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
2002

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
2002

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 *Baum-Welch* 125
4.2.5.4. A Validação do Treinamento: Algoritmo de Validação Cruzada..... 132
4.2.5.5. O Reconhecimento: Algoritmo *Forward* 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 Classficação: 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 *scanner* [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) *threshold* = 217 (método anisotropia) c) *threshold* = 110 (método OTSU)..... 32
Figura 2.15 Inclinação: a) da linha de base (*skew*) b) vertical dos caracteres (*slant*) 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 *grid*, 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 *soft*: 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 externo..... 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 freqüência 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

bmp	*bit map*
CDVDHMM	*Continuous Density Variable Duration Hidden Markov Models*
CEDAR	*Center of Excellence in Document Analysis Recognition*
CENPARMI	*Center of Pattern Analysis and Machine Intelligence*
dpi	*dots per inch*
ETS/Canada	*École de Technologie Supérieure – Montréal - Canada*
jpeg	*Joint Photographers Expert Group*
HMM	*Hidden Markov Models*
HWR	*Handwritten Word Recognition*
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	*Magnetic Ink Character Recognition*
NIST	*National Institute Standards and Technology*
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
tiff	*Tagged Image Format File*
k	Número de pixels vizinhos mais próximos
N	Número máximo de estados do modelo de *Markov*
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 *Markov*
Q	Conjunto de estados do modelo de *Markov*
V	Conjunto de símbolos observáveis
A	Matriz de transição entre os estados da cadeia de *Markov*
B	Matriz de probabilidade de observação de símbolos condicionada pelos estados da cadeia de *Markov*
π	Conjunto de densidde de probabilidade inicial
T	Comprimento da seqüência de observações

t	Intervalo de tempo
O	Seqüência de observações
$\hat{\lambda}$	Modelo estinado
argmax(.)	Função que retorna o índice relacionado ao argumento que maximiza uma função
α	Probabilidade *Forward*
β	Probabilidade *Backward*
i, j, k	Valores inteiros
log P(.)	Logaritmo da probabilidade
w	Número de símbolos ou *codebooks*
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, as 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-escritor, 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.

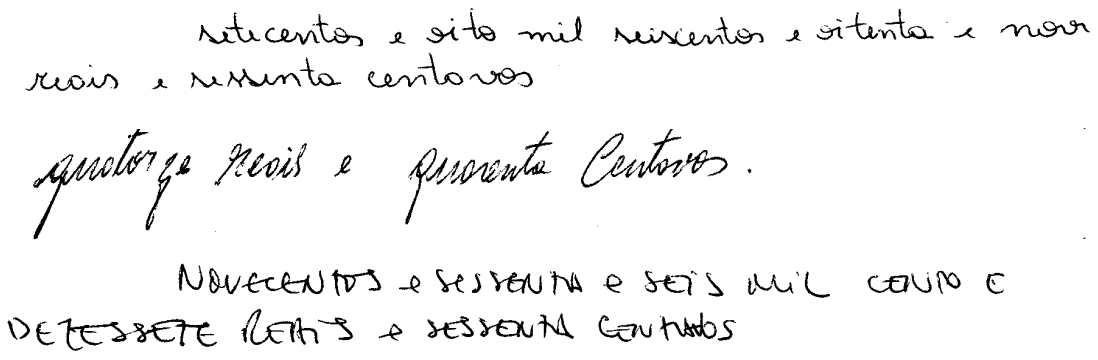

Figura 1.1:Exemplos de estilos de escrita

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 Makov – 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.

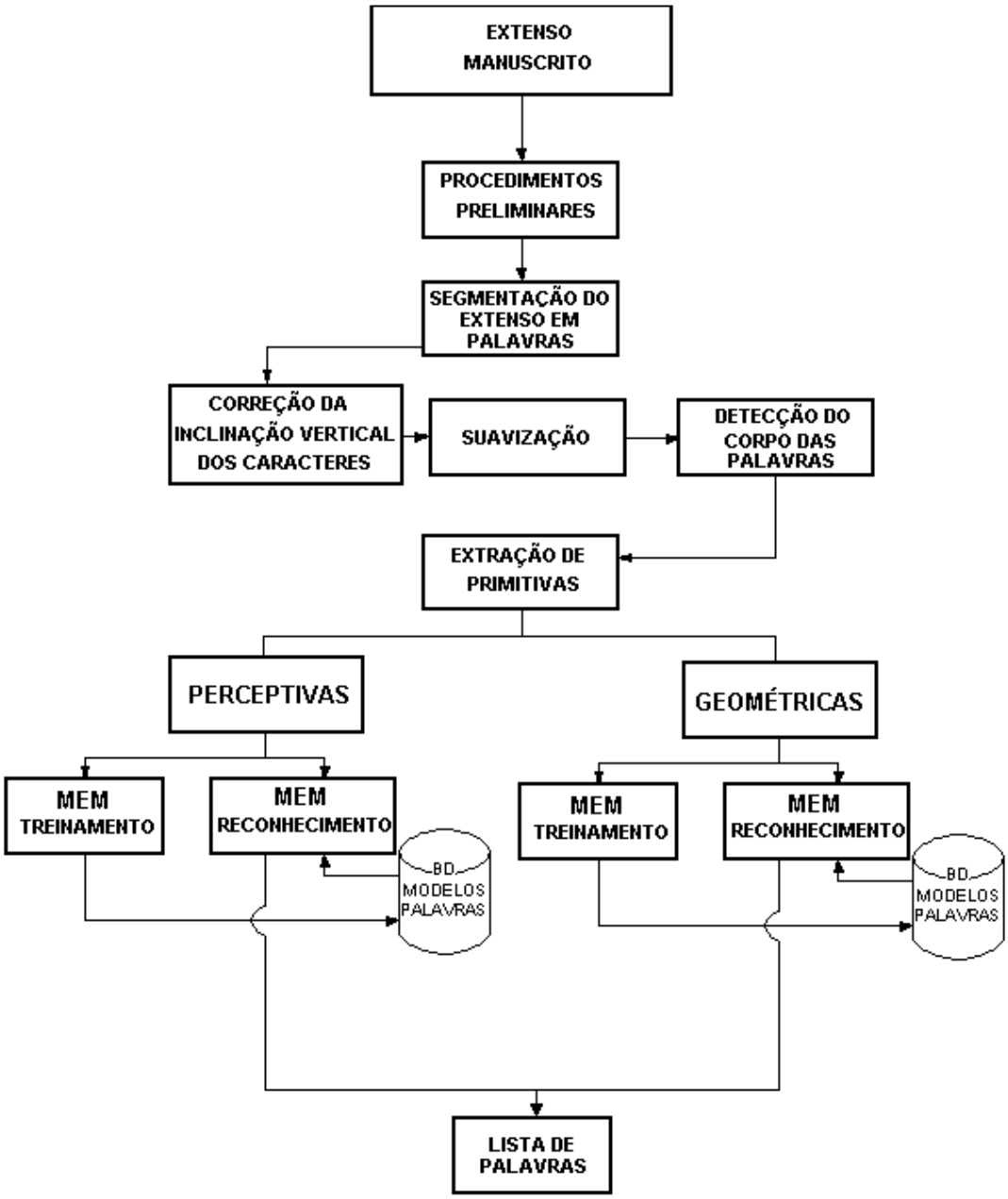

Figura 1.2: Esquema do método de reconhecimento de palavras manuscritas

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.

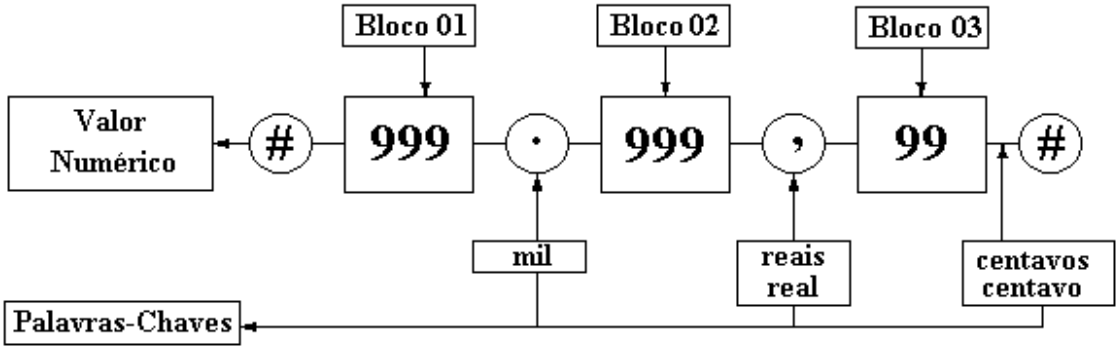

Figura 2.1: Valor numérico e palavras-chaves

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 **mil**har, 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}.

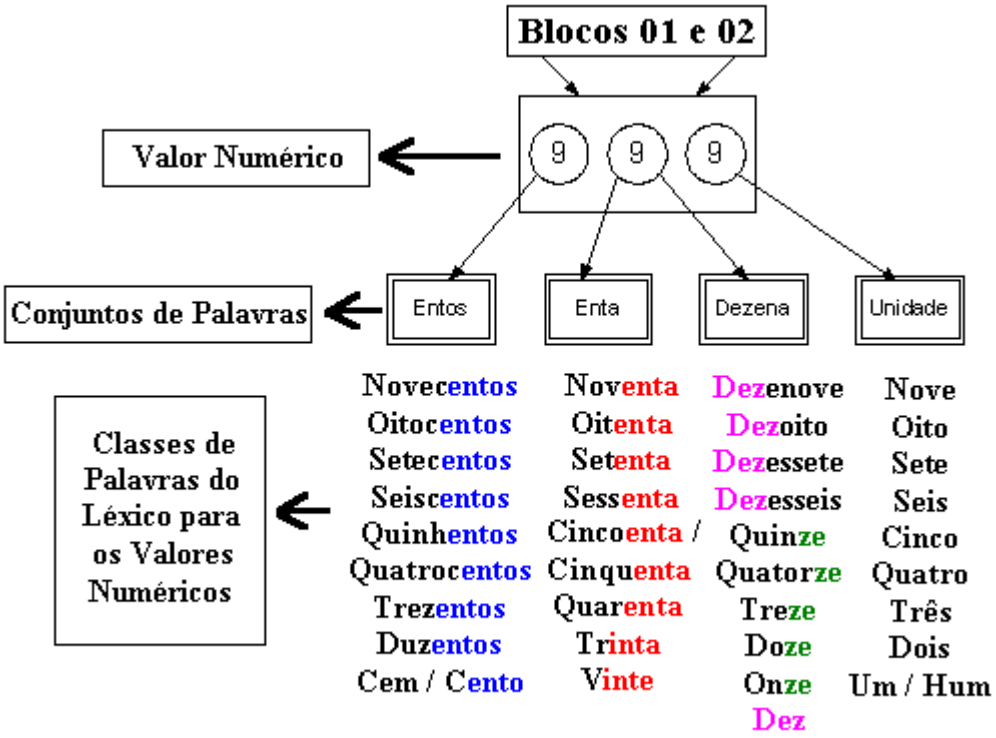

Figura 2.2: Classes de palavras do léxico em português

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, extraindo-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 o 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 seqüencial 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 escrita por estar representado por unidades de linguagem aceitas e compreendidas pelo sistema em questão.

A história da escrita delineia 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.

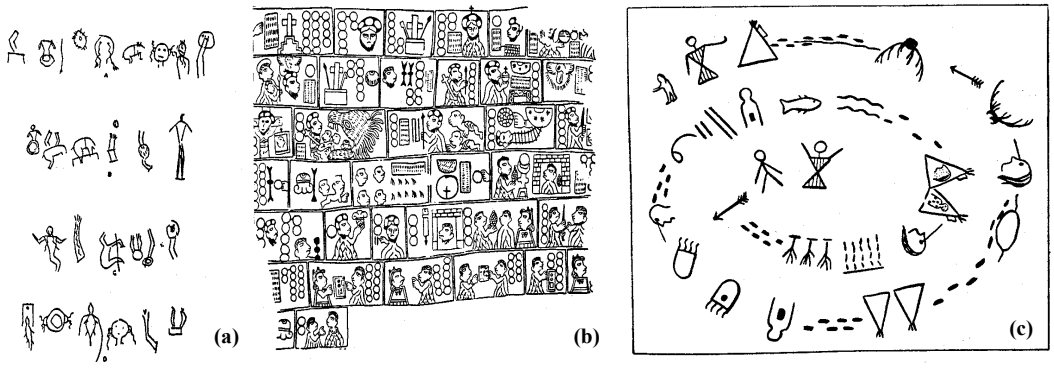

Figura 2.3: Exemplos de escrita pictórica: a) cantos Ojibwa; b) catecismo asteca; c) inscrição antiga [CAGLIARI, 1991]

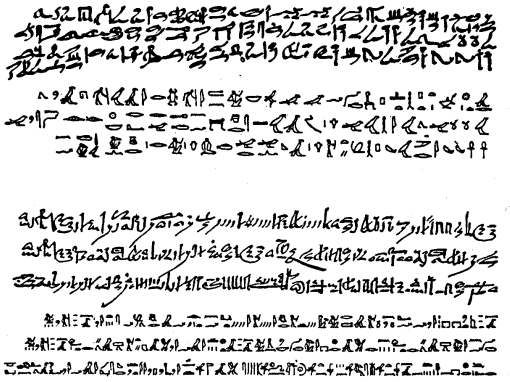

Figura 2.4: Exemplos de escrita ideográfica: escrita egípcia [CAGLIARI, 1991]

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. Haviam 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].

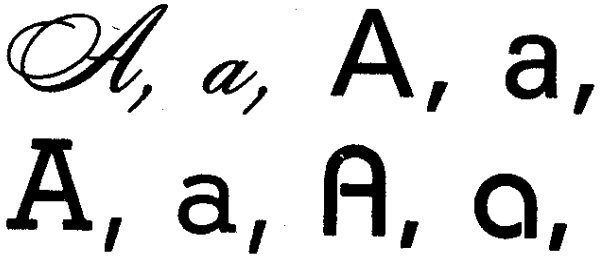

Figura 2.6: Exemplos da letra “a” em diferentes alfabetos [CAGLIARI, 1991]

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.

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

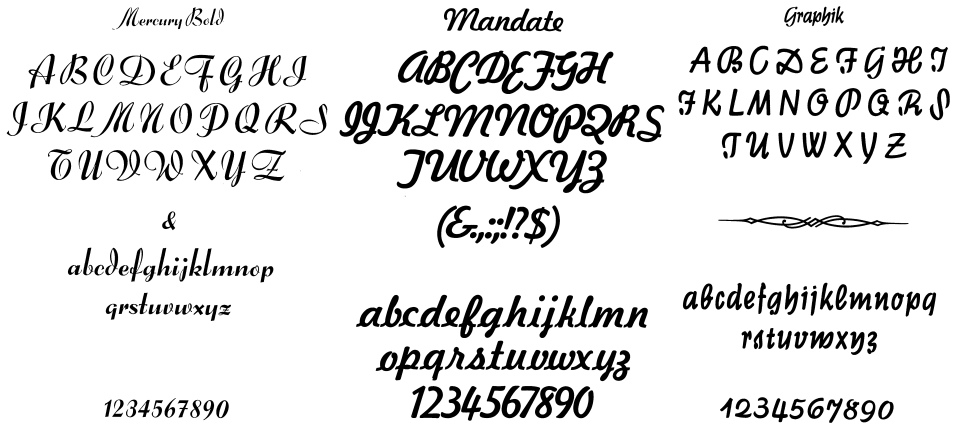

Figura 2.8: Estilos de escrita cursiva [CAVANAGH, 1974] e [SOLO, 1989]

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.

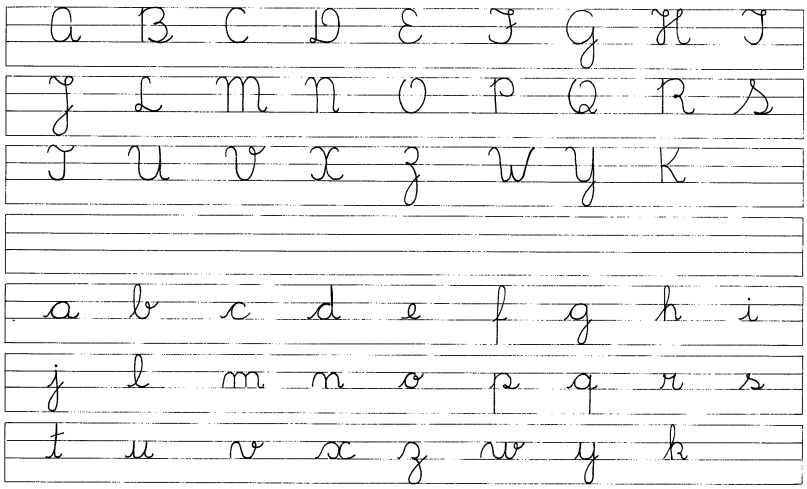

Figura 2.9: Alfabeto e estilo de escrita cursiva para alfabetização no Brasil

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 orgão do sentido que possibilita a leitura, além dos dedos que, através do tato, permitem aos deficientes visuais decifrarem o texto em *braile*. 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].

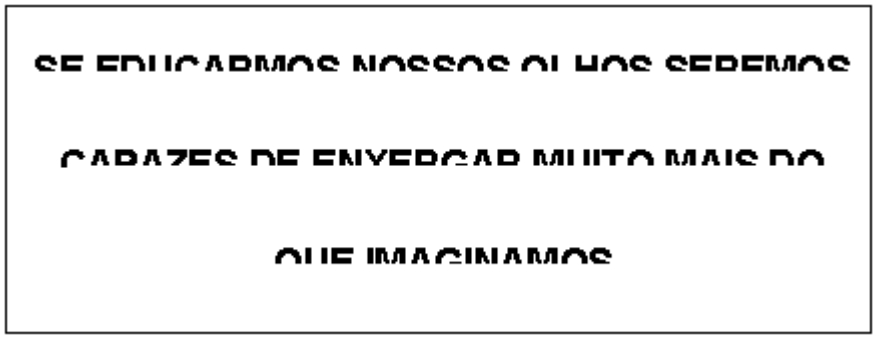

Figura 2.10: Verificação do princípio de fechamento[MACHADO & TEIXEIRA, 1996]

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-letras.

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 [GUILEVIC, 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 à 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 tem 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 psicolingüística, 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.

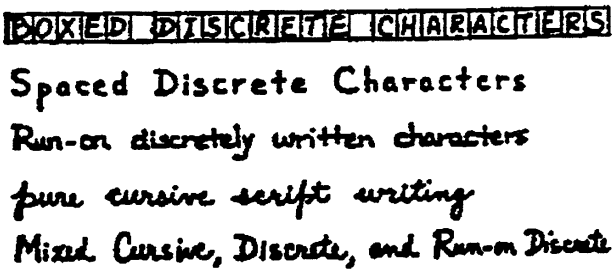

Figura 2.11: Tipos de escrita [TAPPERT et al., 1990]

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 métosos 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];

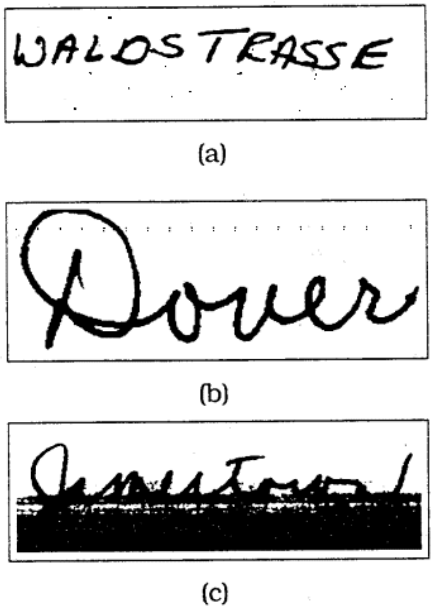

Figura 2.12: Exemplos de ruídos gerados no processo de aquisição da imagem: a) ruído "sal e pimenta", b) ruído gerado pelo alimentador e c) incorreta operação do *scanner* [KIM, 1996]

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. Ápos 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;

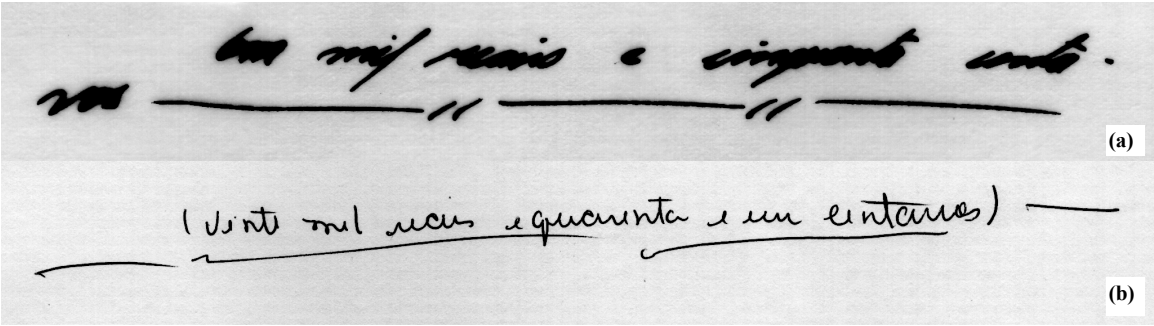

Figura 2.13: Exemplos da qualidade de imagem associado ao meio utilizado:
a) papel vegetal e b) papel sulfite branco

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*;

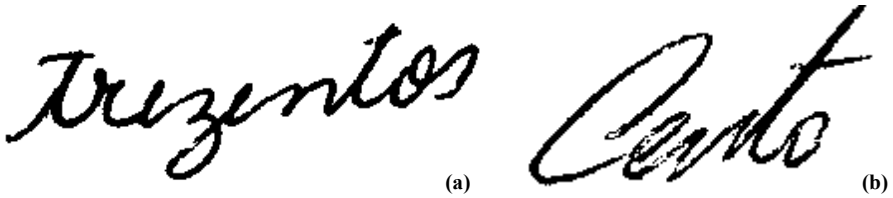

Figura 2.15: Inclinação: a) da linha de base (*skew*) e b) vertical dos caracteres (*slant*)

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;

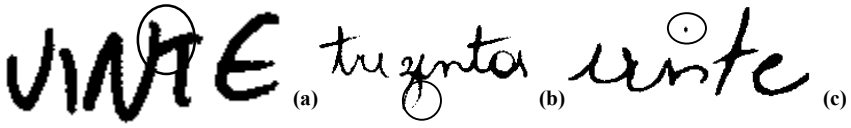

Figura 2.16: Exemplos de: a) traços contíguos, b) traços desconectados e c) posição errônea

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: "cinqüenta" 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: "do**s**e", "quatro**s**entos", "tre**s**e", "**ces**enta" e muitos outros encontrados na base de dado, não recebem tratamento específico;

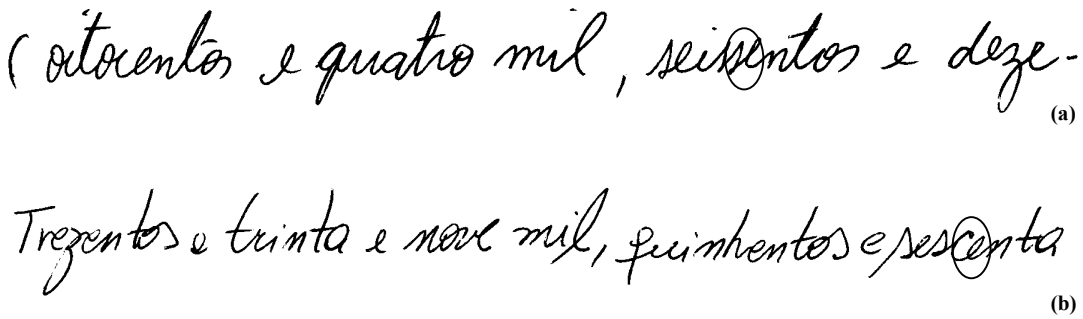

Figura 2.17: Erros de ortografia: a) AE0217a e b) AE0233a

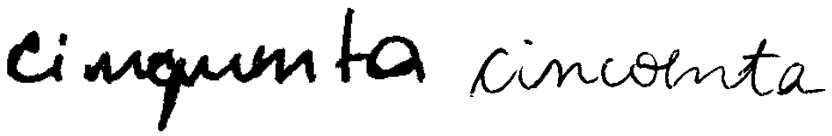

Figura 2.18: Ortografia diferenciada para uma mesma palavra

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-impressos 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 “cinqüenta”. 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 ter 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 freqüência
Local	[BOZINOVIC & SRIHARI, 1989]	Cálculos de percursos com parâmetros heurísticos

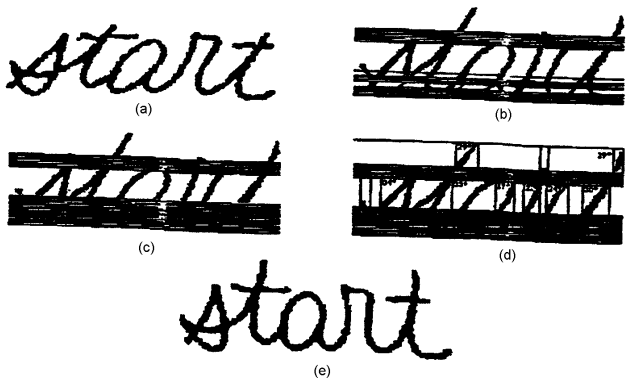

Figura 2.19: Correção da inclinação vertical dos caracteres: a) imagem original; b) remoção das linhas horizontais de percurso longo, c) remoção das faixas de pequeno comprimento, d) janelas isoladas e faixas conectando o centro de gravidade e ângulos da inclinação mostrados a cada janela, e) imagem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

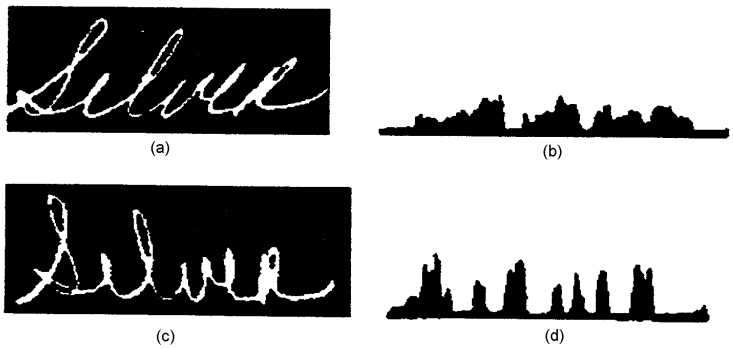

Figura 2.20: Correção através da aplicação do histograma de projeção vertical: a) imagem limiarizada a partir da imagem original, b) histograma de projeção vertical a partir da imagem contida em (a), c) imagem corrigida, d) histograma de projeção vertical da imagem contida em (c) [BUSE et al., 1997]

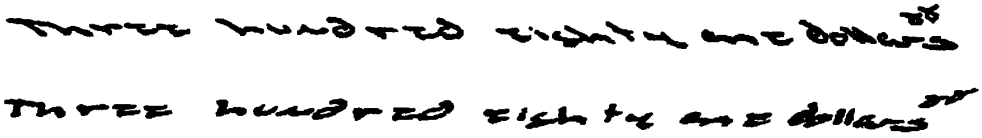

Figura 2.21: Correção de uma imagem com inclinação de 60 graus [GUILLEVIC, 1995]

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.

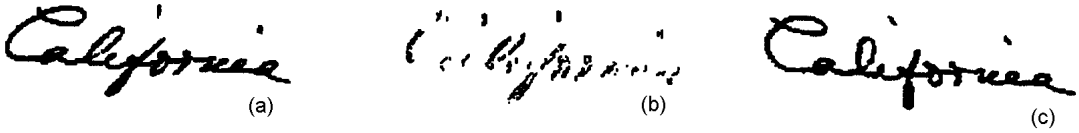

Figura 2.23: Correção da inclinação dos caracteres: a) imagem original, b) pixels de bordas não horizontal, c) imagem corrigida [KIMURA et al., 1993]

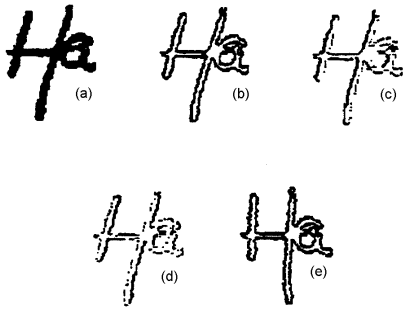

Figura 2.24: Correção vertical da inclinação dos caracteres: a) imagem original, b) contorno, c) e d) linhas verticais inferiores e superiores, e) imagem corrigida com base no ângulo estimado a partir das linhas detectadas em (c) e (d) [KIM & GOVINDARAJU, 1997]

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.

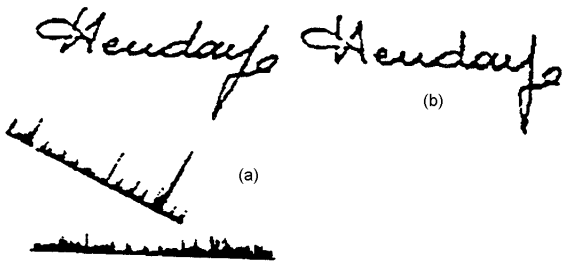

Figura 2.25: Correção da inclinação vertical: a) histograma de projeção vertical de duas retas de ângulos: 30 graus e 0 graus e b) correção da inclinação [EL YACOUBI, 1996]

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 utlizados 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 impíricos, 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".

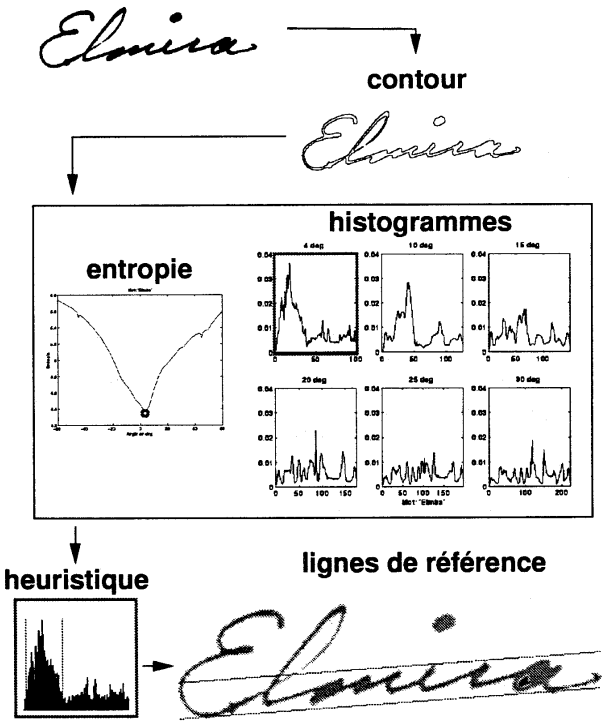

Figura 2.26: Determinação do corpo de palavras [CÔTE, 1997]

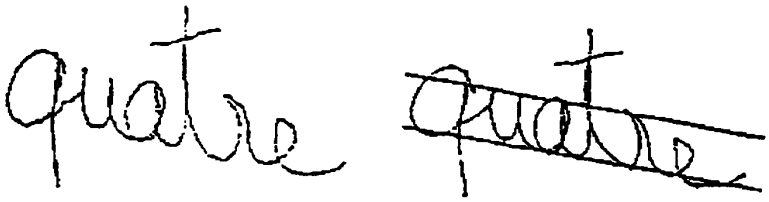

Figura 2.27: Estimação das linhas de referência do corpo das minúsculas [BERCU & LORETTE, 1993]

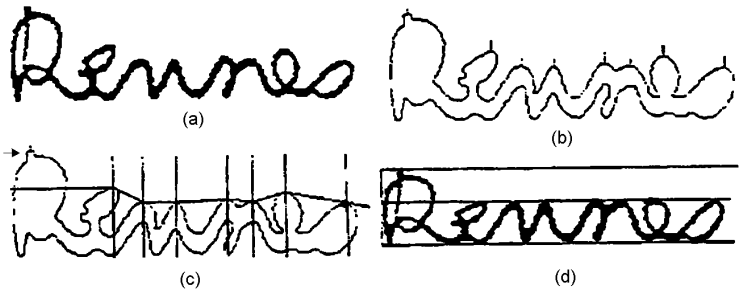

Figura 2.28: Normalização do corpo das minúsculas: a) imagem original, b) detecção dos máximos locais do contorno superior, c) filtragem dos máximos indesejáveis, d) imagem normalizada utilizando uma transformação não linear [EL YACOUBI, 1996]

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 dificíl 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. Provêem 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 *pixels* 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 (*chain code*) 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, *end points*, *branch points*, *crossing points*, pontos côncavos e convexos, curvas simples
Global	[DZUBA et al., 1998]	Alfabeto de 8 traços
Global	[MADHVANATH & GOVINDARAJU, 1998]	Ascendentes, descendentes, comprimento (*Perceptual Features)*

[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 reconstrui 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.

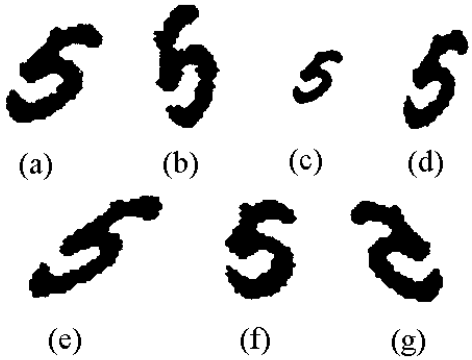

Figura 2.29: Versões transformadas do dígito "5": a) imagem original, b) rotação, c) escala, d) estiramento, e) inclinação, f) correção da inclinação g) espelhamento [TRIER et al., 1996]

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 (*cusp-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 (*rigth 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*).

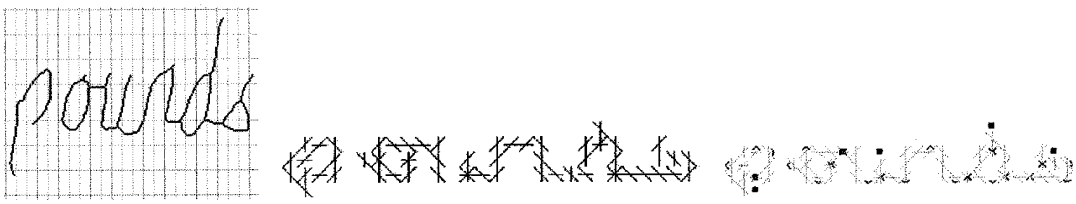

Figura 2.30: Estágios da extração de primitivas: a) esqueleto com *grid*, b) representação dos segmentos de linha parametrizados c) primitivas extraídas [SENIOR & ROBINSON, 1998]

[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 utlizada 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.

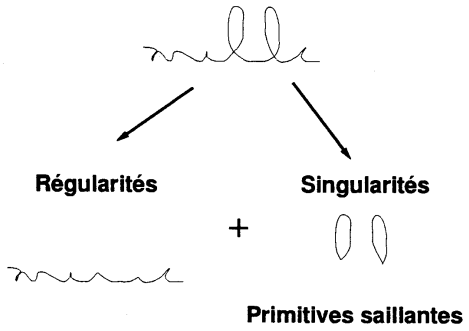

Figura 2.31: Regularidades e singularidades de uma palavra[CÔTE, 1997]

[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.

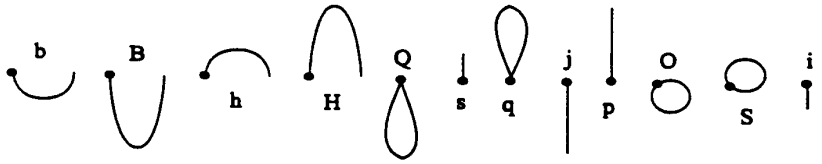

Figura 2.32: Alfabeto de 12 traços [AVILA, 1996]

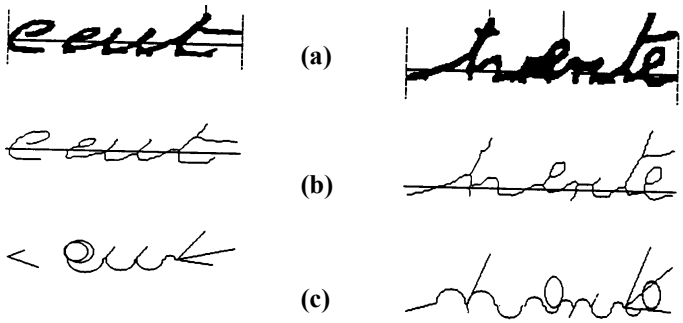

Figura 2.33: Exemplos de descrição de palavras através do alfabeto de 12 traços: imagem original, b) esqueleto, c) alfabeto de 12 traços [AVILA, 1996]

[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.

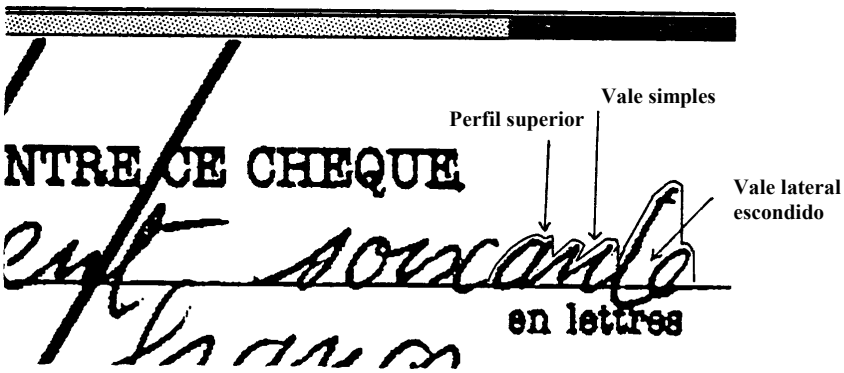

Figura 2.34: Vales e perfil superior [LECOLINET, 1990]

[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:

- Tipo1: 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 arcos 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”).

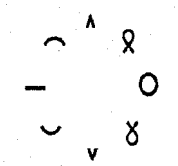

Figura 2.35: Primitivas com base em 8 traços [DZUBA et al., 1998]

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) e 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* = *hndwrtng*),
- 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 ultima 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.

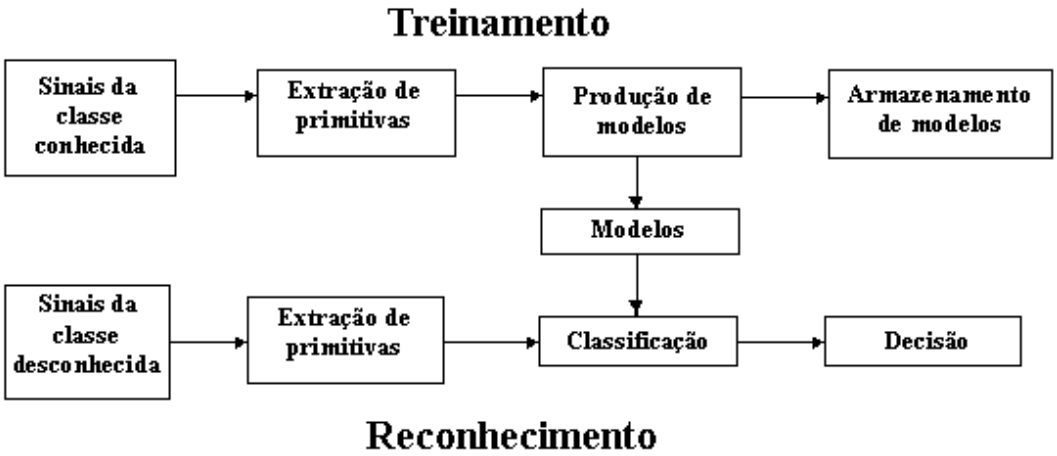

Figura 2.36: Diagrama de um método de reconhecimento de formas

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 Mahalonabis 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 enforques 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 freqüência 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, a 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 reconhecedores 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.

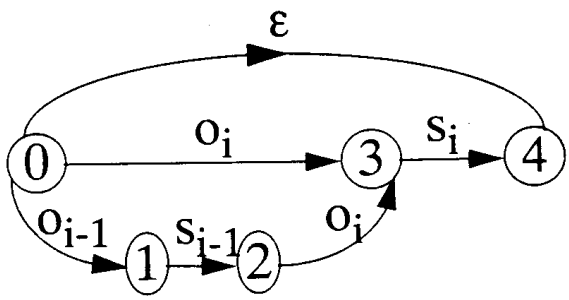

Figura 2.37: MEM para o caracter [EL YACOUBI, 1996]

[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).

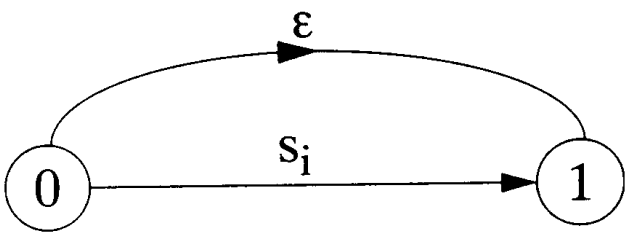

Figura 2.38: MEM para o espaçamento inter-palavras [EL YACOUBI, 1996]

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.

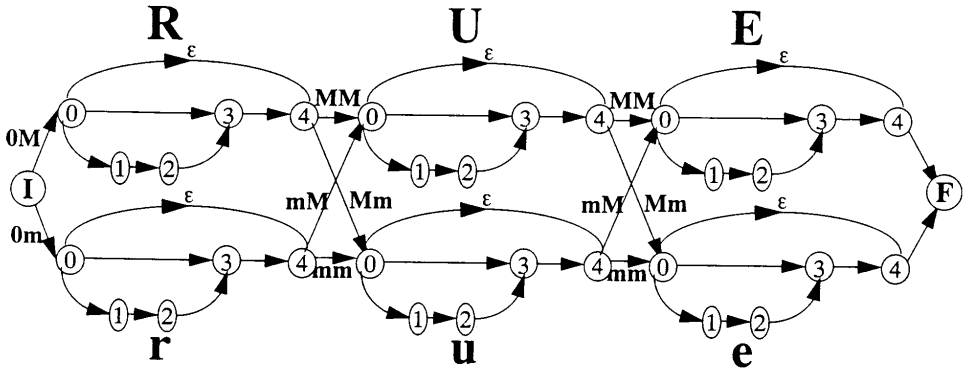

Figura 2.39: Modelo global para o prefixo "rue" [EL YACOUBI, 1996]

[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.

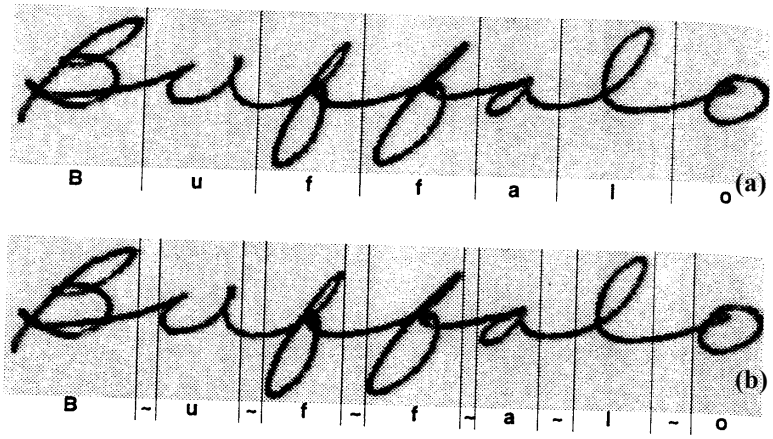

Figura 2.40: Métodos de análise de uma palavra manuscrita:
a) caracter-caracter, b) caracter-ligação-caracter [CHO et al., 1995]

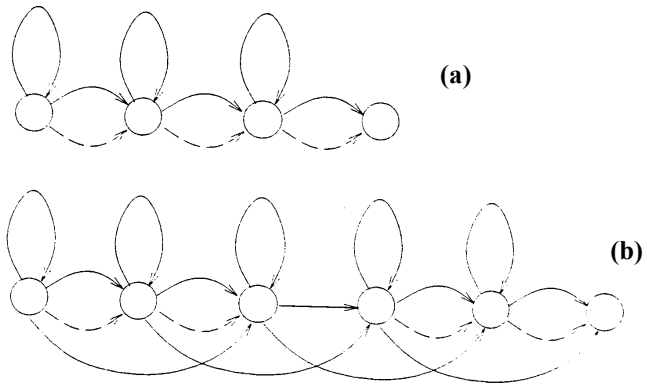

Figura 2.41: Topologias dos modelos:
a) caracter-caracter, b) caracter-ligação-caracter [CHO et al., 1995]

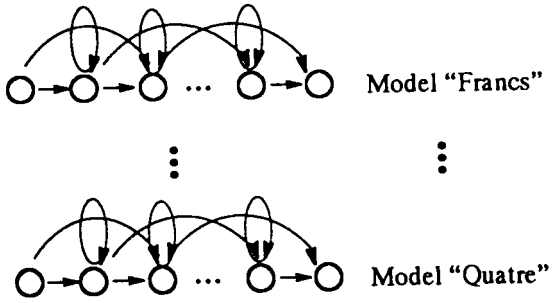

Figura 2.42: MEM para palavras de cheques franceses[KNERR et al., 1996]

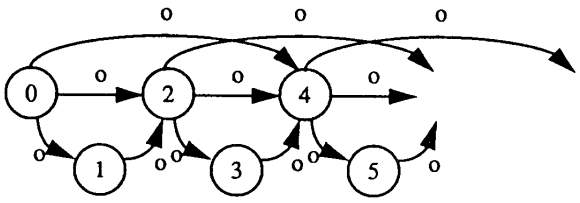

Figura 2.43: MEM para palavras [GILLOUX, 1994]

[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-seqüências 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.

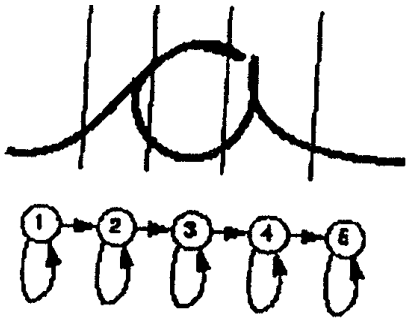

Figura 2.44: Modelo para as letras cursivas [GILLIES, 1992]

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 esta 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 lingüistico 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].

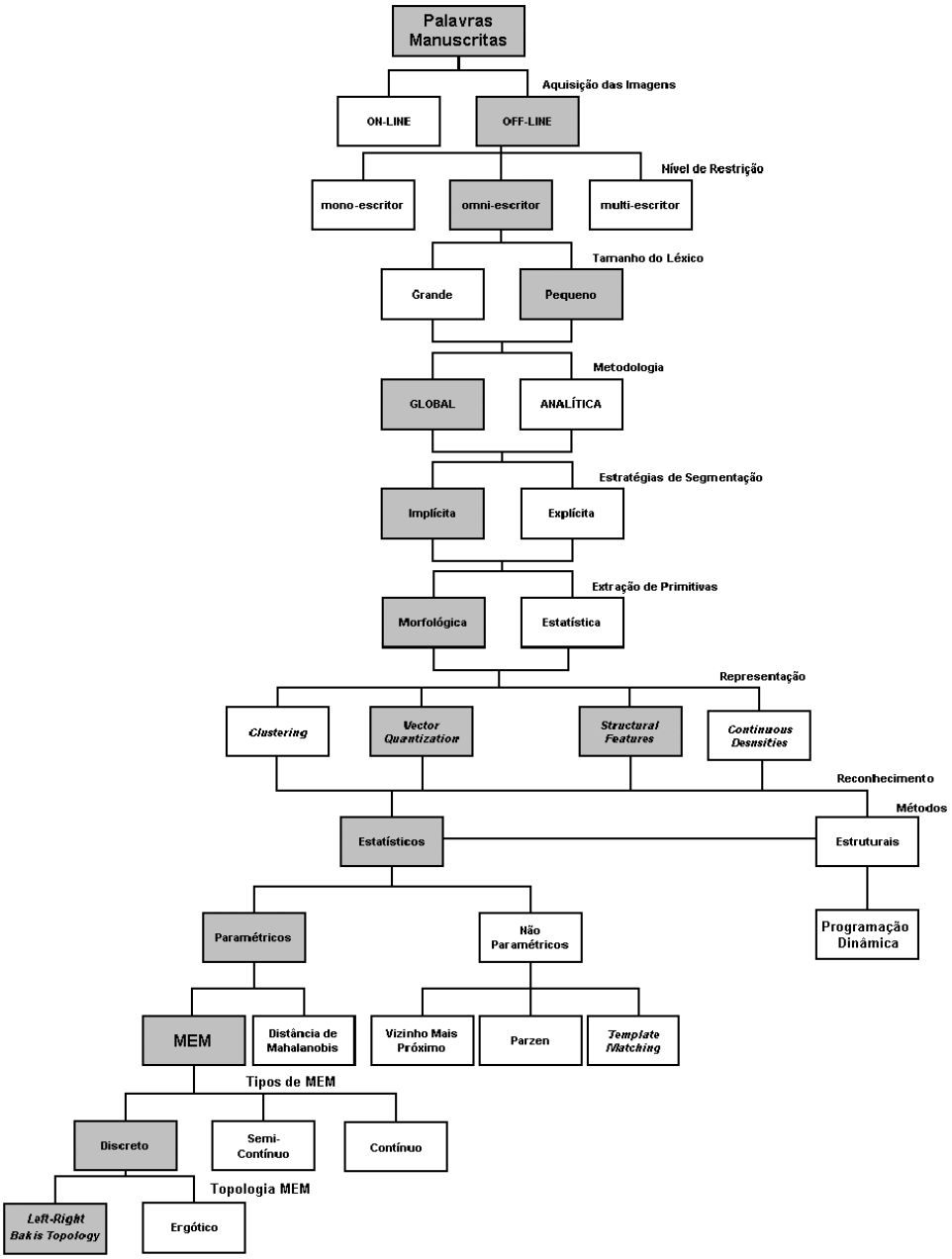

Figura 2.45: Variáveis e metodologias no reconhecimento de palavras manuscritas

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.