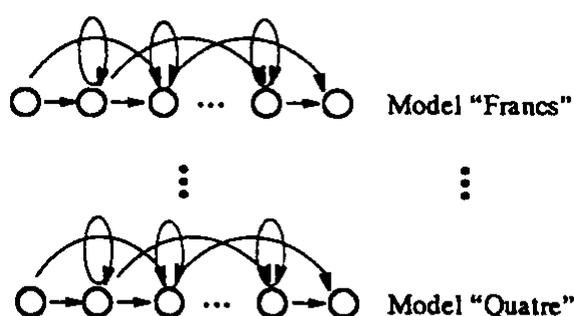


Figura 2.42: MEM para palavras de cheques franceses [KNERR et al., 1996]

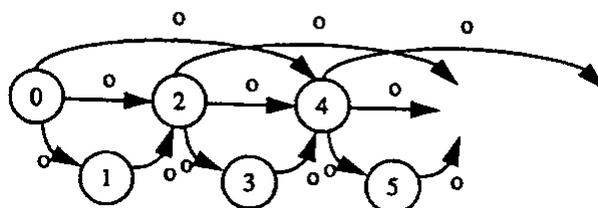


Figura 2.43: MEM para palavras [GILLOUX, 1994]

[GILLIES, 1992] apresenta um método de leitura automática de palavras manuscritas com segmentação explícita das palavras em letras. Para cada coluna da imagem da palavra obtém-se um vetor de características ou *feature vector* que é processado pelo algoritmo de Quantização Vetorial ou *Vector Quantization* que transforma a imagem em uma seqüência de w símbolos, onde w é a largura da imagem original em pixels. Esta seqüência de símbolos é então dividida em um número de sub-sequências com base nas fronteiras entre letras. Assim, é criada um seqüência de observação individual para cada letra. Finalmente, a seqüência de cada letra é dividida

em N espaçamentos iguais, onde N é o número de estados do MEM para cada uma das letras do alfabeto, conforme Figura 2.44.

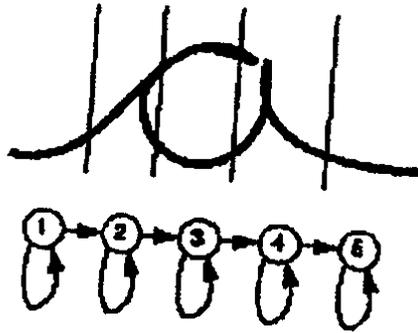


Figura 2.44: Modelo para as letras cursivas [GILLIES, 1992]

Os autores apresentados utilizam a topologia esquerda-direita ou *Bakis Topology*, uma vez que esta topologia permite melhor representar o que ocorre quando se escreve uma palavra em idiomas com características de escrita horizontalizada, da esquerda para a direita.

Os refinamentos de modelagem das palavras, dos caracteres e ligações entre caracteres depende da metodologia adotada por cada um dos autores e está ligada ao método de segmentação empregado na detecção das letras e pseudo-letras. O presente trabalho segue também uma abordagem Global, tal qual [GILLOUX, 1994] e [KNERR et al., 1996].

Finalmente, pode-se dizer que a aplicação de MEM ao problema de reconhecimento de palavras manuscritas não é de todo desconhecido ou novo, mas a busca de “ótimos” conjuntos de primitivas e novas combinações de diferentes abordagens para a solução do problema é que tornam o problema, por sua vez, muito mais interessante e desafiador.

2.4.4. Discussão

Muitas são as estratégias, métodos e técnicas quando o assunto Reconhecimento de Palavras Manuscritas. Por isso, foram apresentados neste capítulo pontos relevantes dos aspectos dos processos de escrita, leitura e reconhecimento prévio realizados pelo ser humano, bem como o estado da arte do reconhecimento de palavras manuscritas e as diferentes etapas necessárias ao seu processamento. Lembra-se que o presente trabalho procura desenvolver um método automático de reconhecimento de palavras manuscritas em cheques bancários brasileiros, levando em consideração os aspectos da escrita e

leitura humana de forma a propiciar caminhos na busca de possíveis soluções. A Figura 2.45 resume as metodologias e variáveis do processo de reconhecimento de palavras manuscritas. Cada uma das etapas e variáveis do caminho seguido pelo presente trabalho encontra-se em destaque nesta mesma Figura.

O entendimento mais detalhado dos processos de escrita e leitura humana permite estudar as características relevantes pelo ser humano quando da realização da leitura. Para [SUEN, 1998] a escrita é o meio mais corrente de comunicação visual para o ser humano. Durante o processo de educação de uma pessoa, esta é apresentada inicialmente às letras isoladas, depois à escrita de forma e finalmente à escrita cursiva. Desta forma, pode-se dizer que as dificuldades do processo de escrita à mão, bem como da leitura de palavras manuscritas são bem conhecidas.

Entretanto, ao pretender-se automatizar o processo de reconhecimento de palavras manuscritas percebe-se que muitos pontos são obscuros, ou mesmos desconhecidos.

De um modo geral, as principais dificuldades encontradas no processo de leitura de palavras manuscritas são:

- a grande variabilidade das formas, tanto ao nível dos caracteres como ao nível das palavras. Esta dificuldade está ligada às características pessoais de cada escritor e
- a ligação entre os diferentes caracteres, ou seja, a escrita cursiva.

O reconhecimento de palavras manuscritas estático representa o processo automático de leitura das palavras manuscritas após a fase da escrita ter sido terminada. Este processo considera a aquisição das imagens das palavras por meio de *scanner* (Figura 2.45).

Todo e qualquer método de reconhecimento de palavras manuscritas deve levar em consideração as dificuldades apresentadas, caracterizando por ser um método de reconhecimento sem restrições, como mostrado na Figura 2.45. Ou ainda, para [SUEN, 1998] um algoritmo para reconhecimento de palavras manuscritas sem restrições deve ser capaz de reconhecer com sucesso a imagem das palavras sem se importar com qual palavra e com qualquer tipo de escrita.

O sucesso completo do processo de reconhecimento de palavras manuscritas depende de uma escolha cuidadosa e precisa das primitivas a serem extraídas das

palavras. O objetivo é representar de forma adequada e suficientemente as palavras a serem processadas.

As dificuldades encontradas no processo de leitura de palavras manuscritas podem ser consideradas também no reconhecimento do extenso manuscrito em cheques bancários brasileiros. Isto, se deve ao fato do extenso manuscrito apresentar uma menor relevância perante outros campos de informação, como por exemplo, o valor numérico, e ainda, ao fato de que as pessoas ao preencherem um cheque fazem-no com pressa e às vezes em condições desconfortáveis.

Usualmente, o extenso manuscrito desempenha um papel secundário no processo de leitura do cheque. De um modo geral, as pessoas lêem primeiro o valor numérico e, somente, em caso de dúvida ou para confirmar o valor numérico realizam a leitura do extenso manuscrito. Para [DZUBA et al., 1998] estes motivos fazem com que as pessoas tendam a falhar na leitura do extenso quando não utilizam informações de contexto.

Como já apresentado o tratamento de palavras manuscritas pode ser realizado tanto ao nível dos caracteres através de uma abordagem Local (*local or analytical approach*) como ao nível das palavras através de uma abordagem Global (*global approach*), como resumido na Figura 2.45. Geralmente, de acordo com [SUEN, 1998], o tratamento global necessita buscar subsídios lingüísticos. E ainda, conforme [DZUBA et al., 1998] o escritor deseja passar a mensagem como um todo para ser entendida e não está interessado em escrever cuidadosamente cada letra.

O enfoque proposto para o tratamento das palavras do extenso manuscrito em cheques bancários brasileiros trabalha com as palavras como um todo, ou seja, sem segmentação explícita das palavras em letras ou pseudo-letras, esquematizado na Figura 2.45. Portanto, ao se considerar este tipo de enfoque define-se um primeiro conjunto de primitiva a ser extraído e modelado. Este conjunto está relacionado à pergunta: **Quais são as primitivas perceptivas relevantes no processo de reconhecimento de palavras manuscritas?**

Para responder é preciso retornar a revisão dos aspectos de escrita, leitura e reconhecimento de palavras apresentada neste capítulo, e ainda, os trabalhos de [MADHVANATH & GOVINDARAJU, 1998] e [SCHOMAKER & SEGERS, 1998]. Assim, o primeiro conjunto é formado, essencialmente, por primitivas estruturais

denominadas de perceptivas - PP: ascendentes, descendentes e laços, sendo o mesmo apresentado no Capítulo 4.

Entretanto, ao analisar-se cuidadosamente o vocabulário do extenso em cheques bancários brasileiros verifica-se a presença de um sub-conjunto de palavras (“um”, “cinco”, “seis”, “nove”, “cem”, “reais”) que não possui este tipo de característica ou caracteriza-se por ser uma palavra curta (“dois”, “três”, “sete”, “oito”, “dez”, “mil”, “real”). Fica então, outra pergunta: **Como representar palavras cursivas sem a presença de primitivas perceptivas?**

A resposta está em analisar este conjunto de palavras que possuem um elevado grau de confusão. Inclui-se neste conjunto também as palavras manuscritas em letra caixa alta, bastão ou *boxed discrete characters*, de acordo com [TAPPERT et al., 1990]. Ressalta-se que o presente trabalho não faz restrições ao tipo da escrita.

Deve-se, também levar em conta as seguintes características dos processos de escrita, leitura e reconhecimento humano de palavras manuscritas que são os pontos de partida para o presente trabalho:

- reconhecimento ideográfico das palavras, ou seja, como um todo,
- capacidade de realizar uma abordagem Global e, se necessário, uma abordagem Local,
- habilidade humana de reconhecer palavras incompletas.

Assim, trabalha-se com concavidades e convexidades, no corpo das palavras, através de dois outros conjuntos: primitivas perceptivas, concavidades e convexidades - PPCC e primitivas perceptivas, concavidades e convexidades rotuladas - PPCCR. Estes conjuntos tem a característica de complementar o primeiro conjunto de primitivas, permitindo modelar de forma mais adequada palavras que não possuem primitivas perceptivas, sendo os mesmos descritos no Capítulo 4.

Sabe-se ainda, da revisão bibliográfica apresentada, que os leitores rápidos utilizam um enfoque global e que quando as palavras não são soletráveis (isto é, quando não se pode identificar cada uma das letras que formam as palavras) é necessário basear o reconhecimento em um conjunto de características especializado [SCHOMAKER & SEGERS, 1998]. Pois, para estes autores as palavras curtas necessitam de um maior número de informações para serem reconhecidas e as palavras longas necessitam de mais informações no final das mesmas.

A idéia principal é permitir um enfoque mais local à extração de primitivas, com o objetivo de enriquecer as informações dos elementos lingüísticos que servem de “pistas” ao estabelecimento de uma relação entre o lingüístico e o conceitual-cognitivo. Em outras palavras, busca-se uma nova chance de ler o que está escrito. Sendo que o interesse é melhorar a representação das regularidades de uma palavra, tal qual [CÔTE, 1997].

A revisão bibliográfica enfatiza a habilidade humana de ler e reconhecer palavras utilizando somente a parte superior das mesmas. Este conhecimento também é utilizado pelos pesquisadores que baseiam seus métodos de reconhecimento de palavras em singularidades e regularidades das palavras, tais como [LU & SHRIDHAR, 1996] e [CÔTE, 1997], ou consideram que as palavras devem ser modeladas como uma seqüência de letras e ligações [CHO et al., 1995]. O presente trabalho explora estas características trabalhando com primitivas geométricas extraídas dos corpos das palavras.

Desta forma, a relação existente entre a revisão bibliográfica apresentada e o presente trabalho está no fato de que os processos humano e automatizado interagem de maneira a fornecer subsídios ao entendimento e repasse de características de um processo para o outro.

O ponto forte do presente capítulo é embasar a escolha de conjuntos de primitivas que tenham por base a percepção humana das palavras manuscritas, e portanto, primitivas perceptivas. Procura-se enfatizar a habilidade humana de ler palavras mal escritas através da exploração da informação contida no corpo das palavras através das concavidades e convexidades.

O processo de leitura fornece o embasamento necessário à definição do método de reconhecimento das palavras manuscritas, a partir do momento que sabe-se que o leitor rápido lê ideograficamente. Assim, segue-se com a idéia da definição de modelos individuais, um para cada palavra do léxico.

Neste capítulo apresenta-se o reconhecimento de palavras através dos MEM, pois é este o método escolhido pelo presente trabalho. Isto, se deve a capacidade de oferecer um modelo probabilístico ao enfoque estrutural e permitir, através de algoritmos, o treinamento dos parâmetros a estimar dos modelos [HUANG et al., 1990] [RABINER & JUANG, 1993].

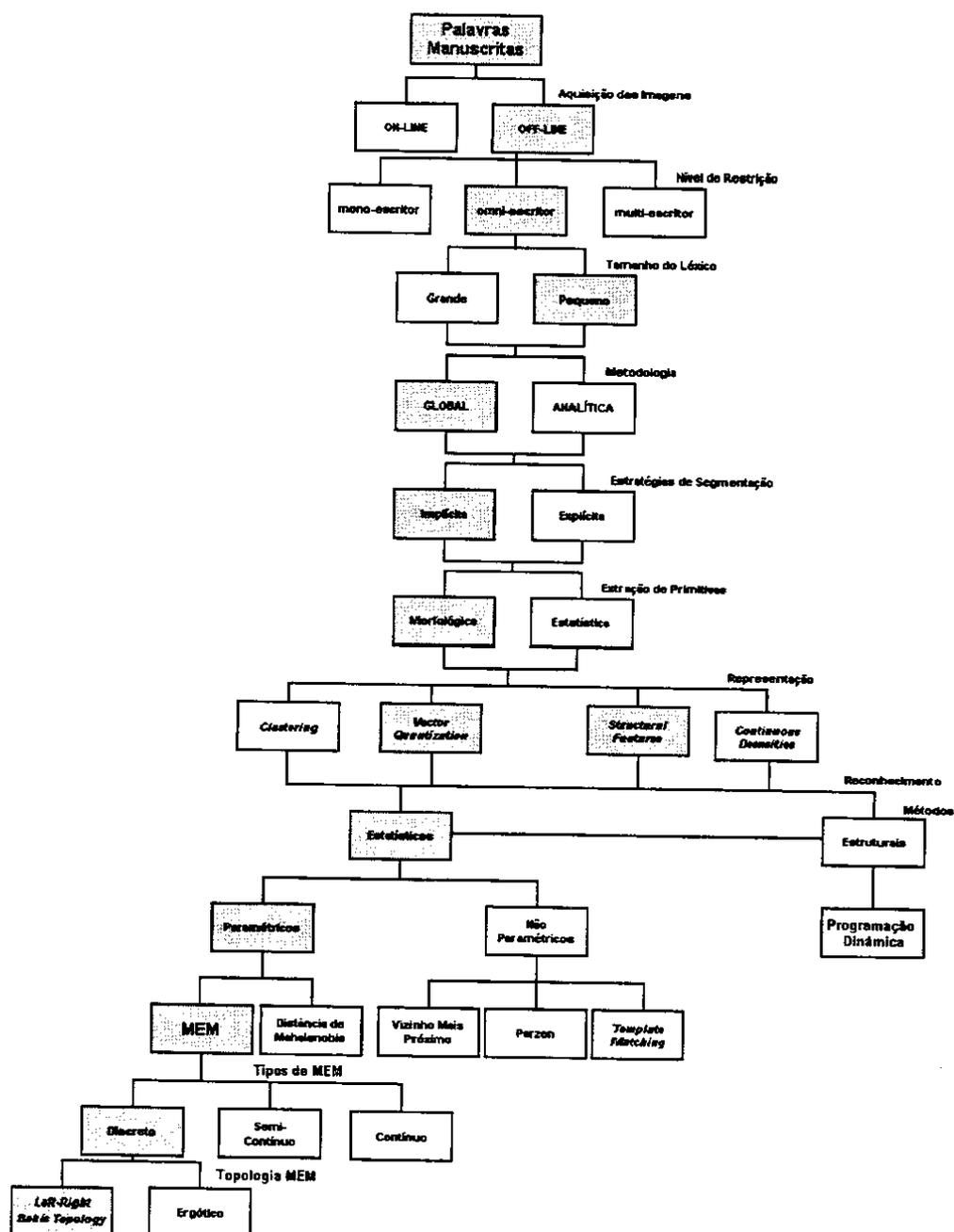


Figura 2.45: Variáveis e metodologias no reconhecimento de palavras manuscritas

O ponto forte dos MEM está primeiro no fato, da integração coerente dos diferentes níveis de modelagem (morfológico, léxico, sintático), e segundo, na existência de algoritmos eficazes para determinar os parâmetros dos modelos, fornecendo a melhor adequação entre o modelo e a base de dados (conhecimento), chamada de base de treinamento.

A Figura 2.45 classifica os MEM como métodos estatísticos não-paramétricos. Esta classificação segue a consideração dos MEM no que se refere aos modelos propriamente ditos. Os MEM são utilizados neste trabalho tomando-se como entrada as seqüências de observações provenientes do processo de extração de primitivas estruturais: perceptivas e geométricas.

A seguir, o Capítulo 3 apresenta as diversas etapas da formação da base de dados de cheques bancários brasileiros, fazendo-se referências às bases internacionalmente reconhecidas. Descreve-se a metodologia empregada no estabelecimento da base, coleta, aquisição digital e armazenamento das imagens, destacando-se as principais características dos cheques coletados.

Capítulo 3

Base de Dados

Este Capítulo descreve o processo de formação da base de dados de laboratório de imagens de cheques bancários brasileiros. Descreve-se a metodologia empregada no estabelecimento da base, coleta, aquisição digital e armazenamento das imagens. A base de dados de imagens de cheques bancários brasileiros constitui CD-ROM's e encontra-se disponíveis no Laboratório de Análise e Reconhecimento de Imagens de Documentos - LARDOC da PUC-PR.

3.1. Bases de Dados

Existem diversas bases de dados de cheques bancários em nível internacional (CENPARMI, CEDAR, NIST, SRTP, ETL, ESSEX, IRONOF), porém as mesmas não se enquadram nas especificações necessárias para atender os objetivos do presente trabalho. Algumas bases apresentam dados isolados que não representam os dados encontrados na realidade dos cheques brasileiros. Outras bases apresentam dados reais, porém em quantidades insuficientes para testes em processos de reconhecimento automático. Outras, ainda, apresentam somente imagens binárias, estando o usuário limitado ao processo de aquisição destas imagens, bem como aos processos e algoritmos de limiarização ou *thresholding* aplicados. Além do fato, de que as bases de dados que possuem os valores numéricos escritos por extenso e as datas, apresentam palavras em língua inglesa ou francesa, não sendo compatível com o estudo de palavras da língua portuguesa.

Deste modo, observa-se que as bases de dados internacionalmente relevantes possuem uma grande quantidade de imagens armazenadas, permitindo uma grande variedade de trabalhos de pesquisa relativos a diferentes etapas do processo de

reconhecimento de documentos manuscritos. A seguir estão apresentadas resumidamente as bases de dados reconhecidas em nível internacional:

- Centre d'Etudes en Reconnaissance des Formes et Intelligence Artificielle (CENERFIA/CENPARMI) - Montreal, Canadá: composta por 2500 cheques manuscritos preenchidos em inglês provenientes de 800 escritores e 1900 cheques manuscritos preenchidos em francês provenientes de 600 escritores. E, ainda, códigos de endereçamento postal (CEP ou *ZIP code*), com aproximadamente 17.000 dígitos obtidos a partir dos códigos de endereçamento postal, adicionalmente a um conjunto de 20.000 números manuscritos sem restrições coletados de 500 estudantes da Concordia University [GUILLEVIC, 1995] e [CÔTE, 1997];
- Center of Excellence in Document Analysis Recognition (CEDAR) - Buffalo, USA: possui dados de palavras manuscritas relativas a imagens de 5632 nomes de cidades, 4938 nomes de estados e 9454 códigos de endereçamento postal (CEP ou *ZIP code*), caracteres alfabéticos e numéricos mixados e dígitos extraídos de imagens digitais de endereços manuscritos [MOHAMED, 1995] e [KIM, 1996];
- United State National Institute of Standards and Technology (NIST) - Maryland, USA: as bases de dados correlatas com a área de reconhecimento de manuscritos são as de número 1, 3 e 7. Estas bases contém números e textos manuscritos (cada texto de um mesmo escritor), compondo um total de 402.953 dígitos, 220.304 caracteres manuscritos em caixa alta e 190.998 caracteres manuscritos minúsculos [HULL & FERINCH, 1993], [HULL, 1994] e [MONTOLIU, 1995];
- Service de Recherche Technique de la Poste (SRTP) - Paris, França, ligado ao Serviço de Correios da França (que também presta serviços bancários): possui aproximadamente 50.000 imagens de endereços extraídos de envelopes e cheques postais, podendo ser considerada a maior base de dados de imagens extraída de situações reais [GILLOUX et al., 1993], [EL YACOUBI, 1996], [AVILA, 1996] e [GRANDIDIER et al., 2000];
- Electrotechnical Laboratory in Japan (ETL) - Japão: a base de dados como um todo é sub-dividida em 9 sub-bases, sendo as bases ETL1 (141.319 imagens) e ETL6 (157.662 imagens) relacionados aos trabalhos de

reconhecimento de dados alfa-numéricos e a base ETL9 relacionada com os trabalhos de reconhecimento de caracteres [HULL & FERINCH, 1993];

- University of Essex (ESSEX) - Inglaterra: base de dados criada a partir da preenchimento à mão de envelopes padrões na Inglaterra [HULL & FERINCH, 1993];
- IReste ON/OFF Dual Database (IRONOFF) - Nantes, França: base de dados manuscrita com dupla entrada de dados (on-line e off-line) constituída por 32.000 caracteres isolados e 50.000 palavras cursivas [VIARD-GAUDIN et al., 1999];
- Empresa de Correios Brasileiros (ECT) - Brasil: possui 50.000 imagens de envelopes postais, sub-divididas em 4 sub-bases de 2 x 20.000 imagens e 2 x 5.000 imagens, com o objetivo do estudo da triagem automática de envelopes.

Como iniciativa brasileira, com cheques bancários brasileiros, pode-se citar a base de dados do Laboratório de Reconhecimento de Padrões e Rede de Computadores (LRPRC) da Universidade Estadual de Campinas, que possui 620 cheques bancários reais provenientes de 62 escritores [KOERICH, 1997].

A Tabela 3.1 apresenta um resumo das características das bases de dados já existentes e alguns trabalhos que fizeram uso das mesmas. Observa-se que as bases de dados possuem uma grande quantidade de imagens armazenadas, tanto para estudos de endereços (nomes de cidades, estados e códigos de endereçamento postal), bem como de cheques bancários e postais (valor numérico e por extenso).

Tabela 3.1: Características e aplicações de trabalhos com bases de dados existentes

Côte, M. (1997) – [CÔTE, 1997]						
BASE de DADOS	Cheques		Escritores		Palavras	
Tipo da Base	Inglês	Francês	Inglês	Francês	Inglês	Francês
CENPARMI	2500	1900	800	600	7837	7135
Aplicação: Sistema PERCEPTO	-	-	-	-	184(A) 2929(T)	-
Guillevic, D. (1995) – [GUILLEVIC, 1995]						
Tese Doutorado	2500	1900	800	600	5322(A) 2515(T)	-

Tabela 3.1: Características e aplicações de trabalhos com bases de dados existentes (continuação)

Mohamed, M. (1995) – [MOHAMED, 1995]										
BASE de Dados	Cidades		Estados		ZIP code		Caracteres		Dígitos	
Tipo da Base	A	T	A	T	A	T	A	T	A	T
CEDAR / CD-ROM 1	5100	532	4465	470	9019	435	19145	2183	5802	707
Montoliu, L. (1995) – [MONTOLIU, 1995]										
Base de Dados	Cheques			Números			Palavras			
Tipo da Base	A	T	A	T	A	T	A	T	A	T
NIST	----			223125 imagens			----			
LIX	1720	1585	10010	9259	16819	15538				

A = treinamento e T = testes.

Deste modo, optou-se pela criação de bases de dados de laboratório de cheques bancários brasileiros para atender as necessidades do presente estudo: reconhecimento de palavras extenso manuscrito.

3.2. Criação da Base de Dados de Laboratório de Cheques Bancários Brasileiros

A criação de uma base de dados, seja essa, de laboratório ou real, não constitui uma tarefa fácil. Pode-se dizer, que esta etapa é a mais trabalhosa e depende do esforço e colaboração de todo o grupo envolvido com atividades de pesquisa nesta área. Portanto, as deficiências e bons exemplos existentes em outras bases permitiram estabelecer os requisitos necessários da base a ser gerada com as finalidades já apresentadas.

Bases de dados reais, aquelas formadas com imagens de cheques reais, encontram uma grande barreira quando do fornecimento dos cheques reais, ou seja, das suas respectivas imagens pelas instituições bancárias. Por razões de proteção e segurança, no que diz respeito tanto aos clientes como das próprias instituições, os centros de pesquisas não conseguem obter os reais documentos bancários. A maioria dos estudos com cheques bancários reais brasileiros apresenta um número reduzido de exemplares.

Por todas estas razões, as pesquisas na área de reconhecimento de palavras manuscritas em cheques bancários dependem das bases de dados fornecidas pelas instituições responsáveis pelo seu manuseio e tratamento ou, na inexistência dessas

bases, trabalham com a criação de bases de laboratório, atendendo então, as necessidades peculiares do processo de reconhecimento de manuscritos.

A seguir apresenta-se cada uma das etapas da criação da base de dados, sendo estas as seguintes [FREITAS et al., 2000a]:

- critérios para o estabelecimento da base de dados,
- vocabulário de cheques bancários,
- geração dos valores dos cheques bancários,
- estatística prévia da base de dados,
- processo de coleta dos cheques,
- armazenamento e recuperação das informações,
- características da base de dados de treinamento.

3.2.1. Critérios de Estabelecimento da Base de Dados

As bases de dados de laboratório estabelecidas atendem aos seguintes critérios:

- valor mínimo de 0,01 (um centavo);
- valor máximo de 999.999,99 (novecentos e noventa e nove mil, novecentos e noventa e nove reais e noventa e nove centavos);
- garantia de repetição de no mínimo 20 vezes do mesmo dígito em cada posição do domínio do conjunto de valores numéricos e
- existência das palavras: real, reais, centavo e centavos.

Visando o reconhecimento as palavras manuscritas do extenso dos cheques bancários divide-se a base de dados original em 3 bases, denominadas como a seguir:

- BASE 1 - Base de Treinamento (A),
- BASE 2 - Base de Validação (V),
- BASE 3 - Base de Teste (T).

A dimensão de cada uma das bases é de 1600 exemplares de cheques de laboratórios, dado que o número de 20 (vinte) repetições para cada uma das 8 (oito) possíveis (xxx.xxx,xx) do maior valor numérico. Portanto, o valor máximo de R\$ 999.999,99 tem por justificativa a própria dimensão das bases de dados a serem coletadas e, ainda, que a inclusão de valores maiores, com cifras de milhão ou bilhão, acarreta somente a inclusão dos modelos destas palavras, sendo as palavras correspondentes aos valores numéricos tratados pelos mesmos modelos já definidos.

3.2.2. Vocabulário de Cheques Bancários

O vocabulário de cheques bancários estudado, considerando o domínio de valores a serem representados e reconhecidos está apresentado na Tabela 3.2. Este vocabulário tem por objetivo representar a maioria das possibilidades de escrita das palavras manuscritas encontradas em cheques bancários brasileiros. Não constituem objetos do presente estudo os erros ortográficos.

Tabela 3.2: Vocabulário de cheques bancários para o domínio estudado

ITEM	VALOR NUMÉRICO	EXTENSO CURSIVO
1	1	hum, um
2	2	dois
3	3	três, treis
4	4	quatro
5	5	cinco
6	6	seis
7	7	sete
8	8	oito
9	9	nove
10	10	dez
11	11	onze
12	12	doze
13	13	treze
14	14	quatorze, catorze
15	15	quinze
16	16	dezesseis
17	17	dezesete
18	18	dezoito
19	19	dezenove
20	20	vinte
21	30	trinta
22	40	quarenta
23	50	cinquenta, cincoenta
24	60	sessenta
25	70	setenta
26	80	oitenta
27	90	noventa
28	100	cem, cento
29	200	duzentos
30	300	trezentos
31	400	quatrocentos
32	500	quinhentos
33	600	seiscentos
34	700	setecentos
35	800	oitocentos
36	900	novecentos
37	1000	mil
38	----	real/reais
39	----	centavo/centavos

Fazem parte do vocabulário as seguintes palavras:

- 37 palavras relativas aos valores numéricos,
- 02 palavras que representam a moeda corrente no Brasil: “real/reais” e “centavo/centavos”.

3.2.3. Geração dos Valores dos Cheques Bancários

A geração aleatória de valores para preenchimento dos cheques se fez necessária, uma vez que, ao se solicitar às pessoas que preenchessem os cheques sem orientação alguma, muitos valores do tipo “um real”, ou “cem reais” seriam efetivados.

Deste modo, sabe-se que estudos de reconhecimento da parte manuscrita requerem do preenchimento de cheques uma distribuição da amostra de caracteres numéricos que possibilitem a uma distribuição da amostra de palavras dos valores por extenso adequada. A amostra de valores numéricos está diretamente ligada à amostra de palavras, pois sabe-se que o extenso tem sua origem no valor numérico.

A geração dos valores a serem preenchidos nos cheques de laboratório considera um gerador de números aleatórios com distribuição uniforme entre os dígitos [0,...,9] com igual probabilidade de ocorrência.

Controla-se o número de 20 (vinte) repetições de cada dígito, em cada posição possível do valor numérico, formatando-se o problema conforme Tabela 3.3, na forma R\$ w3w2w1.y3y2y1.x2x1.

As posições foram denominadas w3, w2, w1, y3, y2, y1, x2 e x1 correspondendo ao valor numérico de oito dígitos. Na Tabela 3.3 a posição x1 foi fixada com o dígito 0, repetidamente vinte vezes. Deve-se observar, também, que as posições w3, w2, w1, y3, y2, y1 foram fixadas com o dígito 0, nas linhas de 1 a 6, formando um conjunto de valores específicos, os quais com a geração aleatória poderiam ser determinados. A exceção é a coluna y1, linha 6 que foi fixada com o dígito 1, para que fossem gerados valores numéricos com a palavra “real”. Pois, somente, quando encontra-se o dígito 1 nesta posição é que existe a grafia da palavra “real”, em todos os demais casos a grafia é “reais”.

Fixou-se a posição x2, nas linhas de 1 a 6, com os dígitos 1,2,...,9, respectivamente, estabelecendo-se assim a possibilidade de grafia dos números de 10 a 99 centavos.

Tabela 3.3: Geração aleatória dos valores numéricos para a base de treinamento

CÓDIGO	POSIÇÃO								
	w3	w2	w1	y3	y2	y1	,	x2	x1
AE001	5	6	1	3	1	0	,	1	0
AE002	3	7	2	8	0	0	,	2	0
AE003	0	8	4	0	0	0	,	3	0
AE004	0	2	0	0	0	0	,	4	0
AE005	1	0	0	0	0	0	,	5	0
AE006	0	0	0	0	0	1	,	6	0
AE007	2	3	6	2	6	8	,	7	0
AE008	1	5	8	0	4	9	,	8	0
AE009	9	6	2	5	8	5	,	9	0
AE010	3	9	5	0	5	4	,	7	0
AE011	4	1	4	2	1	9	,	0	0
AE012	9	1	6	0	7	0	,	3	0
AE013	7	0	8	6	8	9	,	6	0
AE014	7	9	9	1	1	4	,	4	0
AE015	9	6	6	1	1	7	,	6	0
AE016	0	0	7	9	2	3	,	7	0
AE017	2	1	5	7	6	9	,	3	0
AE018	7	5	0	0	6	7	,	8	0
AE019	8	5	4	7	6	4	,	8	0
AE020	6	3	0	1	3	9	,	9	0

3.2.4. Estatística Prévia das Bases de Dados

Com base nos números aleatórios gerados para formação das 3 bases, efetuou-se uma estatística prévia sobre o total de dígitos, considerando o domínio [0,...,9]. A Tabela 3.4 apresenta o total de dígitos aleatórios gerados. Alguns exemplos de valores gerados aleatoriamente estão mostrados na Tabela 3.5.

Tabela 3.4: Total de dígitos aleatórios gerados

TOTAL de DÍGITOS	BASE 1 (Treinamento)	BASE 2 (Validação)	BASE 3 (Teste)
dígito 0	1.926	1.662	1.778
dígito 1	1.244	1.233	1.263
dígito 2	1.230	1.266	1.200
dígito 3	1.220	1.264	1.186
dígito 4	1.181	1.284	1.243
dígito 5	1.260	1.235	1.299
dígito 6	1.156	1.200	1.168
dígito 7	1.205	1.254	1.218
dígito 8	1.196	1.212	1.208
dígito 9	1.182	1.190	1.237
TOTAL de dígitos por base	12.800	12.800	12.800
TOTAL de números	1.600	1.600	1.600

Tabela 3.5: Exemplos de valores gerados

R\$ 0,60	R\$ 846519,86	R\$ 76748,10
R\$ 232268,70	R\$ 992924,35	R\$ 0,16
R\$ 37587,25	R\$ 7496,62	R\$ 55664,00
R\$ 9201,57	R\$ 10,71	R\$ 700,72

Para cada valor gerado aleatoriamente, obtiveram-se os extensos respectivos. A partir desses extensos realizou-se a contagem por caracteres e por palavras do vocabulário que pretende-se obter ao término do processo de coleta dos cheques. A Tabela 3.6 apresenta alguns exemplos de extensos obtidos a partir dos valores numéricos gerados aleatoriamente. As Tabelas 3.7 e 3.8 apresentam a contagem realizada sobre as bases considerando caracteres e palavras, respectivamente. Na Tabela 3.8 indica-se como Base 1 os valores estimados e como Base 1* os valores coletados de palavras manuscritas.

Tabela 3.6: Exemplos de valores numéricos por extensos

Quinhentos mil reais e cinquenta centavos
Cento e vinte e dois mil, setecentos e setenta e quatro reais e noventa centavos
Setenta e três mil e noventa reais
Oitocentos e trinta e dois mil, quatrocentos e noventa e um reais e um centavo

Tabela 3.7: Contagem dos caracteres

LETRAS	BASE 1	BASE 2	BASE 3
a	7.561	7.750	7.612
c	4.649	4.771	4.800
d	1.004	1.006	967
e	21.369	21.902	21.699
h	312	293	311
i	7.647	7.732	7.607
l	1.566	1.561	1.564
m	1.965	2.022	1.975
n	10.405	10.578	10.559
o	9.004	9.197	9.254
q	2.018	2.097	2.085
r	5.583	5.731	5.613
s	10.369	10.628	10.495
t	12.249	12.634	12.370
u	2.241	2.407	2.305
v	3.222	3.251	3.279
z	1.070	1.040	1.070
ê	421	422	412
ü	475	482	484
TOTAL	103.130	105.504	104.461

Tabela 3.8: Contagem das palavras

Palavras	BASE 1	BASE 1*	BASE 2	BASE 3
Um	399	332	464	415
Dois	435	350	417	403
Três	421	336	422	412
Quatro	404	322	425	427
Cinco	423	332	422	457
Seis	398	302	404	362
Sete	383	267	426	420
Oito	390	305	402	402
Nove	396	318	400	420
Dez	53	47	37	32
Onze	60	50	34	55
Doze	37	29	45	48
Treze	47	37	46	43
Quatorze	46	35	47	52
Quinze	50	45	38	47
Dezesseis	35	30	50	48
Dezessete	48	36	45	46
Dezoito	41	34	33	44
Dezenove	56	45	51	57
Vinte	459	382	476	460
Trinta	454	372	510	422
Quarenta	456	338	499	453
Cinqüenta	475	376	482	484
Sessenta	441	356	442	481
Setenta	475	370	465	453
Oitenta	477	363	465	457
Noventa	439	316	435	453
Cem	10	7	7	6
Cento	302	269	302	315
Duzentos	299	249	328	289
Trezentos	298	236	286	309
Quatrocentos	275	218	313	311
Quinhentos	312	240	293	311
Seiscentos	282	219	304	277
Setecentos	299	248	318	299
Oitocentos	288	252	312	305
Novocentos	291	236	304	307
Mil	1.556	1.286	1.551	1.554
Real	10		10	10
Reais	1.572	1.262	1.573	1.574
Centavo	25		18	13
Centavos	1.556	1.094	1.567	1.569
e	7.537	-	7.801	7.653
TOTAL	22.710	11.948	23.269	22.955

3.2.5. Processo de Coleta dos Cheques de Laboratório

Inicialmente criou-se um cheque de laboratório, com base no Modelo Único de cheque bancário brasileiro, instituído pelo Conselho Monetário Nacional, o qual estabelece a distribuição de informações nos cheques, o modelo padrão de

preenchimento, as dimensões mínima e máxima e seus elementos componentes [BANCO CENTRAL DO BRASIL, 1983], conforme descrito no Apêndice A.

Procedeu-se, então, a coleta dos cheques de laboratórios preenchidos. Este processo foi realizado de forma que as pessoas preenchessem em uma folha em branco, sob uma mesa de luz, tendo como referência uma folha com quatro cheques de laboratórios seguidos. As pessoas foram orientadas a proceder da seguinte forma:

- cheque 1: somente o valor numérico;
- cheque 2: somente a 1ª linha do valor por extenso;
- cheque 3: a 2ª linha do valor por extenso;
- cheque 4: a data.

Isto foi necessário, para que fossem eliminadas quaisquer interferências entre o preenchimento das diversas linhas, das próprias linhas do cheque de laboratório e outras marcas existentes. Desta forma, pode-se iniciar o trabalho de pesquisa, considerando o cheque já segmentado, ou seja, processado e extraído somente as informações manuscritas. A Figura 3.1 exemplifica um cheque de laboratório preenchido. Ressalta-se que a coleta de dados de assinaturas em cheques foi realizado em processo e formulário distintos dos aplicados na coleta dos dados aqui em questão.

Todas as pessoas que participaram como voluntárias, forneceram autorização para uso dos dados coletados. Ao participar do processo, as pessoas recebiam o valor numérico e a data (previamente definidos) gerados aleatoriamente e estavam livres para preencher os dados da forma que costumam preencher um cheque real, conforme apresentado na Figura 3.2, para os valores por extenso.

Puderam ser utilizadas, durante o preenchimento, canetas esferográficas (preta ou azul), tinteiro ou de ponta porosa. O preenchimento a lápis não foi permitido, pois o mesmo não é permitido em cheques reais, além de gerar imagens de baixa qualidade.

Segundo as Normas do Banco Central do Brasil [BANCO CENTRAL DO BRASIL, 1983] o preenchimento de um cheque pode ser feito a mão ou a máquina, pelo próprio emitente ou por terceiros. Obviamente, a assinatura deve ser de próprio punho do emitente, admitindo-se a chancela mecânica. No que concerne ao valor por extenso as normas estabelecem que é obrigatória a grafia por extenso do valor numérico correspondente, sendo facultativa a grafia dos centavos somente por extenso.

The image shows four examples of filled-out laboratory checks. Each check has a header 'Pague por este cheque a quantia de' followed by a line for the amount and another line for the date. The amount is written in words, and the date is written in digits. The checks are numbered 1, 2, 3, and 4.

Check 1: Amount: #610.247,31; Date: [blank]

Check 2: Amount: (seiscentos e dezeses mil duzentos e quarenta e [blank]); Date: [blank]

Check 3: Amount: sete reais e trinta e um centavos; Date: [blank]

Check 4: Amount: [blank]; Date: Curitiba, 22 de abril de 99

Figura 3.1: Exemplo de cheque de laboratório preenchido

Todos estes cuidados foram tomados, uma vez que, o processo de coleta dos cheques de laboratório demanda uma parcela de tempo razoável e, portanto, nenhum exemplar de coleta deve ser eliminado ou descartado.

3.2.6. Armazenamento e Recuperação das Informações

As imagens dos cheques de laboratórios coletados foram adquiridas através de scanner FUJITSU M30096GX, em 300 dpi, 256 níveis de cinza, 100% da escala, 80 de brilho e 128 de contraste, formatos BMP e TIFF.

O formato BMP não oferece uma boa compressão das imagens porém facilita o processamento posterior em programas para plataforma Windows. O formato TIFF é mais utilizado pelos programas de plataforma UNIX. Para o armazenamento das

imagens em CD-ROM foi utilizado o formato JPEG, o qual permite um alto grau de compressão das imagens com pouca perda de informação.

duzentos e quinze mil, setecentos e sessenta e
nove reais e trinta centavos

AE0017

(selecemos e cinquenta mil e sessenta e sete reais
e oitenta centavos)

AE0018

Quatrocentos mil, cinquenta e um reais

AE0025

duzentos e dezesseis mil, cento e vinte e um
reais e setenta e um centavos

AE0027

Quateocentos e Quarenta e Quateo mil, Tre
zentos e Sessenta e Sete Reais e Oitenta e um Centavos

AE0028

Figura 3.2: Exemplos de estilos de valores por extenso

Após o processo de aquisição das imagens, essas foram cadastradas segundo uma codificação que permite armazenar e recuperar os cheques, bem como informações específicas encontradas nos cheques.

A codificação utilizada permite que os itens que compõem o cheque bancário possam ser armazenados individualmente e ao mesmo tempo serem identificados como um único elemento, ou seja, o cheque. Os itens que compõem o cheque são:

- Valor numérico;
- Extenso;
- Data .

A estrutura lógica do código para identificação dos itens dos cheques de laboratório é composta por:

- 2 (duas) letras: L1 e L2 e
- 4 (quatro) dígitos : D1 a D4, conforme Figura 3.3.



Figura 3.3: Esquema lógico da codificação dos cheques de laboratório

Por sua vez, cada campo possui a seguinte descrição:

- **Campo L1:** Identifica o tipo de base, podendo ser:
 - ⇒ A = treinamento;
 - ⇒ V = validação e
 - ⇒ T = teste.
- **Campo L2:** Identifica o item do cheque contido na base, podendo ser:
 - ⇒ N = número;
 - ⇒ E = extenso;
 - ⇒ D = data.
- **Campo D1 a D4:** Número de 1 a 1600, identificando o valor numérico e a data do cheque (gerados aleatoriamente). A Tabela 3.9 apresenta e explica alguns exemplos de códigos.

Tabela 3.9: Exemplos da codificação dos cheques de laboratório

L1	L2	D1	D2	D3	D4	Descrição
A	N	0	0	0	1	Cheque da base de treinamento, com valor numérico referente ao primeiro valor numérico gerado.
A	E	1	6	0	0	Cheque da base de treinamento, com valor por extenso referente ao último valor numérico gerado.
T	D	1	2	4	5	Cheque da base de testes, com data referente ao valor numérico gerado na posição 1245.
V	E	0	0	3	3	Cheque da base de validação, com extenso referente ao valor numérico gerado na posição 33.

Atualmente as bases de dados de cheques de laboratório referentes ao extenso manuscrito encontram-se na seguinte situação:

- Base de Treinamento: Coleta realizada, aquisição digital (*scanner*), tamanho de 1,3Gbytes, 256 níveis de cinza. Esta base de dados encontra-se em dois formatos: base dos valores por extenso e base de palavras isoladas,
- Base de Teste: em processo de coleta, com 416 exemplares já coletados,
- Base de Validação: não iniciado processo de coleta.

Neste trabalho utiliza-se 2016 cheques sendo que após o processo de segmentação do extenso em palavras obteve-se um conjunto de 11.948 palavras. Portanto, divide-se este conjunto de palavras isoladas em 3 sub-bases, a saber:

- 60% para treinamento = 7.146 palavras,
- 20% para validação = 2.427 palavras e
- 20% para testes = 2.375 palavras.

3.2.7. Características da Base de Dados de Treinamento

A base de dados de treinamento possui relevante importância no processo de desenvolvimento de um método de reconhecimento de palavras manuscritas. Isto devido ao fato de que é sobre esta base que se realiza o aprendizado dos modelos de reconhecimento, especialmente no caso das aplicações que envolvem modelos estatísticos, como por exemplo: MEM e Redes Neurais. Para que tal tarefa tenha êxito torna-se importante o conhecimento e análise da base de dados de treinamento sob dois aspectos: o tipo de escrita e os símbolos utilizados como iniciais e terminais do extenso manuscrito em cheques bancários brasileiros.

3.2.7.1. Estilo de Escrita

A base de dados de treinamento foi avaliada quanto aos estilos de escrita cursiva encontrados. Para [TAPPERT et al., 1990] existem 5 categorias distintas de escrita cursiva, conforme já apresentado na Figura 2.11. A análise visual dos cheques de laboratórios preenchidos encontrou as seguintes categorias:

- palavras em caracteres disjuntos (caixa alta);
- palavras em caracteres disjuntos com espaçamento regular;
- palavras em escrita cursiva pura, ou seja, todos os caracteres de uma palavra são conectados dois a dois;
- palavras em escrita mixada, ou seja, misturando os demais tipos de escrita, e ainda incluindo a classe de palavras com caracteres disjuntos com a presença

de vínculos entre eventuais caracteres. Observou-se que quando o escritor gera palavras formadas por caracteres disjuntos com a presença de vínculos entre eventuais caracteres, o escritor utiliza aleatoriamente as outras 3 categorias de tipo de escrita cursiva.

A Tabela 3.10 apresenta a distribuição dos tipos de escrita na base de dados de treinamento, mostrando que a escrita cursiva é o tipo mais encontrado nesta base de dados.

Tabela 3.10: Distribuição dos tipos de escrita na base de treinamento

Tipo de Escrita	Exemplo	Porcentagem
Cursiva Pura	<i>reais</i>	72%
Caixa Alta	REAIS	13%
Caracteres Disjuntos	<i>reais</i>	7%
Outros	<i>reais</i>	8%

3.2.7.2. Símbolos Iniciais e Terminais do Extenso dos Cheques

Encontrou-se uma grande variedade de símbolos utilizados como iniciais e terminais no valor numérico por extenso. Porém, inicialmente deve-se explicar o que estão sendo considerados como símbolos iniciais e terminais.

Símbolo inicial e terminal é qualquer tipo de simbologia utilizada no início e/ou no final do valor por extenso em um cheque bancário brasileiro. É possível que as pessoas ao preencherem um cheque iniciem por “(“ ou “#” e terminem, após a grafia do valor numérico por extenso, com “.”, “)”, ou simplesmente um traço longo. A Figura 3.4 exemplifica cheques da base de dados de treinamento com símbolos inicial e terminal.

Observa-se que os escritores não utilizam na sua grande maioria (94%) um símbolo inicial, porém o caracter “(“ é o mais utilizado (75%) para iniciar o preenchimento do extenso quando o fazem. Como símbolo terminal, a maioria (63%) não utiliza qualquer símbolo, mas quando o fazem (55%), utilizam o caracter “.” para terminar o preenchimento.

Estas duas características, tipo de escrita e símbolos iniciais e terminais, permitem conhecer melhor a base de dados de treinamento e avaliar as possíveis dificuldades a serem encontradas, por exemplo, no processo de segmentação do extenso em palavras.

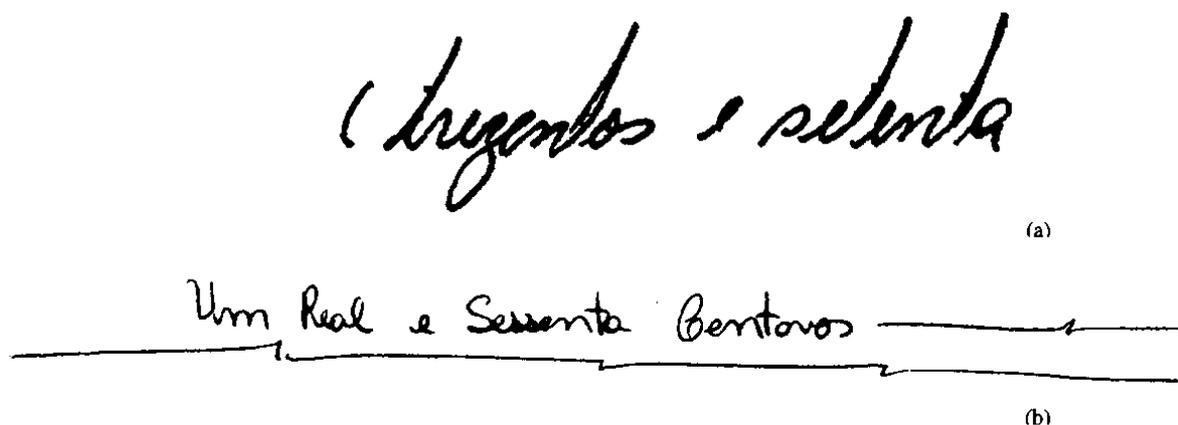


Figura 3.4: Exemplos de símbolos: a) inicial e b) terminal

3.2.8. Considerações Finais

Neste Capítulo descreve-se a criação da base de dados de laboratório, a qual se reveste de grande importância para todos os pesquisadores, visto que estas imagens constituem a matéria prima para a realização dos seus trabalhos.

A criação da base de dados de laboratórios de cheques bancários brasileiros ainda encontra-se em processo de coleta, como mencionado anteriormente. Porém, foram coletados 2.016 cheques de laboratório, perfazendo um total de 15.139 dígitos isolados e 15.636 palavras isoladas, sendo 11.948 palavras do valor numérico por extenso e 3.688 palavras relativas aos meses da data. Ao final do processo, espera-se obter uma base de dados com 4.800 imagens de cheques [FREITAS et al., 2000a]. Estas imagens constituem CD-ROM's e encontram-se disponíveis no Laboratório de Análise e Reconhecimento de Imagens de Documentos - LARDOC da PUC-PR.

Utiliza-se a base de dados descrita nos processos de treinamento, validação e teste de métodos de reconhecimento da parte manuscrita de cheques bancários brasileiros através dos Modelos Escondidos de Markov.

No que concerne ao usuário final, será a partir de bases de dados que os sistemas de automação do processamento de documentos com informações manuscritas poderão ser desenvolvidos, validados e testados. Então, esses sistemas poderão realmente reduzir o tempo de processos cotidianos e melhorar a eficiência global de sistemas bancários.

O Capítulo 4 apresenta a metodologia do trabalho através da descrição e detalhamento das estratégias, métodos e técnicas utilizados nas diversas etapas do reconhecimento das palavras isoladas do extenso manuscrito em cheques bancários brasileiros.

Capítulo 4

Um Método para Reconhecimento de Palavras

Manuscritas do Extenso de Cheques Bancários

Brasileiros

Os capítulos anteriores fornecem os subsídios necessários para que se possa apresentar os materiais e métodos desenvolvidos visando o reconhecimento de palavras manuscritas do extenso manuscrito em cheques bancários brasileiros. A revisão bibliográfica do Capítulo 2 apresenta os aspectos relevantes da escrita, do conhecimento prévio, da leitura e do reconhecimento. Além é claro, do estado da arte da área de Reconhecimento de Palavras Manuscritas. O Capítulo 3 descreve a formação da base de dado utilizada no presente trabalho.

Este capítulo está organizado de acordo com a seqüência de etapas necessárias ao desenvolvimento de um método de reconhecimento, como mostrado na Figura 1.2, sendo estas as seguintes:

- Pré-processamento: procedimentos preliminares, segmentação do extenso manuscrito em palavras isoladas, correção da inclinação vertical dos caracteres e detecção do corpo das palavras,
 - Extração de primitivas: Perceptivas e Geométricas e
 - Treinamento e validação dos modelos das palavras utilizando MEM e
 - Reconhecimento das palavras do extenso manuscrito em cheques bancários utilizando MEM.

com uma iteração, para o primeiro maior valor de pixel vizinho. Como as imagens são binárias (preto ou branco), o maior valor será o do pixel branco.

As vantagens do procedimento proposto são as vantagens herdadas do procedimento de fechamento de imagens em níveis de cinza, de acordo com [FACON, 1995]:

- suaviza as fronteiras pelo exterior;
- preenche buracos;
- emenda partículas próximas e
- torna o conjunto mais regular.

A Figura 4.1 apresenta alguns exemplos de resultados obtidos com a aplicação da filtragem *soft*.

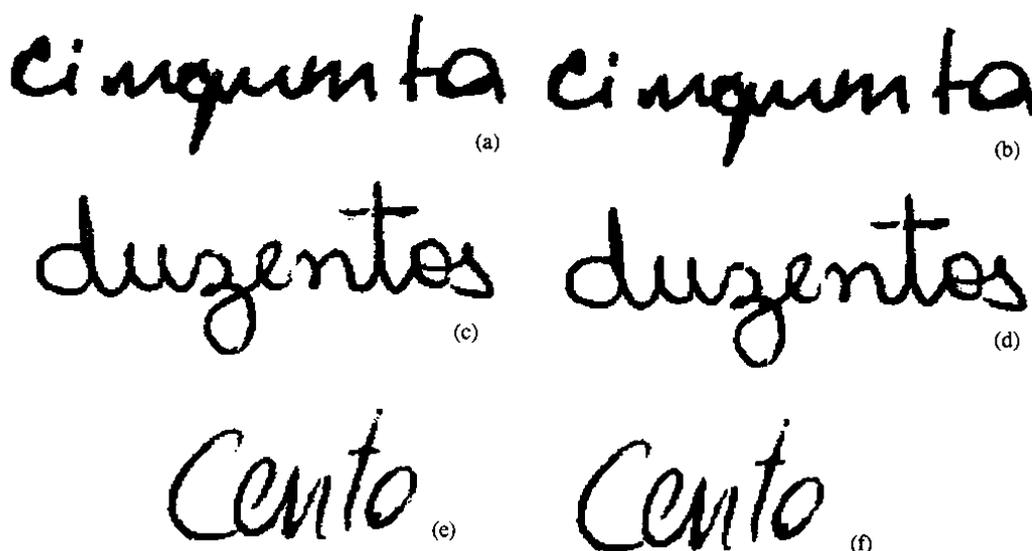


Figura 4.1: Exemplos de filtragem *soft*: a, c, e) imagens originais, b, d, f) imagens filtradas

4.1.2. Segmentação do Extenso Manuscrito em Palavras

A segmentação do extenso manuscrito em palavras possui dois objetivos, a saber:

- o de isolar as palavras para facilitar a realização da etapa de desenvolvimento dos MEM específicos de cada palavra do léxico,
- o de estabelecer uma estratégia de segmentação para o tratamento do extenso completo.

O presente trabalho refere-se ao primeiro objetivo, pois trata do reconhecimento das palavras isoladas e não do extenso como um todo.

Este processo de segmentação tem conhecimento das informações relativas ao tipo da escrita e dos símbolos iniciais e terminais, conforme apresentado no Capítulo 3. Porém inicialmente, o procedimento para segmentar o extenso em palavras não realiza nenhum tratamento específico para reconhecimento dos tipos de símbolos utilizados como elementos iniciais ou terminais do extenso propriamente dito.

A determinação do histograma de projeção vertical do extenso completo auxilia no processo de localização das diversas palavras que compõem o mesmo. Paralelamente, a determinação deste histograma são definidos dois tipos de espaçamento: o entre palavras e o intra-palavra.

O espaçamento entre palavras é por definição o espaço utilizado pelo escritor para separar duas palavras. E, o espaçamento intra-palavra é entendido como o espaço existente dentro de uma palavra, entre dois caracteres disjuntos. A Figura 4.2 apresenta a determinação do histograma de projeção vertical e exemplifica os dois tipos de espaçamentos definidos, isolando então, as diversas palavras contidas no extenso completo.



Figura 4.2: Exemplos de espaçamentos intra e entre palavras

Um primeiro procedimento de segmentação do extenso em palavras permitiu iniciar o estudo dos espaçamentos intra-palavra e entre as palavras. Esse estudo permite que este algoritmo de segmentação possa ser melhorado quando da sua aplicação no tratamento do extenso completo. O algoritmo utilizado para segmentar a base de dados

em palavras, evitando assim, o processo manual, tem por base o cálculo do tamanho médio de todos os espaçamentos horizontais existente em um extenso, obtidos pela análise do histograma de projeção vertical da imagem. Desta forma, os espaçamentos horizontais menores ou iguais a média são considerados, para o extenso em questão, como um espaçamento intra-palavra, não gerando segmentação. Por outro lado, os espaçamentos horizontais maiores que a média são considerados como espaçamento entre palavras, gerando segmentação.

O referido procedimento foi aplicado em toda a base de dados, gerando 11.948 palavras isoladas. Os erros existentes no processo foram os seguintes:

- segmentar uma palavra, devido a existência de um espaço intra palavra maior que a média dos espaçamentos existentes no extenso analisado e
- considerar uma palavra seguida de “e” ou “,” como uma única palavra, devido a existência de um espaço entre palavras menor ou igual a média dos espaçamentos. Este erro ocorreu também, somente em dois casos, onde o escritor realizou a grafia das palavras muito próximas, por exemplo “cinquenta e seis”.

Outro ponto a ser considerado, são os símbolos iniciais e terminais utilizados pelos escritores. Na segmentação realizada não foram considerados como erro a segmentação de uma palavra associada a um símbolo inicial ou terminal. Todos esses casos foram resolvidos com mais uma passada pelo procedimento de segmentação.

Os dois tipos de erros ocorrem em 1076 palavras, representando 9% de erro no total das palavras segmentadas. Com isto, o resultado obtido mostra que o procedimento aplicado é eficiente, evitando o procedimento manual de segmentação do extenso em palavras.

Para o tratamento do extenso como um todo, este procedimento de segmentação do extenso em palavras isoladas deve ser modificado. Portanto, um estudo detalhado do espaçamentos existentes nas imagens dos extenso bancários deverá ser realizado. Com base nesses futuros estudos uma estratégia de segmentação do extenso em palavras poderá ser formulada. Essa estratégia refere-se a representação e modelagem dos espaçamentos, visando desenvolver um processo de segmentação robusto.

Um procedimento de segmentação robusto objetiva minimizar a segmentação excessiva ou *over-segmentation* e evitar segmentações que gerem palavras distintas falsamente unidas como uma única palavra ou *under-segmentation*.

A segmentação do extenso manuscrito em palavras pode parecer uma tarefa fácil para o ser humano, pois ao ler um texto consegue-se localizar palavra a palavra. No caso de dúvidas, basta ao ser humano aplicar o conhecimento prévio que possui e avaliar informações do contexto. Porém, o processo automático de localização de palavras pode se tornar uma grande dificuldade quando da identificação individual de palavras em um texto ou uma simples frase.

Segundo [PAQUET & LECOURTIER, 1993] assumir que os espaços entre palavras são sempre maiores que os espaços intra-palavras pode não ser válido para todos os casos. Fato este verdadeiro, uma vez que o procedimento aplicado resultou em 9% de erro. Ainda segundo os mesmos autores, este é o porque do processo de segmentação ser entendido como um provedor dos separadores das possíveis palavras, o que será confirmado ou não, dependendo da estratégia aplicada.

A estratégia mais comum é a chamada estratégia heurística que tem por base um processo de limiar global ou *global threshold*. Em resumo, os espaços com largura superior ao valor limite são considerados como separadores e os espaços com largura inferior ao valor limite são descartados.

Os estudos que podem ser realizados com os espaçamentos intra e entre palavras devem caracterizar-se por fornecer informações à modelagem dos espaçamentos através dos MEM, em lugar de aplicar-se uma estratégia heurística de segmentação do extenso em palavras. Desta forma, a melhor opção não é segmentar explicitamente, ou seja, o utilizar um valor fixo globalmente no extenso.

Outra estratégia é a recomendada por [PAQUET & LECOURTIER, 1993] na qual nenhuma decisão definitiva é tomada sobre os espaços em branco. E sim, os espaçamentos são considerados como fonte de diferentes formas de segmentação de um mesmo extenso. Neste caso, pode-se considerar a opção pela segmentação implícita, ou seja, que o extenso será tratado como uma única unidade com múltiplas opções de segmentação.

Uma proposta para este método é tratar o extenso manuscrito através das seguintes estratégias de segmentação:

- Segmentação Explícita: avaliar o desenvolvimento de MEM para modelar os espaçamentos intra e entre palavras, desde que os estudos realizados demonstrem a existência de dois conjuntos distintos de espaçamentos sobre a base de dados de treinamento;

- **Estratégia Implícita:** considerar todo e qualquer espaçamento como provável espaçamento entre palavras. Efetuar uma análise de espaçamentos ambíguos. Neste caso, serão considerados como espaçamentos ambíguos, da mesma forma que [PAQUET & LECOURTIER, 1993], aqueles que tanto podem ser modelados como espaçamento intra ou entre palavras.

De um modo geral, estas estratégias de segmentação fornecem ao método um mecanismo de adaptação aos diferentes aspectos de variabilidade da escrita cursiva humana.

4.1.3. Correção da Inclinação Vertical da Escrita

Com base na revisão bibliográfica, muitas técnicas podem ser aplicadas na correção da inclinação vertical dos caracteres, por exemplo, as técnicas apresentadas por [EL YACOUBI, 1996] e [GUILLEVIC, 1995] com base nos histogramas de projeção horizontal. Realizou-se a implementação de ambas as técnicas, iniciando pela técnica proposta em [EL YACOUBI, 1996], conforme descrição no Capítulo 2.

Em um segundo estudo, aplica-se o algoritmo dado por [GUILLEVIC, 1995]. Esta técnica tem como princípio a determinação dos histogramas de projeção horizontal das imagens, como na técnica proposta por [EL YACOUBI, 1996], para vários ângulos a partir do eixo vertical em passos de “delta_teta” graus. [GUILLEVIC, 1995] considera “delta_teta” igual a 5° .

Considerando inicialmente a premissa do valor de “delta_teta” da mesma forma que feito por [GUILLEVIC, 1995] percebe-se que para inclinações pequenas o algoritmo não gera alteração do ângulo de inclinação. Portanto, o algoritmo foi alterado para aceitar qualquer valor de “delta_teta”, permitindo que testes fossem realizados.

Outra dificuldade do algoritmo de [GUILLEVIC, 1995] reside no caso de traços longos que podem influenciar a determinação do ângulo de inclinação uma vez que estes traços possuem um grande concentração de pixels. Portanto, o algoritmo proposto trabalha com *delta_teta* a partir de 1 grau e avalia inclinações entre 0 e ± 90 graus. Desta forma algumas dificuldades encontradas pelo autor foram superadas. A Figura 4.3 apresenta alguns resultados obtidos com esta técnica.

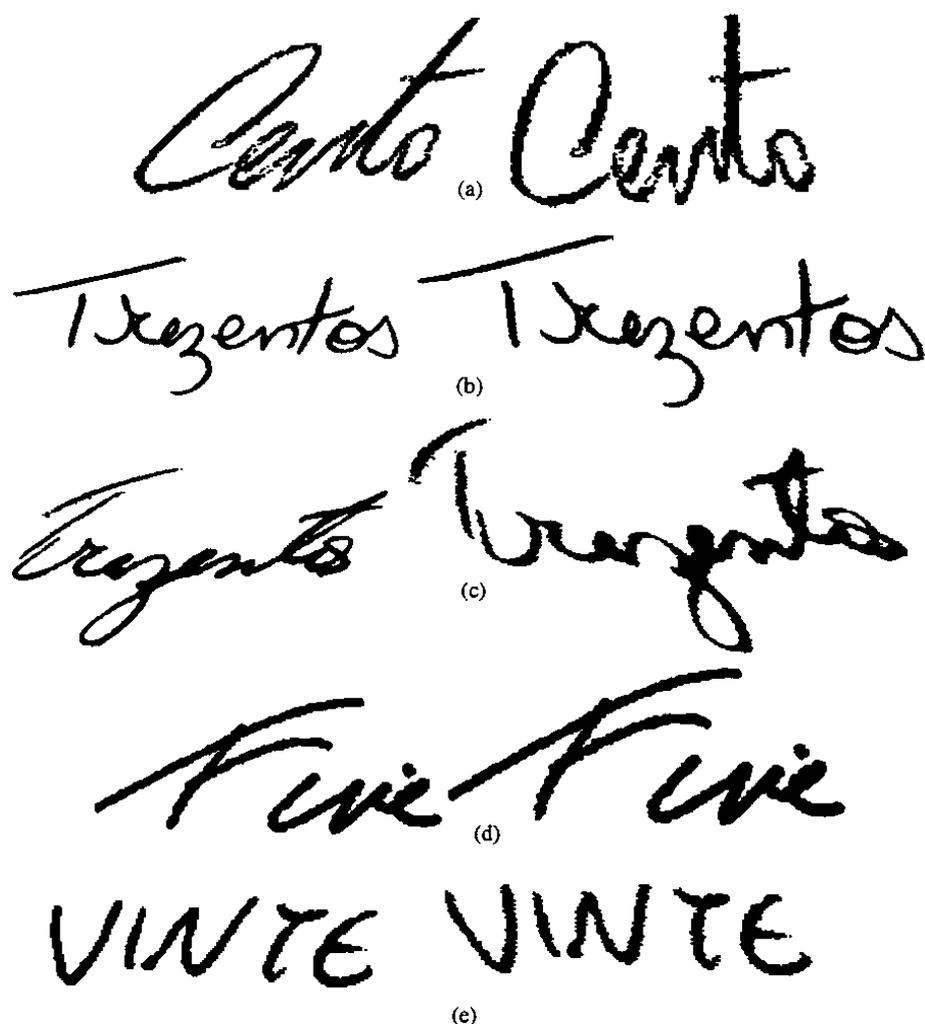


Figura 4.3: Exemplos de correção da inclinação vertical dos caracteres: a) correção de 40 graus, b) correção de 23 graus, c) correção de 58 graus, d) imagem extraída de [GUILLEVIC, 1995] com 7 graus de inclinação (ver resultado obtido por [GUILLEVIC, 1995] na Figura 2.22), e) correção de 29 graus

Ambas as técnicas fazem uma estimativa global da inclinação dos caracteres, sendo apresentadas como desvantagens, o tempo de processamento [EL YACOUBI, 1996] e [GUILLEVIC, 1995], e ainda, a influência na determinação do ângulo de inclinação para os casos de traços longos horizontais com inclinação diferente de 0 graus [GUILLEVIC, 1995].

Deste modo, opta-se por aplicar uma técnica simples, porém rápida, apresentada por [EL YACOUBI et al., 1999] baseada no contorno externo das palavras. Para todos os componentes da imagem extraí-se o contorno externo e a espessura média do traçado destes componentes. Percorre-se, então, o contorno externo com um passo igual a espessura e, para estes pontos do contorno, calcula-se o ângulo de inclinação médio, considerando-se todos os componentes da imagem analisada.

Após a correção da inclinação vertical dos caracteres aplica-se um procedimento de suavização da imagem da palavra de acordo com [STRATHY, 1993]. O objetivo é regularizar a continuidade do contorno das palavras, eliminando pequenos ruídos nas imagens. A Figura 4.4 mostra os resultados obtidos com a aplicação das etapas de correção da inclinação vertical das palavras e de suavização.

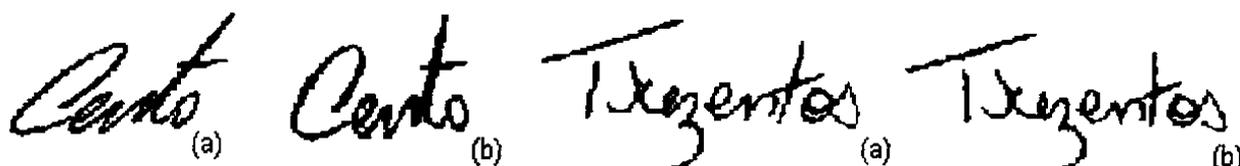


Figura 4.4: Pré-processamento das palavras: a) imagens originais e b) correção da inclinação dos caracteres e suavização do contorno externo

4.1.4. Detecção do Corpo da Palavra

A detecção do corpo da palavra permite estabelecer 3 zonas principais que são a base da relação entre a primitiva extraída e a sua localização na palavra. Portanto, as principais zonas de uma palavra podem ser definidas como:

- **Z1 = zona ascendente:** compreendida entre o limite superior máximo (LSM) da imagem e o limite superior do corpo da palavra (LS);
- **Z2 = zona corpo da palavra:** compreendida entre o limite superior (LS) e inferior (LI) do corpo da palavra;
- **Z3 = zona descendente:** compreendida entre o limite inferior do corpo da palavra (LI) e o limite inferior mínimo da palavra (LIM).

A Figura 4.5 exemplifica as zonas de uma palavra manuscrita em estilo cursivo puro. Deve-se ressaltar que para palavras formadas por caracteres disjuntos contidos em retângulos pré-impresos (bastão ou caixa alta) [TAPPERT et al, 1990] os limites superior máximo e inferior mínimo da palavra coincidem com os limites superior e inferior do corpo da palavra, respectivamente.

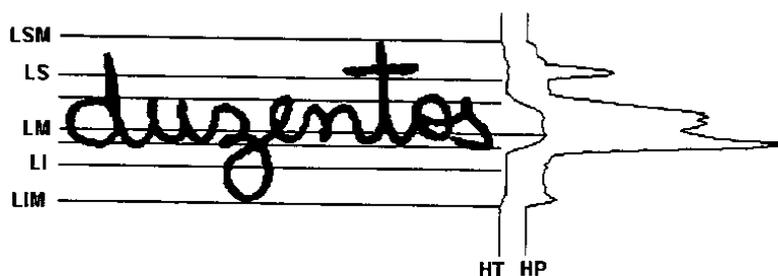


Figura 4.5: Zonas de uma palavra

A determinação das zonas ascendente, corpo da palavra e descendente é realizada através do histograma de projeção horizontal das transições (HT) branco-preto da palavra. Inicialmente, determina-se o pico máximo deste histograma, ou seja, a linha onde mais ocorrem transições branco-preto, denominada com Linha Média (LM). Aplica-se uma suavização no histograma para eliminar inflexões bruscas, seguindo a seguinte formulação:

$$P_x' = \frac{P_{x-2} + P_{x-1} + P_x + P_{x+1} + P_{x+2}}{5} \quad (4.1)$$

Com o histograma suavizado e a partir do pico (LM) buscam-se os limites LS e LI para 70% do valor do pico do histograma de transições. A determinação deste limiar heurístico baseou-se no estudo da diferença entre as coordenadas y dos picos do histograma de transição branco-preto (HT) e do histograma de densidade de pixels (HP), para 1.742 imagens das palavras de "um" até "dez" da base de treinamento.

Ao encontrar-se as coordenadas y dos pontos com 70% do valor de pico de HT, LS e LI são determinados acrescentando-se às coordenadas encontradas 50% da altura do corpo da palavra, conforme Figura 4.5. Isto deve-se ao fato de que nenhum procedimento de correção da linha de base está sendo aplicado, uma vez que, o preenchimento do valor extenso nos cheques conta com o auxílio de duas linhas de referência.

A Figura 4.6 exemplifica a determinação das diferentes zonas para palavras da base de dados de treinamento. Os exemplos sofreram a aplicação de todas as etapas de pré-processamento descritas anteriormente.

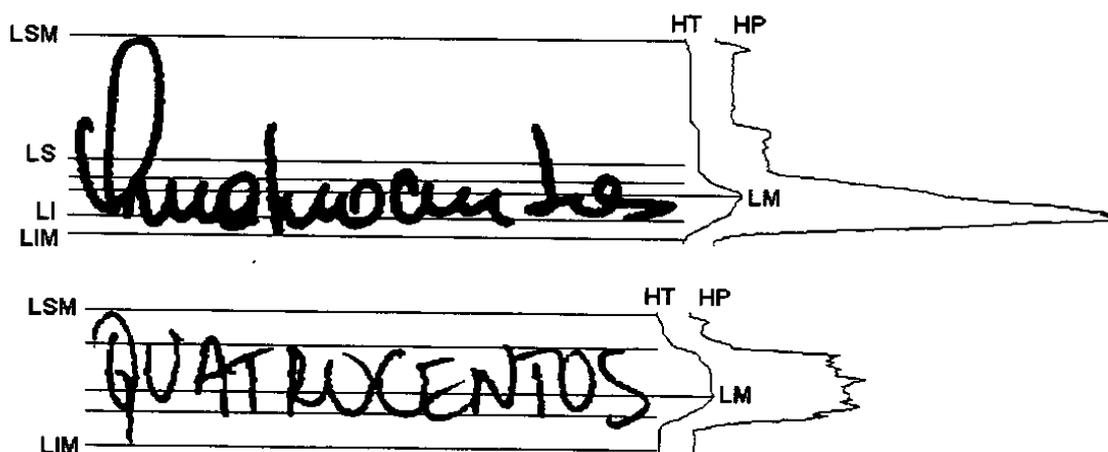


Figura 4.6: Determinação das zonas de palavras "quatrocentos"

O histograma de densidade de pixel não é utilizado para detecção do corpo das palavras uma vez que o mesmo gera distorções quando as palavras possuem uma concentração de pixels em local onde não se encontra o verdadeiro corpo da palavra, por exemplo: barra de corte da letra *t* e traçado irregular na escrita.

A seguir apresenta-se a extração de primitivas globais, compreendendo a seleção, o alfabeto de símbolos e as metodologias de extração de cada uma das primitivas.

4.2. Primitivas Perceptivas

A extração de primitivas caracteriza-se por uma das tarefas mais importantes no sucesso de métodos de reconhecimento de palavras manuscritas. É a partir das primitivas ou características das palavras escolhidas para serem extraídas que se obterá a robustez do método. Busca-se escolher um conjunto de primitivas invariantes, ou seja, que permanecem constantes para um mesmo estilo de escrita. E ainda, cada primitiva desse conjunto deverá ser independente, ou seja, não estar relacionada com as demais. Cada primitiva deve possuir identidade própria e expressar o significado de elementos que são característicos para o reconhecimento.

Existe uma variedade de primitivas que podem ser selecionadas para o processo de extração visando representar as formas a serem reconhecidas, como apresentado no Capítulo 2. O estudo destas primitivas permite compreender a complexidade e riqueza de informações contidas nas palavras manuscritas. Esta riqueza de informações, por sua vez, permite uma redução da complexidade do problema, quando explorada adequadamente.

No Capítulo 2, seção 2.3.5, apresenta-se uma discussão referente a seleção das primitivas a serem extraídas das palavras manuscritas. Leva-se em consideração os aspectos de escrita, leitura e conhecimento prévio, também apresentados no mesmo capítulo. Ressalta-se ainda os trabalhos de [MADHVANATH & GOVINDARAJU, 1998] e [SCHOMAKER & SEGERS, 1998] para o processo de seleção das primitivas globais.

Deve-se ter em mente que o enfoque do presente trabalho trata as palavras como uma unidade para o reconhecimento, ou seja, uma abordagem Global. Para trabalhar

com as palavras manuscritas extraíndo primitivas perceptivas foram selecionados 3 conjuntos de primitivas, a saber:

- Conjunto 01 – PP – Primitivas Perceptivas,
- Conjunto 02 – PPCC – Primitivas Perceptivas, Concavidades e Convexidades,
- Conjunto 03 – PPCCR – Primitivas Perceptivas, Concavidades e Convexidades Rotuladas.

A seguir descreve-se cada um dos conjuntos de primitivas selecionados e os respectivos processo de extração.

4.2.1. Conjunto 01 – Primitivas Perceptivas

Cabe a cada pesquisador buscar um conjunto de primitivas representativo, invariante, independente entre si, de tamanho adequado ao tempo de processamento e adequado ao método de reconhecimento a ser utilizado. Portanto, a seguinte pergunta deve ser respondida: *Quais são as primitivas perceptivas relevantes no processo de reconhecimento de palavras manuscritas?* O trabalho de [MADHVANATH & GOVINDARAJU, 1998] fornece a definição de primitivas perceptivas e o trabalho de [SCHOMAKER & SEGERS, 1998] fornece elementos que auxiliam na escolha das primitivas a serem extraídas.

Deste modo, o primeiro conjunto de primitivas proposto denomina-se Conjunto 01 - PP – primitivas perceptivas, sendo composto por: ascendentes, descendentes e laços fechados nas 3 zonas de uma palavra, conforme Tabela 4.1. Estas primitivas geram informação sobre o tipo de primitiva, o tamanho relativo ao corpo da palavra e a informação posicional (em relação ao eixo x e em relação as zonas da palavra).

Tabela 4.1: Conjunto de Primitivas 01 - PP – Primitivas Perceptivas

Primitivas	Conjunto 01 – PP
Ascendente grande ou pequeno	T, t
Descendente grande ou pequeno	F, f
Laço grande ou pequeno no corpo da palavra	O, o
Laço superior ou inferior	j, l
Vazio	X

Início

```

Pintura_pixels_branco;
Centro_gravidade_laço_fechado (xCG,yCG);
x_laço = xCG;
// A coordenada xCG é utilizada no momento da extração da seqüência de primitivas
y_laço = yCG;
// A coordenada yCG é utilizada para cálculo da altura do laço em relação as zonas
da palavra (ascendente, descendente e corpo)
Cálculo_Coordenadas_Máxima_Mínima_Altura_Laço(xCG, ymáxima, ymínima);
Cálculo_Diâmetro_Laço (xCG,yCG, diâmetro);
Se (diâmetro > 2*espessura média do traço da palavra)
    Então É laço;
Se (É laço)
    Então Se (LS < ymáxima ≤ LSM)
        Então posição é superior;
        Senão Se (LI ≤ ymáxima e ymínima ≤ LS)
            Então posição é corpo da palavra;
            Senão Se (LI < ymínima ≤ LIM)
                Então posição é superior;
Se (posição = corpo)
    Então Tamanho_laço = ymáxima - ymínima;
Se (Tamanho_laço > 50% da altura do corpo)
    Então tamanho é grande;
    Senão tamanho é pequeno;
Fim_Encontra_Laços;

```

Por sua vez, o algoritmo **Encontra_Asc_Desc** está estruturado como a seguir:

```

Início
// Com base nos valores de LS e LI convencionam-se que toda parte da imagem que está
localizada entre LSM e LS é ascendente e toda parte entre LI e LIM é descendente
y = LS;
Para (x = Xesquerda; x ≤ Xdireita; x=x+1) faça //(largura da imagem)
    Encontra_Coordenadas_X_Asc(x, Xascendente);
Fim_para;
y = LS;
Para (x = Xesquerda; x ≤ Xdireita; x=x+1) faça
    Encontra_Coordenadas_X_Desc(x, Xdescendente);
Fim_para;
Cálculo_Altura_Asc (hascendente);
Cálculo_Altura_Desc (hdescendente);
Se (hascendente > 50 % da altura do corpo da palavra)
    Então tamanho_asc é grande;
    Senão tamanho_asc é pequeno;
Se (hdescendente > 50 % da altura do corpo da palavra)
    Então tamanho_desc é grande;
    Senão tamanho_desc é pequeno;
Fim_Encontra_Asc_Desc;

```

A Figura 4.7 exemplifica a extração de primitivas para dois exemplares, da base de dados de treinamento e de reconhecimento, respectivamente. Observa-se que o processo de extração de primitivas para a palavra “duzentos” permite representar as primitivas perceptivas. Porém, o mesmo não ocorre com o exemplo da palavra “quarenta”, percebendo-se, portanto, a influência do estilo da escrita no processo de extração de primitivas.

As primitivas perceptivas são extraídas sobre a imagem das palavras e um procedimento de pseudo-segmentação é aplicado para obter-se a seqüência de observações correspondente, como mostrado na Figura 4.7. A cada transição branco-preto sobre LM realiza-se uma observação e emite-se um conjunto de símbolos correspondentes as primitivas extraídas, caracterizando um grafema, conforme Tabela

4.1. São consideradas somente as transições branco-preto que não se encontram dentro de laços no corpo das palavras. A Figura 4.8 exemplifica um pseudo-segmento retirado da palavra “quarenta” mostrada na Figura 4.7. O caracter # representa um separador entre grafemas.

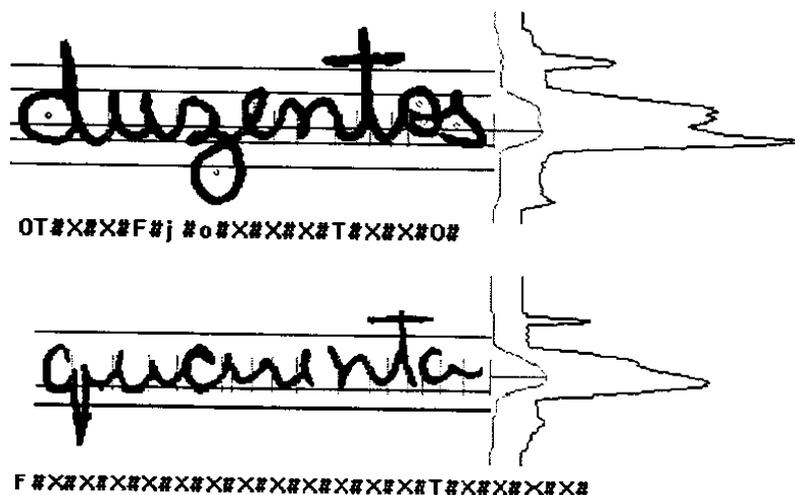


Figura 4.7: Extração de primitivas: conjunto 01 – PP – primitivas perceptivas

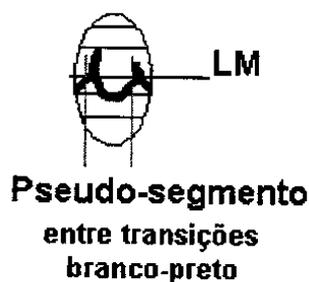


Figura 4.8: Pseudo-segmentação das palavras

Caso, nenhuma primitiva tenha sido extraída no segmento em análise, o símbolo X de vazio é emitido, de acordo com Tabela 4.1. O símbolo emitido independe da seqüência de ocorrência das primitivas no pseudo-segmento, uma vez que, a diferenciação entre as letras "b" , "d" e "p", "q" não é necessária. Pois, as letras "b" e "p" não fazem parte da formação das palavras que compõem o léxico.

4.2.2. Conjunto 02 – PPCC – Primitivas Perceptivas, Concavidades e Convexidades

Ao analisar-se os exemplos da Figura 4.7, palavra “quarenta”, somente os ascendentes e descendentes são extraídos. Percebe-se que a representação da palavra torna-se pobre diante de tantas características existentes e perceptíveis aos olhos humanos.

Ao observar-se também, cuidadosamente, o vocabulário do extenso em cheques bancários brasileiros verifica-se a presença de um conjunto de palavras que não possui naturalmente primitivas perceptivas – ascendentes e descendentes, tais como: “um”, “cinco”, “seis”, “nove”, “cem”, “reais”.

Outro conjunto de palavras caracteriza-se por ser uma palavra curta, por exemplo: “dois”, “três”, “sete”, “oito”, “dez”, “onze”, “doze”, “mil” e “real”. Estas palavras possuem um grau de confusão natural, dependendo do estilo do escritor.

Deve-se ressaltar ainda, um terceiro conjunto de palavras que inclui as palavras manuscritas em letra tipo bastão ou de forma [TAPPERT et al., 1990], as quais não possuem as regiões de ascendentes e descendentes.

Assim a segunda pergunta deve ser respondida: *Como representar palavras cursivas sem a presença de primitivas perceptivas ?*

[CÔTE, 1997] trata as palavras manuscritas como um conjunto de regularidades e singularidades. Classifica as singularidades como os ascendentes e os descendentes. E explora as concavidades e convexidades existentes no corpo das palavras para representação das regularidades, como mostrado na Figura 2.31 do Capítulo 2.

Assim, estuda-se a complementação do conjunto formado pelas primitivas perceptivas. Esta complementação tem por base a representação das concavidades e convexidades existentes no corpo das palavras [LECOLINET, 1990], [CHO et al., 1995], [LU & SHRIDHAR, 1996] e [CÔTE, 1997]. O segundo conjunto de primitivas é então denominado Conjunto 02 – PPCC - primitivas perceptivas, concavidades e convexidades, conforme Tabela 4.2.

As concavidades e convexidades são extraídas somente no corpo das palavras sobre o esqueleto das imagens e nos segmentos onde não ocorrem primitivas perceptivas. Desta forma, as concavidades e convexidades são primitivas complementares, auxiliando na representação das formas que não se classificam como

ascendentes, descendentes e laços fechados. Estas primitivas auxiliam na representação das curvaturas das letras e ligações entre letras, ou ainda, de laços abertos existentes no corpo das palavras.

Tabela 4.2: Conjunto de Primitivas 02 - PPCC - Primitivas Perceptivas, Concavidades e Convexidades

Primitivas	Conjunto 02 - PPCC
Ascendente grande ou pequeno	T, t
Descendente grande ou pequeno	F, f
Laço grande ou pequeno no corpo da palavra	O, o
Laço superior ou inferior	j, l
Côncavo	(
Convexo)
Vazio	X

Os pontos convexos são determinados com o auxílio de uma família de 5 elementos estruturantes (matriz 7x3) e os pontos côncavos com uma família de 10 elementos estruturantes (matriz 9x4). Ambos os procedimentos de extração são aplicados sobre o esqueleto da imagem original pelo processo de *template matching*.

A definição dos elementos estruturantes teve por base a observação dos resultados de extração dos pontos côncavos e convexos para 1.742 imagens das palavras de "um" até "dez" da base de treinamento.

A Figura 4.9 mostra um exemplo da extração de primitivas considerando-se o Conjunto 02 – PPCC para as mesmas palavras da Figura 4.7. Observa-se que este conjunto de primitivas permite representar com mais informações as palavras que não possuem laços fechados, além do fato de caracterizar-se como adequado à extração de ligações e curvaturas naturais existentes em palavras manuscrita.

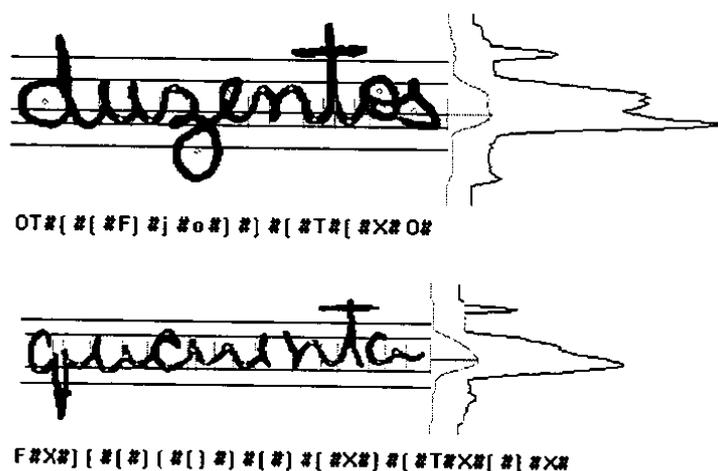


Figura 4.9: Extração de primitivas: conjunto 02 – PPCC

4.2.3. Conjunto 03 – PPCCR – Primitivas Perceptivas, Concavidades e Convexidades Rotuladas

Observa-se na base de treinamento que a proporção entre os grafemas extraídos como vazios, símbolo X, para o Conjunto de primitivas PP em relação aos grafemas com concavidades e/ou convexidades para o Conjunto PPCC é da ordem de 67.3%, fato que justifica a existência do Conjunto PPCC.

Por outro lado, a proporção dos grafemas com concavidades e/ou convexidades para o Conjunto PPCC em relação ao total de grafemas extraídos é da ordem de 50,5%. Portanto, percebe-se que a representação das concavidades e convexidades necessita de uma classificação mais detalhada para extração dos grafemas.

Deste modo, selecionou-se um terceiro conjunto de primitivas capaz de diferenciar grafemas extraídos de pseudo-segmentos compostos pelas letras “C”, “S”, “E” e “Z”, e ainda, “u”, “n”, “r” e “v”.

Este conjunto denominado de Conjunto 03 – PPCCR - primitivas perceptivas acrescido de concavidades e convexidades rotuladas, tem por base tratar as *Convex Deficiencias* através de um procedimento de rotulação de pixels proposto por [PARKER, 1997], de acordo com a Tabela 4.3.

Tabela 4.3: Conjunto 03 – PPCCR - Primitivas Perceptivas, Concavidades e Convexidades Rotuladas

Primitivas	Conjunto 03 - PPCCR
Ascendente grande ou pequeno	T, t
Descendente grande ou pequeno	F, f
Laço grande ou pequeno no corpo da palavra	O, o
Laço superior ou inferior	j, l
Côncavo e Convexo (fechado)	a
Convexo (aberto embaixo)	n
Convexo (aberto a esquerda)	Z
Convexo (aberto a direita)	C
Côncavo (aberto em cima)	u
Ligação inferior	i
Ligação superior	r
C +)	S
C + (E
Vazio	X

A rotulação é realizada a partir da análise de 4 direções (norte, sul, leste e oeste) para cada um dos pixels do fundo ou do *background* da área correspondente a parte

interna da concavidade e/ou convexidade extraída da imagem. Para [PARKER, 1997] existem 6 possibilidades de rotulação dos pixels, como descrito a seguir: a) Rótulo 0: todas as direções são fechadas, Rótulos 1 a 4: aberto em sul, aberto em oeste, aberto em norte, aberto em leste e Rótulo 5: quando o pixel não pode ser rotulado com nenhum dos casos anteriores.

Após a rotulação de todos os pixels da área em questão, os rótulos são contados e classifica-se a área com as seguintes opções: laço aberto, convexo do tipo “n”, côncavo do tipo “u”, côncavo do tipo “(“ ou do tipo “)”, convexo do tipo “C” ou do tipo “Z”, convexo do tipo ligação superior “r” e côncavo do tipo ligação inferior “i”. A Figura 4.10 exemplifica a extração deste conjunto de primitivas – PPCCR – para as mesmas palavras das Figuras 4.7 e 4.9. Observa-se que as palavras passam por um processo de enriquecimento de detalhes e representações ao comparar-se as Figuras 4.7, 4.9 e 4.10.

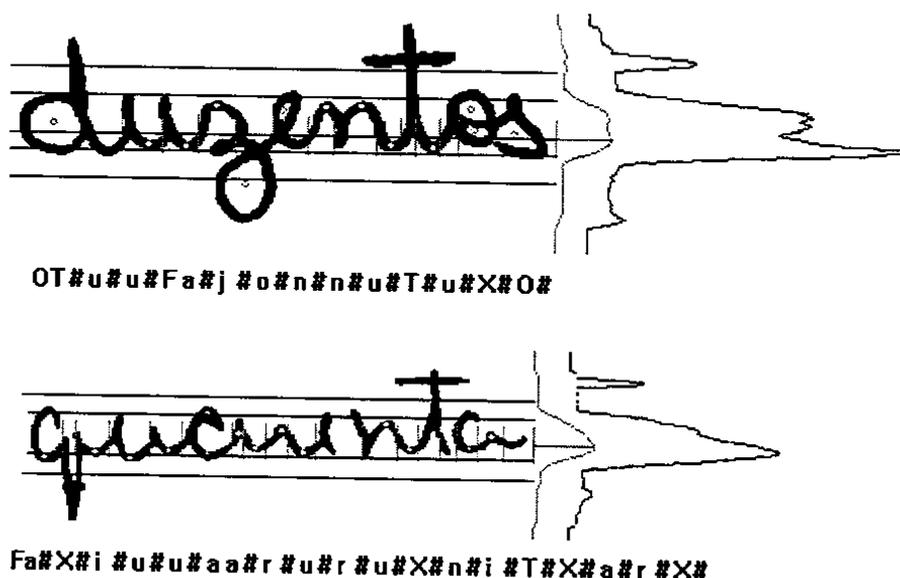


Figura 4.10: Extração de primitivas: conjunto 03 – PPCCR – primitivas perceptivas, concavidades e convexidades rotuladas

Assim, trabalha-se com a evolução de 3 conjuntos de primitivas globais buscando-se melhor representar as diferenças entre as palavras e as formas de escrita de uma mesma palavra. Ressalta-se que o presente trabalho não faz restrições ao tipo da escrita encontrada na base de dados utilizada.

Neste ponto, sabendo-se que para cada pseudo-segmento tem-se a extração de um grafema. Este grafema por sua vez representa a extração de nenhuma, uma ou mais

primitivas. Pergunta-se, então: *Que associações entre as primitivas extraídas são relevantes?*

Diferentes conjuntos de primitivas e enfoques de validação podem ser encontrados na literatura. Estudos como os realizados por [BATTITI, 1994], [AVILA, 1996] e [GRANDIDIER et al., 2000] mostram a preocupação com a representatividade dos grafemas extraídos.

A seguir apresenta-se o estudo dos grafemas e a escolha dos alfabetos de símbolos que atendem as necessidades de representação das formas manuscritas no contexto de cheques bancários brasileiros.

4.2.4. Escolha de Alfabetos de Grafemas

A escolha de um alfabeto de grafemas necessita de dois passos importantes: o conhecimento de métodos de seleção dos grafemas e, principalmente, de uma definição de grafema. Esta definição é variável entre diferentes autores, pois depende diretamente dos métodos utilizados para descrever a palavra.

No presente trabalho, a noção de grafemas está diretamente ligada ao modo através do qual se realiza a extração de primitivas. O grafema corresponde ao traçado das entidades gráficas próximas da letra. Os pontos de pseudo-segmentação sobre a linha média (LM) das palavras, como apresentado anteriormente, definem a formação dos grafemas. Este enfoque assemelha-se ao dos pontos de ancoragem de [AVILA, 1996].

Define-se como conjunto de grafemas "bruto" o conjunto de todas as combinações de primitivas extraídas para os pseudo-segmentos e observadas sobre a base de dados de treinamento. Deve-se codificar o conjunto de grafemas "bruto" a fim de reduzir-se o número de símbolos possíveis para representar as palavras. Pois, para processo Markovianos de treinamento é interessante que o alfabeto de grafemas seja composto por um número limitado de símbolos representativos do problema. A operação de redução do conjunto de grafemas "bruto" permite estabelecer um alfabeto a ser codificado em símbolos, a partir da seqüência de observações extraídas das palavras, visando o processo de treinamento dos modelos através de MEM.

Apresenta-se a seguir o que é um grafema no contexto do presente trabalho e os métodos de estudo e definição dos alfabetos utilizados no trabalho. Os métodos aplicados são a Entropia de Shannon que permite uma medida da informação que cada

grafema carrega consigo e critérios de entropia para estudar as possíveis concatenações entre os grafemas.

4.2.4.1. O que é um Grafema?

Considera-se como um grafema uma entidade que pode corresponder a uma parte de uma letra, uma letra ou de letras conectadas. Anteriormente, apresentou-se que as primitivas são extraídas das palavras considerando-se as transições branco-preto sobre a linha média (LM), sendo que estas transições não podem estar localizadas dentro de laços fechados no corpo das palavras. Desde modo, uma palavra contém um número de grafemas correspondente ao número de transições branco-preto válidas sobre LM detectadas na palavra.

Supondo que os grafemas podem ser representados por um vetor de n elementos binários, onde cada elemento representa a ausência (0) ou a presença (1) de cada primitiva que compõe o Conjunto adotado, ter-se-ia conjuntos distintos de grafemas com as seguintes dimensões:

- Conjunto 01 – PP: $2^9 = 512$,
- Conjunto 02 – PPCC: $2^{11} = 2.048$,
- Conjunto 03 – PPCCR: $2^{18} = 262.144$.

Na realidade, foram observados na base de treinamento as seguintes configurações de grafemas, para um total de 58.432, extraídos de 7.146 amostras:

- Conjunto 01 – PP = 70 grafemas distintos,
- Conjunto 02 - PPCC = 74 grafemas distintos,
- Conjunto 03 - PPCCR = 94 grafemas distintos.

Diante deste fato, precisa-se melhorar a representação das palavras diante da grande variabilidade da escrita manuscrita, reduzindo-se assim o tamanho dos alfabetos de grafemas. Esta melhoria passa, então, pela análise dos conjuntos de grafemas “bruto” extraídos das palavras. Portanto, necessita-se de um processo que consiste, principalmente, em suprimir todas as configurações com baixa frequência de aparição (por exemplo, configurações que ocorrem somente uma vez na base de treinamento).

Suprimir grafemas não significa eliminá-los, mas concatená-los com outros. Pois, o problema que se deseja evitar quando se trabalha com MEM pode ser respondido através da pergunta: *Como treinar modelos sobre grafemas que pouco são encontrados nas seqüências de observações?*

A concatenação de grafemas do conjunto "bruto" com baixa frequência de aparição pode ser realizada ao definir-se um grafema que represente dois ou mais do conjunto "bruto" e supor-se que este novo grafema mais freqüente engloba os de baixa frequência. Portanto, esta concatenação não deve reduzir a qualidade dos alfabetos definidos.

4.2.4.2. Escolha de uma Medida de Semelhança

A definição dos alfabetos parte do princípio que necessita-se de métodos que permitam estabelecer uma medida de semelhança entre as amostras, para assim poder concatenar os grafemas pouco frequentes e definir os alfabetos.

Quatro métodos podem ser utilizados na busca da semelhança entre os grafemas e definição de alfabetos, a saber:

- Distância de Hamming: permitir calcular as similaridades entre componentes presentes e ausentes;
- Distância de Hamming Ponderada: permite calcular a importância de cada uma das características dos grafemas representados por vetores de n elementos binários. Possui a desvantagem da atribuição empírica de pesos às diferentes primitivas consideradas;
- Alfabeto Hierárquico: introduz explicitamente uma hierarquia entre os grafemas de acordo com a frequência das primitivas;
- Entropia de Shannon: este método estabelece uma medida da quantidade de informação aportada por cada um dos grafemas.

No presente trabalho os métodos Hierárquico e da Entropia de Shannon são adotados para estudo dos alfabetos a serem definidos. Esta escolha está ligada diretamente ao entendimento da frequência de ocorrência dos grafemas e da quantidade de informação que cada um carrega consigo em relação aos demais.

A seguir apresenta-se a formulação da Entropia de Shannon e da Informação Mútua decorrente da aplicação no contexto dos grafemas estudados.

4.2.4.3. Entropia de Shannon e Informação Mútua

Seja uma variável aleatória discreta X , com $\{X=x_i\}$, $i \in \{0, \dots, N\}$ conjunto dos eventos de probabilidade de ocorrência P_i . A entropia de Shannon de X é uma medida

de informação que é definida como sendo [COVER & THOMAS, 1991] e [HUANG et al., 1990]:

$$H(P) = -\sum_{i=1}^N P_i \log_2(P_i) \quad (4.2)$$

Esta relação foi estabelecida por C. Shannon em 1948 e $H(P)$ é, então, denominada de Entropia de Shannon. Esta medida de informação possui várias propriedades naturais, como as apresentadas a seguir:

- A entropia é máxima logo os eventos são equiprováveis;
- A entropia é mínima logo um dos eventos está correto. O sistema é perfeitamente conhecido e algum aporte de informação é possível;
- A entropia é uma função positiva, contínua e simétrica de seus argumentos;
- Para N eventos equiprováveis, a entropia é diretamente proporcional a N ;
- A informação média adquirida entre duas etapas se ajusta por:

$$H(P_1 \dots P_N) = H(Q_1, Q_2) + Q_1 H\left(\frac{P_1}{Q_1}, \dots, \frac{P_M}{Q_1}\right) + Q_2 H\left(\frac{P_{M+1}}{Q_2}, \dots, \frac{P_N}{Q_2}\right) \quad (4.3)$$

onde,

$$Q_1 = \sum_{i=1}^M P_i \quad \text{e} \quad Q_2 = \sum_{i=M+1}^N P_i \quad (4.4)$$

são as probabilidades de dois grupos de eventos $1 \dots M$ e $M+1 \dots N$. A informação média do evento total é, portanto, a soma das informações médias associadas aos dois eventos da divisão ponderada pelas suas probabilidades de aparição e da informação de divisão.

O léxico para a aplicação dos extensos manuscritos de cheques bancários brasileiros é composto por 39 palavras diferentes. O valor de $H(P)$ para o presente estudo foi calculado sobre a frequência relativa das palavras base da treinamento. Portanto, a entropia $H(P)$ é:

$$H(P) = -P_0 \log_2(P_0) - \dots - P_{39} \log_2(P_{39}) = 4,77 \text{ bits} \quad (4.5)$$

sendo a unidade de $H(P)$ o bit. O bit é o valor de incerteza associado à seleção de um evento entre dois eventos equiprováveis (probabilidade = 0,5). O valor de entropia quantifica a variabilidade do problema. Este valor é a quantidade de informação mínima necessária a fim de poder resolver o problema de classificação de palavras. O valor da

entropia $H(P)$ permite comparar o problema do presente trabalho com um problema de 27,28 classes equiprováveis, pois $\log_2(27,28) = 4,77$.

Para definir a noção de Informação Mútua, deve-se introduzir duas variáveis aleatórias X e Y [COVER & THOMAS, 1991]. Nota-se $P(x_i)$ a probabilidade $P(X=x_i)$, $P(y_i)$ a probabilidade $P(Y=y_i)$, $P(x_i|y_i)$ e $P(y_i|x_i)$ as probabilidades condicionais $P(X=x_i|Y=y_i)$ e $P(Y=y_i|X=x_i)$.

A entropia da lei condicional é aquela que fornece a informação média aportada por X logo $Y = y_i$ é dado por:

$$H(X | Y = y_i) = - \sum_{i=1}^N P(x_i | y_i) \log_2(P(x_i | y_i)) \quad (4.6)$$

A entropia condicional se deduz como a média obtida sobre Y :

$$H(X | Y) = - \sum_{j=1}^M P(y_j) H(X | Y = y_j) \quad \text{ou} \quad (4.7)$$

$$H(X | Y) = \sum_{i=1}^N \sum_{j=1}^M P(y_j) P(x_i | y_i) \log_2(P(x_i | y_i))$$

A entropia condicional quantifica a variabilidade do problema X quando Y é conhecido. A diferença entre $H(P)$ e $H(X|Y)$ define a quantidade de informação aportada por X sobre Y ou inversamente, pois entropia é simétrica. Esta informação é, então, denominada Informação Mútua entre X e Y e é definida por:

$$I_M(X, Y) = H(X) - H(X | Y) \quad \text{ou} \quad (4.8)$$

$$I_M(X, Y) = H(Y) - H(Y | X)$$

que pode-se escrever sob a forma a seguir:

$$I_M(X, Y) = \sum_{i=1}^N \sum_{j=1}^M P(x_i, y_j) \log_2 \left(\frac{P(x_i, y_j)}{P(x_i)P(y_j)} \right) \quad (4.9)$$

A Equação (4.9) permite avaliar a quantidade de informação aportada individualmente por cada um dos grafemas.

O presente trabalho utiliza a Informação Mútua como uma medida para validar a redução do alfabeto de grafemas. Em um primeiro momento, utiliza-se para obter a informação mútua aportada por cada grafema "bruto" individualmente, identificando-se

os grafemas de maior representatividade das palavras permitindo uma análise global dos conjuntos de grafemas ditos “bruto”. Em um segundo momento, utiliza-se para realizar uma análise local, permitindo a validação dos grafemas concatenados através do método Hierárquico.

A aplicação da Informação Mútua para os grafemas toma como variáveis a palavras M e os grafemas G_k , sendo dada por [AVILA, 1996]:

$$I_M(M, G_K) = \sum_{i=1}^N \sum_{j=1}^A P(M = M_i, G_K = j) \log_2 \left(\frac{P(M = M_i, G_K = j)}{P(M = M_i)P(G_K = j)} \right) \quad (4.10)$$

onde, $j \in \{0, 1, 2, \dots, A\}$ representando o número de aparições do grafema k numa amostra de seqüências de observações da palavra M e $i \in \{0, 1, 2, \dots, N\}$ representando o número de palavras do léxico estudado.

4.2.4.4. Alfabetos Hierárquicos

Os alfabetos hierárquicos são assim denominados pois introduzem explicitamente uma hierarquia entre os grafemas segundo a freqüência das primitivas extraídas. Uma vez que, estes sejam organizados de maneira hierárquica permitem a conversão destes em símbolos, visando a obtenção das seqüências de observações para entrada nos processos de treinamento e reconhecimento utilizando-se MEM.

Para formação de um alfabeto hierárquico deve-se considerar o conhecimento prévio da morfologia da forma a ser reconhecida, isto é, a palavra [AVILA, 1996]. Portanto, os conjuntos de primitivas devem auxiliar na descrição dos traçados das letras.

Desta forma, procura-se estabelecer símbolos que descrevem as letras do alfabeto possíveis de serem encontradas no léxico de palavras do extenso de cheques bancários brasileiros. Deve-se ressaltar que as seguintes letras não fazem parte da formação das palavras estudadas: b, f, g, p, x, k, w, y.

Assim, os alfabetos hierárquicos foram estabelecidos levando-se em consideração a freqüência de ocorrência dos grafemas “brutos” na base de dados de treinamento e as combinações mais apropriadas no que se refere a relação grafemas extraídos com as letras a serem representadas. Efetuaram-se, portanto, concatenações de dois ou mais grafemas dos conjuntos “brutos”.

A Tabela 4.4 apresenta o alfabeto hierárquico para o Conjunto 01 de primitivas - PP. Este alfabeto é composto por 18 símbolos que descrevem os grafemas “brutos” mais frequentes na base de dados de treinamento. [DOLFING, 1998] lembra que não é possível treinar um MEM com menos de 30 exemplares.

Tabela 4.4: Alfabeto Hierárquico – PP – Primitivas Perceptivas

Conversão Grafema-Símbolo	Descrição ¹
Se X então 01	Ausência
Senão o então 02	Laço peq no corpo
Senão O então 03	Laço grd no corpo
Senão T então 04	Asc grd
Senão t então 05	Asc peq
Senão F então 06	Desc grd
Senão f então 07	Desc peq
Senão j então 08	Desc com laço inf
Senão l então 09	Asc com laço sup
Senão (OT ou oT) então 10	Laço peq ou grd no corpo + Asc grd
Senão (Ot ou ot) então 11	Laço peq ou grd no corpo + Asc peq
Senão (Ol ou ol) então 12	Laço peq ou grd no corpo + Asc com laço sup
Senão (OF ou oF) então 13	Laço peq ou grd no corpo + Desc grd
Senão (Of ou of) então 14	Laço peq ou grd no corpo + Desc peq
Senão (Oj ou oj) então 15	Laço peq ou grd no corpo + Desc com laço inf
Senão (TF ou Tf ou tF) então 16	Asc grd + Desc grd ou Asc grd + Desc peq ou Asc peq + Desc grd
Senão tf então 16	Asc peq + Desc peq
Senão (Fl ou fl) então 17	Desc peq ou grd + Asc com laço sup
Senão (Tj ou tj) então 18	Asc peq ou grd + Desc com laço inf
Senão então 18	Todas as demais opções

A Tabela 4.5 apresenta o alfabeto hierárquico para o Conjunto 02 de Primitivas – PPCC. Este alfabeto é formado por 22 símbolos distintos, incluindo-se os 18 símbolos do conjunto PP acrescido de 4 símbolos para representação das concavidades e convexidades no corpo das palavras.

Tabela 4.5: Alfabeto Hierárquico PPCC - Primitivas Perceptivas + Côncavo e Convexo

Conversão Grafema-Símbolo	Descrição
Senão (então 19	Côncavo no corpo (parte inferior)
Senão) então 20	Convexo no corpo (parte superior)
Senão (e) então 21	Côncavo e Convexo no corpo
Senão) e (então 22	Convexo e Côncavo no corpo
Senão então 18	Todas as demais opções

¹ asc (ascendente), desc (descendente), peq (pequeno), grd (grande), inf (inferior), sup (superior).

A Tabela 4.6 mostra o alfabeto hierárquico para o Conjunto 03 de Primitivas – PPCCR. Este alfabeto é formado por 29 símbolos diferentes, sendo os 20 primeiros símbolos provenientes do Conjunto de Primitivas PPCC, acrescido de 9 símbolos para representação das concavidades e convexidades rotuladas.

Tabela 4.6: Alfabeto Hierárquico PPCCR - Primitivas Perceptivas + Concavidades e Convexidades Rotuladas

Conversão Grafema-Símbolo	Descrição
Senão a então 21	Côncavo e Convexo (fechado)
Senão n então 22	Convexo (aberto embaixo)
Senão Z então 23	Convexo (aberto a esquerda)
Senão C então 24	Convexo (aberto a direita)
Senão u então 25	Côncavo (aberto em cima)
Senão i então 26	Ligação inf
Senão r então 27	Ligação sup
Senão S então 28	C +)
Senão E então 29	C + (

A composição destes alfabetos busca um equilíbrio na distribuição dos símbolos, oriundos dos conjuntos de grafemas “brutos”. Os alfabetos definidos proporcionam uma melhor distribuição da informação vinculada a estes símbolos, melhorando assim, o processo de treinamento, uma vez que reduziu-se também o número de símbolos de 70, 74 e 94 para 18, 22 e 29, respectivamente para os conjuntos PP, PPCC e PPCCR.

Fica a seguinte pergunta: *Como validar os alfabetos hierárquicos e garantir uma melhor distribuição da informação aportada pelos grafemas?* A seguir, apresenta-se o método de validação dos alfabetos hierárquicos através da Informação Mútua.

4.2.4.5. Validação dos Alfabetos Hierárquicos através da Informação Mútua

A validação dos alfabetos tem por base uma avaliação direta da quantidade de informação aportada pelos grafemas dos conjuntos “brutos” e dos concatenados pelo método hierárquico.

[GRANDIDIER et al., 2000] utiliza como indicador para a validação do conjunto de primitivas selecionado a perplexidade condicional baseada na noção de entropia. Outro exemplo, é o trabalho de [BATTITI, 1994] onde a Informação Mútua é aplicada para avaliar a informação aportada pelas primitivas e selecionar um sub-conjunto de primitivas representativas a ser utilizado como entrada de dados para um

classificador com Redes Neurais. [AVILA, 1996] aplica a Informação Mútua para melhorar o conjunto de primitivas utilizando um critério de concatenação de grafemas a partir de um conjunto “bruto”, visando o reconhecimento de extensos manuscritos de cheques franceses.

O método hierárquico permite que as concatenações entre grafemas corresponda a concatenação de mais de dois elementos. Assim, as concatenações podem ser validadas utilizando-se o seguinte critério descrito por [AVILA, 1996]:

$$\frac{I_M(M, G'')}{\max(I_M(M, G_1), I_M(M, G_2))} > \alpha \quad (4.11)$$

De outro modo, a relação entre a informação mútua do grafema resultante da concatenação $I_M(M, G')$ e a maior informação mútua dos grafemas isolados deve ser superior ao limiar fixo α . O valor desta relação é superior a 1 se a concatenação vincula mais informação que o melhor dos grafemas isoladamente. Este valor é inferior e próximo de 1 se a concatenação vincula pouca informação.

Realiza-se a validação dos alfabetos hierárquicos aplicando-se a Equação (4.11) sobre os conjuntos dos “brutos” e codificados pelos símbolos das Tabelas 4.4 a 4.6, extraídos da base de treinamento.

A validação das concatenações para o conjunto de primitivas PPCCR está mostrada na Tabela 4.7. Pode-se tomar como exemplo o grafema hierárquico “*Or*” (presença de um laço no corpo da palavra + um ascendente pequeno) corresponde a concatenação de quatro outros: “*tO*”, “*Ot*”, “*ot*” e “*to*”. O resultado “*Or*” não faz diferença se o ascendente encontra-se antes ou depois do laço no corpo da palavra. Fato este, que vem de encontro a não necessidade de diferenciação das letras *d* e *b*, uma vez que a letra *b* não faz parte do alfabeto das palavras do léxico [FREITAS et al., 2001].

Os resultados mostram que os alfabetos hierárquicos são válidos, pois todas as concatenações realizadas apresentam $\alpha > 1$. Isto, significa uma melhoria na distribuição da informação aportada pelos grafemas com uma redução do tamanho dos alfabetos de símbolos.

A Informação Mútua mostra-se, portanto, um método rápido para uma inspeção direta sobre os grafemas dos conjuntos “brutos” e as concatenações realizadas, seja hierarquicamente ou por outro método. Permitindo, se necessário, que outros

procedimentos de concatenação possam ser aplicados, tais como os trabalhos de [BATTITI, 1994], [AVILA, 1996] e [GRANDIDIER et al., 2000].

Tabela 4.7: Validação das Concatenações para o Conjunto de Primitivas PPCCR

Grafemas "brutos"	max $I_M(M, G_k)$	Grafema Hierárquico	$I_M(M, G_k)$	Valor $\alpha > 1$
X	0,0096422	X	9,237108E-02	1,04385485
i	0,1575270	i	1,619584E+00	10,28131051
u	0,3182913	u	3,275907E+00	10,29216633
r	0,4338941	r	4,462884E+00	10,28565265
o	0,4754084	o	4,847420E+00	10,19632804
a,au,a),a,aC	0,4829099	a	4,932612E+00	10,21435262
O	0,4846769	O	4,971346E+00	10,25703102
n,na,na	0,4583982	n	4,715003E+00	10,28582355
T	0,4100424	T	4,211878E+00	10,27181092
t	0,4089634	t	4,203431E+00	10,27825717
l	0,4107125	l	4,219515E+00	10,27364641
)	0,4084629)	4,195261E+00	10,27084957
Z	0,4095092	Z	4,196184E+00	10,24686136
OT,TO,oT,To	0,4094520	OT	4,203069E+00	10,26510800
(0,4086669	(4,198055E+00	10,27255939
F	0,4097471	F	4,200450E+00	10,25132332
OF,FO,oF,Fo	0,4101486	OF	4,207193E+00	10,25772854
j	0,4088597	j	4,194596E+00	10,25925519
Ot,tO,ot,to	0,4102425	Ot	4,198101E+00	10,23321816
C	0,4095130	C	4,201715E+00	10,26027257
f	0,4090797	f	4,202851E+00	10,27391728
Ei,iE,Ea,aE,Er,rE,E	0,4104830	E	4,204452E+00	10,24269458
Ol,lO,ol,lo	0,4102881	Ol	4,207276E+00	10,25444316
jO,Oj,oj,jo	0,4104498	Oj	4,206434E+00	10,24835193
Ft,tF,ft,tf,FT,Tf,TF,FT	0,4104620	TF	4,209219E+00	10,25483236
of,fo,Of,fO	0,4102834	Of	4,209685E+00	10,26043218
lf,fl,Fl,Fl	0,4103748	Fl	4,210608E+00	10,26039611
jt,tj,lj,jl,Tj,jT e grafemas com 3 primitivas*	0,4105242	Tj	4,213850E+00	10,26455931
S	0,4103597	S	4,281129E+00	10,43262533

* {TOF,fot,Fol,fto,loF,FtO,fOl,tFO,ljO,tOf,fol,ofT,Tfo,OfT,jot,FTO,flo,flo,tOj,tOF}

A seguir apresentam-se os experimentos de reconhecimento de palavras manuscritas dos extenso de cheques bancários brasileiros utilizando-se as seqüências de observações codificadas pelos alfabetos de símbolos definidos através de Modelos Escondidos de Markov - MEM.

4.2.5. Reconhecimento de Palavras Manuscritas do Extenso de Cheques Bancários Brasileiros

No Capítulo 2, fez-se uma revisão das atividades de escrita e leitura, além de uma associação destes processos com o conhecimento prévio. Foi apresentada também uma revisão dos aspectos de leitura e reconhecimento. Portanto, neste capítulo, foram

apresentadas as etapas que antecipam o reconhecimento de palavras manuscritas. É como se fosse necessário preparar as palavras para poder aprender sobre as mesmas e depois poder reconhecê-las.

Agora, pode-se apresentar a metodologia proposta para o reconhecimento das palavras isoladas do extenso manuscrito em cheques bancários brasileiros, bem como, detalhar os experimentos com MEM, suas dificuldades e soluções encontradas. Inicialmente, apresenta-se a fundamentação teórica dos MEM e os algoritmos utilizados nas etapas de treinamento, validação e reconhecimento, respectivamente.

A intenção é descrever as etapas de treinamento e validação dos 39 modelos estabelecidos para as palavras com uso de primitivas perceptivas, considerando-se um modelo para cada palavra do léxico. E, posteriormente, apresentar os resultados da etapa de reconhecimento obtidos com os 3 conjuntos de primitivas selecionados: PP, PPCC e PPCCR.

4.2.5.1. Fundamentação Teórica dos Modelos Escondidos de Markov

Os MEM resultam da associação de dois mecanismos ligados entre si: uma cadeia de Markov subjacente tendo um número finito de estados e um conjunto de funções de densidade de probabilidade de emissão de observações onde cada uma está associada com um estado. A cadeia de Markov serve de suporte a este conjunto de funções associadas a um estado.

Para a cadeia de Markov supõe-se que esta esteja em um estado qualquer e uma observação é emitida pela função de densidade de probabilidade condicionada pelo estado corrente. A cadeia de Markov troca então, de estado conforme sua matriz de probabilidade de transição. O observador vê somente os símbolos da observação e não pode diretamente observar os estados da cadeia de Markov, de onde o termo Modelos Escondidos de Markov – MEM, em *inglês Hidden Markov Model* – HMM.

A fundamentação teórica dos MEM está apresentada a seguir pelos seguintes tópicos:

- a) Definição de MEM,
- b) Tipos de Estruturas e
- c) Problemas Básicos.

a) Definição de MEM

Para processos discretos, [HUANG et al., 1990] e [RABINER & JUANG, 1993] definem um MEM denotado por $\lambda = (Q, V, A, B, \pi, T)$ pelos seguintes parâmetros:

T : comprimento da sequência de observações. Este comprimento pode ser variável,

$Q = \{q_1, q_2, \dots, q_N\}$: o conjunto de estados do modelo. Q_t designará o estado da cadeia no instante t ,

N = número de estado do modelo,

$V = \{v_1, v_2, \dots, v_m\}$: conjunto de símbolos de observações possíveis. O_t designará o símbolo da observação ao instante t ,

M = número de símbolos da observação,

$A = \{a_{ij} \mid 1 \leq i, j \leq N\}$: matriz de transição entre estados. O elemento a_{ij} corresponde a probabilidade que a cadeia de Markov transite do estado q_i para o estado q_j :

$$a_{ij} = pr(Q_{t+1} = q_j / Q_t = q_i) \quad \forall i, j \in [1 \dots N], \forall t \in [1, T-1] \quad (4.1)$$

com as seguintes restrições:

$$a_{ij} \geq 0 \quad \forall i, j \quad e \quad \sum_{j=1}^N a_{ij} = 1 \quad (4.2)$$

$B = \{b_{jk} \mid 1 \leq j \leq N; 1 \leq k \leq M\}$: matriz de probabilidade de observação de símbolos condicionada pelos estados da cadeia. O elemento b_{jk} representa a probabilidade de observar o símbolo V_k quando o modelo se encontra no estado q_j :

$$b_j = pr(O_t = v_k / Q_t = q_j) \quad 1 \leq j \leq N \quad 1 \leq k \leq M \quad (4.3)$$

com as seguintes restrições:

$$b_{jk} \geq 0 \quad \forall j, k \quad e \quad \sum_{k=1}^M b_{jk} = 1 \quad \forall j \quad (4.4)$$

$\pi = \{\pi_i \mid 1 \leq i \leq N\}$: conjunto de densidades de probabilidade inicial, π_i representa a probabilidade que o processo de Markov inicie no estado q_i :

$$\pi_i = pr(Q_1 = q_i) \quad 1 \leq i \leq N \quad (4.5)$$

$$\pi_i \geq 0 \quad \forall i \quad e \quad \sum_{j=1}^N \pi_i = 1 \quad (4.6)$$

A avaliação desses processos de Markov é inteiramente determinada pela probabilidade inicial e pela probabilidade das transições sucessivas. O modelo de Markov assim definido é dito de primeira ordem.

b) Tipos de Estruturas de MEM

A determinação do MEM ótimo é uma tarefa que depende estritamente da estrutura do modelo e do número de estados escolhidos. De um modo geral, de acordo com [HUANG et al., 1990] e [RABINER & JUANG, 1993], existem 3 tipos de estruturas para os MEM:

- os modelos sem restrição, ditos ergóticos,
- os modelos seqüenciais e
- os modelos paralelos.

Os exemplos típicos desses modelo estão representados na Figura 4.11. Nos métodos sem restrição, Figura 4.11a, todas as transições possíveis entre os estado da cadeia são autorizadas. Isto é possível, se não se fizer restrição a nenhum dos valores a_{ij} igual nulo.

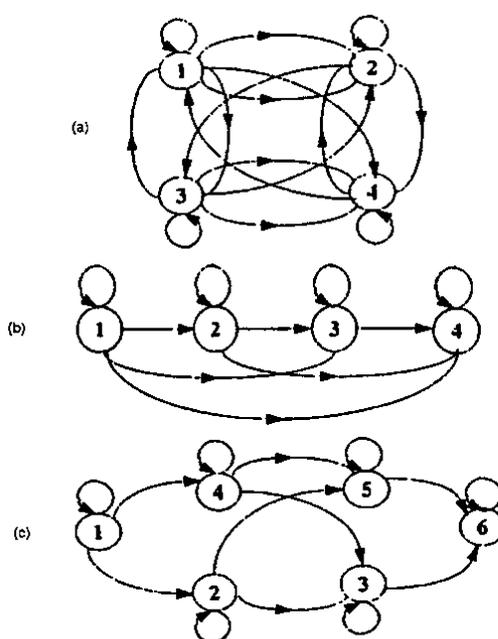


Figura 4.11: Tipos de estrutura dos modelos de Markov: a) modelo sem restrições, b) modelo seqüencial, c) modelo paralelo [EL YACOUBI, 1996]

Os modelos seqüenciais e os modelos paralelos fazem parte dos modelos que seguem a *Topologia de Bakis*, ou ainda, demoninados de "left-right" ou "esquerda-

direita". Para esses modelos, a matriz de transição entre estados é triangular superior. Os modelos sequenciais, Figura 4.11b, funcionam segundo uma evolução em série do modelo através de seus estados, mesmo que qualquer um desses estados possam ser saltados no curso do processo. Para os modelos paralelos, Figura 4.11c, muitos caminhos através da rede de Markov são permitidos, sabendo que cada um desses caminhos pode saltar um ou vários estados do modelo.

Cada uma das estruturas dos modelos da Figura 4.11 podem ser generalizadas por incluir um número de estados arbitrário. Entretanto, o número de parâmetros a estimar em um modelo de Markov é da ordem de N^2 (para a matriz A) mais $N \times M$ (para a matriz B). Por consequência, se N é muito grande, uma determinação coerente e precisa das matrizes A e B ótimas vem a ser muito difícil de realizar por uma base de treinamento de tamanho fixado.

Não existem meios teóricos pré-determinados de maneira a escolher o número N de estados necessários no modelo. Para [EL YACOUBI, 1996] a razão deste fato é que os estados nem sempre são fisicamente ligados aos fenômenos observáveis.

Além da variabilidade do tipo da estrutura do MEM, esses ainda podem ser classificados segundo o endereço onde são colocadas as observações, conforme [HUANG et al., 1990]:

- modelos de estados e
- modelos de arcos.

A Figura 4.12a mostra um exemplo de modelo de estados e a Figura 4.12b mostra um exemplo de modelo de arcos. No caso dos modelos de arcos, as duas probabilidades, a_{ij} e $b_j(k)$ consideradas no modelo de estados são substituídas por $pr_{ij}(k)$ que corresponde a probabilidade de transição do estado q_i ao estado q_j emitindo o símbolo da observação v_k .

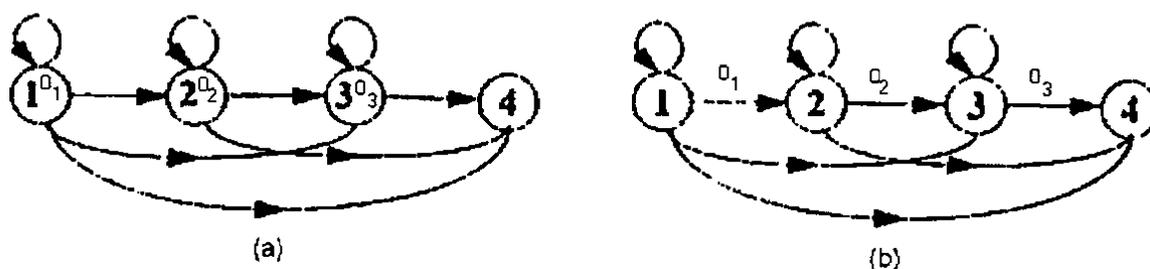


Figura 4.12: Modelo de estados e modelos de arcos: a) modelo de estados e b) modelo de arcos [EL YACOUBI, 1996]

c) Problemas Básicos

Os MEM são aplicados para responder a três problemas básicos [RABINER & JUANG, 1993]:

- Dada uma seqüência de observação e o modelo λ , computar eficientemente a probabilidade da seqüência de observações. Este é o problema da avaliação da probabilidade de observação,
- Dado um conjunto de seqüências de observações relativas as palavras da base de dados de treinamento, produzir um ou mais MEM para cada palavra. Este é o problema de treinamento dos parâmetros dos modelos.
- Dado um conjunto de MEM e uma seqüência de observações correspondente a uma palavra desconhecida, dar uma interpretação a esta seqüência atribuindo esta palavra a um dos modelos (ou classes) do sistema. Este é o problema de reconhecimento ou de classificação.

Existem duas maneiras principais de verificar a evolução da probabilidade de observação: por evolução direta e por evolução dos algoritmos de *Forward-Backward* [RABINER & JUANG, 1993]. Estes algoritmos serão apresentados a seguir.

O problema de treinamento é mais complexo. Felizmente, existem algoritmos potentes que garantem a convergência na direção dos modelos ótimos.

O problema de reconhecimento é resolvido calculando-se a probabilidade que cada modelo tem, produzindo-se à seqüência de observações correspondente a palavra desconhecida e atribuindo-se à esta palavra o modelo que fornece a maior verossimilhança. Esta é a classificação por máxima verossimilhança.

A seguir descreve-se o conjunto de condições iniciais para o estabelecimento dos modelos das palavras do léxico dos cheques bancários brasileiros, o treinamento, a validação e os experimentos de reconhecimento propriamente ditos.

2.4.5.2. Condições Iniciais

Considerando-se que a coleta da base de dados de laboratório não está terminada e o total de 11.948 palavras isoladas, optou-se pela divisão do conjunto de seqüências em três sub-conjuntos: 60% para treinamento, 20% para validação e 20% para reconhecimento.

A Tabela 4.8 apresenta o total de seqüências de observações obtidas para cada uma das palavras e a dimensão dos sub-conjuntos de treinamento, validação e reconhecimento, respectivamente.

Tabela 4.8: Conjuntos de Seqüências de Observações (¹ PC = palavras-chaves)

	Palavras	Total	T	V	R
U N I D A D E	Um	332	199	67	66
	Dois	350	209	71	70
	Três	336	201	68	67
	Quatro	322	193	65	64
	Cinco	332	199	67	66
	Seis	302	181	61	60
	Sete	267	160	54	53
	Oito	305	182	62	61
	Nove	318	190	65	63
	%	23,97	23,98	23,89	24,00
	D E Z E N A	Dez	47	28	10
Onze		50	29	11	10
Doze		29	17	7	5
Treze		37	22	8	7
Quatorze		35	20	8	7
Quinze		45	26	10	9
Dezesseis		30	17	7	6
Dezessete		36	21	8	7
Dezoito		34	20	8	6
Dezenove		45	26	10	9
%		3,25	3,16	3,58	3,16
E N T A	Vinte	382	229	77	76
	Trinta	372	223	75	74
	Quarenta	338	202	69	67
	Cinqüenta	376	225	76	75
	Sessenta	356	213	72	71
	Setenta	370	221	75	74
	Oitenta	363	217	74	72
	Noventa	316	189	64	63
	%	24,05	24,06	23,98	24,08
E N T O S	Cem	7	4	2	1
	Cento	269	161	55	53
	Duzentos	249	149	51	49
	Trezentos	236	141	48	47
	Quatrocentos	218	130	45	43
	Quinhentos	240	143	49	48
	Seiscentos	219	131	45	43
	Setecentos	248	148	51	49
	Oitocentos	252	151	51	50
	%	18,25	18,23	18,42	18,15

Tabela 4.8: Conjuntos de Sequências de Observações (¹ PC = palavras-chaves) - Continuação

P C ¹	Palavras	Total	T	V	R
		Mil	1.286	771	258
	Reais/Real	1.262	757	253	252
	Centavos/Centavo	1.094	656	220	218
	%	30,48	30,56	30,12	30,61
	Total	11.948	7.146	2.427	2.375

A seguir descreve-se cada um dos algoritmos e critérios adotados nas etapas de treinamento, validação e reconhecimento das palavras.

2.4.5.3. O Treinamento: Algoritmo *Baum-Welch*

Na fase de treinamento são necessárias técnicas que permitem captar a informação estatística sobre a variabilidade das formas que representam uma mesma palavra. O interesse dos MEM resulta da existência de algoritmos eficazes que permitem a convergência dos parâmetros dos modelos. Partindo de um valor inicial dos parâmetros, esses algoritmos iterativos alinham os modelos sobre observações conhecidas (base de dados de treinamento) e calculam as estatísticas quando do momento dessas produções.

A partir dessas estatísticas, determina-se novos valores dos parâmetros que melhoram a probabilidade total da emissão das observações da base de dados de treinamento. Esta base de dados deve possuir uma dimensão que possibilite que cada fenômeno (forma) particular de variabilidade seja suficientemente representado.

O treinamento Markoviano está ligado a um contexto Bayesiano onde procura-se o modelo λ que melhor considera os dados, isto quer dizer é o mais provável em função dos dados. Procura-se portanto, $\hat{\lambda}$ tal que:

$$\hat{\lambda} = \arg \max (\Pr(\lambda / O_i, i = 1, \dots, t)) \quad (4.7)$$

sendo $\hat{\lambda}$ o modelo estimado e O_i os dados do treinamento.

Pela regra de Bayes:

$$\Pr(\lambda / O_i, i = 1, \dots, t) = \frac{\Pr(O_i / \lambda) \times \Pr(\lambda)}{\Pr(O_i)} \quad (4.8)$$

a otimização desta equação é realizada sob os modelos λ , onde obtém-se, portanto:

$$\hat{\lambda} = \arg \max (\Pr(O_i / \lambda) \times \Pr(\lambda)) \quad (4.9)$$

pois, os dados são por definição independentes:

$$\hat{\lambda} = \arg \max (\Pr(\lambda) \times \prod_i \Pr(O_i / \lambda)) \quad (4.10)$$

na qual $\Pr(\lambda)$ é um meio pelo qual se faz preferir um modelo a outros. Em geral, a escolha é realizada utilizando-se,

$$\hat{\lambda} = \arg \max \prod_i \Pr(O_i / \lambda) \quad (4.11)$$

o que resulta em fazer um treinamento por Máxima Verossimilhança.

O treinamento supõe que todos os modelos são a priori equiprováveis. É difícil, então, estabelecer um critério para estimar a probabilidade *a priori* do modelo.

Dois algoritmos podem ser utilizados na etapa de treinamento dos modelos: Baum-Welch e Viterbi. [EL YACOUBI, 1996] afirma que no processo de treinamento o algoritmo de Viterbi não leva em conta os eventos correspondentes ao alinhamento ótimo entre as formas observáveis e o modelo. O algoritmo de Baum-Welch leva em consideração todos os alinhamentos possíveis e maximiza a verdadeira probabilidade de emissão das observações. Por consequência, a desigualdade $P^V \leq P^{BW}$ é verdadeira, sendo a igualdade satisfeita somente quando a seqüência de estados que produz as observações é única.

A seguir descreve-se o algoritmo de Baum-Welch utilizado no escopo do treinamento dos modelos, por máxima verossimilhança, para o presente trabalho. A aplicação deste algoritmo está ligada ao fato do mesmo considerar todos os alinhamentos possíveis e maximizar a verdadeira probabilidade de emissão das observações. Isto é importante para o treinamento de formas que envolvem tantos fatores de variabilidade como os encontrados nas palavras manuscritas. Não apresenta-se o algoritmo de Viterbi, uma vez que não se aplica o mesmo, podendo ser encontrado em [HUANG et al., 1990] e [RABINER & JUANG, 1993].

O Algoritmo de Baum-Welch no Treinamento

Num primeiro instante, supõem-se de que somente uma seqüência de observações será processada. Isto por duas razões, por um lado para simplificar o problema, e por outro lado porque a extensão para múltiplas seqüências de observações é direta.

A idéia principal sobre a qual está baseada o algoritmo de Baum-Welch é que sendo dada uma estimação qualquer λ do modelo, e uma seqüência de observações O , as melhores estimativas dos parâmetros do novo modelo λ^{NM} são dadas por [HUANG et al., 1990] e [RABINER & JUANG, 1993]:

$$a_{ij}^{NM} = \frac{P(\text{transição do estado } q_i \text{ ao estado } q_j / O, \lambda)}{P(\text{transição do estado } q_i \text{ a um estado qualquer} / O, \lambda)} \quad (4.12)$$

$$b_{jk}^{NM} = \frac{P(\text{emissão do símbolo } v_k \text{ ao estado } q_j / O, \lambda)}{P(\text{emissão de um símbolo qualquer a partir do estado } q_j / O, \lambda)} \quad (4.13)$$

$$\pi_i^{NM} = P(\text{seqüência de observação começando pelo estado } q_i / O, \lambda) \quad (4.14)$$

O interesse do algoritmo de Baum-Welch reside no fato de que as fórmulas de reestimação de A, B e π acima, garantem o acréscimo de $P(O/\lambda)$ até que um ponto de convergência seja atingido, ponto para o qual os parâmetros não mudam mais. Este procedimento de controle do treinamento encontra-se apresentado através do Processo de *Cross-Validation* descrito posteriormente.

O algoritmo de Baum-Welch é composto por dois algoritmos: *Forward* e *Backward* que permitem calcular as probabilidades exigidas pelas Equações (4.12) a (4.14). A seguir apresenta-se esses dois algoritmos para posteriormente retornar ao cálculo das equações mencionadas, que por sua vez permitem obter os parâmetros do modelo em treinamento.

a) O Algoritmo de *Forward*

O algoritmo de Baum-Welch necessita do cálculo da probabilidade chamada *Forward*, que são as probabilidades conjuntas de emitir a seqüência de observação parcial O_1, O_2, \dots, O_t e de ocupar o estado q_i no instante t . Designa-se as probabilidades *Forward* por $\alpha_t(i)$, então:

$$\alpha_t(i) = P(O_1 O_2 \dots O_t, q_t = i / \lambda) \quad (4.15)$$

A probabilidade que deseja-se calcular $P(O/\lambda)$ é obtida sob a seguinte forma (a notação desta probabilidade P^{BW} fará referência ao algoritmo de Baum-Welch):

$$P^{BW} = \sum_{i=1}^N \alpha_T(i) \quad (4.16)$$

porque $\alpha_t(i)$ é a probabilidade de emitir O em conclusão ao estado q_i . Os α_t podem ser calculados recursivamente pelo algoritmo a seguir:

<p>Início</p> <p style="text-align: center;">1) INICIALIZAÇÃO</p> $\alpha_1(i) = \pi_i b_i(O_1) \quad 1 \leq i \leq N \quad (4.17)$ <p style="text-align: center;">2) INDUÇÃO</p> $\alpha_{t+1}(j) = \left[\sum_{i=1}^N \alpha_t(i) a_{ij} \right] b_j(O_{t+1}) \quad 1 \leq t \leq T-1 \quad e \quad 1 \leq j \leq N \quad (4.18)$ <p style="text-align: center;">3) TÉRMINO</p> $P(O/\lambda) = \sum_{i=1}^N \alpha_T(i) \quad (4.19)$ <p>Fim_Forward;</p>
--

b) Algoritmo de *Backward*

Para calcular essas quantidades, as probabilidades *Forward* $\alpha_t(i)$ são completadas pelas probabilidades *Backward*, de acordo com o algoritmo seguinte:

<p>Início</p> <p style="text-align: center;">1) INICIALIZAÇÃO</p> $\beta_T(i) = 1 \quad 1 \leq i \leq N \quad (4.20)$ <p style="text-align: center;">2) INDUÇÃO</p> $\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(O_{t+1}) \beta_{t+1}(j) \quad T-1 \geq t \geq 1 \quad 1 \leq i \leq N \quad (4.21)$ <p>Fim_Backward;</p>
--

Assim, o numerador da equação (4.12) é dado por:

$$P(\text{transição do estado } q_i \text{ ao estado } q_j, O/\lambda) = \sum_{t=1}^T \alpha_t(i) \beta_{t+1}(j) b_j(O_{t+1}) a_{ij} \quad (4.22)$$

sendo:

$\alpha_t(i)$ fornece a probabilidade de estar no estado q_i no instante t ;

$a_{ij}b_j(O_{t+1})$ calcula o número de transições a um estado q_j emitindo o símbolo O_{t+1} a partir desse estado;

$\beta_{t+1}(j)$ conta o número de ocupações do estado q_j no instante $t+1$ e completando em seguida a seqüência de observações.

Essa probabilidade deve ser calculada sobre todos os instantes aos quais é possível de realizar uma transição, isto é, de $t=T-1$ para $t=1$.

c) Obtenção dos Parâmetros do Modelo

Para calcular o denominador da equação (4.12) recorda-se que:

$$\begin{aligned}\alpha_t(i) &= P(O_1 O_2 \cdots O_t, q_t = i / \lambda) \\ \beta_t(i) &= P(O_{t+1} O_{t+2} \cdots O_T / q_t = i, \lambda)\end{aligned}\quad (4.23)$$

Segue-se que $\alpha_t(i)\beta_t(i)$ é a probabilidade de ocupar o estado q_i ao instante t no modelo λ . Portanto, o denominador da equação (4.12) escreve-se:

$$P(\text{Transição do estado } q_i \text{ ao estado qualquer, } O/\lambda) = \sum_{t=1}^{T-1} \alpha_t(i)\beta_t(i) \quad (4.24)$$

e a equação completa pode ser escrita como:

$$a_{ij}^{NM} = \frac{\sum_{t=1}^{T-1} \alpha_t(i)\beta_{t+1}(j)b_j(v_{t+1})a_{ij}}{\sum_{t=1}^{T-1} \alpha_t(i)\beta_t(i)} \quad (4.25)$$

Um raciocínio similar leva a escrever a equação (4.13) sob a forma:

$$b_{jk}^{NM} = \frac{\sum_{t=1}^{T-1} \alpha_t(j)\beta_t(j)}{\sum_{t=1}^{T-1} \alpha_t(j)\beta_t(j)} \quad (4.26)$$

O somatório do numerador da equação (4.26) se lê: "somar sobre todos os t para os quais O_t é o símbolo v_k ". Finalmente, a reestimação de π é dada por:

$$\pi_i^{NM} = \frac{\alpha_1(i)\beta_1(i)}{P^{BW}} \quad (4.27)$$

A modelagem das 39 palavras do léxico dos cheques bancários brasileiros segue o MEM proposto pela Figura 4.13. A topologia em questão caracteriza-se por um modelo do tipo discreto, esquerda-direita (*left-right* ou *Bakis Topology*) e com passo

igual a 2. Isto, deve-se ao fato de que as palavras são assim melhor modeladas e não se caracterizam, por exemplo, como modelos ergóticos, onde as transições para cada estado está na relação 1:N, onde N é o número de estados que compõem o modelo.



Figura 4.13: Topologia MEM para cada uma das palavras

Assim, o treinamento dos modelos MEM seguindo-se uma abordagem Global deve encontrar respostas às seguintes questões:

- Qual o número de estados adequado para cada um dos modelos-palavras?
- As seqüências de observações (extração de primitivas) são adequadas à representação das palavras?
- Os conjuntos de seqüências para treinamento atendem às necessidades inerentes desta etapa?

Com base no tamanho médio das seqüências de observações estabeleceu-se uma estratégia de busca do número de estados adequado ao modelo durante a fase de treinamento. Busca-se por um modelo que tenha de 3 ao tamanho mínimo das seqüências de observações.

Considera-se que 3 é o número mínimo de estados que um modelo pode ter seguindo um topologia esquerda-direita com passo 2. Por outro lado, sabe-se que modelos com número de estados maior do que o tamanho mínimo das seqüências de observações não conseguem ser treinados por falta de observações, não proporcionando a propagação das probabilidades no modelo.

A estratégia de busca do número adequado de estados dos modelos é auxiliada pelo processo de Validação Cruzada que permite, a cada iteração com a base de dados de treinamento, validar o modelo estimado através da base de dados de validação.

Sabe-se de [RABINER & JUANG, 1983] que seqüências de observações de tamanho muito grande fazem com que o número de estados nos MEM também sejam elevados, gerando problemas nos MEM durante a fase de treinamento. O problema mais

comum é que as probabilidades de transição entre estados e de observação de símbolos tendam a zero.

A topologia adotada para os modelos evita que as probabilidades tendam a zero, uma vez que os modelos não são do tipo ergótico. A cada estado tem-se a possibilidade de transitar para no máximo dois estados seguintes ou permanecer em si próprio. Portanto a matriz A é inicializada de maneira a atender esta especificação.

Outro detalhe da topologia é que o último estado não permite o auto-ciclo, ou seja, não se permite a permanência neste estado. Assim, o modelo não converge rapidamente para o último estado permanecendo aí até o fim das observações. Evita-se que o último estado interfira de maneira prejudicial no treinamento do modelo.

Outro fato positivo, é que os alfabetos de símbolos não são de grandes dimensões e foram validados anteriormente ao uso na etapa de treinamento.

Observa-se que as palavras com conjuntos menores de amostras realizam menos iterações, ou seja, o processo de aprendizado não é muito longo. Assim, a tendência é que a diferença entre as probabilidades de treinamento e validação seja maior. Deste modo, estes conjuntos geram maiores erros no reconhecimento, quando comparado com os demais conjuntos.

Os MEM tem a capacidade de absorver a variabilidade da escrita cursiva humana, conforme já mencionado, para tal necessita de um conjunto de dados de treinamento proporcional a esta variabilidade. Então, quanto maior a variabilidade, maior deverá ser o conjunto de dados na fase de treinamento dos modelos. Sabe-se que para as palavras do conjunto da dezena (“dez”, “onze”, “doze”, “treze”, “quatorze”, “quinze”, “dezesesseis”, “dezessete”, “dezoito” e “dezenove”) apresentam poucos exemplares, compondo 3,25% da base de treinamento, e portanto, sendo o processo de treinamento dos respectivos modelos prejudicado em relação às demais palavras.

O treinamento dos modelos considera o algoritmo de Baum-Welch realizando o processo de Validação Cruzada sendo a inicialização das matrizes A , B e π de acordo com as seguintes especificações:

- matriz A (matriz de transição entre estados): $A[i][j]$, onde $i = 1, \dots, n^\circ$ de estados e $j = 1, \dots, n^\circ$ de estados, sendo $A[i][j] = 1/n^\circ$ de estados,

- matriz B (matriz de probabilidade transição de observação de símbolos condicionada aos estados): $B[i][k]$, onde $i = 1, \dots, n^\circ$ de estados e $j = 1, \dots, n^\circ$ de símbolos, sendo $B[i][k] = 1/n^\circ$ de símbolos,
- vetor π (matriz de probabilidade inicial): $\pi[i]$ para $i = 1, \dots, n^\circ$ de estados, onde $\pi[1] = 1$, indicando que há somente uma entrada no modelo que ocorre pelo primeiro estado.

Para [RABINER & JUANG, 1983] a inicialização das matrizes A, B e π não é um processo fácil e muito menos existe um único caminho a ser seguido. A razão para inicializar as matrizes A e B como equiprováveis é dar a chance para que as transições entre estados e as transições de observação de símbolos tenham condições de evoluir na etapa de treinamento. A inicialização randômica das matrizes pode favorecer um determinado estado ou transição de observação de símbolos prejudicando o treinamento dos modelos, tendo sido este fato observado como experimento na prática.

2.4.5.4. A Validação no Treinamento: Algoritmo de Validação Cruzada

O treinamento dos modelos considera o algoritmo de Baum-Welch [HUANG et al., 1990] e [RABINER & JUANG, 1993] associado a um processo de controle do treinamento através do algoritmo de Validação Cruzada ou *Cross-Validation*.

Os objetivos do processo de Validação Cruzada são o de monitorar o desempenho geral durante o processo de treinamento dos modelos e permitir a parada do processo de treinamento quando a melhoria da reestimação dos parâmetros do modelo não é mais possível. Este processo é realizado sobre dois conjuntos de dados: treinamento e validação. Portanto, o algoritmo de Validação Cruzada pode ser descrito em duas etapas, como apresentado a seguir.

Assim, aplica-se o algoritmo de Baum-Welch, sobre o conjunto de dados de treinamento e calcula-se a probabilidade do modelo de treinamento. Em seguida, aplica-se o algoritmo de *Forward* sobre o conjunto de dados de validação e também obtém-se a probabilidade do modelo. Desta forma, o processo de *Cross-Validation* compara as probabilidades obtidas com os conjuntos O^v de treinamento e O^u de validação realizando a reestimação dos modelos. O modelo final a ser considerado é aquele que possui demonstra a convergência do processo de Validação Cruzada. Este processo é realizado até que um critério de erro ϵ seja atingido.

Início

Modelo_Inicial (A, B, π);

//Para a Base de Dados de Treinamento

Forward (cálculo dos valores de $\alpha_t(i)$);

Backward (cálculo dos valores de $\beta_t(i)$);

Reestimação_Modelo($\bar{\lambda} = \bar{A}, \bar{B}, \bar{\pi}$);

Se ($pr(O^u/\lambda) - pr(O^v/\lambda) \geq \epsilon$)

Então $\lambda = \bar{\lambda}$

Forward (cálculo dos valores de $\alpha_t(i)$);

Senão $\lambda^* = \arg \max_{\lambda} \sum_{u=1}^{u^v} \log P(O^u | \lambda)$;

//Para a Base de Dados de Validação

Forward (cálculo dos valores de $\alpha_t(i)$);

Cálculo_ $pr(O^v/\lambda)$;

Fim_Validação_Cruzada;

O objetivo do processo de Validação Cruzada é o de monitorar o desempenho geral durante o processo de treinamento dos modelos. A probabilidade do modelo de treinamento é crescente ao longo das diversas iterações do processo. Por outro lado, a probabilidade do modelo sobre os dados de validação acompanha a curva de desenvolvimento da treinamento até o ponto onde inicia-se o declínio da curva de validação. Portanto, este processo compara as probabilidades de treinamento e de validação realizando a reestimação dos modelos. O modelo final do treinamento a ser considerado é aquele que possui a melhor probabilidade de validação, considerando-se que a curva de validação está no pico máximo. Todo este processo é realizado até que um critério de erro seja atingido. Para o estudo em questão o critério aplicado foi que a proporção entre as probabilidades, para as bases de dados de treinamento e validação, nos instantes $i-1$ e i seja inferior a 10^{-3} .

2.4.5.5. O Reconhecimento: Algoritmo *Forward*

No problema do reconhecimento consideram-se dados [HUANG et al., 1990] e [RABINER & JUANG, 1993]:

- uma seqüência de observações desconhecidas e um conjunto de modelos $\lambda_1, \lambda_2, \dots, \lambda_{N_c}$ (N_c sendo o número de palavras ou de classes possíveis, e cada

modelo λ_i é definido por um conjunto de parâmetros (A, B, π) que lhe é próprio),

e

- uma descrição desconhecida O , composta de uma seqüência de T observações O_1, O_2, \dots, O_T , onde cada O_i é um dos M símbolos do alfabeto $\{v_1, v_2, \dots, v_m\}$.

Pretende-se, então, dar uma interpretação para a descrição, determinando-se a classe de pertinência.

Para o reconhecimento necessita-se calcular a probabilidade $P(O/\lambda_i)$ que o modelo λ_i produz a seqüência O para $i = 1, 2, \dots, N_c$. A classe k reconhecida verifica o critério seguinte:

$$P(O/\lambda_k) = \arg \max_{i, 1 \leq i \leq N_c} (P(O/\lambda_i)) \quad (4.28)$$

Esta é a regra de máxima verossimilhança e para calcular essas probabilidades aplica-se o algoritmo de *Forward*, apresentado anteriormente.

A obtenção do modelo para cada uma das palavras constitui a determinação das matrizes A , B e π definidas como: matriz de transição entre estados, matriz de probabilidade transição de observação de símbolos condicionada aos estados e matriz de probabilidade inicial, respectivamente.

A Figura 4.14 exemplifica a metodologia aplicada para decisão de reconhecimento das palavras isoladas. O esquema explica como em [RABINER & JUANG, 1983] que para cada modelo $\{A, B, \pi\}$ correspondente a cada palavra do léxico, toma-se a seqüência de observações O extraída de uma palavra conhecida X e faz-se o cálculo das probabilidades (algoritmo de *Forward*) por todos os possíveis modelos.

Seleciona-se, então, a probabilidade de maior valor (TOP1) e verifica-se qual o modelo λ correspondente. Desta forma, resulta que a seqüência O é reconhecida pelo modelo λ e, portanto, se o modelo λ corresponde a palavra X tem-se o reconhecimento correto da mesma. Os resultados obtidos apresentam as taxas de reconhecimento para os experimentos realizados, considerando-se as 5 melhores hipóteses de reconhecimento (TOP1, TOP2, ..., TOP5).

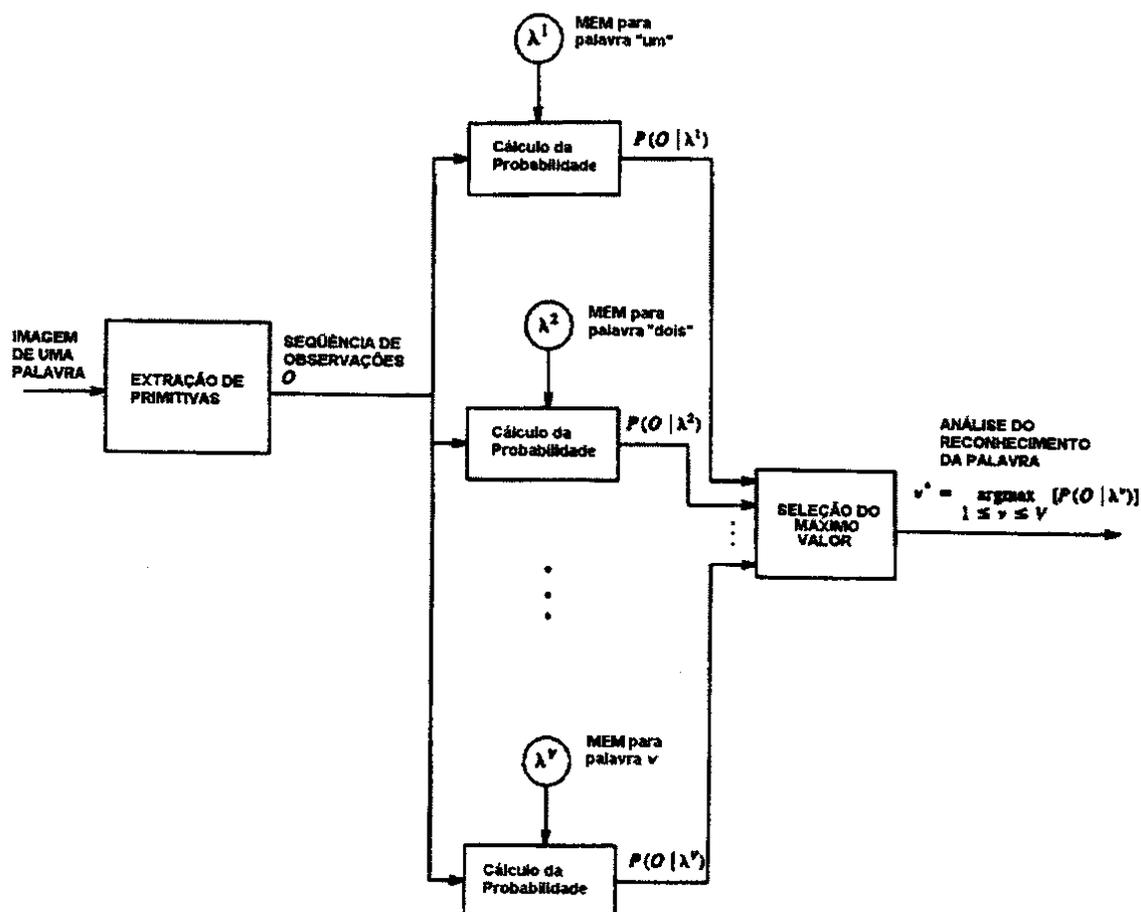


Figura 4.14: Reconhecimento de palavras isoladas MEM [RABINER & JUANG, 1983]

2.4.5.6. Vantagens e Desvantagens dos MEM

Algumas das propriedades de interesse dos MEM devem ser ressaltadas, tais como:

- Os MEM possuem uma representação complementar nos seus dois conjuntos de parâmetros A e B. As probabilidades de observação de símbolos representam a morfologia do fenômeno estudado. O outro conjunto de parâmetros, as probabilidades de transições, representa a interação temporal (ligada a ordem) entre os componentes de uma forma. Finalmente, a otimização conjunta destes desses parâmetros faz com que os MEM sejam particularmente adaptados à modelagem de sinais variantes no tempo, assim como a fala ou a escrita,

- Seja na fala ou em palavras manuscritas, o problema de reconhecimento torna-se difícil pelo fato da existência de várias fontes de variabilidade, conforme já mencionado neste capítulo. Para tratar esse problema de maneira eficaz, é necessário

uma grande quantidade de exemplos de treinamento para uma modelagem correta dos fenômenos observados. O ajuste manual entre as formas e as classes associadas, em uma base qualquer de treinamento, é portanto, uma tarefa extremamente enfadonha, na verdade impossível. Os MEM possuem esta capacidade, colocando-se inicialmente hipóteses sobre a estrutura do modelo e utilizando-se em seguida a probabilidade do modelo gerado a partir da base de treinamento (critério da máxima verossimilhança) para melhorar as estimativas de parâmetros,

- O treinamento dos parâmetros com o auxílio dos MEM e das técnicas de reestimação apóia-se sobre uma base matemática garantindo a convergência para um ponto crítico ótimo,

- O treinamento ajusta-se convenientemente ao aumento do número de exemplos da base de dados, necessitando simplesmente de um aumento linear dos cálculos,

- O reconhecimento de formas muito variáveis, tais como a fala ou a escrita, requerem a ponderação com incertezas. Os MEM sendo modelos probabilísticos, e portanto, são os candidatos naturais à representação destas formas,

- O reconhecimento implica numa pesquisa dos espaços dos estados visando uma solução ótima ou quase ótima. As pesquisas baseadas em MEM não propagam erros e permitem integrar diferentes fontes de conhecimento a cada decisão.

A utilização dos MEM pelos problemas reais encontra algumas dificuldades apresentadas a seguir:

- Os MEM se apoiam sobre a hipótese de independência das probabilidades para determinar a probabilidade de observação de uma seqüência de símbolos. Isto decorre da hipótese que as observações extraídas de um sinal são independentes entre elas, e que a geração de um símbolo de observação não é função do estado corrente da cadeia de Markov. É claro que para os sinais reais, tais como a fala ou a imagem, esta hipótese é inadequada e não pode se aproximar da realidade. Todavia, tem a vantagem de simplificar os cálculos,

- [LEVINSON et al., 1982] apud [EL YACOUBI, 1996] estudaram vários exemplos concretos para julgar o comportamento das técnicas de modelagem por MEM sobre dados ideais, e verificaram que os algoritmos de estimação de parâmetros são corretos. A lição mais importante que pode-se tirar é que mesmo sobre condições ideais (quando os dados são gerados por um MEM conhecido a

priori e que tem-se, portanto, a vantagem de conhecer o verdadeiro modelo), os modelos calculados sobre a base de treinamento podem não representar de maneira fiel a estrutura inerente dos dados tratados. Por consequência, uma grande atenção e validações empíricas são requeridas para uma utilização destas técnicas.

As vantagens e desvantagens da aplicação dos Modelos Escondidos de Markov possibilitam entender melhor o problema do reconhecimento e as dificuldades encontradas, principalmente, durante o processo de treinamento dos modelos.

Os experimentos realizados, bem como, os resultados obtidos e a análise de erros estão apresentados a seguir.

4.2.5.7. Experimentos Realizados

Os modelos definidos para as 39 palavras do léxico apresentam número de estados diferentes como mostrado na Figura 4.15. Observa-se que alguns modelos mantêm o número de estados independentemente do conjunto de primitivas extraído.

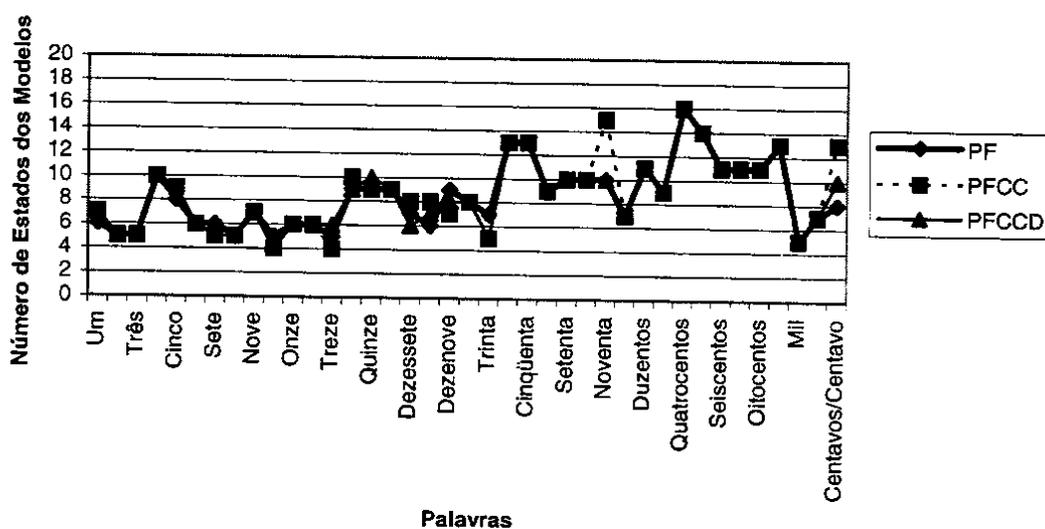


Figura 4.15: Número de estados dos modelos das palavras

As classes de palavras possuem uma similaridade em relação ao número de estados dos modelos. Isto está relacionado ao fato de que as palavras possuem uma mesma origem denunciada em seus prefixos e sufixos, além de possuírem o número de letras e, portanto, de pseudo-segmentos também similares para um mesma classe de palavras.

A Tabela 4.9 apresenta as taxas de reconhecimento para cada uma das 39 palavras do léxico estudado considerando-se os 3 conjuntos de primitivas: PP, PPCC e PPCCR.

Tabela 4.9: Taxas de Reconhecimento das Palavras para os Conjuntos de Primitivas

	Palavras	PP (%)	PPCC (%)	PPCCR (%)
U N I D A D E	Um	80,30	83,33	92,42
	Dois	64,29	65,71	72,86
	Três	47,68	55,22	59,70
	Quatro	70,31	78,13	81,25
	Cinco	63,64	74,98	77,27
	Seis	35,00	50,00	51,67
	Sete	43,40	51,77	43,40
	Oito	73,77	73,77	73,77
	Nove	61,90	75,37	76,19
	D E Z E N A	Dez	66,67	55,56
Onze		70,00	40,00	50,00
Doze		40,00	40,00	20,00
Treze		85,71	57,14	71,43
Quatorze		28,57	42,86	28,57
Quinze		55,56	77,78	55,56
Dezesseis		50,00	33,33	16,67
Dezessete		28,57	57,14	28,57
Dezoito		12,24	48,98	66,67
Dezenove		55,56	70,31	88,89
E N T A	Vinte	52,63	72,37	63,16
	Trinta	63,51	35,14	47,30
	Quarenta	59,70	62,69	65,67
	Cinquenta	78,67	84,92	78,67
	Sessenta	20,29	37,67	36,62
	Setenta	32,43	48,65	37,84
	Oitenta	50,00	47,22	55,56
	Noventa	52,38	71,43	65,08
E N T O S	Cem/Cento	47,35	58,18	58,18
	Duzentos	67,35	69,39	69,39
	Trezentos	31,91	46,81	53,19
	Quatrocentos	58,14	69,07	67,44
	Quinhentos	81,25	78,23	79,17
	Seiscentos	28,87	39,98	44,19
	Setecentos	55,10	57,67	57,14
	Oitocentos	54,00	70,80	66,00
	Novocentos	76,60	68,09	80,85
P C	Mil	59,92	83,27	84,82
	Reais/Real	44,14	68,65	67,46
	Centavos/Centavo	80,83	86,30	83,03
	Média	57,22	67,34	67,66

Observa-se que as taxas de reconhecimento são crescentes para os 3 conjuntos de primitivas em um total de 15 palavras, a saber: “um”, “dois”, “três”, “quatro”, “cinco”, “seis”, “nove”, “dezoito”, “dezenove”, “quarenta”, “oitenta”, “trezentos”, “seiscentos”, “novecentos” e “mil”.

Para a palavra “oito” a taxa de reconhecimento independe do conjunto de primitivas extraído. As palavras “doze”, “cem/cento” e “duzentos” obtiveram taxas iguais para os conjuntos PP e PPCCR.

As palavras “dez”, “onze”, “treze”, “dezesesseis”, “trinta” e “quinhentos” obtiveram uma maior taxa de reconhecimento para o conjunto PP. As 14 palavras restantes obtiveram a maior taxa com o conjunto PPCC.

Desde modo, nota-se que o conjunto PPCCR (67,66%) não traz grande contribuição quando comparado ao conjunto PPCC (67,34%). Porém, ao compará-lo com o conjunto PP (57,22%) percebe-se um aumento significativo da taxa de reconhecimento. O mesmo ocorre entre os conjuntos PP e PPCC.

Pode-se concluir num primeiro instante que os conjuntos PPCC e PPCCR são similares e que o esforço computacional de rotulação dos pixels em busca de uma melhor representação das letras “C”, “S”, “E” e “Z”, e ainda, “u”, “n”, “r” e “v” não corresponde ao esperado. A aplicação do conjunto de primitivas PPCC incluindo somente a extração das concavidades e convexidades no corpo das palavras sobre a imagem do esqueleto é suficiente para melhorar a performance do método. A Tabela 4.10 apresenta os resultados obtidos com as 5 melhores hipóteses de reconhecimento das palavras isoladas (TOP1, TOP2,..., TOP5). Outros autores apresentam seus resultados considerando 10, 16 ou até 20 hipóteses para léxicos menores que 39 palavras [GUILLEVIC, 1995], [AVILA, 1996] e [KNERR et al., 1996].

Tabela 4.10: Taxa de Reconhecimento dos Conjuntos de Primitivas Perceptivas

Conjunto de Primitivas	TOP1	TOP2	TOP3	TOP4	TOP5
PP	57,22	73,16	80,32	84,78	88,10
PPCC	67,34	79,97	85,43	88,87	91,12
PPCCR	67,66	80,42	86,65	89,94	92,21

Outro experimento realizado com as palavras considera a probabilidade de ocorrência de cada uma das palavras na base de dados de treinamento. Assim,

$$k = \arg \max ; [\text{pr}(O / \lambda_i) \cdot p(\lambda_i)] \quad (4.29)$$

onde k é o índice i que maximiza a função ($i = 1, \dots, N_c$), N_c é o número de palavras do léxico (39 palavras), O é a seqüência de observações da palavra e $p(\lambda_i)$ é a freqüência relativa da palavra i na base de dados de treinamento. Realiza-se este experimento para considerar da freqüência de ocorrência na base de dados de treinamento de cada palavra sobre a probabilidade calculada para a palavra desconhecida e, posteriormente, computar as hipóteses de reconhecimento. Os resultados obtidos estão apresentados na Tabela 4.11.

Verifica-se um acréscimo da taxa de reconhecimento para os 3 conjuntos de primitivas, indicando que a freqüência de ocorrência das palavras na base de treinamento deve ser levada em conta no momento da formulação das hipóteses de reconhecimento.

Tabela 4.11: Taxa de Reconhecimento dos Conjuntos de Primitivas Perceptivas para Equação (4.29)

Conjunto de Primitivas	TOP1	TOP2	TOP3	TOP4	TOP5
PP	61,52	76,34	83,24	86,85	90,27
PPCC	69,70	79,94	85,42	88,87	91,11
PPCCR	70,61	82,44	88,08	90,53	92,84

4.2.5.8. Análise de Erros

A análise de erro é realizada sobre a matriz de confusão das palavras da base de teste. Observa-se nesta matriz a confusão existente no reconhecimento de uma palavra por outra de uma mesma classe ou por outra palavra de classe diferente. Isto, auxilia a compreensão da similaridade das formas, das dificuldades do processo de treinamento e, por conseqüência, dos processos de extração de primitivas.

Inicialmente, apresenta-se a matriz de confusão para cada um dos conjuntos de primitivas e, posteriormente, realiza-se uma análise resumo identificando os erros ocorridos intra-classe e inter-classes.

A Tabela 4.12 apresenta a matriz de confusão para o conjunto de primitivas PP. As palavras estão indicadas pelos valores numéricos para facilitar a apresentação da matriz, sendo as letras M, R e C representação das palavras-chaves: “mil”, “ reais/real” e “centavos/centavo”, respectivamente. As Tabelas 4.13 e 4.14 apresentam as matrizes de confusão para os conjuntos PPCC e PPCCR de primitivas.

Para uma correta análise de erros deve-se abstrair das formas das palavras e ter em mente a representação utilizada, ou seja, os conjuntos de primitivas selecionados. Caso contrário, torna-se difícil entender a similaridade de formas e as palavras que são fontes de confusão.

As matrizes de confusão mostram que a similaridade das formas pode ser maior do que a esperada. [OLLIVIER, 1999] afirma que a similaridade de formas deve ser simétrica em relação a matriz de confusão. Ao buscar-se esta simetria na matriz obtida para o conjunto de primitivas perceptivas percebe-se que para alguns casos isto é verdadeiro, como por exemplo: “seis”, “nove” e “um”, “setenta” e “oitenta”, “trezentos” e “duzentos”. Por outro lado, existem casos onde a similaridade simétrica de formas não é verdadeira, como nos seguintes exemplos: “três” e “um”, “dezoito” e “duzentos”, “sessenta” e “noventa”, “novecentos” e “seiscentos”, “seiscentos” e “setecentos”.

As palavras-chaves devem ser analisadas em separado das demais, pois são palavras diferentes entre si na forma agrupadas somente por interesse e, portanto, o grau de confusão entre as mesmas é muito baixo (apenas alguns exemplares) independente do conjunto de primitivas globais utilizado. Observa-se, no entanto, que a palavra “mil” tem sua fonte de confusão na palavra “vinte” (PP e PPCC) e “seis” (PPCCR). Para a palavra “reais/real” a confusão está na palavra “nove” (PP e PPCC) e “seis” (PPCCR). Para a palavra “centavos/centavo” a confusão torna-se evidente com a palavra “seiscentos” (PP), com as palavras “sessenta”, “setenta”, “oitenta” e “seiscentos” (PPCC) e com as palavras “oitenta”, “quinhentos” e “seiscentos” (PPCCR).

Para auxiliar a análise de erros as Tabelas 4.15 a 4.17 resumem as matrizes de confusão, para os 3 conjuntos de primitivas (PP, PPCC e PPCCR), de forma a classificar os erros cometidos pelo método de reconhecimento intra-classe e inter-classes e, ainda, especificar em que classes estão sendo reconhecidas as palavras encontradas na matriz de confusão como inter-classes.

As matrizes de confusão e a classificação dos erros permite analisar três fatores de influência na taxa de reconhecimento das palavras manuscritas no contexto do léxico em português para cheques bancários, sendo estes os seguintes:

- a similaridade das formas,
- a frequência da palavra na base de treinamento,
- os conjuntos de primitivas selecionados.

Tabela 4.15: Análise de Erros - PP

	Palavras	Tipo de Erro		Erro Inter-Classes de Palavras				
		Intra	Inter	Unidade	Dezena	Enta	Entos	PC
U N I D A D E	Um	12	1	-	1	0	0	0
	Dois	13	12	-	10	1	0	1
	Três	22	13	-	5	4	3	1
	Quatro	2	17	-	11	5	0	1
	Cinco	16	8	-	2	2	2	2
	Seis	30	9	-	3	1	0	5
	Sete	18	12	-	7	0	2	3
	Oito	9	7	-	5	0	1	1
	Nove	10	14	-	4	2	0	8
	TOTAL	58,67	41,33	-	51,61	16,13	8,60	23,66
D E Z E N A	Dez	2	1	1	-	0	0	0
	Onze	2	1	0	-	0	1	0
	Doze	3	0	0	-	0	0	0
	Treze	0	1	1	-	0	0	0
	Quatorze	2	3	1	-	1	0	1
	Quinze	1	3	3	-	0	0	0
	Dezesseis	3	0	0	-	0	0	0
	Dezesseite	2	3	0	-	1	2	0
	Dezoito	1	4	1	-	0	3	0
	Dezenove	1	3	1	-	0	2	0
TOTAL	47,22	52,78	42,11	-	10,53	42,11	5,26	
E N T A	Vinte	10	26	15	2	-	5	4
	Trinta	3	24	3	2	-	14	5
	Quarenta	4	23	5	6	-	12	0
	Cinqüenta	3	13	2	1	-	10	0
	Sessenta	22	35	6	2	-	25	2
	Setenta	16	34	10	4	-	17	3
	Oitenta	10	26	6	9	-	7	4
	Noventa	8	22	0	0	-	18	4
TOTAL	27,24	72,76	23,15	12,81	-	53,20	10,84	
E N T O S	Cem/Cento	0	29	10	1	17	-	1
	Duzentos	6	10	2	2	3	-	3
	Trezentos	26	6	1	2	1	-	2
	Quatrocentos	10	8	1	2	4	-	1
	Quinhentos	5	4	0	1	3	-	0
	Seiscentos	29	13	1	0	9	-	3
	Setecentos	12	10	0	2	8	-	0
	Oitocentos	10	13	1	4	6	-	2
	Novecentos	9	2	0	0	2	-	0
TOTAL	52,97	47,03	16,84	14,74	55,79	-	12,63	
P C	Mil	0	103	58	3	37	5	-
	Reais/Real	1	140	116	9	10	5	-
	Centavos/Centavo	0	42	4	4	15	19	-
	TOTAL	0,35	99,65	62,46	5,61	21,75	10,18	-

Tabela 4.16: Análise de Erros - PPCC

	Palavras	Tipo de Erro		Erro Inter-Classes de Palavras				
		Intra	Inter	Unidade	Dezena	Enta	Entos	PC
U N I D A D E	Um	8	3	-	1	1	1	0
	Dois	11	13	-	8	2	0	3
	Três	19	11	-	4	3	2	2
	Quatro	1	13	-	9	4	0	0
	Cinco	3	14	-	1	7	4	2
	Seis	19	11	-	6	2	0	3
	Sete	16	10	-	4	3	0	3
	Oito	10	6	-	4	1	1	0
	Nove	3	13	-	1	3	0	9
	TOTAL	48,91	51,09	-	40,43	27,66	8,51	23,40
D E Z E N A	Dez	4	0	0	-	0	0	0
	Onze	2	4	3	-	0	1	0
	Doze	3	0	0	-	0	0	0
	Treze	2	1	1	-	0	0	0
	Quatorze	1	3	1	-	0	2	0
	Quinze	0	2	1	-	1	0	0
	Dezesseis	2	2	0	-	0	2	0
	Dezesseite	3	0	0	-	0	0	0
	Dezoito	0	3	2	-	0	1	0
	Dezenove	3	0	0	-	0	0	0
TOTAL	57,14	42,86	53,33	-	6,67	40,00	0,00	
E N T A	Vinte	5	16	9	0	-	6	1
	Trinta	18	30	5	2	-	21	2
	Quarenta	6	19	3	6	-	10	0
	Cinquenta	4	9	0	2	-	7	0
	Sessenta	21	24	0	4	-	19	1
	Setenta	14	24	1	4	-	18	1
	Oitenta	17	21	2	5	-	10	4
	Noventa	7	11	0	2	-	8	1
	TOTAL	37,40	62,60	12,99	16,23	-	64,29	6,49
E N T O S	Cem/Cento	1	22	6	2	13	-	1
	Duzentos	6	9	2	1	4	-	2
	Trezentos	22	3	0	1	2	-	0
	Quatrocentos	5	8	1	2	5	-	0
	Quinhentos	6	4	0	0	4	-	0
	Seiscentos	32	10	1	1	7	-	1
	Setecentos	13	8	0	2	5	-	1
	Oitocentos	9	6	0	0	6	-	0
	Novecentos	9	6	0	1	5	-	0
	TOTAL	57,54	42,46	13,16	13,16	67,11	-	6,58
P C	Mil	1	42	26	1	13	2	-
	Reais/Real	2	77	57	7	8	5	-
	Centavos/Centavo	1	29	2	5	12	10	-
	TOTAL	2,63	97,37	57,43	8,78	22,30	11,49	-

Tabela 4.17: Análise de Erros - PPCCR

	Palavras	Tipo de Erros		Erro Inter-Classes de Palavras				
		Intra	Inter	Unidade	Dezena	Enta	Entos	PC
UNIDADE	Um	4	1	-	1	0	0	0
	Dois	7	12	-	7	1	1	3
	Três	18	9	-	3	0	3	3
	Quatro	1	11	-	5	4	0	2
	Cinco	5	10	-	1	6	2	1
	Seis	17	12	-	4	1	1	6
	Sete	16	14	-	6	6	0	2
	Oito	11	5	-	3	0	2	0
	Nove	6	9	-	1	4	0	4
	TOTAL	50,60	49,40	-	37,35	26,51	10,84	25,30
DEZENAS	Dez	3	0	0	-	0	0	0
	Onze	2	3	2	-	0	1	0
	Doze	4	0	0	-	0	0	0
	Treze	1	1	0	-	0	0	1
	Quatorze	1	4	0	-	1	1	2
	Quinze	1	3	2	-	1	0	0
	Dezesseis	3	2	0	-	0	2	0
	Dezessete	2	3	0	-	1	2	0
	Dezoito	1	1	1	-	0	0	0
	Dezenove	1	0	0	-	0	0	0
TOTAL	52,78	47,22	29,41	-	17,65	35,29	17,65	
CENTAS	Vinte	9	19	11	0	-	5	3
	Trinta	19	20	2	5	-	12	1
	Quarenta	4	19	3	3	-	11	2
	Cinquenta	5	11	1	0	-	10	0
	Sessenta	18	27	1	3	-	22	1
	Setenta	25	21	1	4	-	15	1
	Oitenta	14	18	3	3	-	9	3
	Noventa	10	12	0	0	-	12	0
	TOTAL	37,40	62,60	12,99	16,23	-	64,29	6,49
ENTOS	Cem/Cento	2	21	4	1	14	-	2
	Duzentos	6	9	1	3	4	-	1
	Trezentos	17	5	0	1	4	-	0
	Quatrocentos	6	8	0	4	3	-	1
	Quinhentos	7	3	0	0	3	-	0
	Seiscentos	35	7	1	0	5	-	1
	Setecentos	11	10	0	1	9	-	0
	Oitocentos	10	7	0	0	6	-	1
	Novocentos	6	3	0	0	3	-	0
TOTAL	57,80	42,20	8,22	13,70	69,86	-	8,22	
PCC	Mil	1	38	25	1	12	0	-
	Reais/Real	3	79	58	10	6	5	-
	Centavos/Centavo	1	36	5	5	10	16	-
	TOTAL	3,16	96,84	57,52	10,46	18,30	13,73	-

Alguns aspectos da similaridade das formas das palavras já foram mencionados anteriormente, deseja-se neste momento discutir a similaridade de formas entre palavras de uma mesma classes e entre as classes de palavras.

Na Figura 1.3, Capítulo 1, apresenta-se o conjunto de palavras do léxico o qual tem sua origem no valor numérico grafado no cheque bancário. Agora, pode-se observar o mesmo conjunto de palavras levando-se em consideração a similaridade de formas entre as palavras de uma mesma classes. Esta similaridade torna-se evidente com os sufixos e prefixos destacados na Figura 1.3, a saber: “dez-”, “-ze”, “-enta” e “-entos”.

As Tabelas 4.15 a 4.17 mostram que a similaridade de formas entre as classes de palavras ocorre da seguinte forma:

- unidade com dezena e dezena com unidade,
- "enta" com "entos" e vice-versa,
- palavras-chaves com unidade, porém deve-se observar que esta similaridade é válida para as palavras "mil" e "reais/real". No caso da palavra "centavos/centavo" a similaridade está com as classes "enta" e "entos".

Outro fator de influência é a frequência da palavra na base de dados de treinamento. A Tabela 4.8 demonstra que a classe da dezena representa aproximadamente 3% dos conjuntos de treinamento e validação. Este fato, não é diferente com outros autores, tais como: [AVILA, 1996] e [OLLIVIER, 1999]. Sabe-se, então, que o processo de treinamento destas palavras não será adequado, confirmando suas deficiências nos experimentos de reconhecimento.

As Tabelas 4.15 a 4.17 permitem também uma análise dos conjuntos de primitivas selecionados. Observa-se que a % de erros de reconhecimento cometidos entre as classes de palavras diminui para as classes da unidade, da dezena e das palavras-chaves, quando compara-se os 3 conjuntos de primitivas PP, PPCC e PPCCR.

As classes "enta" e "entos" apresentam um acréscimo desta % de erro (inter-classes diferentes). Deste modo, pode-se concluir que o conjunto de primitivas perceptivas PP trabalha bem com as palavras da classe da dezena. Isto, explica-se uma vez que esta classe possui baixa frequência na base de treinamento/validação sendo, portanto, melhor o treinamento quando realizado com um conjunto menor de símbolos.

O conjunto PPCC (primitivas perceptivas, concavidades e convexidades) trabalha bem com a classe "enta" e palavras-chaves. A exceção entre as palavras-chaves é a palavra "mil" que obtém os melhores resultados com o conjunto PPCCR.

Finalmente, o conjunto PPCCR (primitivas perceptivas, concavidades e convexidades rotuladas) trabalha bem para as classes de palavras da unidade e "entos". Para a classe "entos" pode-se ter o direito da incerteza entre os conjuntos PPCC e PPCCR. Esta classe gera confusões não somente pela similaridade com a classe "enta", mas também, por ser o conjunto de palavras mais longas do léxico em português e possuir como sufixos outras palavras do léxico, por exemplo: "quatro", "seis", "sete", "oito", "nove". A Figura 4.16 apresenta alguns erros de reconhecimento de palavras ocorridos com os experimentos realizados.

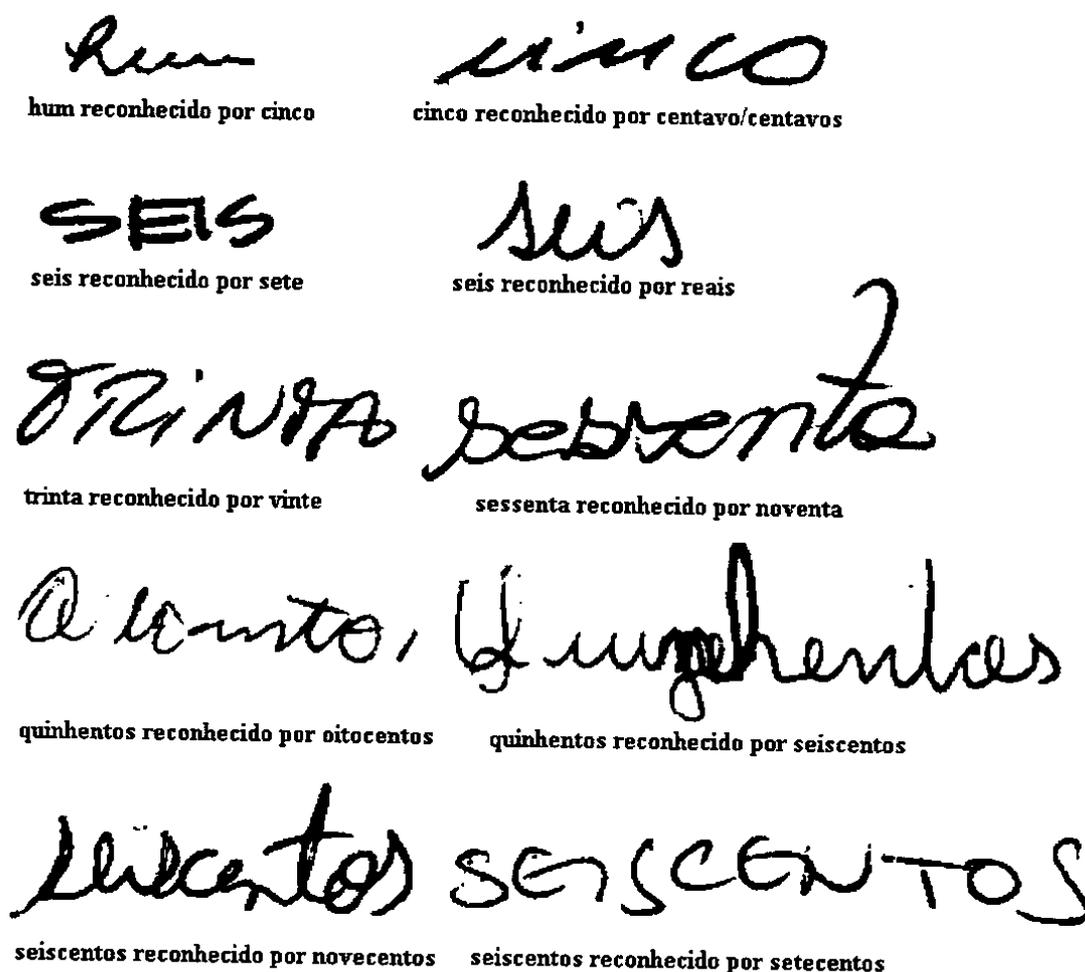


Figura 4.16: Exemplos de confusão no reconhecimento de palavras - conjunto PPCCR

Finalmente, conclui-se que os resultados obtidos são promissores, principalmente, quando comparados aos de outros autores e trabalhos já realizados, conforme apresentado no Apêndice B.

Como resultado da análise de erros percebe-se a necessidade de uma melhor representação das classes com ausência de primitivas perceptivas (ascendentes, descendentes e laços no corpo) e das palavras curtas, tais como: "cinco", "seis", "sete",

“oito”, “nove”, “dez”, “onze”, “doze”, “cem/cento” e “mil”. Deve-se melhorar também a representação das palavras longas e com sufixos similares (“enta” e “entos”). E, ainda, das palavras que estão escritas em caixa alta não possuidoras de primitivas perceptivas. Assim, propõe-se um conjunto de primitivas geométricas apresentado a seguir neste capítulo.

4.3. Primitivas Geométricas

A implementação de um conjunto de primitivas geométricas tem por base a melhoria da representação das palavras curtas e das palavras sem presença de primitivas perceptivas (ascendentes, descendentes e laços no corpo). Pois, para [SCHOMAKER & SEGERS, 1998] as palavras curtas necessitam de um maior número de informações para serem reconhecidas e as palavras longas necessitam de mais informações no final das mesmas (Capítulo 2).

A idéia principal é permitir um enfoque mais local à extração de primitivas, com o objetivo de enriquecer as informações dos elementos lingüísticos que servem de “pistas” ao estabelecimento de uma relação entre o lingüístico e o conceitual-cognitivo (Capítulo 2). Em outras palavras, busca-se uma nova chance de ler o que está escrito. Sendo que o interesse é melhorar a representação das regularidades de uma palavra, tal qual [CÔTE, 1997].

Assim, a utilização de primitivas geométricas em métodos de reconhecimento de palavras manuscritas explica-se diante do fato que estas apresentam tolerância às distorções e às variações de estilo e que estas podem descrever as propriedades globais ou locais de palavras e caracteres [HEUTTE, 1994].

Uma análise das partes côncavas e convexas das palavras permite evidenciar as propriedades topológicas e geométricas das formas a reconhecer. Para [HEUTTE, 1994] existem dois tipos de métodos que permitem extrair as concavidades, a saber:

- métodos com base na rotulação das áreas brancas das imagens ou do fundo ou *background* das imagens,
- métodos com base no tratamento do contorno das formas.

Opta-se pelo método composto pela rotulação dos pixels do fundo da imagem do corpo das palavras, de maneira a extrair para cada coluna da imagem um vetor de características ou *feature vector*. Métodos baseados no contorno das formas foram

descartados, visto que a extração de primitivas globais trabalha com a forma em si propriamente dita.

Outros autores desenvolveram conjuntos de primitivas com base no estudo das concavidades, rotulando os pixels do fundo da imagem, tais como: [LECOLINET, 1990], [GILLIES, 1992]. O primeiro autor trabalha com a extração de primitivas geométricas, rotulando os pixels do fundo das imagens e extraíndo medidas estatísticas para valores numéricos. O segundo autor trabalha rotulando as concavidades e convexidades no corpo de imagens de palavras manuscritas de endereços postais.

4.3.1. Vetor de Características

A composição do vetor de características tem por base a rotulação de cada pixel do fundo da imagem mapeando-se 13 situações diferentes. Deste modo, este vetor armazena a quantidade de cada um dos 13 rótulos encontrados para cada uma das colunas da imagem no corpo das palavras, considerando-se dois tipos de verificação sobre a imagem, a saber:

- verificação dos 4 pixels vizinhos, norte-sul-leste-oeste (NSLO), para cada pixel branco do corpo da palavra e
- verificação de falso laço com possibilidade de escape em 4 direções de sondagem para os pixels determinados como fechado para as 4 direções principais (NSLO).

A Figura 4.17 exemplifica a rotulação dos pixels brancos do corpo da palavra “cinco” retirada da base de dados de treinamento, seguindo-se a rotulação especificada na Tabela 4.18.

Apresenta-se na Figura 4.17 a imagem original da palavra “cinco”, o resultado da correção da inclinação vertical dos caracteres e a detecção do corpo da palavra. Um esquema da rotulação dos pixels do fundo da imagem mostra as direções fechadas (com seta), abertas (sem seta) e com escape (seta \perp) para alguns pixels da imagem. Mostra-se a extração do vetor de características para uma das colunas da palavra “cinco” exemplificada. Cada posição do vetor contabiliza o total de cada tipo de rótulo encontrado na coluna analisada.

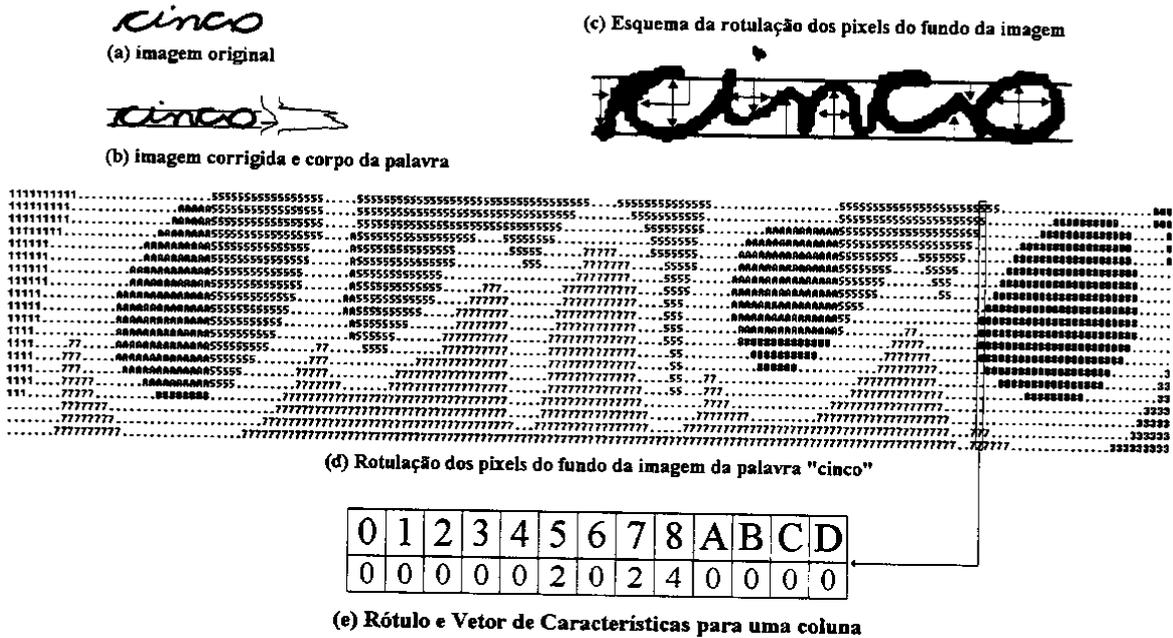


Figura 4.17: Exemplo da rotulação do corpo da palavra para primitivas geométricas

Tabela 4.18: Rótulos do Vetor de Características

Rótulo	Tipo	Busca
0	Aberto à direita e acima	(NSLO)
1	Aberto à esquerda e acima	(NSLO)
2	Aberto à esquerda a abaixo	(NSLO)
3	Aberto à direita e abaixo	(NSLO)
4	Aberto à direita	(NSLO)
5	Aberto acima	(NSLO)
6	Aberto à esquerda	(NSLO)
7	Aberto abaixo	(NSLO)
8	Laço fechado	(NSLO)
A	Escape para direita e acima	Sondagem
B	Escape para esquerda e acima	Sondagem
C	Escape para direita e abaixo	Sondagem
D	Escape para esquerda e abaixo	Sondagem

A distribuição dos rótulos extraídos para a base de dados de treinamento mostram uma representação adequada das formas a reconhecer, conforme Figura 4.18. Observa-se que a palavra “um” é a que possui menor quantidade de pixels rotulados por 8, visto que a mesma não possui laços fechados no corpo como parte da morfologia básica. Por outro, lado a palavra “dois” é a palavra que mais apresenta laços fechados, sendo a proporção de rótulos A a D inferior as demais palavras. O rótulo A apresenta a

maior proporção (20% a 34%) entre os rótulos **A** a **D**. Este rótulo caracteriza o escap para direita e acima, tal qual as letras **c** da palavra “cinco” na Figura 4.17.

Os rótulos **5** e **7** apresentam uma a maior proporção entre os demais rótulos (de 25% a 45%). Estes rótulos são justamente os que permitem descrever as regularidades das palavras, como especificado na Tabela 4.18.

Assim, os vetores de características passam a ser a entrada de dados para o método de classificação que permitirá estabelecer os símbolos correspondentes a cada coluna do corpo da palavra.

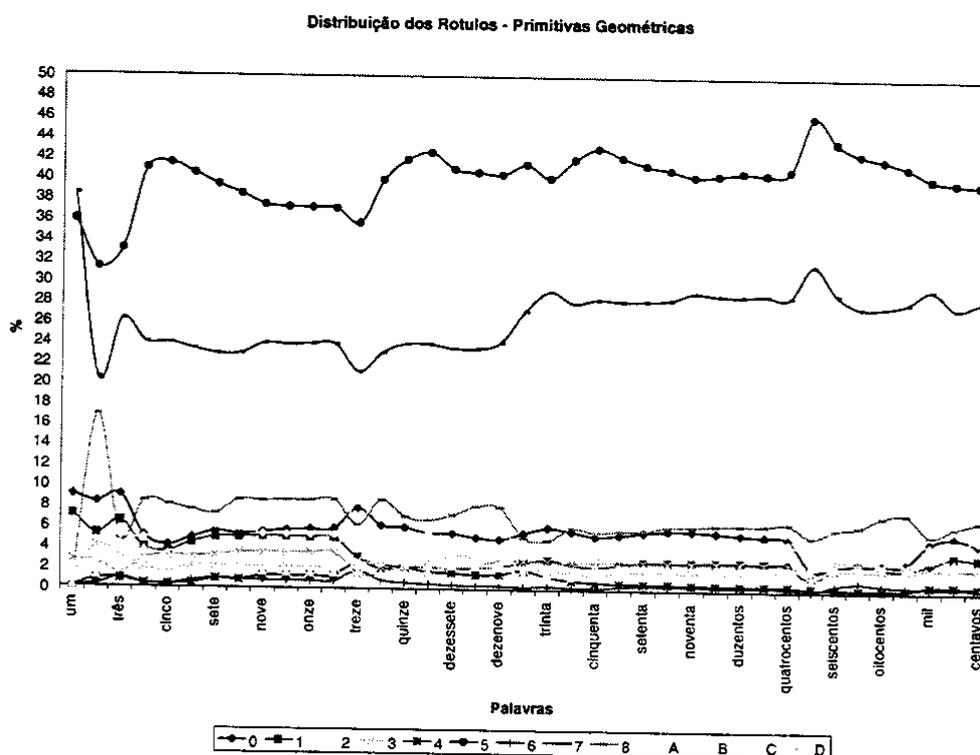


Figura 4.18: Distribuição de frequência dos rótulos

4.3.2. Método de Classificação: Quantização Vetorial

O Capítulo 2 fornece os subsídios necessários a escolha de um método de classificação visto que para cada coluna da imagem do corpo da palavra deseja-se emitir um símbolo correspondente a descrição das concavidades e convexidades rotuladas anteriormente.

Percebe-se, portanto, uma diferença em relação ao conjunto de primitivas globais denominado de Conjunto 03 – PPCCR – primitivas perceptivas, concavidades e convexidades rotuladas. Para este conjunto efetua-se a classificação juntamente com o processo de extração das primitivas, rotulando-se cada pseudo-segmento. Agora, estes

processos estão separados, visando-se uma classificação mais adequada e discriminativa das formas a reconhecer.

O método escolhido é denominado *k*-Vizinhos Mais Próximos (*k*-Nearest-Neighbor - *k*NN) que utiliza o conceito de classificação através de funções de distância e caracteriza-se por ser um método não-paramétrico (Capítulo 2). A motivação para aplicação deste método está no estabelecimento de uma medida de similaridade entre vetores de características, considerando-se os vetores em um espaço Euclidiano e determinando-se a proximidade entre estes vetores.

O método dos *k*-Vizinhos Mais Próximos trabalha calculando as distâncias entre uma entrada de dados, vetor de características, $x \in V$ de um espaço vetorial e um número de símbolos ou *codebooks* $w_i \in V$. A distância Euclidiana mínima entre o vetor x e os *i*-ésimos símbolos é dada por [NESCHEN, 1995]:

$$D_i = \min_l \|x - w_i^l\| \quad l = 1, 2, \dots, N_i \quad (4.30)$$

Entende-se por *codebook* o conjunto de símbolos que permitem a classificação de cada vetor de características em um símbolo, o qual possui a menor distância em relação aos centróides dos símbolos que compõem o próprio *codebook*.

O problema da classificação inicia-se pela definição das fronteiras ou *clustering* entre os símbolos do *codebook* e, posteriormente, pela efetiva classificação dos vetores de primitivas nos símbolos do *codebook* correspondentes aos de mínima distância. Usualmente, a decisão é tomada considerando-se os símbolos dos *k*-vizinhos mais próximos. Esta aproximação é denominada de Quantização Vetorial ou *Vector Quantization* [NESCHEN, 1995].

O problema da definição das fronteiras entre os símbolos do *codebook* pode ser tratado por critérios heurísticos ou por critérios de minimização ou de maximização de parâmetros. As fronteiras são determinadas a partir do momento que se conhece o centróide (x,y) de cada símbolo do *codebook* w_i e a área de influência de cada um no espaço, conforme Figura 4.19.

Para que as regiões não resultem sobrepostas uma pesquisa da "representação ótima" dos centróides dos símbolos deve ser efetuada. Para isto, existem vários algoritmos que auxiliam na definição dos centróides dos símbolos [NESCHEN, 1995]: *Simple Cluster-Seeking Algorithm*, *Maxmin-Distance Algorithm*, *K-Means Algorithm* e *Isodata Algorithm*.

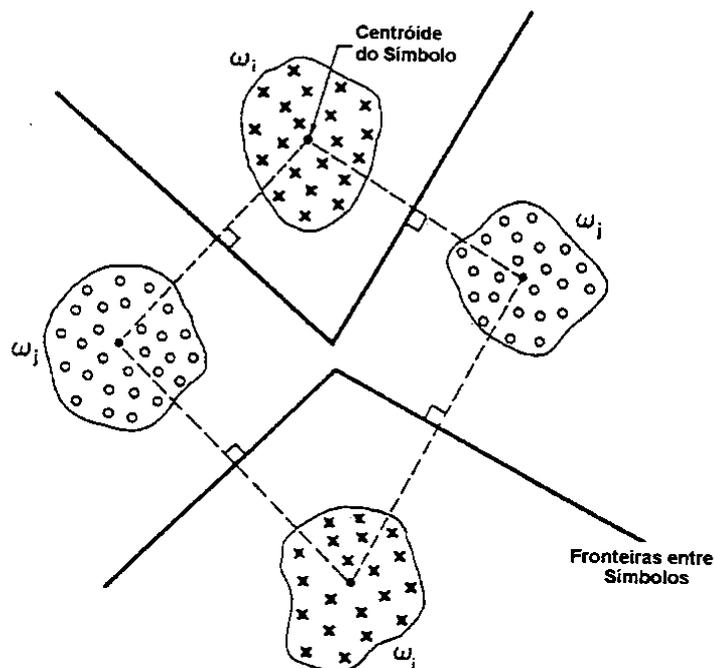


Figura 4.19: Fronteiras para dois símbolos compostas cada qual por duas amostras

O Algoritmo *K-Means* é um procedimento iterativo de classificação que trabalha com a minimização de uma medida de distância dada pela soma dos quadrados das distâncias de todos os vetores de características no domínio da aplicação ao centróide do símbolo. O algoritmo *K-Means* é composto pelos seguintes passos de acordo com [RABINER & JUANG, 1993]:

Passo 1: Escolher K iniciais símbolos com centróides $z_1(1), z_2(1), \dots, z_k(1)$. Estes centróides são arbitrários e normalmente utiliza-se as primeiras K amostras do conjunto de vetores de características da base de treinamento. Para evitar a associação inicial dos centróides dentro de uma mesma classe de palavras, realiza-se um procedimento randômico para escolha do valor inicial dos centróides, evitando-se também que diferentes centróides recebem valores iguais;

Passo 2: Distribuir as amostras $\{x\}$ entre o domínio dos K símbolos, utilizando a seguinte relação:

$$x \in S_j(k) \text{ se } \|x - z_j(k)\| < \|x - z_i(k)\| \quad (4.31)$$

Para todo $i = 1, 2, \dots, K, i \neq j$, onde $S_j(k)$ denota o conjunto de amostras de vetores de características cujo os centróides dos símbolos são $z_j(k)$.

Passo 3: Com base no resultado do Passo 2, calcular os novos centróides $z_j(k+1), j = 1, 2, \dots, K$, tal que a soma do quadrado das distâncias de todos os vetores de

características em $S_j(k)$ para os novos centróides dos símbolos seja minimizada. Em outras palavras, os novos centróides $z_j(k+1)$ são calculados até que a taxa de erro seja minimizada. Adota-se um critério de erro $< 10^{-3}$. O centróide $z_j(k+1)$ que minimiza esta taxa de erro é a média dos $S_j(k)$. Então, o novo centróide é dado por:

$$z_j(k+1) = \frac{1}{N_j} \sum_{x \in S_j(k)} x \quad j = 1, 2, \dots, K \quad (4.32)$$

Onde N_j é o número de amostras de $S_j(k)$. O nome “K-means” é obviamente derivado da maneira na qual os centróides dos símbolos são sequencialmente recalculados.

Passo 4: Se $z_j(k+1) = z_j(k)$ para $j = 1, 2, \dots, K$, o algoritmo converge e o procedimento de definição dos centróides dos símbolos pode ser finalizado. Caso contrário, deve-se retornar ao Passo 2.

O comportamento do algoritmo *K-Means* é influenciado pelo número de centróides especificados, pela escolha inicial dos valores dos centróides, pela ordem em que estes são escolhidos e, também, pelas propriedades geométricas dos dados, ou melhor, das palavras. Na prática, realiza-se diversos experimentos com diferentes valores de K .

A definição do melhor conjunto de símbolos somente é possível após a execução dos experimentos de reconhecimento com a obtenção das respectivas taxas e matrizes de confusão. Portanto, apresenta-se a seguir as características das etapas de treinamento-validação e reconhecimento dos modelos MEM.

4.3.3. Reconhecimento das Palavras Manuscritas com MEM

As seqüências de observações que servem de entrada para a etapa de treinamento e validação dos modelos MEM advém da conversão, para cada conjunto de símbolos em teste, de cada vetor de primitivas em um símbolo.

A etapa de treinamento dos modelos segue os mesmos algoritmos já apresentados no Capítulo 2, Baum-Welch e Validação Cruzada. Estes procedimentos geram, portanto, para cada palavra dois modelos: λ_{PE} (perceptivo) e λ_{GE} (geométrico). Inicialmente, compara-se as taxas de reconhecimento isoladamente, para depois compor as hipóteses de reconhecimento através da combinação das probabilidades de reconhecimento da palavra desconhecida por ambos os modelos ($\log pr(\lambda_{PE}) + \log pr(\lambda_{GE})$). Esta estratégia está representada na Figura 4.20.

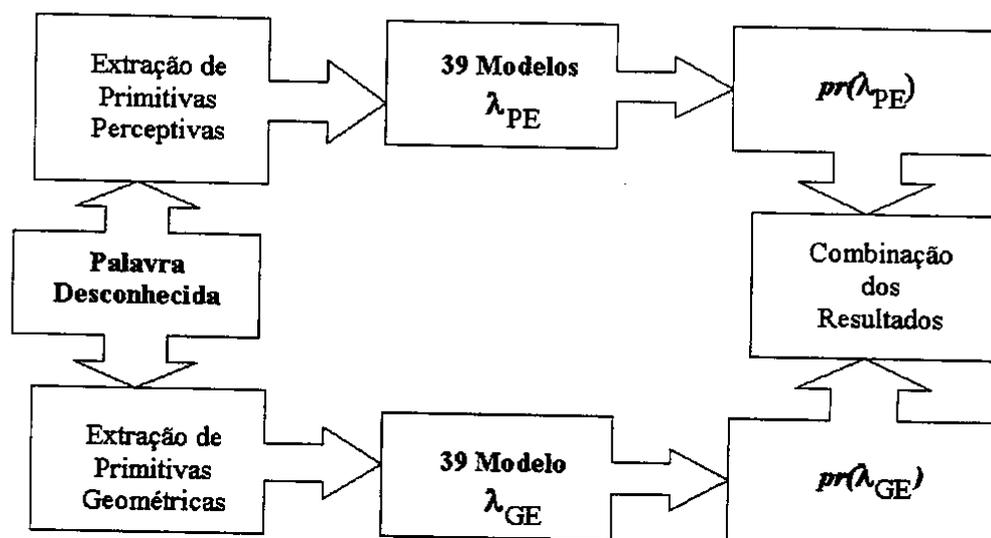


Figura 4.20: Esquema de reconhecimento: primitivas perceptivas e geométricas

A utilização de Múltiplos Conjuntos de Símbolos ou *Multiple Codebooks* para os MEM não é possível, uma vez que os processos de extração das primitivas perceptivas e geométricas são diferentes. Este fato impossibilita a combinação de observações obtidas em instantes t diferentes.

O processo paralelo de definição do conjunto de símbolos e obtenção dos modelos está ainda em realização. Porém, alguns problemas podem ser adiantados referentes ao tamanho das seqüências de observações, ao número de símbolos de cada conjunto testado e ao número de amostras para o treinamento.

Deve-se lembrar que para cada coluna da imagem é emitido um símbolo e, portanto, as seqüências de observações tornam-se longas. A média do tamanho das seqüência para as 39 palavras é de 244 colunas, sendo o menor valor 130 para a palavra “dez” e o maior valor 425 para a palavra “quatrocentos”. Sabe-se da teoria dos MEM que para seqüências longas os cálculos dos algoritmos *Forward* e *Backward* podem fazer com que a probabilidade tenda a zero. Deste modo, deve-se recorrer ao procedimento de Reescalonamento ou *Scaling* [RABINER & JUANG, 1993].

Sabe-se também, que um conjunto composto por um número pequeno de símbolos não é discriminativo das formas a reconhecer. Por outro lado, um número elevado de símbolos associado com a existência de poucas amostras de treinamento pode fazer com que o modelo não possa ser treinado adequadamente.

Mesmo tendo-se consciência de todos estes prováveis problemas acredita-se que as primitivas geométricas contribuam positivamente no processo de reconhecimento das palavras.

Constata-se que os resultados obtidos neste trabalho para os diferentes experimentos realizados vão de encontro aos objetivos estabelecidos para o reconhecimento de palavras isoladas do extenso manuscrito de cheques bancários brasileiros. Desta forma, apresenta-se no Apêndice B um comparativo dos resultados obtidos por outros trabalhos realizados no mesmo escopo.

A seguir, no Capítulo 5 apresenta-se a conclusão deste trabalho, bem como as possibilidades de trabalhos futuros visando a continuidade das pesquisas na área de reconhecimento de palavras manuscritas na PUCPR.

Capítulo 5

Conclusão

O presente trabalho modela e avalia computacionalmente o reconhecimento de palavras manuscritas no contexto de cheques bancários brasileiros. Para isso, foram consideradas características perceptivas relacionadas à escrita e a leitura, e ao uso de Modelos Escondidos de Markov – MEM para processamento da informação.

Considerando-se todos estes aspectos, as primitivas selecionadas respondem as seguintes questões:

- **Quais são as primitivas perceptivas relevantes no processo de reconhecimento de palavras manuscritas?**
- **Como representar palavras cursivas sem a presença de primitivas perceptivas?**

Assim, explora-se os ascendentes, os descendentes, os laços, as concavidades e as convexidades existentes nas formas a reconhecer. Estuda-se 3 conjuntos de primitivas denominados de PP - primitivas perceptivas, PPCC - primitivas perceptivas, côncavo, convexo e PPCCR - primitivas perceptivas, concavidades e convexidades rotuladas.

Pode-se concluir que as concavidades e convexidades exercem um importante papel no desempenho do método, uma vez que permitem representar palavras cursivas sem a presença de primitivas perceptivas, bem como ressaltar as regularidades existentes nas mesmas. Este fato, torna-se evidente diante da modelagem das palavras cursivas através dos MEM.

Os MEM possuem a capacidade de oferecer um modelo probabilístico ao enfoque estrutural e permitem um treinamento automático dos parâmetros a estimar dos modelos. Além do que, possibilitam modelar eficazmente diferentes fontes de conhecimento, a saber: morfológico – a forma a reconhecer, léxico – o vocabulário de

palavras e sintático – o modelo para cada palavra. Estes fatos vem ao encontro dos aspectos de escrita e leitura apresentados neste trabalho.

Desta forma, o desempenho alcançado com o método desenvolvido é promissor, visto que os resultados obtidos com os 3 conjuntos de primitivas são comparáveis aos obtidos por outros autores (Apêndice B). As taxas de reconhecimento obtidas são 56,4% (PP), 64,8% (PPCC) e 70,6% (PPCCR-FR), respectivamente. Além disto, os principais problemas são apontados pela análise de erros. O presente trabalho não inclui observações ou descrição da performance do método uma vez que não se tem preocupação com o tempo de execução.

As questões abordadas na análise de erros referem-se à similaridade das formas das palavras encontradas na matriz de confusão reconhecidas como intra-classe e inter-classes de palavras, à frequência das palavras na base de treinamento e ao desempenho dos conjuntos de primitivas selecionados. Todas estas questões foram exploradas apontando os pontos favoráveis e as dificuldades encontradas no processo de reconhecimento.

Sendo assim, os estudos realizados e os resultados obtidos possibilitam que outros trabalhos possam ser realizados, uma vez que o mesmo contribui para aplicação dos Modelos Escondidos de Markov no reconhecimento de palavras manuscritas provenientes de cheques bancários. Além, da contribuição de sua aplicação em problemas de reconhecimento de palavras do idioma português.

Ressalta-se a contribuição do presente trabalho através da criação da base de dados de cheques bancários brasileiros de laboratório.

Para que as bases de dados não tenham proporções infinitas, reduz-se a variabilidade da escrita cursiva através das etapas pré-processamento, a saber: aquisição das imagens, correção da linha de base, correção da inclinação vertical dos caracteres, detecção e normalização do corpo das palavras. Sabe-se que muitos outros procedimentos podem vir a ser desenvolvidos e aplicados visando-se melhorar os resultados obtidos.

Os trabalhos futuros envolvem a continuidade da modelagem das palavras através das primitivas geométricas melhorando a representação das regularidades de uma palavra, tal qual [CÔTE, 1997]. Busca-se uma nova chance de ler o que está escrito com um enfoque mais local. Para [HEUTTE, 1994] a utilização de primitivas geométricas em métodos de reconhecimento de palavras manuscritas explica-se diante

do fato que estas apresentam tolerância às distorções e às variações de estilo e que estas podem descrever as propriedades globais ou locais de palavras e caracteres.

Outro trabalho é o tratamento do extenso completo, incluindo-se a gramática de grafia dos valores numéricos por extenso para o idioma português no processo de reconhecimento. Uma abordagem para o extenso é considerar as palavras “mil”, “reais/real”, “centavos/centavo” como palavras-chaves no extenso, como apresentado em [FREITAS et al., 2000b], [FREITAS et al., 2000c], [FREITAS et al., 2001].

A modelagem dos prefixos e sufixos das classes das palavras caracteriza-se por outro trabalho a ser realizado, levando-se em consideração as classes da Figura 1.3. Os estudos podem envolver, ainda, a modelagem das palavras de maneira que o reconhecimento seja realizado através da concatenação de modelos, como por exemplo: palavra “novecentos” = modelo “nove” + modelo “entos”, palavra “oitenta” = modelo “oito” + modelo “enta” e palavra “dezesesseis” = modelo “dez” + modelo “seis”.

O presente trabalho coloca-se, ainda, como mais uma contribuição na área de Análise e Reconhecimento de Documentos, podendo vir auxiliar outras pesquisas, tais como: reconhecimento de palavras manuscritas em formulários, projetos e anotações de campo, identificação dos procedimentos de recebimento e despacho de documentos para recuperação e consultas, conversão de desenhos e anotações manuscritas para recuperação ou união com Projetos Auxiliados por Computador – PAC.

Referências Bibliográficas

- [AGARWAL et al, 1995] AGARWAL, A.; GRANOWETTER, L.; HUSSEIN, K.; GUPTA, A. *Detection of Courtesy Amount Block on Bank Checks*. Montreal, ICDAR'95, IEEE 1995, p.748-751.
- [AGAZZI & KUO, 1993] AGAZZI, O. E. & KUO, S.S. *Hidden Markov models based optical character recognition in the presence of deterministic transformations*. Pattern Recognition, Vol.26, No.12, 1993, p.1813-1826.
- [AVILA, 1996] ÄVILA, M. *Optimisation de Modeles Markoviens pour la Reconnaissance de L'ecrit*. Thèse de doctorat, Université de Rouen, France, 1994. 215p.
- [BANCO DO BRASIL, 1983] Banco Central do Brasil. *Manual e Normas de Instruções*. Brasília, DF. 1983.
- [BATITTI, 1994] BATITTI, R. *Using Mutual Information for Selecting Features in Supervised Neural Net Learning*. IEEE Transactions on Neural Networks, Vol.5, no.4,1994, p.537-550.
- [BERCU & LORETTE, 1993] BERCU, S.; LORETTE, G. *On-line handwritten word recognition: an approach based on hidden markov models*. IWFHR-3, 1993, p.385-390.
- [BERTILLE et al, 1994] BERTILLE, J.M.; GILLOUX, M.; YACOUBI, A. *Localisation et reconnaissance conjointes de noms de voies dans les lignes distribution des adresses postales*. SRTP/RD/ Traitement automatique ligne distribution. TRANSITION No. 7, 1994, p.16-25.
- [BOZINOVIC & SRIHARI, 1989] BOZINOVIC, R.M. & SRIHARI, S.N. *Off-line cursive script recognition*. IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol.11, No. 1, 1989, p.68-83.
- [BUNKE et al, 1995] BUNKE, H.; ROTH, M.; SCHUKAT-TALAMAZZINI, E.G. *Off-line cursive handwriting recognition using hidden markov models*. Pattern Recognition, Vol.28, No. 9, 1995, p.1399-1413.
- [BURR, 1983] BURR, D.J. *Designing a handwriting reader*. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-5(5). 1983, p.554-559.

- [BUSE et al., 1997] BUSE,R.; LIU,Z.; CAELLI, T. *A structural and relational approach to handwritten word recognition*. IEEE Transactions on Systems, Man, and Cybernetics, Part B, Cybernetics, Vol.27, No. 5, 1997.
- [CAGLIARI, 1991] CAGLIARI, L.C. *Alfabetização & lingüística*. São Paulo: Editora Scipione. 1991, p.96-166.
- [CAVANAGH ,1974] CAVANAGH, J.A. *Lettering and alphabets. 85 complete alphabets*. New York: Dover Publications, Inc., 1974. 121p.
- [CHEN & KUNDU, 1993] CHEN, M.Y. & KUNDU, A. *An alternative to variable duration HMM in handwritten word recognition*. IWFHR'93. 1993, p.82-91.
- [CHEN et al, 1995a] CHEN, S.; HARALICK, R.M.; PHILLIPS, I.T. *Simultaneous Word Segmentation from Document Images Using Recursive Morphological Closing Transform*. Montreal, ICDAR'95,IEEE 1995, p.761-764.
- [CHEN et al, 1995b] CHEN, M.Y. & KUNDU, A.; SRIHARI, S.N. *Variable duration hidden Markov model and morphological segmentation for handwritten word recognition*. IEEE, Vol.4, No.12, 1995, p.1675-1686.
- [CHO et al., 1995] CHO, W.; LEE, S.W.; KIM, J.H. *Modeling and recognition of cursive words with hidden Markov models*. Pattern Recognition, Vol.28, No.12, 1995, p.1941-1953.
- [CONNELL, 1996] CONNELL, S. *A comparison of hidden markov model features for the recognition of cursive handwriting*. Master Dissertation, Michigan State University, USA, 1996. 154p.
- [CONGEDO et al., 1995] CONGEDO, G.; DIMAURO, G.; IMPEDOVO, S.; PIRLO, G. A. *Structural Method with Local Refinig for Handwritten Character Recognition*. Montreal, ICDAR'95, IEEE 1995, p.853-856.
- [CÔTE, 1997] CÔTÉ, M. *Utilisation d'un modèle d'accès lexical et de concepts perceptifs pour la reconnaissance d'images de mots cursifs*. Thèse de doctorat, École Nationale Supérieure des Télécommunications, France,1997. 171p.
- [COVER & THOMAS, 1991] COVER, T.M., THOMAS, J.A. *Elements of Information Theory*. Wiley Series in Telecommunications, 1991.
- [DOLFING, 1998] DOLFING, J.G.A. *Handwritten recognition and verification – a hidden markov approach*. 1998.
- [DZUBA et al., 1998] DZUBA, G.; FILATOV, A.; GERSHUNY, D.; KIL, I. *Handwritten word recognition – the approach proved by practice*. IWFHR'98, Taejon, Korea, august, 1998, p.99-122.

- [EL YACOUBI et al., 1994] EL YACOUBI, A.; BERTILLE, J.M.; GILLOUX, M. *Towards a more effective handwritten word recognition system*. IWFRH'94, 1994, p.378-385.
- [EL YACOUBI, 1996] EL YACOUBI, A. *Modélisation Markovienne de l'écriture manuscrite. Application à la reconnaissance des adresses postales*. Thèse de doctorat, Université de Rennes, France, 1996. 307p.
- [EL YACOUBI et al., 1999] EL YACOUBI, A., GILLOUX, M., SABOURIN, R., SUEN, C.Y. *Unconstrained handwritten word recognition using hidden markov models*. IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol.2, no.8, 1999, p.752-760.
- [FACON, 1995] FACON, J. *A morfologia matemática através de exemplos*. 2ª. SBAI-Simpósio Brasileiro de Automação Inteligente. CEFET-PR. Curitiba, set. 1995.
- [FARAG, 1979] FARAG, R. *Word-level recognition of cursive script*. IEEE Transactions on Computers. C-28(2). 1979, p.172-175.
- [FAVATA & SRIHARI, 1992] FAVATA, J. & SRIHARI, S. *Recognition of general handwritten words using a hypothesis generation and reduction methodology*. Proc. United State Postal Service Advanced Technology Conference. Washington DC, nov. 1992, p.237-253.
- [FILATOV et al, 1995] FILATOV, A.; GITIS, A.; KIL, I. *Graph-based Handwritten Digit String Recognition*. Montreal, ICDAR'95, IEEE 1995, p.845-848.
- [FREITAS et al, 2000a] FREITAS, C. O. A.; MORITA, M.; SOARES, L. E. O.; JUSTINO, E.; EL YACOUBI, A.; LETHELIER, E.; BORTOLOZZI, F.; SABOURIN, R. *Brazilian Bank Check Databases*. XXVI Conferência Latino-Americana de Informática. CLEI'2000, Mexico, 2000, p.134-152.
- [FREITAS et al., 2000b] FREITAS, C. O. A.; EL YACOUBI, A.; BORTOLOZZI, F.; SABOURIN, R. *Brazilian Bank Check Handwritten Legal Amount Recognition*. 13º Brazilian Symposium on Computer Graphics and Image Processing. SIBGRAPI2000, Brasil, 2000, p.97-104.
- [FREITAS et al., 2000c] FREITAS, C. O. A.; EL YACOUBI, A.; BORTOLOZZI, F.; SABOURIN, R. *Isolated Word Recognition in Brazilian Bank Check Legal Amounts*. 4º Workshop on Document Analysis and Systems. DAS2000, Brasil, 2000, p.279-290.
- [FREITAS et al., 2001] FREITAS, C. O. A.; BORTOLOZZI, F.; SABOURIN, R. *Handwritten word recognition: an approach based on mutual information for feature set validation*. Accepted to Sixth International Conference on Document Analysis and Recognition, ICDAR'2001, september, Seattle, Washington, USA, 2001.

- [GILLIES, 1992] GILLIES, A.M. Cursive word recognition using hidden markov models. Proc. of 5th USPS Advanced Technology Conference, 1992, p.557-562.
- [GILLOUX et al., 1993] GILLOUX, M.; LEROUX, M.; BERTILLE, J.M. *Strategies for handwritten words recognition using hidden markov models*. Proc. of the Second International Conf. On Document Analysis and Recognition. ICDAR'93, 1993, p.299-304.
- [GILLOUX, 1994] GILLOUX, M. *Hidden markov models in handwriting recognition*. IWFHR'94, 1994, p.264-288.
- [GILLOUX et al, 1995] GILLOUX, M.; LEROUX, M.; BERTILLE, J.M. *Strategies for cursive script recognition using hidden markov models*. Machine Vision and Applications, Springer-Verlog, Vol.8, 1995, p.197-205.
- [GRANDIDIER et al., 2000] GRANDIDIER, F., SABOURIN, R., SUEN, C.Y., GILLOUX, M. *Une nouvelle stratégie pour l'amélioration des jeux de primitives d'un système de reconnaissance de l'écriture*. CIFED2000, França, may, 2000, p.210-228.
- [GUILLEVIC, 1995] GUILLEVIC D. *Unconstrained Handwriting Recognition Applied to the Processing of Bank Cheques*. Thesis of Doctor's Degree in the Department of Computer Science at Concordia University.Canada.1995. 183p.
- [HAN & SETHI, 1997] HAN, K. & SETHI, I.K. *An off-line cursive handwritten word recognition system and its application to legal amount interpretation*. International Journal of Pattern Recognition and Artificial Intelligence. Vol.11, No. 5, 1997, p.757-770.
- [HARMON, 1962] HARMON, L.D. *Handwriting reader recognizes whole words*. Eletronics, 35. 1962, p.29-31.
- [HEUTTE, 1994] HEUTTE, L. *Reconnaissance de caractères manuscrits: application à la lecture automatique des chèques et des enveloppes postales*. Docteur Thèse de L'Université de Rouen. Rouen, France, dec., 1994. 239p.
- [HEUTTE et al., 1998] HEUTTE, L.; PAQUET, T.; MOREAU, J.V.; LECOUTIER, Y.; OLIVER, C. *A strutural / statistical feature based vector for handwritten character recognition*. Pattern Recognition Letters 19, 1998, p.629-641.
- [HULL & FERINCH, 1993] HULL,J.J. & FERINCH, R.K. *Large database organization for document images*, In: Fundamentals in Handwriting Recognition, Edited by Sabastiano Impedovo, Nato ASI Series, Vol. 124, 1993, p.397-414.
- [HULL, 1994] HULL, J.J. *A database for handwritten text recognition research*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.16, No. 5, 1994, p.550-554.

- [HULL, 1988] HULL, J.J. *A computational theory of visual word recognition*. PhD Thesis, University of New York at Buffalo, USA, 1988.
- [HUANG et al., 1990] HUANG, X.D.; ARIKI, Y.; JACK, M.A. Hidden Markov models for speech recognition. Information Technology Series; 7. Edinburgh University Press. 1990, 276p.
- [IMPEDOVO et al., 1997] IMPEDOVO, S.; WANG, P.S.P.; BUNKE, H. *Automatic bankcheck processing*. Series in Machine Perception and Artificial Intelligence. World Scientific Publishing Co. Pte. Ltd., 1997, 381p.
- [KATO, 1990] KATO, M. *O aprendizado da leitura*. São Paulo: Martins Fontes, 1990, 121p.
- [KIM, 1996] KIM, G. *Recognition of offline handwritten words and its extension to phrase recognition*. PhD Thesis, University of New York at Buffalo, USA, mar., 1996. 138p.
- [KIM & GOVINDARAJU, 1995] KIM, G.; GOVINDARAJU, V. *Handwritten word recognition for real-time applications*. International Conference on Document Analysis and Recognition (ICDAR), Vol.1, 1995, p.24-27.
- [KIM & GOVINDARAJU, 1997] KIM, G.; GOVINDARAJU, V. *A lexicon driven approach to handwritten word recognition for real-time applications*. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.9, No.4, 1997.
- [KIMURA et al., 1993] KIMURA, F.; SHRIDHAR, M.; CHEN, Z. *Improvements of a lexicon directed algorithm for recognition of unconstrained handwritten words*. International Conference on Document Analysis and Recognition. ICDAR'95. vol 1, aug., 1995, p.24-27.
- [KNERR et al., 1996] KNERR, S.; ANISIMOV, V.; BARET, O.; PRICE, D. & SIMON, J.C. *The A2iA recognition system for handwritten checks*. A2iA Technical Report, No.11, 1996. 49p.
- [KOERICH, 1997] KOERICH, A. L. *Uma proposta para o processamento de imagens de cheques bancários: extração e armazenamento de informações*. Dissertação de mestrado na Universidade Estadual de Campinas. Campinas, SP, fev., 1997, 190p.
- [KOSKINEN et al, 1994] KOSKINEN, L.; HUTTUNEN, H.; ASTOLA, J. *Text enhancement method based on soft morphological filters*. SPIE - Document Recognition, Vol. 2181, 1994, p.243-253.
- [LECOLINET, 1990] LECOLINET, E. *Segmentation d'images de mots manuscrits: application à la lecture de chaînes de caractères majuscules alphanumériques et à la lecture de l'écriture cursive*. Thèse de doctorat, Université Pierre et Marie Curie (Paris VI), France. 1990. 283p.

- [LEPLUMEY et al., 1995] LEPLUMEY, I.; CAMILLERAPP, J.; QUEGUINER, C. *Kalman Filter Contributions Towards Document Segmentation*. Montreal, ICDAR'95, IEEE, 1995, p.765-769.
- [LEROUX et al., 1991] LEROUX, M.; SALOME, J.C.; BADARD, J. *Strategie de controle pour la reconnaissance du montant litteral des cheques postaux*. 1991, p.695-706.
- [LI & YEUNG, 1997] LI, X. & YEUNG, D.Y. *On-line handwritten alphanumeric character recognition using dominant points in strokes*. Pattern Recognition, Vol. 30, No. 1, 1997, p.31-44.
- [LIU & SRIHARI, 1994] LIU, Y.; SRIHARI, S.N. *Document image binarization based on texture analysis*. SPIE, Vol. 2181, 1994.
- [LU & SHRIDHAR, 1996] LU, Y. & SHRIDHAR, M. *Character segmentation in handwritten words – an overview*. Pattern Recognition, Vol. 29, No. 1, 1996, p.77-96.
- [MACHADO & TEIXEIRA, 1996] MACHADO, A.M.B. & TEIXEIRA, E.A. *Aprendendo leitura dinâmica*. São Paulo: Makron Books, 1996, 87p.
- [MADHVANATH & GOVINDARAJU, 1992] MADHVANATH, S.; GOVINDARAJU, V. *Using holistic features in handwritten word recognition*. Proc. United State Postal Service Advanced Technology Conference. Washington DC, Vol.1, nov. 1992, p.183-199.
- [MADHVANATH & GOVINDARAJU, 1998] MADHAVANATH, S. & GOVINDARAJU, V. *Preceptual features for off-line handwritten word recognition: a framework for prediction, representation and matching*. Advances in Pattern Recognition, august, 1998, p.524-531.
- [MARTI & BUNKE, 1998] MARTI, U.V. & BUNKE, H. *Towards general cursive script recognition*. Proc. of Sixth International Workshop on Frontiers in handwriting Recognition - IWFHR, Korea, 1998, p.379-388.
- [MOHAMED, 1995] MOHAMED, M.A. *Handwritten word recognition using generalized hidden markov models*. PhD Thesis, University of Missouri-Columbia, USA, 1995.
- [MONTOLIU, 1995] MONTOLIU, L. *Architecture multi-agents et réseaux connexionnistes. Application à la lecture de chèques manuscrits*. Thèse de doctorat, Ecole Polytechnique, France, 1995. 233p.
- [MOREAU et al., 1991] MOREAU, J.V.; PLESSIS, B.; BOUGEOIS, J.L.P. *A postal check reading system*. Proc. of ICDAR '91, France, 1991, p.758-766.

- [LEPLUMEY et al., 1995] LEPLUMEY, I.; CAMILLERAPP, J.; QUEGUINER, C. *Kalman Filter Contributions Towards Document Segmentation*. Montreal, ICDAR'95, IEEE, 1995, p.765-769.
- [LEROUX et al., 1991] LEROUX, M.; SALOME, J.C.; BADARD, J. *Strategie de controle pour la reconnaissance du montant litteral des cheques postaux*. 1991, p.695-706.
- [LI & YEUNG, 1997] LI, X. & YEUNG, D.Y. *On-line handwritten alphanumeric character recognition using dominant points in strokes*. Pattern Recognition, Vol. 30, No. 1, 1997, p.31-44.
- [LIU & SRIHARI, 1994] LIU, Y.; SRIHARI, S.N. *Document image binarization based on texture analysis*. SPIE, Vol. 2181, 1994.
- [LU & SHRIDHAR, 1996] LU, Y. & SHRIDHAR, M. *Character segmentation in handwritten words – an overview*. Pattern Recognition, Vol. 29, No. 1, 1996, p.77-96.
- [MACHADO & TEIXEIRA, 1996] MACHADO, A.M.B. & TEIXEIRA, E.A. *Aprendendo leitura dinâmica*. São Paulo: Makron Books, 1996, 87p.
- [MADHVANATH & GOVINDARAJU, 1992] MADHVANATH, S.; GOVINDARAJU, V. *Using holistic features in handwritten word recognition*. Proc. United State Postal Service Advanced Technology Conference. Washington DC, Vol.1, nov. 1992, p.183-199.
- [MADHVANATH & GOVINDARAJU, 1998] MADHAVANATH, S. & GOVINDARAJU, V. *Preceptual features for off-line handwritten word recognition: a framework for prediction, representation and matching*. Advances in Pattern Recognition, august, 1998, p.524-531.
- [MARTI & BUNKE, 1998] MARTI, U.V. & BUNKE, H. *Towards general cursive script recognition*. Proc. of Sixth International Workshop on Frontiers in handwriting Recognition - IWFHR, Korea, 1998, p.379-388.
- [MOHAMED, 1995] MOHAMED, M.A. *Handwritten word recognition using generalized hidden markov models*. PhD Thesis, University of Missouri-Columbia, USA, 1995.
- [MONTOLIU, 1995] MONTOLIU, L. *Architecture multi-agents et réseaux connexionnistes. Application à la lecture de chèques manuscrits*. Thèse de doctorat, Ecole Polytechnique, France, 1995. 233p.
- [MOREAU et al., 1991] MOREAU, J.V.; PLESSIS, B.; BOUGEOIS, J.L.P. *A postal check reading system*. Proc. of ICDAR '91, France, 1991, p.758-766.

- [MORITA, 1998] MORITA, M.E. *Estudo para a melhoria da correção da inclinação da linha de base de palavras manuscritas*. Dissertação de Mestrado, Curitiba: CPGEI/CEFET-PR, 1998, 137p.
- [NESCHEN, 1995] NESCHEN, M. *Vector quantization classifiers for handwritten character recognition*. Proc. of the ZEUS-95 Workshop, Sweden, may, 1995, p.1-14.
- [OLIVEIRA, 1998] OLIVEIRA, L.E.S. *Estudo sobre a extração das estruturas lógica e física a partir de imagens de cheques bancários brasileiros*. Dissertação de Mestrado, Curitiba: CPGEI/CEFET-PR, 1998, 120p.
- [OLLIVIER, 1999] OLLIVIER, D. *Une approche économisant les traitements pour reconnaître l'écriture manuscrite: application à la reconnaissance des montants littéraux de chèques bancaires*. Thèse de doctorat, Université de Paris XI Orsay, France, 1999. 251p.
- [OTSU, 1979] OTSU, N. *A threshold selection method from gray-level histograms*. IEEE Transactions Systems, Man. and Cybernetics. SMC 9, Vol.1, 1979, p.63-66.
- [PAQUET & LECOURTIER, 1993] PAQUET, T. & LECOURTIER, Y. *Automatic reading of the literal amount of bank checks*. Machine Vision and Application, No.3, 1993, p.151-162.
- [PARKER, 1997] PARKER, J.R. *Algorithms for Image Processing and Computer Vision*. Ed. John Wiley & Sons, Inc. 1997, p.310-315.
- [PARRAMÓN, 1991] PARRAMÓN, J.M. *Como dibujar letras y logotipos*. Barcelona: Parramón Ediciones, 1991, 112p.
- [RABINER & JUANG, 1993] RABINER, L. JUANG, B.H. *Fundamentals of speech recognition*. Prentice Hall Inc., London, UK. 1993, 506p.
- [ROSA, 1991] ROSA, V.S. *Letras & cartazes*. Porto Alegre: Mercado Aberto, 1991, 79p.
- [SANTOS, 1997] SANTOS, J.E.B. *Estudo sobre métodos e técnicas para a segmentação de imagens de cheques bancários*. Dissertação de Mestrado, Curitiba: CPGEI/CEFET-PR, 1997, 74p.
- [SENIOR & ROBINSON, 1998] SENIOR, A.W. & ROBINSON, A.J. *An off-line cursive handwriting recognition system*. IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol.20, No.3, mar.1998, p.309-321.
- [SAION & BELAÏD, 1997] SAION, G. & BELAÏD, A. *High performance unconstrained word recognition system combining HMMs and markov Random Fields*. International Journal of Pattern Recognition and Artificial Intelligence, Vol. 11, No. 5, 1997, p.771-787.

- [SCHOMAKER & SEGERS, 1998] SCHOMAKER, L. & SEGERS, E. *A method for the determination of features used in human reading of cursive handwriting*. IWFHR'98, 1998, p.157-168.
- [SOLO, 1989] SOLO, D.X. *Bold script alphabets. 100 complete fonts*. New York: Dover Publications Inc., 1989, 100p.
- [STRATHY, 1993] STRATHY, N.W. *A method for segmentation of touching handwritten numerals*. Master's thesis, Concordia University, Montreal-Canada, 1993, 120p.
- [SUEN & LI, 1997] SUEN, C.Y. & LI, Z.C. *Crucial combinations for the recognition of handwritten letters*. BSDIA'97, Curitiba-PR, nov. 1997, p.19-39.
- [SUEN, 1998] SUEN, C.Y. *Réflexions sur la reconnaissance d'écriture cursive*. CIFED'98. Quebec, Canada, may, 1998, p.1-8.
- [TAPPERT et al., 1990] TAPPERT, C.C.; SUEN, C.Y.; WAKAHARA, T. *The state of art in on-line handwriting recognition*. IEEE Trans. on Pattern Analysis and Machine Intelligence. aug. 1990, p.787-808.
- [TEBEROSKY & TOLCHINSKY, 1996] TEBEROSKY, A. & TOLCHINSKY, L. *Além da alfabetização. a aprendizagem fonológica, ortográfica, textual e matemática*. trad: Stela Oliveira. São Paulo: Editora Ática, 1996, 295p.
- [TREVISAN, 1992] TREVISAN, E.M.C. *Leitura: coerência e conhecimento prévio*. Editora da UFSM, Santa Maria. 1992, 117p.
- [TRIER et al., 1996] TRIER, O.D.; JAIN, A.K.; TAXT, T. *Feature extraction methods for character recognition - a survey*. Pattern Recognition, Vol.29, No. 4, 1996, p.641-662.
- [VIARD-GAUDIN et al., 1999] VIARD-GAUDIN, C.; LALLICAN, P.M.; KNERR, S.; BINTER, P. *The IRESTE On/OFF (IRONOFF) dual handwriting database*. International Conference on Document Analysis and Recognition. ICDAR'99, 1999, p.455-458.

Apêndice A

Estrutura dos Cheques Bancários Brasileiros

O cheque bancário brasileiro é um instrumento que os bancos colocam a disposição dos clientes para movimentar fundos em seu poder e efetuar pagamentos sem a necessidade do manuseio do papel moeda. Um cheque representa uma ordem incondicional em direito que: é emitida por uma pessoa em favor de outra; é assinado pela pessoa que emite; e exige que o banco pague na apresentação uma determinada quantia em dinheiro [BANCO DO BRASIL, 1983].

Este Apêndice apresenta a estrutura de um cheque, suas características em relação ao conteúdo e a disposição dos seus elementos componentes.

A.1. Modelos para Cheques: Distribuição das Informações e Preenchimento dos Campos

O Conselho Monetário Nacional instituiu um modelo único de cheque visando uma maior segurança e eficiência nos serviços que os bancos prestam à coletividade. A Figura A1 mostra a localização dos elementos dentro da área de um cheque bancário. A Figura A2 apresenta o modelo padrão de preenchimento dos campos e áreas do cheque, segundo o padrão estabelecido pelas normas do Banco Central do Brasil.

A.2. Dimensões do Cheque

A resolução Nº 855 editada pelo Banco Central [BAN83] estabelece que os cheques devem ter como dimensões:

- comprimento de 175 mm, com tolerância de mais ou menos 1mm;
- largura de 80 mm, com tolerância de 1mm para mais e de até 4 mm para menos.

Desta forma, um cheque digitalizado em 300 dpi pode assumir uma dimensão de no mínimo 2057 pixels de comprimento e 897 pixels de largura; e uma dimensão de no máximo 2079 pixels de comprimento e 957 pixels de largura, como representado na Figura A3.

Número da Compensação, Número do Banco, Número da Agência, Número da Conta, Número do Cheque e Dígitos Verificadores		Valor Numérico
Valor por Extenso		
Nome do Favorecido		
Logotipo do Banco	Local e Data de Emissão	
Nome do Banco (Nome, Número, Endereço e Telefone da Agência)	Assinatura, Nome do Cliente, Número do CPP/CGC e outras informações	
Linha MICR CMC-7		

Figura A1: Modelo para distribuição das informações [KOERICH, 1997]

IT Estado Recibido Recibido Info-credor Saldo	COMP BANCO AG CI CONTA CI CHEQUE No CI DE
	Número do cheque e quantidade
	A _____ de _____ de 19 ____
	Campo restrito a identificação do sacado
	Área destinada a magnetização

Figura A2: Modelo padrão de preenchimento dos campos e áreas de um cheque [OLIVEIRA, 1998]

A.3. Elementos Componentes

Os cheques podem ser divididos em quatro faixas compreendendo os dados descritos como segue e apresentadas na Figura A4:

a) Primeira Faixa:

- número da Câmara de Compensação (COMP);
- número da inscrição do Banco no Serviço de Compensação (BANCO);
- código da agência sacada (AG);
- posição facultativa: dígito verificador correspondente ao código da agência sacada();

A.4. Formulários dos Cheques

Os formulários dos cheques podem ser impressos em qualquer cor, com tinta indelével e de tonalidade firme, sendo permitido o uso de impressões por perfuração ou filigrama, desde que não estejam localizados na faixa de magnetização e não produzam mossas ou rebarbas que prejudiquem a leitura mecânica do cheque.

Número da Compensação, Número do Banco, Número da Agência, Número da Conta, Número do Cheque e Dígitos Verificadores		Valor Numérico
Valor por Extenso		
Nome do Favorecido		
Logotipo do Banco	Local e Data de Emissão	
Nome do Banco (Nome, Número, Endereço e Telefone da Agência)	Assinatura, Nome do Cliente, Número do CPP/CGC e outras informações	
Linha MICR CMC-7		

Figura A4: Faixas de um cheque bancário brasileiro [OLIVEIRA, 1998]

Os cheques podem ser confeccionados em folhas planas ou em formulários contínuos, devendo, nesse último caso, serem destacados por guilhotina de modo que se obtenham bordas retas e lisas nas extremidades inferior e direita, utilizadas como referência no processo de leitura mecânica. A confecção dos cheques bancários brasileiros devem seguir rigorosamente as especificações do documento “Modelo-Padrão do Cheque”, constante do CADOC (Cadastro de Documentos) como modelo n°. 38058-0 (Res. 885), do Banco Central do Brasil.

Os cheques bancários brasileiros apresentam grande diversidade de cores, desenhos e estilos. A estrutura de *layout* de um cheque é considerada complexa, uma vez que caracteriza-se pela presença de desenhos e figuras ilustrativas que fazem parte do padrão de fundo do cheque. O padrão de fundo do cheque pode assumir várias cores e padrões, porém, os elementos que são impressos sobre o fundo, como linhas e seqüência de caracteres, apresentam um posicionamento ordenado, sendo impressos em uma cor única, geralmente, em preto.

Alguns trabalhos podem ser citados no que se refere ao estudo da estrutura do cheque bancário brasileiro [OLIVEIRA, 1998], bem como na extração das informações pré-impressas e padrão de fundo [KOERICH, 1997].

Apêndice B

Comparativo de Métodos de Reconhecimento de Palavras Manuscritas

A comparação de resultados não é uma tarefa simples, porém é de suma importância para posicionar o trabalho realizado diante da comunidade científica que pesquisa na área de reconhecimento de palavras manuscritas em cheques bancários.

A dificuldade na realização da comparação de resultados está, em primeiro plano, no fato que não se dispõe de trabalhos em idioma português, somente inglês e francês. Em segundo plano, no fato de que os trabalhos são realizados com bases de dados diferentes, sendo a comparação de difícil realização até mesmo entre trabalhos sobre a mesma base de dados. Quando os métodos de reconhecimento são os mesmos, por exemplo HMM, existem diferenças no treinamento, nos algoritmos aplicados, na topologia dos modelos, nos critérios de aceitação e rejeição, entre outros. Existe, portanto, uma grande variedade de parâmetros nas condições dos testes de reconhecimento realizados por diferentes autores.

O presente Capítulo apresenta os resultados obtidos com o reconhecimento de palavras isoladas do extenso manuscrito de cheques bancários brasileiros e realiza uma comparação com outros trabalhos no mesmo contexto.

A seleção de outros trabalhos foi realizada com base no contexto de cheques bancários, na abordagem Global adotada para o reconhecimento das palavras, nas primitivas extraídas e nos métodos utilizados para o reconhecimento. Apresenta-se, inicialmente, as características das bases de dados sobre as quais os trabalhos foram realizados, os procedimentos de pré-processamento aplicados, as primitivas extraídas, os métodos de treinamento e reconhecimento adotados e os resultados obtidos.

Tabela B.1-a: Características das Bases de Dados - Palavras do Idioma Inglês

BD Inglês	Côte	Guillevic
Origem da base de dados	CENPARMI	CENPARMI
Nº. de cheques	2.500	2.500
Nº. de palavras	7.837	7.837
Nº. de escritores	800	800
Base de treinamento	184	5.322
Base de validação	Não utiliza	Não utiliza
Base de testes	2.929	2.515
Tamanho do léxico	32	32
Número de letras por palavra	3 a 9	

Tabela B.1-b: Características das Bases de Dados - Palavras do Idioma Francês

BD Francês	Avila	Guillevic	Ollivier
Origem da base de dados	SRTP	CENPARMI	LIX / SRTP
Nº. de cheques	Não indicado	1.861	1.705 / 11.066
Nº. de palavras	6.771	6.135	16.497 / 44.901
Nº. de escritores	Omni-escritor	600	Omni-escritor
Base de treinamento	3.623	1.345	8.113 / 40.400
Base de validação	Não utiliza	Não utiliza	Não utiliza
Base de testes	3.148	516	8.384 / 4.490
Tamanho do léxico	27	25	29
Número de letras por palavra	2 a 9		

Tabela B.1-c: Características das Bases de Dados - Palavras do Idioma Português

BD Português	Freitas
Origem da base de dados	PUCPR/PPGIA/LARDOC
Nº. de cheques	2.016
Nº. de palavras	11.948
Nº. de escritores	2.016
Base de treinamento	7.146
Base de validação	2.427
Base de testes	2.375
Tamanho do léxico	39
Número de letras por palavra	2 a 12

Observa-se que o léxico em português apresenta como fator de maior complexidade o seu próprio tamanho (39 palavras). Isto, torna-se evidente através da Tabela B.2 que apresenta os léxicos nos três diferentes idiomas a serem consideradas no presente Capítulo. Nota-se que o léxico em português apresenta uma palavra para cada valor numérico a ser grafado por extenso. Este fato não ocorre com as palavras em inglês para valores entre 200 a 900, pois estes valores são grafados por extenso

utilizando-se uma composição de palavras (“two hundred” – “nine hundred”). Por outro lado, observa-se no léxico em francês que este mesmo tipo de composição ocorre para os valores numéricos entre 17 a 19, 70 a 90, e também, para os valores entre 200 e 900. Em francês, a grafia do valor 90 é expressa como: “quatre vingt dix”. Estas observações explicam o fato da diferença entre os tamanhos dos léxicos considerados pelos diferentes autores.

Tabela B.2: Léxico em Português, Inglês e Francês

Valor Numérico	Português	Inglês	Francês
0	-	-	Zéro
1	Um, Hum	One	Un
2	Dois	Two	Deux
3	Três	Three	Trois
4	Quatro	Four	Quatre
5	Cinco	Five	Cinq
6	Seis	Six	Six
7	Sete	Seven	Sept
8	Oito	Eight	Huit
9	Nove	Nine	Neuf
10	Dez	Ten	Dix
11	Onze	Eleven	Onze
12	Doze	Twelve	Douze
13	Treze	Thirteen	Treize
14	Quatorze, Catorze	Fourteen	Quatorze
15	Quinze	Fifteen	Quinze
16	Dezesseis	Sixteen	Seize
17	Dezessete	Seventeen	-
18	Dezoito	Eighteen	-
19	Dezenove	Nineteen	-
20	Vinte	Twenty	Vingt
30	Trinta	Thirty	Trente
40	Quarenta	Forty	Quarante
50	Cinquenta, Cincoenta	Fifty	Cinquante
60	Sessenta	Sixty	Soixante
70	Setenta	Seventy	-
80	Oitenta	Eighty	-
90	Noventa	Ninety	-
100	Cem, Cento	Hundred	Cent
200	Duzentos	-	-
300	Trezentos	-	-
400	Quatrocentos	-	-
500	Quinhentos	-	-
600	Seiscentos	-	-
700	Setecentos	-	-
800	Oitocentos	-	-
900	Novocentos	-	-
1000	Mil	Thousand	Mille
Moeda Corrente	Reais, Real	Dollars	Francs, Frs
Fração	Centavos, Centavo	-	Centimes, Cts
Outras palavras	-	And, Only	et

Deve-se, também, levar em consideração a distribuição das palavras nas bases de dados utilizadas pelos diferentes autores na etapa de treinamento como mostrado na Tabela B.3. Observa-se que as palavras correspondentes aos valores numéricos de 10 a 19 apresentam uma baixa frequência independentemente da base de dados, com exceção da base de dados utilizada por [CÔTE, 1997]. Verifica-se também, uma alta frequência das palavras “dollars” – “francs” – “reais / real”. Isto, deve-se ao fato da obrigatoriedade da grafia da moeda corrente no país de origem dos cheques.

Na base de dados de treinamento do presente trabalho as palavras “mil”, “reais/real” e “centavos/centavo” representam 30,56% do total da base de dados de treinamento. Este fato pode ser explicado, uma vez que, as bases de dados utilizadas contam com extensos que foram grafados com base em valores numéricos que necessitam destas palavras (Capítulo 3).

Deve-se levar em consideração as características inerentes ao presente trabalho descritas no Capítulo 2, referentes aos seguintes fatores: tipo de aquisição das imagens, tipo de escrita, número de escritores, tamanho do léxico, qualidade das imagens, inclinação vertical, traços contíguos/desconectados, ortografia diferenciada e linhas guia e linhas de separação. Estes fatores sofrem influência da população de escritores que geram as amostras para as bases de dados, sobre as quais os métodos são desenvolvidos e testados. Como mostrado no Capítulo 2 a escrita cursiva é diferente para cada pessoa e, ainda, sofre influência do processo de alfabetização utilizado no país de origem.

Com base nestas características apresenta-se a seguir os procedimentos de pré-processamento aplicados nas imagens das palavras pelos diferentes autores.

B.2. Pré-Processamento das Palavras

Entende-se por pré-processamento as etapas que antecedem o reconhecimento de palavras manuscritas, ou ainda, as etapas que preparam as imagens para a extração de primitivas. Portanto, descreve-se as técnicas adotadas e/ou desenvolvidas para realização das etapas necessárias ao objetivo pretendido, estando organizado pelos seguintes itens: Procedimentos preliminares, Correção da linha de base, Correção da inclinação vertical dos caracteres e Normalização do corpo das palavras.

Tabela B.3: Distribuição das Palavras nas Bases de Dados de Treinamento

Léxico Inglês	Côte	Guillevic	Léxico Francês	Avila	Ollivier	Léxico Português	Freitas
-	-	-	Zéro	0,14	0,10	-	-
One	3,80	4,41	Un	0,55	0,80	Um, Hum	2,78
Two	4,35	3,67	Deux	5,74	5,20	Dois	2,92
Three	4,35	4,04	Trois	3,48	3,30	Três	2,81
Four	3,26	3,95	Quatre	7,26	7,20	Quatro	2,70
Five	2,72	3,95	Cinq	3,42	3,40	Cinco	2,78
Six	1,09	3,79	Six	1,41	2,20	Seis	2,53
Seven	2,17	4,05	Sept	1,49	2,10	Sete	2,24
Eight	2,17	3,62	Huit	1,63	2,30	Oito	2,55
Nine	2,72	3,67	Neuf	1,79	2,60	Nove	2,66
Ten	4,35	1,43	Dix	3,23	3,30	Dez	0,39
Eleven	2,72	1,41	Onze	0,30	0,30	Onze	0,41
Twelve	4,35	1,03	Douze	0,55	0,60	Doze	0,24
Thirteen	4,35	1,21	Treize	0,30	0,40	Treze	0,31
Fourteen	2,72	1,19	Quatorze	0,28	0,40	Quatorze	0,28
Fifteen	2,17	1,24	Quinze	0,77	0,90	Quinze	0,36
Sixteen	3,80	1,05	Seize	0,25	0,50	Dezesseis	0,24
Seventeen	1,63	1,23	-	-	-	Dezessete	0,29
Eighteen	3,26	1,20	-	-	-	Dezoito	0,28
Nineteen	1,63	1,13	-	-	-	Dezenove	0,36
Twenty	4,35	1,88	Vingt	7,26	6,70	Vinte	3,20
Thirty	4,35	1,72	Trente	2,24	2,20	Trinta	3,12
Forty	1,63	1,72	Quarante	2,01	1,90	Quarenta	2,83
Fifty	2,72	1,69	Cinquante	2,15	2,50	Cinquenta	3,15
Sixty	2,72	1,61	Soixante	3,12	3,90	Sessenta	2,98
Seventy	3,26	1,99	-	-	-	Setenta	3,09
Eighty	2,17	1,79	-	-	-	Oitenta	3,04
Ninety	2,72	1,43	-	-	-	Noventa	2,64
Hundred	2,17	15,41	Cent	20,62	17,10	Cem, Cento	2,36
-	-	-	-	-	-	Duzentos	2,09
-	-	-	-	-	-	Trezentos	1,97
-	-	-	-	-	-	Quatrocentos	1,82
-	-	-	-	-	-	Quinhentos	2,00
-	-	-	-	-	-	Seiscentos	1,83
-	-	-	-	-	-	Setecentos	2,07
-	-	-	-	-	-	Oitocentos	2,11
-	-	-	-	-	-	Novencentos	1,97
Thousand	4,35	6,21	Mille	3,51	4,70	Mil	10,79
Dollars	4,35	6,06	Francs, Frs	24,68	19,80	Reais, Real	10,59
And	4,35	10,17	et	0,80	2,30	-	-
Only	3,26	1,07	-	-	-	-	-
-	-	-	Centimes, Cts	1,02	3,30	Centavos, Centavo	9,18

A etapa de segmentação do extenso manuscrito em palavras não está apresentada no presente estudo, uma vez que esta segmentação está relacionada ao tratamento do extenso como um todo.

O procedimento de normalização das imagens não é normalmente utilizado pela maioria dos autores, tendo sido aplicado somente por [OLLIVIER, 1999]. O autor justifica a aplicação deste procedimento a normalização através de duas vantagens do procedimento: redução da quantidade de informação a tratar e redução da variabilidade das amostras. Portanto, a redução do número de pixels na imagem implica, para o trabalho de [OLLIVIER, 1999] com redes neurais, em:

- redução do tamanho do vetor a ser comparado pelo método de reconhecimento adotado e
- redução do número de neurônios e de conexões em uma rede neural.

Desde modo, [OLLIVIER, 1999] afirma ser possível reduzir os tempos de treinamento e reconhecimento das palavras manuscritas.

A Tabela B.4 apresenta os procedimentos que foram adotados pelos autores selecionados. A composição dos diferentes procedimentos está relacionada ao direcionamento do trabalho de cada um dos autores, bem como, as características intrínsecas das bases de dados utilizadas.

Para obtenção de todos os detalhes de cada um dos procedimentos aplicados pelos autores recomenda-se consultar os respectivos trabalhos.

Tabela B.4: Pré-Processamento de Palavras Manuscritas

Procedimentos	Avila	Côte	Guillevic	Ollivier	Freitas
Pré-processamento	Sim	Sim	Sim	Sim	Sim
Correção da linha de base	Sim	Não	Sim	Não	Não
Correção da inclinação vertical dos caracteres	Sim	Não	Sim	Não	Sim
Detecção do corpo das palavras	Sim	Sim	Sim	Não	Sim
Normalização do corpo das palavras	Sim	Não	Não	Não	Não
Normalização do tamanho das imagens	Não	Não	Não	Sim	Não
Esqueletização	Sim	Não	Não	Não	Sim ¹

¹ Somente para a extração das primitivas globais do tipo côncavo e convexo no corpo das palavras

B.3. Extração de Primitivas

Apresenta-se os conjuntos de primitivas selecionados pelos diferentes autores para serem extraídas das palavras manuscritas. Assim, pode-se considerar que diferentes tipos de formas precisam de diferentes tipos de primitivas, que por sua vez precisam de diferentes tipos de classificadores. Apesar de todos os trabalhos processarem palavras manuscritas no contexto de cheques bancários, diferentes primitivas foram selecionadas buscando-se conjuntos invariantes e que permitam a reconstrução, tal qual descrito no Capítulo 2 [TRIER et al., 1996].

As aplicações na área de reconhecimento de palavras manuscritas combinam os diferentes tipos de primitivas visando obter métodos mais robustos através de um conjunto de primitivas com alta representatividade das formas das palavras.

A Tabela B.5 apresenta as primitivas selecionadas pelos diferentes autores, sendo que a escolha das mesmas está diretamente relacionada com os métodos de treinamento e de reconhecimento adotados.

Ressalta-se que [AVILA, 1996] utiliza primitivas primárias para definição de um alfabeto de 12 traços. [OLLIVIER, 1999] utiliza como método de treinamento e reconhecimento as redes neurais, portanto seu trabalho descreve a arquitetura da rede neural adotada, não enfatizando a extração de primitivas.

Deve-se relacionar ainda, a abordagem adotada por cada autor no que se refere a aplicação ou não da segmentação das palavras em letras ou pseudo-letras. Todos os resultados mostrados no presente estudo dos autores selecionados adotam a abordagem global, ou seja, sem segmentação explícita das palavras em letras ou pseudo-letras. Alguns autores trabalharam também com a abordagem analítica, porém estes resultados não se encontram apresentados neste trabalho.

B.4. Metodologias de Reconhecimento

Os autores selecionados trabalham com os diferentes métodos de treinamento e reconhecimento das palavras manuscritas. Estes métodos encontram-se descritos no Capítulo 2 e foram empregados pelos diferentes conforme Tabela B.6.

Tabela B.5: Extração de Primitivas

Primitivas	Avila	Côte	Guillevic	Freitas
1) Perceptivas: ascendente, descendente e laço	X	X	X	X
2) Secundárias ou condicionais: barras do “t”, laços associados com ascendentes ou descendentes (“d” ou “p”)		X		
3) Concavidades e convexidades		X		X
4) Número de transições preto-branco sobre as linhas do corpo das palavras		X		
5) Comprimento da palavra			X	
6) Traços horizontais e verticais			X	
7) Diagonais leste e oeste			X	

Tabela B.6: Metodologias de Reconhecimento

Métodos	Avila	Côte	Guillevic	Ollivier	Freitas
Rede Neural		X		X	
HMM	X				X
Não-Paramétricos (kPPV)¹			X	X	

¹ kPPV significa Métodos dos k-Vizinhos mais Próximos

B.5. Resultados Obtidos

Os resultados obtidos pelos autores consideram todas as etapas e métodos anteriormente apresentados. Os resultados apresentados são referentes aos experimentos realizados com as bases de dados de testes e com a escolha da melhor opção de reconhecimento, ou seja, para o modelo que apresenta a maior probabilidade de reconhecimento.

As Tabelas B.7 e B.8 mostram os resultados obtidos pelos autores para o reconhecimento de palavras isoladas no contexto de cheques bancários. Na Tabela B.7 estão expostos os resultados obtidos pelos autores que trabalharam com as bases de dados de palavras do idioma inglês. As colunas denominadas *Guillevic ADS* e *Guillevic AD* correspondem aos resultados para dois conjuntos distintos de primitivas, denominados de ADS – conjunto formado por todas as primitivas e AD - conjunto formado somente por ascendentes e descendentes. Os resultados obtidos por [GUILLEVIC, 1995] utilizando o conjunto de primitivas ADS é melhor que quando empregado o conjunto AD. Observa-se também, que os resultados de [CÔTE, 1997] e

[GUILLEVIC, 1995] para o conjunto ADS são muito próximos. Isto, deve-se ao fato das primitivas globais selecionados pelos autores serem de mesma conceituação e formalismo de extração.

As palavras do léxico em inglês que apresentam maior grau de confusão para [GUILLEVIC, 1995] são as seguintes: “one” e “two”, “one” e “five”, “one” e “six”, “one” e “nine”. Para o conjunto de primitivas ADS as palavras com taxa de reconhecimento inferior a 50% são as seguintes: “sixteen”, “sixty”, “seven”, “seventeen”, “nineteen”, “ninety”. Observa-se que estas palavras possuem os mesmos prefixos entre si (“six”, “seven” e “nine”) mostrando que as primitivas extraídas não são suficientes para a representação dos sufixos (“teen” e “ty”).

Tabela B.7: Resultados Obtidos – Idioma Inglês

Palavras	Côte	Guillevic ADS	Guillevic AD
One	51,0	62,5	86,5
Two	81,0	77,5	23,6
Three	79,1	62,4	49,5
Four	82,2	65,5	60,0
Five	85,6	60,2	30,6
Six	30,1	70,5	36,8
Seven	42,2	52,6	54,6
Eight	76,6	87,6	72,4
Nine	46,8	63,0	43,5
Ten	58,5	52,9	52,9
Eleven	71,0	55,6	57,8
Twelve	73,5	57,1	42,9
Thirteen	59,6	58,8	35,3
Fourteen	74,2	56,7	43,3
Fifteen	72,0	59,5	40,5
Sixteen	81,0	39,1	25,0
Seventeen	69,0	45,5	45,5
Eighteen	79,2	74,1	63,0
Nineteen	81,8	36,7	48,3
Twenty	85,9	70,0	45,0
Thirty	63,5	73,3	55,6
Forty	76,9	58,7	58,7
Fifty	71,9	60,5	47,4
Sixty	66,7	37,5	37,5
Seventy	84,2	59,2	59,2
Eighty	84,1	86,0	81,4
Ninety	86,4	47,5	50,0
Hundred	78,2	95,1	95,1
Thousand	85,6	80,9	79,6
Dollars	84,6	72,5	73,2
And	71,6	90,4	81,6
Only	53,4	72,7	36,4
Média	73,6	72,6	63,9

Para o conjunto AD as palavras com taxa de reconhecimento inferior a 50% são as seguintes: “two”, “three”, “five”, “six”, “nine”, “twelve”, “thirteen”, “fourteen”, “fifteen”, “sixteen”, “seventeen”, “nineteen”, “twenty”, “fifty”, “sixty”, “ninety” e “only”. Este maior número de palavras com taxas inferiores a 50 % se traduz em uma taxa média para este conjunto AD inferior a obtida com o conjunto ADS. Estas palavras representam isoladamente parcelas aproximadamente de 1% da base de dados de treinamento, sendo este fato mais evidente nas palavras da classe “teen”.

Para [CÔTE, 1997] as palavras que apresentam maiores problemas de reconhecimento, com taxa inferior a 50%, são as seguintes: “six”, “seven” e “nine”.

Percebe-se que as dificuldades inerentes ao problema do reconhecimento de classes de palavras com prefixos e sufixos semelhantes independe do idioma do léxico estudado.

A Tabela B.8 apresenta os resultados obtidos com as bases de dados de palavras em idioma francês. [GUILLEVIC, 1995] apresenta somente os resultados das palavras isoladas obtidos com o conjunto AD (ascendentes e descendentes), o qual apresenta uma taxa média de reconhecimento de 78,3%, uma vez que, o mesmo revela uma significativa melhora quando comparados com os resultados obtidos com o conjunto ADS (todas as primitivas), média de 71,9%. [GUILLEVIC, 1995] comenta que o fato do conjunto AD ser mais representativo deve-se a dois fatores: primeiro que o tamanho do léxico de palavras francesas (25 palavras) é menor do que o léxico de palavras inglesas (32 palavras), e segundo, que as palavras em francês possuem mais ascendentes e descendentes distintos entre todas as palavras do léxico.

[GUILLEVIC, 1995] mostra que o conjunto AD resulta em taxas de reconhecimento para as palavras “huit” e “dix” inferiores a 30%, taxas estas muito abaixo das taxas das demais palavras. [GUILLEVIC, 1995] explica que a palavra “huit” é confundida com a palavra “cent” e que a palavra “dix” é confundida com as palavras “six” e “deux”.

[AVILA, 1996] não apresenta a matriz de confusão para as palavras estudadas, porém, observa-se que as palavras “zéro”, “six”, “sept”, “huit”, “treize”, “quatorze”, “quinze”, “seize”, “trente”, “centimes” e “et” apresentam taxas de reconhecimento inferior a 30%. As palavras com maior grau de confusão entre si são: “deux” confundida com “dix” e com “trois”, “quarante” confundida com “quatre”, e ainda, “six” confundida com a palavra “dix”.

Os resultados obtidos por [OLLIVIER, 1999] mostram que os dois métodos aplicados, apesar de serem diferentes, traduzem resultados semelhantes, com uma grande diferença de tempo de processamento: as redes neurais são mais rápidas. As redes neurais apresentam grande grau de confusão entre as seguintes palavras: “seize” e “treize”, “quatorze” e “quinze”, “six” e “dix”, “quarante” e “soixante”, “cinquante” e “soixante”, “deux” e “francs”. O método dos kPPV apresentam grande grau de confusão entre as seguintes palavras: “cinquante” e “quarante”, “soixante” e “quarante”, “cent” e “et”, “deux” e “francs”.

Tabela B.8: Resultados Obtidos – Idioma Francês

Palavras	Avila	Guillevic AD	Ollivier Redes Neurais	Ollivier kPPV
Zéro	20,0	-	9,0	3,0
Un	33,3	86,7	43,0	60,0
Deux	50,2	73,1	67,0	61,0
Trois	36,2	76,0	52,0	40,0
Quatre	75,2	95,2	78,0	83,0
Cinq	51,6	92,2	67,0	69,0
Six	14,6	64,4	42,0	36,0
Sept	23,5	94,0	40,0	24,0
Huit	13,7	27,4	58,0	32,0
Neuf	41,7	71,9	64,0	66,0
Dix	67,3	18,2	67,0	66,0
Onze	25,0	79,5	30,0	24,0
Douze	26,7	82,9	29,0	29,0
Treize	11,1	73,0	18,0	8,0
Quatorze	9,1	97,8	37,0	32,0
Quinze	17,1	86,7	51,0	21,0
Seize	11,1	51,4	29,0	13,0
Vingt	68,3	85,2	80,0	68,0
Trente	21,3	64,4	50,0	36,0
Quarante	45,3	91,2	64,0	77,0
Cinquante	56,6	95,0	77,0	83,0
Soixante	59,3	84,5	68,0	67,0
Cent	81,7	87,3	87,0	87,0
Mille	59,8	80,3	74,0	66,0
Francs	78,0	89,2	88,0	88,0
Centimes	9,6	-	45,0	51,0
Et	21,4	95,2	63,0	89,0
Média	62,2	78,3	75,0	75,0

A seguir apresenta-se uma discussão dos resultados obtidos evidenciando-se as principais diferenças entre os experimentos apresentados.

B.6. Análise dos Resultados

Menciona-se no início deste Capítulo que a comparação de resultados não é uma tarefa simples devido a diversos fatores. Apresenta-se os resultados obtidos pelos diferentes autores selecionados e pode-se perceber que diferentes características de cada estudo em particular.

A principal diferença são as origens das palavras: inglesa, francesa e portuguesa. Pode-se dizer, num primeiro momento que o francês e português possuem a mesma origem, ou seja, são idiomas latinos. Desde modo, pode-se esperar que exista uma semelhança entre as palavras, o que de fato ocorre. Por outro lado, o francês expressa alguns valores numéricos por escrito utilizando mais de uma palavras (90 – “quatre vingt dix”), o que torna o léxico a ser reconhecido em cada uma das línguas diferente, considerando-se principalmente o tamanho do léxico. Por sua vez, o inglês é um idioma de origem não latina tornando a comparação de resultados ainda mais difícil.

As palavras da classe “dezena” apresentam taxas de reconhecimento baixas independente do idioma do léxico, os resultados obtidos estão relacionados com a frequência destas palavras nas bases de dados e, especialmente, na base de dados de treinamento.

O reconhecimento das palavras curtas enfrenta problemas independentemente do idioma do léxico, revelando que as observações de [SCHOMAKER & SEGER, 1998] são verdadeiras. Assim, o reconhecimento de palavras curtas exige um maior número de informações a serem extraídas.

O reconhecimento de palavras longas da classe “entos”, para o caso do léxico em idioma português, revela não somente a similaridade com a classe de palavras “enta” como também confirma as observações de [SCHOMAKER & SEGER, 1998]. Para os autores, palavras longas necessitam de maiores informações no final das mesmas para serem reconhecidas. Ressalta-se que esta classe de palavras não existe nos léxicos dos idiomas em francês e inglês, uma vez que os valores numéricos são grafados pela combinação de outras palavras existentes no léxico.

Assim, levando-se em consideração todos estes fatores intrínsecos a cada idioma, a cada léxico e a cada trabalho realizado, apresenta-se na Tabela B.9 as médias dos resultados obtidos pelos diferentes autores. Neste Tabela % é a taxa de

reconhecimento, **TL** é o tamanho do léxico e **BT** é a dimensão da base de dados de treinamento utilizada nos experimentos realizados.

Sabe-se, ainda, que muitos outros trabalhos poderiam ter sido selecionados para realização da comparação apresentada por este estudo. Ressalta-se a importância de alguns sistemas desenvolvidos no mesmo contexto do presente trabalho, descritos em [IMPEDOVO et al., 1997], sendo estes os seguintes:

- O sistema A2iA apresentado por Knerr et al.,
- O sistema proposto por Dimauro et al.,
- O sistema proposto por Leroux et al. e
- O sistema proposto por Saion & Belaïd.

A Tabela B.10 mostra os resultados obtidos com os experimentos realizados pelos respectivos sistemas envolvendo o reconhecimento de palavras isoladas do extenso manuscrito de cheques bancários em idioma francês e italiano.

Tabela B.9: Resumo dos Resultados Obtidos

Autor	Inglês			Francês			Português		
	%	TL	BT	%	TL	BT	%	TL	BT
Côte	73,6	32	184	-			-		
Guillevic ADS	72,6	32	5.322	-			-		
Guillevic AD	63,9	32	5.322	-			-		
Avila	-			62,2	27	3.623	-		
Guillevic AD	-			78,3	25	1.345	-		
Ollivier - Redes Neurais	-			75,0	29	40.400	-		
Ollivier - kPPV	-			75,0	29	40.400	-		
Freitas - PP	-			-			56,4	39	7.136
Freitas - PPCC	-			-			64,8	39	7.136
Freitas - PPCCR	-			-			67,7	39	7.136
Freitas - PPCCR-FR	-			-			70,6	39	7.136

Tabela B.10: Resultados de Sistemas de Reconhecimento

Autores	Idioma	%	TL	BT
Sistema A2iA	Francês	77,1	26	10.000
Dimauro et al.	Italiano	69,0	50	*
Leroux et al.	Francês	74,0	26	23.066
Saion & Belaïd	Francês	72,4	26	4.653

* apresenta somente o tamanho da base de dados de testes