

O que a Lingüística tem a ver com o teclado de microcomputadores

José Marcelino Poersch

ILA/PUCRS

RESUMO

Através do estudo da frequência dos caracteres gráficos em Língua Portuguesa em combinação com o estudo dos reflexos dos dedos na digitação e de seu esforço para acessar as diversas teclas, são sugeridas mudanças no clássico teclado QWERTY de microcomputadores e de outras máquinas eletrônicas de escrever com o objetivo de veicular, na digitação, o máximo de informação com o mínimo de custo.

Essa nova distribuição, por certo, enfrentará o conservadorismo exagerado que obstaculiza a promoção dos avanços científicos e tecnológicos no mundo cultural de maneira semelhante à lei da inércia que dificulta mudanças de movimento no mundo físico. No entanto, se ficar comprovado, experimentalmente que essa nova disposição empresta efetivamente ao digitador maior eficiência do que o teclado convencional, a Lingüística terá dado aos cientistas e aos engenheiros da computação uma significativa contribuição e é de se supor que a tecnologia, num futuro não muito distante, passe a produzir esse novo teclado.

ABSTRACT

Through the study of the frequency of graphic characters and they study of finger reflexes in typing and their effort to access the various keys, changes in the present QWERTY keyboart of microcomputers and of other electronic typewriters are suggested, in order to convey, in typing, a greater amount of information with a lower cost.

Correlations are established between the orders of frequency of graphic characters and the orders of the quickness with which fingers can access the keys in such a way that the most frequent characters are assigned to the keys the most easily pressed. It is presumed that with this layout, better performance with lower cost will be achieved. If a future experiment proves that this new layout effectively provides the operator with more efficiency than the conventional keyboard, linguistics will have given computer scientists and engineers a significant contribution.

1. INTRODUÇÃO

O presente artigo constitui um relato da pesquisa "Contribuição da Lingüística Computacional na disposição dos caracteres gráficos em teclados de microcomputadores", pesquisa desenvolvida no Centro de Pesquisas Lingüística da PUCRS (Instituto de Letras e Artes, Curso de Doutorado em Lingüística) com o apoio financeiro parcial (serviços técnicos) da Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS) e (material permanente e de consumo) do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). A equipe técnica contou com a participação do prof. Dr. José Marcelino Poersch, coordenador do projeto, e das assessorias técnicas da Dra. Josênia Vieira da Silva (assessora lingüística), dos mestres Bertilo Frederico Becker (programador e supervisor de digitação), Arthur Vargas Lopes (assessor informático) e Fernando Lang da Silveira (assessor estatístico). Como auxiliares de pesquisa atuaram os graduandos Marcelo Poersch, Carmen Lucia Pacheco de Araújo, Maria Inês Becker, Marlise Goulart e Tânia Regina Pacheco Fernandes.

A investigação em referência insere-se na linha de pesquisa "Lingüística Computacional", do curso de Doutorado em Lingüística da PUCRS e contou com o apoio técnico do Centro Brasileiro de Lingüística Computacional. O projeto global, no qual este se enquadra, objetiva contribuir com os cientistas e engenheiros da computação na elaboração de programas básicos usados em editores de texto e colaborador no desenvolvimento de estudos da Lingüística para a análise automática de textos.

A pesquisa em pauta vem constituída de três etapas; as duas primeiras já foram executadas e a terceira está sendo implementada.

a. Frequência dos caracteres gráficos em língua portuguesa

Nesta primeira etapa levanta-se a distribuição de frequência dos caracteres gráficos em língua portuguesa e reavalia-se a afirmação de lingüistas de que a tipologia do texto não constitui variável na frequência dos grafemas, visto não estar a distribuição dessa frequência vinculada com o significado. Por outro lado, são oferecidos,

ao analista, dados do português que podem ser comparados com dados de outros idiomas, notadamente do inglês, do francês e do alemão.

b. ARDOM: uma nova configuração para teclados de microcomputadores

Na segunda etapa estuda-se, num primeiro momento, a possibilidade de combinar a prontidão de reflexos dos diferentes dedos da mão e a distância que as teclas se encontram dos dedos colocados em posição de descanso com a produtividade dos caracteres em língua portuguesa. Depois, correlacionando-se a ordem de frequência dos caracteres e a ordem de facilidade de acesso das teclas, consegue-se denominar as teclas mais prontamente acessadas pelos caracteres mais frequentes (teclado ARDOM)

c. Eficiência de teclados de microcomputadores

Através de uma pesquisa experimental, oferecem-se subsídios empíricos para avaliar a conjectura de que, em termos de operacionalidade e de eficiência, os usuários do teclado ARDOM, emparelhadas as demais variáveis, tornam-se melhores digitadores do que os usuários do clássico teclado QWERTY.

Os resultados parciais desta investigação foram apresentados no XI Encontro Nacional de Lingüística (Rio - 1988), no IV Encontro da Associação Portuguesa de Lingüística (Lisboa - 1988) e no IX Congresso Mundial de Lingüística Aplicada (AILA 90 - Grécia) e divulgados nos Actas do IV ENAPL (Poersch, 1988a) e na revista VERITAS (Poersch & Becker, 1990).

2. MICROCOMPUTADORES E O AVANÇO TECNOLÓGICO

Na era da modernidade procuram-se, num aperfeiçoamento constante, novas técnicas que imprimam maior rapidez e eficiência aos

mecanismos que substituem cada vez mais as atividades semi-automáticas do homem.

Na virada do atual século, os computadores ocupam um espaço cada vez mais respeitável no cenário das ciências sociais e humanas; eles tornam-se uma realidade cotidiana para um número crescente de pessoas, facilitando e tornando mais eficiente um sem número de atividades como, por exemplo, a educação. Parghaly (1989:235) observou que "o advento do computador pessoal no final dos anos 70 e sua disponibilidade nas escolas e universidades levou os educadores a pensar a melhor maneira de utilizar essa tecnologia para fins instrucionais".

Assim, o microcomputador aparece como o elemento determinante no processo da modernização. Em vários ramos da atividade humana, o serviço vagaroso e passível de falhas está sendo substituído por um trabalho mais rápido e mais eficiente da máquina. Os avanços da Inteligência Artificial e das Ciências Cognitivas tendem a munir a máquina de inteligência e, desta maneira, executar atividades próprias do homem.

Analisando o fato de que "os computadores só fazem aquilo que mandamos e não, necessariamente, aquilo que desejamos que eles façam (Tremblay & Bunt, 1983)", justificam-se as pesquisas cada vez mais freqüentes que o homem vem implantando na área computacional para transformar seus "desejos" em realidade tecnológica. Entre seus desejos está o de dotar os computadores de habilidades semelhantes as suas. É neste contexto que a lingüística oferece subsídios à ciência da computação para a construção dessas habilidades.

Na área das ciências da linguagem identificam-se três utilidades básicas do computador:

- a) Processador de palavras (editor de textos); programa específico que possibilita digitar, ordenar, corrigir, alterar, formatar e imprimir textos;
- b) Processador de dados: programa que permite introduzir na memória do computador um grande número de dados, organizá-los e tratá-los estatisticamente segundo os desejos de um pesquisador;
- c) Banco de dados: permite armazenar, na memória do computador, informações de todos os tipos; oportuniza, posteriormente, uma recuperação rápida ao usuário e um intercâmbio com outros bancos.

Destas três utilidades, a primeira é que ocupará o centro de interesse do presente artigo. A editoração de textos constitui um dos recursos que o computador oferece para facilitar as atividades de quem lida concomitantemente com produção e com reprodução de textos. Bons editores de texto, além de ótimos recursos para a editoração do texto, devem poder utilizar-se de um teclado realmente eficiente, para minimizar ao máximo o esforço desse processamento.

3. OS TECLADOS DE MICROCOMPUTADORES

Com o avanço da tecnologia, procuram-se instrumentos cada vez mais perfeitos, mais ágeis e mais fáceis de serem operados. Constatase que os teclados dos microcomputadores, em sua versão padrão, apresentam dificuldades para a produção de textos em português, principalmente no que se refere a caracteres gráficos específicos quais sejam o cedilha e as vogais acentuadas (em número de 12). Diversas tentativas de solução foram propostas, experimentadas e integradas aos editores de texto, programas responsáveis pelos recursos que um microcomputador pode oferecer para editar um texto. Na maioria dos casos, utiliza-se a sobreposição de caracteres (o c cedilha é obtido sobrepondo uma vírgula ao c ($\text{ç} = \text{c} + \text{,}$) mediante o toque de três, quatro ou cinco teclas para a produção de um desses caracteres específicos (ã : letra a + utilidades + sobreposição + caixa alta + caracter (^)). Além desse exagerado consumo de energia (tempo), ainda existe o inconveniente de o visor nem sempre representar o caracter da maneira como ele deve aparecer na impressão, ou o visor registrar mais caracteres do que deverão aparecer na impressão o que prejudica enormemente o controle do emparelhamento da margem direita do texto. Outros aparelhos, por possuírem teclas para funções especiais, podem ser programados para produzir esses caracteres, mediante um retrocesso, como nos teclados tradicionais das máquinas datilográficas. No entanto, sempre será com mais de dois toques para produzir uma letra acentuada. Isso gera um consumo desnecessário de esforço.

Uma firma californiana, sediada em Sunnyvale, recebeu da Apple Computer a autorização para produzir o "Diplomata", programa especial que dota os editores de texto da possibilidade de produzir caracteres específicos de línguas diferentes do inglês. O "Diplomata"

consiste num dispositivo eletrônico embutido na unidade central de processamento (CPU) dos microcomputadores Apple IIe, dotando-os, mediante o simples toque de uma tecla, da possibilidade de utilizar, instantaneamente, além do teclado padrão para o inglês (Standard ANSI Keyboard), teclados específicos para a edição de textos em línguas diferentes. A característica básica do "Diplomata" é a sua comutabilidade, qualidade que permite uma conversão instantânea entre dois ou mais conjuntos de caracteres. Esse dispositivo permite gerar todas as letras bem como os demais caracteres gráficos (como sinais de acentuação) mediante o simples toque de uma única tecla e exibi-los no visor com uma forma idêntica àquela com que deverão aparecer na versão impressa.

Analisando esse programa, verifica-se que a distribuição dos caracteres no teclado apresentam várias falhas ou deficiências: permutação desnecessária e injustificável de letras (A pelo Q, W pelo Z), da inclusão de caracteres não pertencentes ao alfabeto do português (K, W, Y), da exclusão de algumas vogais acentuadas (í, á), do privilegiamento de certas vogais acentuadas (â) em detrimento de outras (á, ê, ó) que são bem mais produtivas.

As deficiências do "Diplomata" trouxeram a idéia de proceder a um levantamento criterioso da frequência dos caracteres gráficos em português para, mediante a eliminação de caracteres desnecessários ou de baixa produtividade na editoração de textos em língua portuguesa, abrir espaço para a inserção de outros caracteres específicos de nossa língua e dispô-los segundo sua frequência de ocorrência.

A par desse problema, verifica-se um outro: a disposição dos caracteres gráficos num teclado padrão.

Pergunta-se quais foram os critérios que nortearam os engenheiros de máquinas de escrever a distribuir os caracteres gráficos pelo teclado, na forma como eles se encontram dispostos atualmente. Sabe-se que essa disposição deve obedecer a certos critérios que devem levar em conta a distância em que as teclas se localizam dos dedos da mão posicionada (em seu ponto de repouso) na parte central da segunda carreira de teclas, de baixo para cima, e a energia diferenciada que cada dedo possui (o indicador e médio parecem possuir, mais energia do que o anular e o mínimo). Partindo do pressuposto de que os caracteres mais produtivos devem corresponder às posições mais fáceis de serem acessadas, fácil se torna deduzir que o levantamento de frequência dos caracteres gráficos muito pode contribuir para a elaboração dessa distribuição.

O teclado de um microcomputador é igual ao de uma máquina de escrever eletrônica padrão (teclado QWERTY), acrescido de alguns outros comandos. Esse teclado é assim denominado devido à disposição das seis primeiras letras da carreira superior do teclado. Essa disposição dos caracteres do teclado não foi a primeira a ser elaborada e utilizada e, como todas as outras disposições, levava em conta a frequência de ocorrência dos caracteres da língua inglesa. Os outros teclados, contudo, não se perpetuaram porque as letras mais frequentes e de ocorrência sucessiva (tais como D e E) estavam dispostas lado a lado no teclado e, devido à rapidez de toque dos datilógrafos, as teclas se acavalavam na hora de atingir a folha. Assim, para diminuir a rapidez de toque, bem como para evitar danos às máquinas de escrever, Glidden, Sholes et Soulé (Grupo, 1984 a e b), três norte-americanos, separando justamente os caracteres de maior frequência de ocorrência, criaram o conhecido teclado QWERTY.

Colocam-se, então, duas questões que devem merecer toda nossa consideração. Como pode um editor de textos para o português bem servir a um digitador se possuir um teclado baseado na frequência dos caracteres da língua inglesa? E como pode-se aceitar o fato de tais caracteres estarem dispostos de forma a aumentar o espaço de tempo entre um toque e o toque imediatamente seguinte a fim de evitar o acavalamento das teclas, se um microcomputador funciona por impulsos elétricos, independente da sequência ou proximidade de alavancas mecânicas?

Com base na problemática assim estabelecida supõe-se que o levantamento da frequência dos grafemas do português, bem como a frequência dos dígramas (sequência de dois grafemas) pode conduzir-se à elaboração de uma disposição mais racional, ou melhor, mais eficiente do teclado se se fizer corresponder às teclas mais rápidas de serem acionadas os caracteres gráficos mais produtivos da língua, de modo a permitir um ganho razoável de tempo.

A análise desses dois problemas leva-nos a concluir que a lingüística pode trazer uma significativa contribuição à tecnologia da ciência da computação através de uma pesquisa que objetive estudar a frequência dos caracteres gráficos e dos dígramas do português. E o que acentua ainda mais a interdisciplinariedade é o fato de que o lingüista se utiliza do cientista da computação para a elaboração do programa de levantamento de dados, e de um microcomputador para o efetivo levantamento automático dos dados a partir de um texto.

4. OS ASPECTOS QUANTITATIVOS DO FENÔMENO LINGÜÍSTICO

As observações que fazemos na vida comum, bem como as observações mais sistemáticas da ciência, revelam certas regularidades na natureza.

As leis da ciência são formulações que expressam essas regularidades com a máxima precisão possível. Entretanto, nem todas as leis da ciência são universais, pois, em vez de afirmar que tal regularidade ocorre em todos os casos, algumas leis afirmam que só ocorrem em certa percentagem de casos. Se a percentagem for especificada ou se se formular uma afirmação quantitativa sobre a relação entre um evento e outro, então formula-se uma lei estatística.

Às vezes, quando damos uma explicação dos fatos, as únicas leis que se aplicam são leis estatísticas e não leis universais. Na falta de leis universais conhecidas, muitas vezes as explicações estatísticas são o único tipo de explicação disponível.

Se, em determinado campo, encontramos ordem suficiente para fazer comparações e afirmar que, em algum aspecto, uma coisa está acima de outra, então, em princípio, há possibilidade de medição. O primeiro passo consiste em formular regras de comparação e, depois, se for possível, regras quantitativas.

A quantificação dos conceitos leva, antes de tudo, à formulação de um vocabulário mais eficiente; com isso elimina-se a necessidade de recorrer a um vocabulário muito vasto e sujeito a todos os matizes do subjetivismo, com suas variações às vezes imprevisíveis.

O mais importante, porém, é que os conceitos quantitativos nos permitem formular leis quantitativo-explicativas dos fenômenos e permitem a previsão de novos fenômenos ou estados de fenômenos. O enfoque estatístico é especialmente útil quando devemos analisar grande quantidade de dados.

O que a pessoa diz, segundo Miller (1951), encontra-se restrinvido ou controlado de diversas maneiras: pelo público, pela gramática da língua, por suas próprias necessidades e experiências. Antes de começar a considerar algumas dessas restrições particulares impostas ao que a pessoa diz, necessitamos, sem dúvida, contar com uma orientação estatística geral que nos indique que classe de expressões deve ser levada em conta.

Um tipo de restrição imposta ao falante é a estrutura da língua que utiliza. A fala e a escrita (conduta verbal) estão abertas a todas as influências que afetam qualquer tipo de conduta, e uma análise do contexto completo de um ato comunicacional deve incluir as necessidades do falante, suas percepções, o público com que conta e sua bagagem cultural.

Para controlar a amplitude do contexto verbal e avaliar quantitativamente sua influência, é necessário possuir dados estatísticos referentes às frequências relativas de ocorrência das unidades verbais.

Segundo Malmberg (1971), todo tratamento científico dos fenômenos lingüísticos, toda conclusão a respeito deles, toda descrição de estados de língua pressupõem, talvez em sua forma mais simples, o auxílio da estatística.

A imagem da língua só é completa se pudermos medir a frequência dos tipos e das combinações; por um lado, no vocabulário tal como ela aparece no dicionário e, por outro lado, na língua viva, ou nos textos impressos. Assim, a descrição de um determinado fonema de um sistema fonológico e a descrição de suas possibilidades de combinação com outros fonemas deve ser completado por uma pesquisa sobre a frequência dos ditos elementos e das combinações em questão, comparada à de outros fonemas. A descrição qualitativa deve ser completada por uma pesquisa quantitativa.

O desenvolvimento humano e o avanço das civilizações dependem, além de outras variáveis, do progresso alcançado nos meios de receber, de comunicar e de registrar o conhecimento.

A comunicação, segundo Cherry (1971), implica essencialmente uma linguagem, um simbolismo, quer seja um dialeto falado, uma inscrição em pedra, um sinal do Código Morse, ou uma série de pulsos binários num computador moderno.

As formas de comunicação sofreram inúmeros processos de mudanças desde os pictogramas, ideogramas e hieroglifos, até a sua forma atual. Neste processo de mudança, sempre imperou a preocupação com a economia dos símbolos usados, procurando a redundância, sem, contudo, descuidar a eficiência da comunicação.

Com a introdução do código MORSE, percebeu-se o aspecto estatístico da economia da linguagem. Morse se deu conta de que as variáveis letras da língua inglesa não são usadas com igual frequência; uma visita a uma tipografia e uma contagem das quantidades de tipos usados forneceu-lhe uma estimativa das frequências relativas das

letras. Morse concebeu o seu código de modo que as letras mais usadas correspondessem aos símbolos de ponto-e-traço mais curtos.

A lei da frequência de letras de Morse exibe uma tendência, mas é puramente descritiva; a forma precisa dessa e de outras relações foi explorada sistematicamente por Zipf (1949), que fez uma coleta e estudo de aspectos estatísticos da fala e da escrita. Ele foi um dos precursores do tratamento estatístico dos fenômenos estatísticos. Ele mostrou que a complexidade dos fonemas é inversamente proporcional a sua frequência e que os fonemas surdos, nas línguas onde existe a distinção surdo/sonoro, são duas vezes mais frequentes que os sonoros. Também foi Zipf quem estabeleceu o princípio do "menor esforço" que ele considerou válido não só para os sons mas também para outros elementos da língua, particularmente para as palavras.

Tais resultados tendem a mostrar que a frequência dos fonemas segue leis determinadas, que são funções dos seus caracteres físicos (acústicos e fisiológicos) e da reação do ouvido a esses estímulos. Esses estudos tem demonstrado que os fonemas não se combinam em unidades superiores por obra do acaso, mas em virtude de princípios determinados.

Segundo Malmberg (1971), tão importante quanto o conhecimento dos diversos fonemas que aparecem no vocabulário ou na língua viva é o das leis que, em cada língua, condiciona a construção de unidades superiores: sílabas, morfemas, palavras.

Inúmeras análises estatísticas de línguas faladas e, em escala bem menor, de línguas escritas, são de interesse dos lingüistas, psicólogos e engenheiros da comunicação. Essas análises do comportamento lingüístico oportunizam o surgimento de leis definidas.

Atualmente, a análise estatística é um importante método de estudo lingüístico. Além da frequência de letras e palavras, já foram feitas observações acerca da frequência de sílabas, de partes da oração, dos hábitos de acentuação e inflexão de interlocutores ao telefone.

O conhecimento da frequência dos fonemas é de importância capital para fins práticos. Os primeiros a chamar a atenção dos lingüistas para a frequência relativa dos fonemas, ou das letras, foram os estenógrafos.

Em Guiraud (1959, pág. 31) lemos que "a linguagem é um sistema de signos e, como tal, é submetido às leis das probabilidades;... A frequência dos diferentes fonemas é estabelecido sobre um com-

promisso entre a economia da transmissão e aquela da recepção; assim a redação de um telegrama tende para o menor número compatível com a compreensão da mensagem. A frequência não tem, portanto, de forma alguma, um caráter, arbitrário; ela é determinada pela função, pela natureza do signo e pelas suas coordenadas físico-psicológicas".

Sendo que, num sistema escrito, as letras mantêm correspondência com os sons, fácil fica concluir que a frequência daquelas também se sujeita às leis probabilísticas. (Poersch, 1987).

Herdan (1966, pág. 15) afirma que "as proporções das formas lingüísticas pertencentes a um nível particular de compreensão, ou a um estágio de codificação lingüística (fonológica, gramatical, métrica) permanecem sensivelmente constantes para uma dada língua, num dado tempo de seu desenvolvimento e para um número suficientemente grande de observações".

Como as letras não estão diretamente ligadas a um significado e, portanto, não dependem da variável escolha individual, os dados estatísticos de Zipf comprovam a constância de sua distribuição em amostras das mais variadas.

O fato marcante da estabilidade das frequências relativas dos símbolos parece ser uma característica comum das formas lingüísticas. Segundo Herdan (1966, pág. 16) "existe uma ampla similitude entre os membros da comunidade falante, não apenas quanto ao sistema fonêmico, ao vocabulário e à gramática, mas também quanto à frequência de uso de fonemas particulares, itens lexicais e determinadas formas de estruturas gramaticais; em outras palavras, uma semelhança não só no que é usado, mas também no quantas vezes é usado".

5. EDITORES DE TEXTO

Os editores de texto são programas (software) que, na forma de disquetes, são comercializados para serem adaptados a computadores a fim de que o usuário possa escrever e imprimir textos via computador. Esses programas permitem que o autor do texto faça uma primeira escritura e a melhore, com facilidade, através de recursos mais ou menos sofisticados: troca de letras e/ou palavras, apagamento de letras, palavras e linhas, reorganização de parágrafos, inserção

de textos, paginação e distribuição do texto na página de forma a atender certas exigências convencionadas (formatação).

Segundo França (1985:64), o editor de texto permite que seu usuário "armazene informações em arquivos de seu micro e conserve essas informações, se for necessário. Os editores servem para criar programas, arquivos de dados e arquivos de texto". Profissionais como escritores, jornalistas, pesquisadores e professores, que precisam valer-se da palavra escrita com muita frequência, têm no editor de textos uma útil ferramenta de trabalho, visto que "ao escrever um texto na memória do computador, ao invés de fazê-lo no papel, o autor está livre das restrições lineares das canetas de escrever e pode "brincar" com um texto até estar inteiramente satisfeito. Só então ele solicita ao computador que imprima o texto no papel através de uma impressora nele acoplada (Jones & Fortescue, 1987:49)".

Nas áreas das humanidades - educação, literatura, lingüística, ciências sociais, jornalismo -, certamente uma das aplicações mais importantes dos computadores é a editoração de textos (artigos, livros, material instrucional). Grande parte dos profissionais dessas áreas já conhecem as virtudes da produção de textos via computador. O trabalho dos escritores, jornalistas, e acadêmicos tem-se tornando mais rápido e mais eficiente depois da adoção de editores de textos, sem contar com os benefícios que esses recursos têm trazido aos setores burocráticos (correspondência oficial).

Disso pode-se inferir que os editores de textos não só constituem um bem social cujo aperfeiçoamento será sempre importante, como também criam um mercado em crescimento que interessa a qualquer comunidade, gerando fontes de riqueza e de empregos.

"O mercado de editores de texto vive um de seus melhores momentos. A necessidade de atualização constante dos software tem levado produtores nacionais e representantes de produtos estrangeiros a lançar no mercado novos editores, cada vez mais sofisticados e com recursos de Desktop Publishing (Damasio, 1989:40)". O desempenho desse mercado pode ser medido pelo número de lançamentos das últimas feiras de informática e pela expectativa de alguns produtores em lançar continuamente novas versões.

Há, no entanto, que ressaltar os percalços que devem ser enfrentados na transposição do texto do papel para a memória do computador. Faz-se necessária a utilização de um programa específico para esse tipo de atividade, visto que o computador só responde a deter-

minados comandos quando lhe são fornecidos os passos exatos. Dessa forma, inserido no computador o programa desejado, ele responderá unicamente de acordo com a informação desejada. Caso sua atitude frente à solicitação não corresponde àquilo que se deseja, deve-se voltar aos passos do programa e verificar os possíveis erros cometidos pelo programador ou pelo digitador.

Um dos problemas enfrentados pelos editores de textos para a língua portuguesa são os grafemas acentuados (Poersch, 1987). A maior parte dos editores de textos produzem a acentuação mediante o acionamento de mais de uma tecla. A solução apontada por Poersch é o de elaborar um dispositivo que consiga incluir os grafemas acentuados mais frequentes no próprio teclado, eliminando caracteres não utilizados (ou pouco utilizados) na editoração de textos em língua portuguesa.

Um outro problema, já referido anteriormente, diz respeito à disposição dos caracteres gráficos nos teclados de microcomputadores. Duas pesquisas (Poersch, 1988b; Poersch, 1990) trouxeram subsídios lingüísticos e fisiológicos (ergométricos) para a reelaboração dos teclados clássicos, adaptando-os à produção de textos em língua portuguesa. Esse novo teclado (ARDOM) associa os caracteres mais frequentes às teclas mais facilmente acessadas.

6. EFICIÊNCIA DE TECLADOS

Os antigos fabricantes de máquinas de escrever estavam obcecados pelas semelhanças entre os teclados dessas máquinas com os teclados de piano. A maioria das máquinas apresentavam de oito a dez carreiras já que precisavam de teclas distintas para as letras maiúsculas antes da invenção da tecla de câmbio. A maioria desses primeiros teclados estavam ordenados numa estrita ordem alfabética.

O "layout" do teclado propriamente dito tem sido o centro de discórdia entre seus projetistas. Quando surgiram as primeiras máquinas de escrever, no século XIX, havia tantos tipos de teclados quanto eram os fabricantes; de modo geral, o mais usado era o de teclas agrupadas no centro do teclado.

Christopher Latham Sholes, o inventor americano responsável pela primeira produção de máquinas de escrever em 1872, descobriu que o arranjo alfabético produzia acasalamento de teclas quando utilizado por datilógrafos velozes. Sholes consultou seu irmão, professor, que apresentou a solução de separar as letras que apareciam juntas com maior frequência nas palavras. Criaram, então, o teclado QWERTY de nossos dias. O objetivo principal dessa configuração era para evitar o acasalamento das teclas. Acreditando na resistência do usuário a novas configurações, procurou apresentar razões científicas para esse teclado dizendo que o teclado QWERTY exigia o menor esforço possível no movimento das mãos, ao datilografar. Essa afirmação foi, posteriormente, colocada em choque; verificou-se que a datilografia de certas palavras exigia uma grande "caminhada" dos dedos pelo teclado. Ainda hoje em dia a maioria dos digitadores acredita nessa propaganda enganadora. Por outro lado, não há razão alguma para que um teclado eletrônico se preocupe com o problema do acasalamento e tenha forçosamente aquele layout, a não ser para continuar mantendo-se dentro de padrões convencionais – um exemplo típico de como um padrão global "de fato" pode tornar-se obsoleto e até indesejável e, ainda assim, continuar imutável (Grupo Educação e Cultura, 1984 b: 322).

Augusto Dvorak (Olson & Jasinski, 1986: 241), professor de estatística na Universidade de Washington, projetou o teclado que leva seu nome nos incios da década de 30. Dvorak interessava-se particularmente com a eficiência da velocidade e do movimento e com a redução de erros tipográficos e da fadiga.

Na carreira principal do teclado, Dvorak localizou as cinco vogais na esquerda e as cinco consoantes mais frequentes, na direita. A carreira superior continha as sete letras mais comuns seguintes, enquanto na carreira de baixo localizavam-se as letras menos usadas.

Dvorak afirmava que 70% da digitação processava-se na carreira principal e que 35% das palavras usadas em inglês podia ser digitada utilizando somente a carreira principal, sem quase nenhum movimento dos dedos. Os proponentes do sistema Dvorak sustentaram que os progressos em velocidade e exatidão variavam entre 35 a 50% para digitadores hábeis. Vários alunos de Dvorak venceram competições nas décadas de 30 e 40. A datilógrafa mais veloz do mundo, Bárbara Blackburn (170 palavras por minuto, de acordo com Guinness Book of World Records), usava o teclado Dvorak.

Outra afirmação sobre a superioridade do teclado Dvorak relaciona-se com a redução da "caminhada" dos dedos e, possivelmente, com a fadiga. Bárbara Blackburn afirmava que os dedos de um datilógrafo usando o teclado QWERTY por oito horas "viajavam" entre 15 e 16 milhas enquanto os dedos de um datilógrafo usando o teclado Dvorak "viajavam" somente uma milha.

A adoção do teclado Dvorak foi prejudicada por vários motivos; entre eles destacamos os seguintes: seu surgimento durante o período da Grande Depressão (1930), a padronização, por parte do governo, do teclado QWERTY durante a II Guerra Mundial, a posição privilegiada do teclado QWERTY no mundo dos negócios e, a última mas não a menos importante, a resistência dos usuários em se reciclar.

Norman & Gentner (1984: 66-72) juntamente com David Rumehart e outros colegas de Universidade da Califórnia, em San Diego, analisaram os movimentos dos dedos de datilógrafos. Tanto peritos quanto novatos digitavam em teclados conectados a computadores que, por sua vez, registravam o tempo de cada batida. Esses testes permitiram ao grupo construir um modelo computacional que simulava um perito digitador humano e estimava a rapidez de digitação baseado nas frequências dos movimentos exigidos pelos diferentes dedos. Baseado nesse modelo, os pesquisadores concluíram que, em termos de rapidez de digitação, o teclado Dvorak é realmente superior ao QWERTY, embora a diferença seja pequena (talvez 5 a 10% mais rápido).

Embora se verifique o aspecto altamente científico da construção do teclado ARDOM, baseado na lógica de designar as teclas mais facilmente acessadas para os caracteres mais frequentes, sabemos da resistência natural que todo usuário possui em relação ao novo, em relação a substituir padrões aprovados e enraizados na sociedade. Para se poder exigir para o novo teclado o apoio e a simpatia de engenheiros de fabricação e de usuários de teclados necessário se torna provar, antes de mais nada, a efetiva superioridade, em termos de eficiência e de economia de tempo, dessa nova configuração. Nada melhor para isso que a apresentação de dados empíricos, coletados através de um experimento, que comprovem uma significativa diferença entre o desempenho de digitadores que usem um ou outro "design".

7. A FREQUÊNCIA DOS CARACTERES GRÁFICOS EM LÍNGUA PORTUGUESA

7.1. Especificação do problema

7.1.1. Estabelecimento dos objetivos

O objetivo operacional básico (imediatos) é levantar, em textos escritos do português do Brasil, a frequência dos caracteres gráficos e dos dígrafos em posição inicial, medial e final de palavras.

Os objetivos aplicativos (mediatos) são os seguintes: sugerir mudanças no atual teclado padrão QWERTY de microcomputadores e de outras máquinas eletrônicas de escrever tomando em consideração a redução de esforço e o aumento de rapidez; fornecer dados confiáveis aos linguistas aplicados para a elaboração científica de cartilhas de alfabetização; reanalisar e reavaliar a afirmação de linguistas que pressupõem que o tipo de texto não constitui variável (interveniente) na frequência de grafemas e de dígrafos, visto estes não estarem relacionados com o significado; comparar os dados levantados no português com os dados encontrados em outros idiomas, notadamente no inglês, no francês e no alemão.

7.1.2. Formulação das hipóteses de trabalho

- 7.1.2.1. Os caracteres gráficos em textos de língua portuguesa apresentam percentagens de frequência diferentes. Esta hipótese será avaliada com base na listagem da frequência percentual dos caracteres.
- 7.1.2.2. O tipo e o tamanho das amostras não influem na distribuição da frequência dos caracteres gráficos. Esta hipótese será analisada com base no coeficiente de correlação calculada entre as ordens percentuais dos caracteres em textos diferentes;
- 7.1.2.3. Certos diagramas são mais produtivos do que outros. Esta variação será estudada em três posições (inicial, medial e final) e será apreciada com base no percentual das frequências obtidas.

7.1.3. Operacionalização das variáveis

A variável independente corresponde aos caracteres gráficos de textos em língua portuguesa, classificados como segue: caracteres grafêmicos, caracteres supragrafêmicos, caracteres intergrafêmicos, caracteres numéricos e outros caracteres.

Os caracteres grafêmicos formam dois grupos: consonantais e vocálicos. Os consonantais são: b, c, ç, d, f, g, h, j, k, l, m, n, p, q, r, s, t, w, v, x, y, z. Os vocálicos são: a, e, i, o, u. Os caracteres supragrafêmicos são: acento grave (`), acento agudo (´), acento circunflexo (^), til (~) e trema (¨). Os caracteres supragrafados são: á, à, â, ã, é, ê, í, ó, ô, õ, ú, û. Entre os caracteres intergrafêmicos, relacionam-se os seguintes: vírgula (,), ponto (.), ponto-e-vírgula (;), dois pontos (:), ponto de interjeição (!), ponto de exclamação (?), barra (/), abrir parênteses ((), hífen (-), aspas ("), apóstrofe ('), travessão (-) e ponto de reticência (...).

A variável dependente corresponde à frequência de ocorrência dos diversos tipos de caracteres. A frequência relativa servirá para ordená-los, isto é, fornecer-lhes um número de ordem (rango). A ordenação será analisada em diversos agrupamentos os quais serão selecionados de acordo com os objetivos propostos.

A variável interveniente corresponde ao tipo de texto, ao assunto tratado e ao tamanho da amostra. Para calcular a influência dessa variável interveniente, calcular-se-á o coeficiente de correlação dos dados de diversas amostras entre si, tomados dois a dois: tipos diferentes de discurso, assuntos diferenciados, texto curto (porém completo) X parte igual de um texto maior, amostra inicial, medial e final de textos, conjunto de textos curtos: provérbios e dizeres de pára-choques de caminhão.

A análise da variável interveniente tem por objetivo avaliar a hipótese (7.1.2.2) de que o tipo da amostra não influencia no resultado porque os caracteres gráficos independem de significado. Também existe a possibilidade de comparar esses dados com os dados obtidos por outros pesquisadores, principalmente, por aqueles que trabalharem com sistemas linguísticos diferentes do português.

7.2. Implementação da pesquisa

7.2.1. População e amostra

O universo da pesquisa inclui todos os textos escritos em língua portuguesa no Brasil. Baseado em pesquisas correlatas (Malmberg, 1971); Guiraud, 1959; Herdan, 1966; Miller, 1951), sabe-se que uma amostra suficientemente significativa está por volta de duzentas mil palavras de texto corrido; isso corresponde, aproximadamente, a trezentos e trinta páginas datilografadas em espaço simples, com sessenta colunas. Em termos de caracteres, corresponde a um milhão e duzentos mil.

Embora pesquisadores ligados a esse campo, baseados no pressuposto de que os grafemas e suas seqüências diádicas não estão relacionados com o significado, afirmem que não existem variáveis intervenientes a influir no resultado final das freqüências, pretende-se reavaliar essa afirmativa. Para tanto, a amostra global foi constituída de amostras parciais que cobriram os aspectos de: tipologia discursiva, estilo utilizado, assunto tratado, objetivo proposto e autor, entre outros.

A amostra global foi constituída de doze amostras parciais, totalizando 437.719 caracteres; portanto, um terço da amostra prevista acima. Se, no entanto, os coeficientes de correlação apresentados pelas freqüências das diversas amostras entre si, bem como pelas freqüências de cada uma com as freqüências totais, forem significativamente elevados, teremos comprovada a suficiência da amostra. No caso contrário, ter-se-á a necessidade de ampliar o corpus inicial.

As doze amostras parciais foram as seguintes:

- I - Discurso dissertativo correspondendo a um artigo científico completo;
- II - Discurso literário em verso;
- III - Discurso literário, em prosa, parte final de um texto;
- IV - Discurso literário, em prosa, parte inicial de um texto;
- V - Discurso literário, em prosa, parte medial de um texto;
- VI - Discurso dissertativo, parte de um texto técnico-científico;
- VII - Discurso literário, em prosa dirigido a leitores infantis;
- VIII - Discurso descritivo-expositivo, material instrucional (didático), sobre geografia;

IX - Discurso expositivo, amostras feitas com minitextos, abordando os mais variados assuntos;

X - Discurso narrativo-jornalístico, informativo;

XI - Texto técnico-científico (expositivo), sobre filosofia;

XII - Discurso expositivo de idéias, produção infantil.

7.2.2. Preparação do corpus e levantamento dos dados

O corpus total foi constituído de doze corpora para atender as explicitações da segunda hipótese (7.1.2.2). Esse corpus foi digitado num TK 3000 e os dados foram levantados automaticamente, primeiro para cada amostra parcial, depois para o corpus total. Para tanto foi utilizado um programa especialmente elaborado em linguagem C.

7.2.3. Apresentação dos resultados primários

Pela aplicação dos programas de contagem dos caracteres gráficos e dos digramas sobre os doze textos digitados (amostras) verificou-se que o número total de caracteres analisados foi de 437.719, distribuídos por diversos tipos, conforme se pode verificar na TABELA I.

TABELA I - Computação geral dos caracteres digitados nas doze amostras utilizadas.

Amostras	Total de caracteres	Grafemas	Inter-Grafemas	Algarismos	Outros caracteres	Espaços
I	76 143	55 529	2 183	731	2 965	16 327
II	40 051	29 993	937	0	972	8 495
III	39 108	28 757	1 510	10	1 630	8 135
IV	38 104	28 458	1 168	30	1 232	7 849
V	38 383	28 652	1 271	8	1 401	7 760
VI	20 212	14 889	741	451	1 327	3 698
VII	39 220	28 951	1 556	11	1 602	8 081
VIII	39 301	29 626	1 103	261	1 427	7 652
IX	38 792	27 913	1 477	16	1 599	8 703
X	38 642	29 765	1 160	263	1 443	7 832
XI	38 569	29 016	1 326	143	1 525	7 454
XII	11 743	8 699	308	27	424	2 445

A seguir relacionaremos os dígrafos mais frequentes, consideradas suas posições iniciais, mediais e finais. Esses dígrafos aparecem elencados em ordem decrescente de suas frequências.

TABELA II - Os trinta dígrafos iniciais mais frequentes.

Dígrafos	F	Dígrafos	F	Dígrafos	F
DE	4088	MA	1245	NÃ	801
CO	2937	PO	1209	IN	778
QU	2863	NO	1194	EM	765
SE	2473	RE	1142	SO	705
DO	1592	CA	1120	OS	654
ES	1481	PE	1033	EN	636
PA	1404	TE	989	AS	616
UM	1369	DI	919	VE	613
DA	1309	ME	900	FA	556
PR	1258	NA	875	SU	536

TABELA III - Os trinta dígrafos mediais mais frequentes.

Dígrafos	F	Dígrafos	F	Dígrafos	F
DE	3652	PO	926	DI	675
CO	2175	MA	905	OS	650
QU	2140	NO	898	IN	580
SE	2035	RE	888	AS	572
DO	1447	CA	766	EN	507
UM	1225	NA	765	FA	446
DA	1172	EM	743	SO	431
ES	1093	PE	730	SU	416
PA	1033	ME	701	VE	407
PR	990	TE	700	OU	401

TABELA IV - Os trinta dígrafos mais frequentes.

Dígrafos	F	Dígrafos	F	Dígrafos	F
DE	3783	TE	1459	OU	925
OS	3721	IA	1439	TA	888
DO	3306	SE	1433	UM	871
AS	3111	ES	1321	NA	837
AO	2641	AR	1205	RO	760
EM	2074	ER	1170	AL	741
UE	2058	MA	1089	CA	723
DA	1941	IS	988	MO	701
RA	1878	NO	936	AM	684
TO	1696	OR	929	EU	650

7.2.4. Tratamento estatístico

Os caracteres gráficos foram subdivididos em cinco categorias: caracteres grafêmicos, caracteres intergrafêmicos, caracteres numéricos, outros caracteres e espaços. Os caracteres supragrafêmicos não são computados nessa relação por se apresentarem acavalados com os caracteres grafêmicos (vocálicos). As frequências absolutas e percentuais desses caracteres gráficos se encontram na TABELA V.

TABELA V - Distribuição geral das frequências dos caracteres gráficos em Língua Portuguesa.

CARACTERES	FREQUÊNCIAS	
	ABSOLUTAS	PERCENTUAIS
Grafêmicos: Total	324.951	72,69
Consoantes	167.965	51,69
Vogais	156.986	48,31
Intergrafêmicos	14.317	3,20
Numéricos	1.719	0,38
Outros	16.151	3,62
Espaços	89.890	20,18
TOTAL	447.028	

Verifica-se que, entre os caracteres grafêmicos, as consoantes, embora tipicamente sejam mais numerosos do que as vogais (22 para 5), sua ocorrência total somente apresenta uma frequência de 51,69% contra 48,31% das vogais.

As tabelas a seguir (VI, VII e VIII) fornecem as frequências absolutas e percentuais, respectivamente, dos caracteres grafêmicos, dos caracteres supragrafêmicos e dos caracteres intergrafêmicos.

TABELA VI - Frequência dos caracteres grafêmicos no corpus total

IDENTIFICAÇÃO	F	%	POSTOS
A	44.400	13,66	1
B	3.290	1,01	19
C	11.968	3,68	12
Ç	1.819	0,55	21
D	16.700	5,14	8
E	40.446	12,45	2
F	3.500	1,08	18
G	4.321	1,33	16
H	3.208	0,99	20
I	21.808	6,71	5
J	732	0,22	24
K	48	0,01	26
L	9.209	2,83	13
M	14.612	4,50	10
N	17.481	5,38	7
O	35.867	11,04	3
P	8.498	2,61	14
Q	3.545	1,09	17
R	21.406	6,59	6
S	25.307	7,79	4
T	14.987	4,61	9
U	14.465	4,45	11
V	4.886	1,50	15
W	137	0,04	25
X	878	0,27	23
Y	32	0,01	27
Z	1.401	0,43	22

TABELA VII - Frequência dos caracteres supragrafêmicos no corpus total.

IDENTIFICAÇÃO	F	%	POSTOS
Á	1.334	13,96	3
Â	228	2,38	9
Ã	170	1,77	11
Ä	2.762	28,91	1
É	1.749	18,31	2
Ê	637	6,66	6
Ë	1.081	11,31	4
Ó	698	7,30	5
Ô	88	0,92	12
Õ	273	2,85	8
Ú	315	3,29	7
U	217	2,27	10

TABELA VIII - Frequência dos caracteres intergrafêmicos no corpus total

IDENTIFICAÇÃO	F	%	POSTOS
,	5.015	35,02	1
.	4.223	29,49	2
;	243	1,69	10
:	359	2,50	6
!	128	0,89	11
?	289	2,01	9
/	29	0,20	13
(337	2,35	7
)	394	2,75	5
-	2.022	14,12	3
"	316	2,20	8
⇒	25	0,17	14
~	813	5,67	4
...	124	0,86	12

As tabelas VI, VII e VIII bastam, por si só, para confirmar a primeira hipótese (7.1.2.1): os caracteres gráficos, em textos de língua portuguesa apresentam diferentes percentagens de frequência. Devido a essa diversidade de frequências, os caracteres podem ser relacionados em ordem decrescente. Entre os caracteres grafêmicos destacam-se os vocálicos pela sua alta frequência: A (13,66%), E (12,45%), O (11,04%), I (6,71%) e U (4,45%).

Os grafemas consonantais mais frequentes são: S (7,79%), R (6,59%), N (5,38%), D (5,14%), T (4,61%) e M (4,50%). Verifica-se que a frequência dos grafemas estrangeiros (K, W e Y) é inexpressiva: 0,05%. Convém salientar que os dez grafemas mais frequentes cobrem 73,37% do total das ocorrências e que os cinco mais frequentes correspondem a 51,65%, isto é, mais do que a metade de todas as ocorrências grafêmicas.

Retornando à TABELA VII, através das frequências percentuais e dos postos ocupados por cada grafema supragrafado, constata-se que (Ã) ocupa o primeiro posto com uma ocorrência de 28,91%. Contrariamente, o (Ó) ocupa um dos últimos postos com somente 2,85% de ocorrências. Outros grafemas acentuados a ocuparem altos postos na frequência são (É), (Á), (Í) e (Ó). Esses dados podem ser nitidamente visualizados no histograma da Fig. 1.

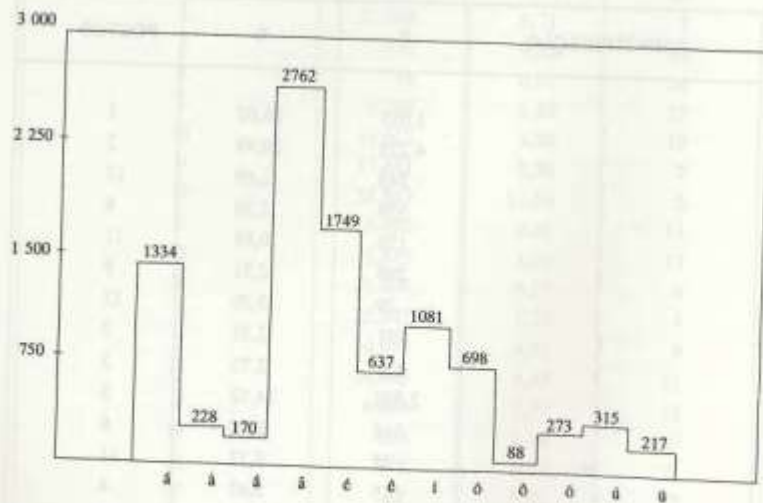


FIGURA 1 - Distribuição da frequência dos caracteres grafêmicos supragrafados.

Para avaliar a segunda hipótese (7.1.2.2), que pretende verificar a relação entre a distribuição das frequências e as amostras, calculou-se, inicialmente, a correlação simples entre as frequências dos caracteres grafêmicos e as amostras, duas a duas (TABELA IX).

TABELA IX - Coeficientes de correlação simples entre as frequências dos caracteres grafêmicos e as amostras, duas a duas.

Amostras	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	Total
I	-	.98	.97	.97	.98	.99	.97	.99	.94	.99	.99	.97	.99
II		-	.96	.97	.98	.98	.97	.98	.95	.98	.97	.99	.98
III			-	.99	.99	.97	.99	.98	.99	.98	.97	.97	.99
IV				-	.99	.97	.99	.99	.98	.99	.98	.97	.99
V					-	.98	.99	.99	.98	.99	.98	.97	.99
VI						-	.96	.99	.95	.99	.99	.97	.99
VII							-	.98	.98	.98	.97	.97	.98
VIII								-	.96	.99	.99	.97	.99
IX									-	.97	.96	.96	.97
X										-	.99	.97	.99
XI											-	.97	.99
XII												-	.98

Os altos coeficientes de correlação evidenciam que a tipologia textual não constitui variável interveniente na distribuição da frequência, o que queríamos confirmar. Isso, em outros termos, significa que a amostra global é suficientemente ampla, não necessitando de um corpus mais extenso. A significância desses dados fica abaixo do nível 0,01 visto que o valor crítico para esse nível é 0,48.

Se observarmos os coeficientes de correlação entre os caracteres intergrafêmicos e as diversas amostras (TABELA X) verificaremos que o comportamento não é o mesmo. Os coeficientes, embora oscilem entre moderados e altos, nos levam a acreditar que, nesses caracteres, a tipologia textual constitui variável que altera os resultados. Para examinarmos a relevância dessa variável, calculamos o qui-

quadrado calculado foi de 4.141,07. Considerando que o valor crítico de alfa, para o nível de significância de 0,05 é 0,532, conclui-se que existe diferença entre os textos no que concerne visualizado ao compararmos os números da ordem de frequência decrescente (postos), nas diversas amostras (Tabela X).

TABELA X - Coeficientes de correlação simples entre as frequências dos caracteres intergrafêmicos e as amostras, duas a duas.

Amostra	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	Total
I	-	.94	.71	.93	.84	.91	.89	.96	.83	.95	.94	.92	.95
II		-	.80	.98	.91	.91	.85	.97	.81	.97	.97	.98	.98
III			-	.84	.95	.55	.86	.68	.92	.69	.80	.76	.86
IV				-	.94	.87	.88	.94	.83	.96	.97	.96	.98
V					-	.71	.89	.83	.86	.84	.89	.87	.94
VI						-	.67	.96	.62	.95	.89	.93	.87
VII							-	.81	.83	.80	.91	.81	.92
VIII								-	.71	.98	.95	.96	.94
IX									-	.72	.85	.83	.88
X										-	.96	.97	.95
XI											-	.97	.98
XII												-	.96
Total													-

TABELA XI - Ordem decrescente das frequências dos caracteres intergrafêmicos, em cada amostra.

Caracteres	AMOSTRAS											
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
,	1	1	2	1	2	1	2	1	2	1	1	1
:	2	2	1	2	1	2	1	2	1	2	2	2
;	6	9	10	10	10	6	12	7	7	9	9	7
!	7	6	6	6	6	8	5	8	5	10	7	4
?	14	13	7	9	9	10	6	11	10	13	13	11
/	13	5	5	8	5	10	7	11	4	12	10	8
(10	10	14	14	12	10	13	10	6	11	13	12
)	5	7	12	11	12	4	10	5	8	6	6	5
.	4	8	13	12	12	3	11	4	9	7	5	6
*	3	3	4	3	4	5	3	3	3	3	3	3
→	9	13	9	5	8	7	8	11	11	4	4	9
-	12	11	11	13	11	10	13	9	13	8	11	10
...	8	4	3	4	3	9	4	6	14	5	8	12
...	11	12	8	7	7	10	9	11	12	13	12	12

TABELA XII - Frequência percentual dos caracteres intergrafêmicos em cada amostra.

Caracteres	Amostras											
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
,	34,12	45,78	20,66	36,90	27,69	46,42	24,80	50,31	29,11	47,5	38,26	50,0
:	20,93	28,81	38,07	26,79	29,18	16,46	31,23	22,39	56,73	22,58	30,09	32,46
;	3,98	0,74	0,33	0,59	0,55	3,50	0,19	1,45	0,60	0,94	0,60	1,62
!	3,66	2,13	2,98	3,59	3,85	0,67	3,98	0,63	1,15	0,68	2,11	1,94
?	-	-	2,64	1,28	1,65	-	3,02	-	0,27	-	-	0,32
/	0,09	2,56	5,09	1,62	5,19	-	2,63	-	3,65	0,17	0,30	0,64
(0,32	0,53	-	-	-	-	-	0,27	0,94	0,25	-	-
)	4,48	0,85	0,26	0,34	-	9,85	0,32	2,90	0,60	2,32	2,48	1,94
.	6,22	0,85	0,26	0,34	-	9,98	0,32	2,99	0,60	2,32	3,69	1,94
*	19,60	11,41	9,53	11,55	11,25	9,17	24,74	15,86	5,55	12,32	15,76	7,79
→	2,61	-	1,32	6,16	2,43	3,50	1,79	-	0,27	6,12	5,20	0,64
-	0,09	0,21	0,33	0,25	0,15	-	-	0,45	0,20	1,12	0,15	0,64
...	3,66	5,86	16,22	8,30	15,42	0,40	5,78	2,71	1,12	3,62	1,20	0,64
...	0,18	0,21	2,25	2,22	2,59	-	1,15	-	-	-	0,15	-

Finalmente, resta avaliar a última hipótese (7.1.2.3); ela se refere às frequências dos digramas. As tabelas V, VI e VII, que nos listam os digramas em ordem decrescente de ocorrência, confirmam a hipótese em referência. Vale a pena observar que o digrama DE ocupa o primeiro posto nas três posições. Verifica-se, outrossim, que os primeiros postos, tanto em posição inicial quanto medial, são ocupados pelos seguintes digramas: DE, CO, QU, SE, DO, ES, PA, UM, DA. Estes encontros são realmente os mais produtivos. Também chama a atenção o fato de que, entre os encontros consonantais, listam-se os grupos PR, TR, GR, CR, CH, BR, FR, LH, PL, CL, e que esses grupos ocupam postos semelhantes, um em relação ao outro, nas duas posições onde eles tem condições de aparecer: posição inicial e medial. Os únicos grafemas consonantais que aparecem em posição final são: L, S, M, R, N, Z.

7.3. Discussão dos resultados

Embora o objetivo imediato – objetivo satisfatoriamente atingido – não extrapole o plano meramente descritivo, existem diversas contribuições no plano teórico. Será, no entanto, o plano aplicativo que merecerá maior atenção em etapas subsequentes.

O objetivo imediato, vinculado à primeira e à terceira hipótese (7.1.2.1 e 7.1.2.3), foi atingido, como demonstra a análise estatística da distribuição de frequência. Tanto as frequências percentuais dos caracteres gráficos quanto dos digramas puderam ser devidamente ordenados.

A segunda hipótese – aquela relacionada com a estrutura sintática-semântico-estilística – fornecem dados suficientes para atingir o terceiro objetivo mediato. Chegou-se à conclusão de que o aspecto “estrutura textual” não constitui variável interveniente para alterar os dados gerais, no que se refere aos caracteres grafêmicos. Todos os coeficientes de correlação, calculados entre os diversos tipos de amostras, dois a dois, apresentaram-se muito fortes. As principais comparações feitas foram as seguintes:

- a – Texto Científico X Texto Literário (amostras I e VII);
- b – Discurso Dissertativo X Narrativo X Descritivo (amostras I, X e VIII);

- c – Texto Completo X Parte de Texto (amostras I e VI);
- d – Assuntos Diferentes: Lingüística, filosofia, história e geografia (amostras I, XI, III e VIII);
- e – Objetivo: informativo X didático (amostras X e VIII);
- f – Texto Completo X Texto Constituído de minitextos (amostras I e IX);
- g – Parte inicial, parte medial e parte final (amostras IV, V e III);
- h – Proza X Verso (amostras V e II);
- i – Produção adulta X produção infantil (amostras V e XII);
- j – Texto endereçado a adultos X texto endereçado a crianças (amostras V e VII).

No entanto, chama a atenção o fato de as correlações estabelecidas no plano dos caracteres intergrafêmicos não apresentarem comportamento similar. Foram verificadas correlações moderadas entre algumas amostras. A análise do qui-quadrado mostrou ser significativa a influência do tipo de amostra na distribuição da frequência desses caracteres. Faz-se, portanto, necessária uma melhor investigação desse campo. Talvez até se consiga encontrar, nos caracteres intergrafêmicos, determinantes capazes de discriminar diversas amostras entre si.

Por outro lado, os dados de frequência aqui levantados e computados deverão oportunizar a comparação com dados de frequência de outros idiomas, dados já amplamente investigados e divulgados para o inglês, o francês e o alemão, entre outros.

Os resultados finais permitem partir para outras investigações e cálculos com os quais poderão ser alcançados os objetivos médios: contribuir na solução de problemas relacionados com editores de texto e com a disposição de caracteres em teclados de microcomputadores e oferecer subsídios ao ensino da linguagem, na área da alfabetização, principalmente no que concerne à gradação de material a ser apresentado ao aprendiz.

Uma das tarefas centrais será a maneira de aplicar os resultados da frequência, em conjunção com a facilidade de acesso dos dedos às diferentes teclas, para um reordenamento dos teclados de máquinas eletrônicas de digitação. Além desse estudo de frequência, com o auxílio de um fisiólogo, deverá ser avaliada a prontidão de reflexos dos diferentes dedos da mão e do esforço (trabalho) exigido aos mesmos dedos para impulsionarem teclas diferentes daque-

las onde eles normalmente se posicionem. Os caracteres mais frequentes devem ocupar as teclas mais fáceis de serem acessadas; também devem ser tomadas em consideração as seqüências grafêmicas mais frequentes. A cada tecla deve ser fornecido um número de ordem segundo a rapidez com que puderem ser acessados. Essa rapidez dependerá da capacidade de resposta de cada dedo a um estímulo enviado pelo cérebro e da distância que as teclas se encontram dos dedos escolhidos para acioná-las. No final desse estudo, procurar-se-á uma correlação positiva perfeita entre a frequência de ocorrência e a facilidade de acesso. Os caracteres mais frequentes devem corresponder às teclas mais facilmente impulsionadas de modo a se obter o maior rendimento com o mínimo de custo.

Para a gradação do material de cartilhas de alfabetização, os dados levantados nesta pesquisa não poderão ser aplicados isoladamente. Far-se-á necessário um estudo de frequência vocabular, primeiro da língua portuguesa em geral e, depois dos vocábulos regionais. Estudos nesse sentido foram feitos, entre outros, por Sebastião (1983). Dos vocábulos selecionados privilegiar-se-ão aqueles que forem constituídos pelas letras e pelos encontros grafêmicos mais produtivos. Ainda merecerão atenção, nesta seleção de material, os problemas oriundos da correspondência entre grafemas e fonemas, amplamente descritos e analisados por Silva (1981) e Lemle (1982: 41-60), e os originados pela interferência de línguas em contato (Bisol, 1986: 71-92).

Uma segunda tarefa consistirá em equacionar a gradação do material de alfabetização com a produtividade dos diversos caracteres grafêmicos, isoladamente ou em conjunção com os digramas mais frequentes. Deverão ser privilegiados os grafemas mais produtivos, não ignorando, no entanto, outras variáveis intervenientes como a facilidade (gráfica) de produzi-los.

O produto final desta pesquisa servirá de sugestão e não de imposição. Como existe a lei da inércia no mundo físico, o conservadorismo exagerado tem dificuldades em promover o avanço científico e tecnológico no mundo cultural. Considerando, no entanto, que esta sugestão se estriba em dados científicos, é de se supor que a tecnologia, veja o alcance desta sugestão e dela faça uso num futuro não muito distante.

7.4. Conclusão

O objetivo operacional básico – levantar, em textos escritos no português do Brasil, a frequência dos caracteres gráficos e dos digramas em posição inicial, medial e final de palavras – foi satisfatoriamente atingido. As três hipóteses operacionais foram confirmadas.

A primeira – os caracteres gráficos em textos de língua portuguesa apresentam percentagens de frequência diferentes – foi avaliada à luz da frequência percentual e dos postos ocupados por cada um dos caracteres grafêmicos.

A segunda – a tipologia das amostras não influi na distribuição dos caracteres grafêmicos – foi avaliada e confirmada pelo exame dos coeficientes de correlação simples entre a frequência dos caracteres grafêmicos e as amostras, duas a duas.

A terceira – certos digramas são mais produtivos do que outros – foi analisada e confirmada através das listas de digramas, apresentados em ordem decrescente de sua frequência.

O atingimento dos objetivos aplicativos constituirá uma etapa posterior, uma investigação e um estudo aditado à presente pesquisa. Nesse estudo deverá receber atenção especial o primeiro desses objetivos: Sugerir mudanças no atual teclado padrão QWERTY de microcomputadores e de outras máquinas eletrônicas de processamento de textos (editoração). Para alcançar esse objetivo, os resultados aqui apontados deverão ser cotejados com levantamentos ergonômicos – reflexos dos diferentes dedos da mão e quantidade de trabalho exigido para acionar as diferentes teclas do teclado.

Por outro lado, deverão ser encontradas aplicações dos dados da frequência dos caracteres gráficos na elaboração científica de cartilhas de alfabetização.

Num terceiro momento, deverá ser feito um estudo comparativo dos dados aqui apresentados com os dados que outros pesquisadores já levantaram, na análise de outros idiomas.

Ao concluir, convém que se saliente o aspecto descritivo que acompanhou a primeira fase desta pesquisa – mero levantamento de dados. A segunda fase deverá privilegiar o aspecto aplicativo: teclados de microcomputadores e material de alfabetização.

8. DIFICULDADE DE ACESSAMENTO DAS TECLAS; POR UM NOVO TECLADO

Na etapa anterior analisou-se a frequência dos caracteres gráficos em textos de português. Levantamento semelhante há havia sido feito para outros idiomas, notadamente o inglês, o francês e o alemão. Teve-se também a oportunidade de reavaliar a afirmação de que o tipo de textos não influi na distribuição da frequência.

Os dados dessa primeira etapa serviram para reafirmar a contribuição que a lingüística pode emprestar aos cientistas da computação e aos técnicos de teclados eletrônicos, para oferecer subsídios para uma nova disposição dos caracteres gráficos em teclados nacionais e para implantar nos microcomputadores um dispositivo de comutar o teclado padrão com um teclado especial que favoreça a digitação de caracteres específicos, como acentuação, mediante o toque de uma única tecla. Desta maneira poderá ser obtida uma significativa economia de tempo e energia na editoração de textos em língua portuguesa.

O novo teclado aqui preconizado fundamenta-se num alto coeficiente de correlação entre a frequência dos caracteres gráficos e a facilidade de acessamento das teclas. É dessa afirmação que se conclui pela importância da segunda etapa desta pesquisa. Para se estabelecer a correlação entre as duas variáveis é preciso que ambas sejam ordenadas, de forma crescente ou decrescente de modo que os altos valores de uma correspondam aos altos valores da outra. Em outros termos: às teclas mais facilmente acessadas devem corresponder os caracteres mais frequentes.

Para ordenar as teclas segundo sua facilidade de acessamento, dois aspectos devem ser analisados: rapidez de reflexo dos diversos dedos no acessamento das teclas e distância das teclas dos dedos em sua posição de descanso. Disso decorre a importância de um estudo específico que registre o tempo médio que cada dedo leva para acionar as teclas que lhe estão designadas, uma vez recebido um estímulo externo ou interno.

De posse desse novo teclado, necessário se torna verificar se, praticamente, ele traz vantagens sobre aquele que pretende ser substituído. Para tanto é preciso projetar e executar um experimento com um grupo experimental e um grupo de controle aos quais são solici-

tadas atividades de treinamento. O grupo experimental receberia um treinamento em digitação utilizando o novo teclado; o grupo de controle seria treinado com o teclado QWERTY. A diferença de desempenho final, medido através de um pós-teste, diria qual dos dois teclados se mostra mais eficiente para a digitação de textos. Daqui decorre a relevância de executar não somente esta segunda etapa, mas também partir para uma terceira etapa.

8.1. Delimitação do problema

8.1.1. Estabelecimento dos Objetivos

Através do estudo da frequência dos caracteres gráficos em língua portuguesa e do estudo dos reflexos dos dedos na digitação e do esforço para acessar as diversas teclas, apresentam-se mudanças no atual teclado QWERTY de microcomputadores e de outras máquinas eletrônicas de escrever, a fim de veicular, na digitação, o máximo de informação com o mínimo de custo.

8.1.2. Formulação das hipóteses de trabalho

8.1.2.1. Hipótese central

Existe a possibilidade de dispor os caracteres gráficos num teclado de microcomputador conjugando os resultados da distribuição de frequência dos caracteres gráficos com os resultados da prontidão de acessamento às teclas pelos dedos correspondentes.

8.1.2.2. Primeira hipótese preliminar

Os caracteres gráficos, em textos de língua portuguesa, apresentam percentagens diferentes de frequência.

8.1.2.3. Segunda hipótese preliminar

Os dedos das mãos, para acessar as teclas de máquinas eletrônicas de escrever, apresentam graus diferentes de dificuldade.

8.1.3. Operacionalização das variáveis

8.1.3.1. Primeira variável independente: frequência dos caracteres gráficos.

Cada carácter gráfico, quer seja grafêmico, supragrafêmico ou intergrafêmico, ocupa um número de ordem decrescente segundo sua maior ou menor frequência em textos de língua portuguesa.

8.1.3.2. Segunda variável independente: prontidão de acessamento das teclas

Cada tecla recebe um número de ordem decrescente de acordo com o maior ou o menor grau de trabalho exigido para ser acessada pelo respectivo dedo da mão. Esse grau de dificuldade será calculado pela capacidade de reflexos de cada dedo e pela distância em que os dedos se encontram das teclas correspondentes.

8.1.3.3. Variável dependente: denominação das teclas.

A escolha dos caracteres gráficos que designarão cada tecla dependerá da ordem de facilidade de acessamento da tecla. Os caracteres mais frequentes corresponderão às teclas de mais fácil acesso.

8.2. Implementação da pesquisa

8.2.1. Dados aproveitáveis da etapa anterior

Os caracteres gráficos, em textos de língua portuguesa, apresentam percentagens diferentes de frequência. Essa hipótese foi confirmada na primeira etapa desta investigação. O número de ordem e as respectivas percentagens dos caracteres grafêmicos, dos caracteres supragrafêmicos e dos caracteres intergrafêmicos foram apresentados nas tabelas VI, VII e VIII (pág.).

8.2.2. Acessamento das teclas

A presente investigação avalia a hipótese de que os dedos das mãos, para acessar as teclas de máquinas eletrônicas de escrever, uti-

izam graus diferentes de trabalho (esforço). Para tanto foi necessário que se fizesse um levantamento dos reflexos exigidos por cada dedo para acionar cada tecla que lhe fosse designada: supõe-se que cada dedo responda diferentemente ao receber um estímulo para pressionar determinada tecla. Além disso foi necessário considerar a distância em que as teclas se encontram dos dedos designados para acioná-las.

Pensou-se inicialmente que essa investigação devesse ser solicitada a um fisiólogo ou a um psicólogo. Chegou-se, no entanto, à conclusão de que os dados necessários poderiam ser obtidos através de um programa específico em que sujeitos interagiriam com o teclado de um microcomputador. Esse programa preveria a simulação, no visor do micro, das duas mãos em posição de descanso, cada dedo se posicionando numa tecla, na segunda carreira de cima para baixo. Randomicamente, o programa indicaria, numa mão simulada, os dedos que deveriam acionar a tecla sobre a qual repousam. O sujeito, então, deveria pressionar a tecla correspondente. O programa registraria, automaticamente, o tempo levado para acessar a tecla, calcularia, outrossim, as médias para cada dedo. O programa também teria a capacidade de armazenar os dados parciais de cada sujeito, podendo, no final, fornecer as médias globais.

O algoritmo do programa de medição apresentou-se da seguinte maneira:

- 1 - Desenha a mão
- 2 - Pede parâmetros:
Número de teclado: NT
Nome do sujeito: NOME
Tempo médio entre teclas: TM
- 3 - Espera inicial
- 4 - Repete NT vezes:
Escolhe dedo (randomicamente)
Espera tempo TM
Verifica se teclou antes
Indica a tecla (dedo)
Inicia cronometragem
Espera resposta
Encerra a cronometragem
Verifica se foi usado o dedo correto
Computa estatísticas
Apaga a indicação da tecla
- 5 - Mostra as estatísticas
- 6 - Fim

Esta pesquisa sobre o tempo de reação de cada dedo orientou-se pelos seguintes parâmetros:

- a. A amostra foi constituída por 40 sujeitos distribuídos em quatro grupos: 10 adolescentes (entre 10 e 15 anos) com prática em digitação; 10 adolescentes sem prática; 10 adultos (mais de 20 anos) com prática; 10 adultos sem prática.
- b. Para a escolha aleatória dos dedos foi tomada a mesma semente para que fosse produzida a mesma seqüência para cada sujeito.
- c. Exigiu-se de cada sujeito um mínimo de 200 toques sendo que os vinte primeiros serviam para aquecimento ou familiarização com a rotina.
- d. De cada sujeito solicitou-se, além do nome, a data de nascimento e a informação de experiência progressiva em digitação. De cada um o computador fazia uma listagem completa do teste, número de toques válidos, tempo total e tempo médio para cada dedo.

O programa METRAT (medida de tempo de reflexo no acessamento das teclas), que se encontra disponível no CEPLIN/PUCRS, contém o seguinte manual de operação.

a. Ativação:

Com o equipamento a nível de sistema, digitar o nome do programa. Por exemplo:

A > METRAT (Enter)

ou

A > METRAT B: ARQUIVO.TST T(Enter)

Ao chamar o programa METRAT sem especificar o nome do arquivo, o mesmo vai ser exigido pelo programa como primeira providência, através do MENU principal. Ao chamá-lo com a especificação de arquivo, então existe uma opção adicional, ilustrada no exemplo acima com a letra T após o nome do arquivo, indicando o chamado ARQUIVO.TST, na unidade de disquete B deve ser truncado, isto é, os dados que eventualmente já contenha devem ser descartados. No caso de se omitir essa opção, os dados serão acrescentados ao final do arquivo (se estiver vazio, é claro que o início e o fim do arquivo coincidem).

b. Opções do Menu Principal:

Seleção do Arquivo
Aquisição dos Dados
Apresentação dos Resultados
Simulação do Teste
Resumo do (s) Teste(s)
Fim

Quando o programa está a nível de menu principal, então uma das ações estará em "bloco reverso". Ao teclar (Enter), essa será a ação executada. Mas ao teclar uma das teclas de setas, para opção, que pode ser ativada com (Enter).

c. Seleção do Arquivo de Dados

Tendo ou não sido especificado na ativação do programa, o nome e a forma de utilização do arquivo de dados podem ser alterados ou indicados. Um menu específico abre-se na parte de baixo da tela quando essa opção é ativada.

d. Aquisição dos dados

Ao ser ativada esta opção, um menu de identificação do sujeito se abre na tela, pedindo:

- O Nome do sujeito
- A Data de nascimento
- A Indicação de experiência em digitação ou datilografia (Sim ou Não)
- O Número de toques do experimento
- O Número de toques de "aquecimento" (treino)
- A Opção ou não por pré-indicação do dedo.

Esta última opção permite avaliar o efeito de se saber com antecedência qual o dedo que deve atuar.

Em seguida, o programa desenha duas mãos na tela, cada dedo identificado com a tecla sobre a qual deve posicionar-se. Mensagens de como deve ser feito, seguem. E, no fim da tela, a mensagem "Aper-te qualquer tecla para prosseguir", que aparece também em outros momentos, como por exemplo ao se constatar um erro, permite ao sujeito (re) concentrar-se antes de seguir adiante.

e. Apresentação dos resultados

Sob esta opção serão mostrados os valores calculados pelo programa.

f. Simulação do Teste

Esta opção refaz um teste específico, escolhido segundo um pequeno menu secundário. Repete, a partir dos dados adquiridos, todos os detalhes dessa, inclusive o cômputo de tempo, mostrando na tela a mesma seqüência de eventos.

g. Resumo do (s) Teste (s)

Tem o mesmo efeito da simulação, menos no aspecto de mostrar os eventos na tela. Permite recalcular rapidamente os valores estatísticos (totais e médias), de um ou mais sujeitos cumulativamente.

O resumo final dos dados de todos os exercícios dos 40 sujeitos encontram-se na TABELA XIII.

TABELA XIII - Dados finais da aplicação do programa METRAT

Discriminação dos dados	Número de toques	Tempo total	Tempo médio	Nº de Ordem
Mínimo esq.	865	57764	66.779191	1
Anular esq.	572	54108	94.594406	8
Médio esq.	717	83377	116.285914	10
Índic. esq.	996	85401	85.743976	5
Polegar esq.	1083	86012	79.420129	3
Polegar dir.	669	55479	82.928251	4
Índic. dir.	805	70040	87.006211	6
Medio dir.	905	90417	99.908287	9
Anular dir.	815	76095	93.368098	7
Mínimo dir.	884	63144	71.429864	2
Totais	8311	721837	86.853207	

A partir destes resultados, combinados com a distância em que as teclas se encontram dos dedos em posição de descanso foi elaborado o quadro abaixo (Figura 2). Nesse quadro deu-se a cada tecla um número de ordem crescente considerando o esforço exigido para o acessamento das mesmas.



FIGURA 2 - Numeração das teclas segundo a facilidade de acessamento das mesmas.

8.2.3. Elaboração do novo teclado

A hipótese central da presente investigação afirma que: existe a possibilidade de dispor os caracteres gráficos num teclado de micro-computador conjugando os resultados da distribuição de freqüência dos caracteres gráficos com os resultados da prontidão de acessamento às teclas pelos dedos correspondentes.

Analisando os teclados de máquinas digitadoras, verifica-se que a maioria dispõe de quarenta e seis teclas, totalizando uma disponibilidade de 92 caracteres, se computarmos tanto a caixa baixa (a) quanto a alta (b). A Figura 2, apresentada anteriormente, mostra essa possibilidade.

A partir desse número deve-se selecionar os caracteres que compõem o teclado para a utilização prioritária de editores de texto. As vinte e sete letras, minúsculas e maiúsculas já totalizam 54 caracteres. Dos caracteres supragrafêmicos selecionamos quatro vogais acentuadas para figurar nesta lista, devido a sua alta incidência: Â, Á, Ê, Í.

Acrescenta-se a isso os caracteres supragrafêmicos, separadamente: til, acento agudo, acento grave, acento circunflexo e trema. São mais treze. Dos catorze caracteres intergrafêmicos eliminamos a reticência por sua baixa produtividade e por poder ser obtida pela digitação de três pontos. Os caracteres numéricos, em número de dez reduzem-se a oito, pois o "um" e o "zero" podem ser obtidos, respectivamente pela letra L (minúscula) e O (maiuscula). Com isso totalizamos 88 caracteres. Sobram quatro. Para preencher essas quatro vagas, selecionamos, por sua importância, os sinais seguintes: &, %, * e +.

Os caracteres assim selecionados encontram-se na tabela XIV em ordem decrescente de sua freqüência, tanto bruta quanto percentual.

TABELA XIV - Caracteres ordenados segundo sua freqüência.

Nº de Ordem	Sinal	F	%	Nº de Ordem	Sinal	F	%
1	A	40.304	11,726	31		895	0.260
2	E	38.697	11.259	32	X	878	0.255
3	O	35.867	10.435	33	-	813	0.236
4	S	25.307	7.363	34	J	732	0.212
5	R	21.406	6.228	35)	394	0.114
6	I	20.727	6.030	36	:	359	0.104
7	N	17.481	5.086	37	(337	0.098
8	D	16.700	4.859	38	9	335	0.097
9	T	14.987	4.360	39	"	316	0.091
10	M	14.612	4.251	40	?	289	0.084
11	U	14.465	4.208	41	~	273	0.079
12	C	11.968	3.482	42	:	243	0.070
13	L	9.209	2.679	43	:	228	0.066
14	P	8.498	2.472	44	"	217	0.063
15	,	5.015	1.459	45	7	174	0.050
16	V	4.886	1.421	46	6	164	0.047
17	G	4.321	1.257	47	2	143	0.041
18	.	4.223	1.228	48	W	137	0.039
19	Q	3.545	1.031	49	!	128	0.037
20	F	3.500	1.018	50	3	101	0.029
21	B	3.290	0.957	51	8	97	0.028
22	H	3.208	0.933	52	5	97	0.028
23	-	2.762	0.803	53	4	76	0.022
24	-	2.022	0.588	54	K	48	0.013
25	Ç	1.819	0.529	55	&	33	0.0096
26	È	1.749	0.508	56	%	32	0.0093
27	Z	1.401	0.407	57	Y	32	0.0093
28	Á	1.334	0.388	58	/	29	0.0084
29	í	1.081	0.314	59	.	25	0.0072
30	⇒	1.013	0.294	60	*	12	0.0034
				61	+	-	-

A figura 3 apresenta o convencional e clássico teclado QWERTY, assim denominado por exibir, em sua segunda carreira de cima para baixo e da esquerda para a direita, teclas denominadas por essas seis letras. Esse teclado QWERTY somente apresenta uniformidade quanto à disposição dos caracteres grafêmicos e numéricos. Os caracteres supra e intergrafêmicos apresentam muita variação na distribuição. Apresentamos, no quadro abaixo, a disposição mais comum.



FIGURA 3 - Disposição mais comum do teclado QWERTY.

Para elaborar uma redistribuição dos caracteres gráficos nesse teclado, objetivo central do presente trabalho, procurou-se estabelecer uma correlação perfeita entre a freqüência de ocorrência dos caracteres gráficos e a facilidade de acessamento das diversas teclas. Alguns detalhes foram observados nessa redistribuição.

a - Os caracteres numéricos permaneceram nos lugares convencionais pelos seguintes motivos. Em primeiro lugar, a baixa freqüência desses caracteres em textos lingüísticos justifica plenamente a sua colocação na primeira carreira de cima para baixo, carreira que fica mais afastada da carreira de repouso. Em segundo lugar, a maior ou menor freqüência desses caracteres muito depende do texto utilizado. Assim, nos textos que serviram de corpus para o seu levantamento, verificamos que, devido aos dados bibliográficos (anos 19--), que os algarismos "1" e "9" aparecem com mais freqüência. Então, por motivos mnemotécnicos, achou-se conveniente manter a disposição clássica.

b - Alguns caracteres, como já foi assinalado anteriormente, não ocupam espaço especial por poderem ser produzidos ou substituídos por outros: o algarismo "um" e "zero" e o ponto de reticência. Essa economia é muito importante para poderem ser introduzidos outros sinais.

c - Para as três letras que não constam do alfabeto português e, por esse fato, apresentam uma frequência de ocorrência muito baixa, foram reservados as três últimas teclas da primeira carreira superior.

d - Achou-se importante reservar um espaço especial para as quatro vogais acentuadas mais frequentes: Â, Á, Ê, Í. Sabe-se que, normalmente, as vogais acentuadas são obtidas por, no mínimo, três toques.

Depois de tomadas essas providências preliminares, fizemos corresponder os caracteres mais frequentes às teclas mais facilmente acessadas. Com isso obteve-se o teclado apresentado na figura 4.

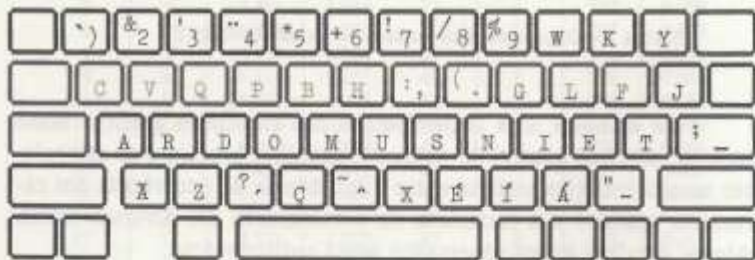


FIGURA 4 - Distribuição preliminar dos caracteres gráficos no teclado.

Essa distribuição preliminar foi submetida a algumas alterações para atender a certas conveniências.

a - Visto que a letra "Q", em português, vem seguida da vogal "U", achou-se mais prático deslocar o "Q" da posição 19 para 22 a fim de ser acessado com o mesmo dedo. As letras intermediárias entre 19 e 22 se reajustam segundo sua frequência:

19 = F 20 = B 21 = H

b - Na segunda carreira de cima para baixo, nas posições 15 e 18, onde se encontram, na caixa baixa, a vírgula e o ponto, convém, novamente por motivos mnemotécnicos, fazer corresponder, respectivamente, na caixa alta, o ponto e vírgula e os dois pontos.

c - Na quarta carreira, o "X" da tecla 32 passou para a tecla 30 reunindo, assim, o "Z", o "X" e o "Ç". O "Ê" e o "Í" passam para as teclas 28 e 24 para serem acessadas pelos mesmos dedos que acessam as letras "E" e "I". O "Á" passa para a tecla 29 formando trio com "Ê" e "Í". Os demais caracteres supragrafêmicos foram mantidos em caixa baixa porque isso diminui, de um toque, o esforço para a sua digitação.

d - As demais alterações são pouco significativas e precisaram ser efetuadas devido aos arranjos finais.

Com a introdução dessas alterações foi produzida a redistribuição final que vem apresentada na figura 5 e corresponde ao novo teclado, resultado e objetivo desta pesquisa.

O teclado, que resulta dessa redistribuição, foi batizado de ARDOM devido aos cinco primeiros caracteres que se encontram na carreira de descanso.



FIGURA 5 - Redistribuição final dos caracteres gráficos: teclado ARDOM.

8.2.4. Discussão dos resultados

Tínhamos colocado como primeira hipótese preliminar que "Os caracteres gráficos, em textos de língua portuguesa, apresentam percentagens diferentes de frequência". Essa hipótese foi confirmada na primeira etapa desta investigação, como atestam as Tabelas VI, VII e VIII (p.). A segunda hipótese preliminar, "Os dedos das mãos, para acessar as teclas de máquinas eletrônicas de escrever, utilizam graus diferentes de trabalho" também foi confirmada; essa confirmação nos permitiu enumerar as teclas segundo a rapidez de acesso das mesmas, conforme se verifica na figura 2 (p.).

Foi igualmente verificado que existe a possibilidade de dispor os caracteres gráficos num teclado de microcomputador conjugando os resultados da distribuição de frequência dos caracteres gráficos com os resultados da aplicação do programa METRAT (Medida de tempo de reflexo de acesso às teclas), isto é, a prontidão de acesso às teclas pelos dedos correspondentes. Essa correlação pode ser analisada na Tabela XV.

TABELA XV - Correspondência entre a ordem de frequência dos caracteres gráficos e a ordem de facilidade de acesso às teclas.

Nº de Ordem	Sinal	Tecla	Nº de Ordem	Sinal	Tecla
1	A	1	31		32a
2	E	2	32	X	30
3	O	3	33	-	26b
4	S	4	34	J	34
5	R	5	35)	31b
6	I	6	36	:	18b
7	N	7	37	(32b
8	D	8	38	9	40a
9	T	9	39	"	33b
10	M	10	40	?	37b
11	U	11	41	-	33a
12	C	12	42	:	15b
13	L	13	43		35a
14	P	14	44	*	35b
15	,	15a	45	7	38a
16	V	16	46	6	45a

Nº de Ordem	Sinal	Tecla	Nº de Ordem	Sinal	Tecla
17	G	17	47	2	39a
18	.	18a	48	W	36
19	Q	22	49	!	38b
20	F	19	50	3	42a
21	B	20	51	8	41a
22	H	21	52	5	44a
23	Ã	23	53	4	37a
24	-	26a	54	K	43
25	Ç	25	55	&	39b
26	Ê	24	56	%	40b
27	Z	27	57	Y	46
28	Á	29	58	/	41b
29	Í	28	59	,	42b
30	⇒	31a	60	*	44b
			61	+	45b

É assim, através do estudo da frequência dos caracteres gráficos em língua portuguesa, do estudo dos reflexos dos dedos na digitação e do esforço para acessar as diversas teclas, apresentamos mudanças no atual teclado QWERTY de microcomputadores e de outras máquinas eletrônicas de escrever, a fim de possibilitar, na digitação, o maior rendimento.

Não queremos afirmar, antecipadamente, que esse novo teclado produza, efetivamente, um rendimento maior do que o produzido pelo clássico teclado QWERTY. Temos, no entanto, razões de plausibilidade para que isso aconteça. Necessário se fará, numa terceira etapa, realizar uma pesquisa experimental que avalie esta última hipótese. Se ela for confirmada, teremos argumentos mais convincentes para sugerir, aos engenheiros de fabricação de teclado de microcomputadores ou de máquinas eletrônicas de escrever, a utilização de teclados que ostentam essa nova disposição.

Sugerimos, para a etapa seguinte, a realização de uma pesquisa experimental que procure avaliar a seguinte hipótese "o treinamento de sujeitos mediante o uso do teclado ARDOM conduz a um desempenho em digitação superior ao desempenho conseguido por sujeitos treinados com o teclado QWERTY.

9. EFICIÊNCIA DO TECLADO ARDOM

9.1. Definição do problema

Através da revisão bibliográfica foi possível constatar que a clássica configuração dos teclados (teclado QWERTY) não pode servir adequadamente nem a um digitador de textos da língua portuguesa – pelo fato de o teclado QWERTY não ter sido projetado segundo dados da língua portuguesa –, nem a digitadores em geral – pelo fato de os teclados de máquinas eletrônicas de escrever não serem afetados pelos problemas mecânicos de acavalamentos próprios de máquinas manuais e pelo fato de não ter sido tomado em consideração a prontidão de acessamento às teclas pelos diferentes dedos.

Portanto, existem razões lógicas para se pleitear plausibilidade para a afirmação de que um teclado configurado tecnicamente, isto é, no qual às teclas mais facilmente acessadas pelos dedos são atribuídos os caracteres mais produtivos da língua, permite um desempenho, na editoração de textos, mais eficiente do que os teclados que seguiram os outros critérios para seu "design".

9.1.1. Formulação da hipótese

Em condições normais e semelhantes de aprendizagem da digitação, sujeitos treinados mediante o uso do teclado ARDOM apresentam um desempenho final superior àqueles que utilizam o teclado QWERTY.

9.1.2. Operacionalização das variáveis

Com base na formulação da hipótese, a variável independente corresponde ao tipo de teclado utilizado no treinamento: teclado ARDOM ou teclado QWERTY. A variável dependente corresponde ao escore final de desempenho. O escore de eficiência desse desempenho corresponderá à razão entre o número de caracteres digitados corretamente e o tempo levado para a digitação. Portanto, os critérios de eficiência são velocidade e correção.

As variáveis intervenientes são aquelas que podem influir no resultado final; devem ser emparelhadas para que o experimento se realize dentro de "condições normais e semelhantes de aprendizagem da digitação". Entre outras, as mais importantes se referem às seguintes: semelhança na aparelhagem e no contexto (mesmo tipo de máquina, mesmo tamanho e configuração geral do teclado, mesmo ambiente físico e psicológico, mesmo monitor) e homogeneização da amostra.

9.2. População e amostra

Um levantamento preliminar em escolas de datilografia mostrou que uma idade média para o aprendizado da digitação está por volta dos doze anos. Escolheram-se, portanto, como população alvo do presente experimento, sujeitos com essa idade. Segundo informações estatísticas, são necessários dois grupos para o experimento: o grupo A utilizará o teclado ARDOM; o grupo Q trabalhará com o teclado QWERTY. Cada grupo deverá ser integrado por, aproximadamente, vinte sujeitos, igualmente emparelhados segundo o nível de idade/escolaridade, o sexo e a experiência pregressa. A influência de outras variáveis, mais ou menos imponderáveis, será reduzida pela comparação dos dados do pré e do pós-teste. O emparelhamento do nível idade/escolaridade será realizado tomando alunos de sexta-série cuja idade varie entre onze anos e meio e doze anos e meio. O emparelhamento da variável sexo se dará tomando um número igual de sujeitos de cada sexo. Quanto à variável "experiência pregressa", só participarão do experimento os sujeitos que ainda não tiveram contato com máquinas de escrever ou com computadores.

9.3. Seleção dos instrumentos

Serão necessários três instrumentos: um para a coleta dos dados primários (amostragem), outro para a avaliação do desempenho em digitação e mais outro para ministração do treinamento.

O instrumento que servirá para a realização da amostragem, constará de uma ficha que reunirá dados sobre os alunos inscritos para o Curso de digitação.

Para a avaliação do desempenho em digitação será desenvolvido um programa interativo (digitador/máquina) que apresentará ao sujeito um texto a ser digitado, que controlará o tempo utilizado na digitação e o número de erros cometidos, e que apresentará, para cada sujeito o resumo dos dados incluindo o escore final (razão entre o número de caracteres digitados corretamente e o tempo levado para a digitação).

O algoritmo de acompanhamento dos exercícios de treinamento em digitação consta de duas partes (a. estrutura dos dados, b. rotinas)

a) Estrutura dos dados

O curso de digitação, em qualquer de seus modos de teclado, baseia-se em um conjunto de textos a serem digitados em forma de exercícios. Esses textos estão organizados em uma hierarquia de módulos, lições e exercícios. Cada texto está registrado em um arquivo de nome 'TEXTOijk.ijk', onde os valores de i, j, k indicam, respectivamente, o módulo, a lição e o exercício a que esse texto corresponde. Isso permite, com um dígito para cada valor, uma gama de até 10 vezes 10 vezes 10 = mil exercícios. Por exemplo, TEXTO731.731 é o exercício de número 1, da lição número 3, do módulo número 7.

Além desses textos, que podem ser os mesmos para os dois grupos de pesquisa, existem ainda os arquivos dos alunos: cada sujeito terá registrados os dados de avaliação dos exercícios que fizer. Os dados relevantes, para cada sujeito, são: o grupo de pesquisa a que pertence, o último exercício já efetuado, e as contagens de palavras, toques, palavras erradas e toques errados, de cada um dos exercícios já efetuados. A estrutura desses dados, expressa na linguagem 'C', é a seguinte:

```
struct C
  char grupo;
  int  m_l_ex, /* modulo-lição-exercicio */
      , palavras [10][10][10]
      , toques   [10][10][10]
      , p_eros   [10][10][10]
      , t_eros   [10][10][10]
      , tempos   [10][10][10]
      ;
} arquivo;
```

Quando começa um exercício de um sujeito (ou a aula de digitação de seu curso, pois é disso que se trata na verdade), é aberto o arquivo correspondente, é atualizado o número da lição, e então, à medida em que ele digitar o texto da lição, este é conferido com o texto do arquivo correspondente. São contadas as palavras, os toques, os erros eventualmente cometidos, e o tempo total desse exercício. Todos esses dados são então passados ao arquivo do sujeito, e é apresentada a opção de passar ao próximo exercício ou encerrar a sessão.

b) Rotinas

Além de uma série de rotinas de apresentação na tela, de controles burocráticos do teste, detalhados em outra oportunidade, e das estruturas auxiliares de dados para isso usadas, os seguintes módulos são relevantes para mostrar a técnica de acompanhamento de cada exercício:

```
compara(x,y)
  char *x,*y; /* palavras a serem comparadas,
              uma do texto, e outra da digitação */
  {
    if( *x && *y ) /* se ambas as palavras têm conteúdo */
      if( *x == *y) /* compara primeiros elementos das
                    palavras */
        return (1+compara(x+1,y+1));
        /* se iguais, compara os restantes
           das palavras, contando */
      else return (max(compara(x,y+1), compara(x+1,y)));
        /* se diferentes, procura o maior
           número de caracteres iguais que
           possam ser achados */
    else return 0; /* se terminou uma das palavras, nada
                  pode ser contado como igual */
  }
```

A separação de palavras no texto oferecido e no texto digitado pelo sujeito é muito semelhante. A diferença principal é que no caso do texto, ele está todo disponível de imediato, a partir do arquivo correspondente, ao passo que o texto digitado é recebido do teclado durante o exercício. Por isso, duas rotinas distintas devem tratar esses casos. São chamadas de 'palavra' e 'word', respectivamente. Só a primeira será mostrada neste relatório, para fins de concisão.


```

palavra( )
{
    static char pal[25] /* espaço de memória para a palavra
        a ser separada */
        , *q; /* ponteiro para dentro desse espaço,
            onde serão colocadas as letras */
    q = pal; /* início da palavra */
    if( *p = '\32' ) { /* fim do texto? */
        tem = 0; /* indica que terminou o exercício */
        return NULL;
    }
    while(!isspace(*p) && *p != '\32')
        /* enquanto for parte de palavra */
        *(q++) = *(p++); /* copia para 'pal' */
    while(isspace(*p) && *p != '\32') p++;
        /* ignora brancos entre palavras */
    *q = '\0'; /* fim da palavra */
    return pal;
}

```

Todo o trabalho de examinar o texto digitado pelo sujeito, em comparação com o texto fornecido para o exercício é feito por uma rotina estanque, relativamente simples:

```

caminha_texto( )
{
    char *x, *y; /* ponteiros para as palavras achadas */
    long h; /* para guardar a hora inicial */
    int m,l,ex /* módulo, lição e exercício atuais */
        , err /* número de toques errados em palavra */
        , pal = 0, toq = 0, p_err = 0, t_err = 0, i;
        /* contagens para este exercício */
    p = texto; /* ponteiro para o início do texto dado */
    q = respo; /* " para área do texto-resposta */
    clr( ); /* limpa a tela */
    for(i=0;i<tam_lin; ++i) printf(" "); printf("\n");
        /* escreve régua de tamanho-da-linha */
    h = hora( ) /* início do exercício */
    m = atual / 100;
    l = (atual % 100) / 10;
    ex = atual % 10;
    while(tem) {

```

```

x = palavra( );
y = word( );
pal ++; /* conta a palavra */
toq += 1 + strlen(x) /* acumul, comprimento */
t_err += (err = strlen(x) - compara(x,y));
if(err) p_err ++; /* conta palavra errada */
}
arquivo.palavras [m][l][ex] = pal;
arquivo.toques [m][l][ex] = toq;
arquivo.p_errores [m][l][ex] = p_err;
arquivo.t_errores [m][l][ex] = t_err;
arquivo.tempos [m][l][ex] = (int) ((hora( )-h)/100L);

```

O principal esforço burocrático do algoritmo é feito pelo módulo principal 'main()', mas esses detalhes só serão relevantes no momento do relatório final, para dar suporte técnico à reprodução dos exercícios fora do ambiente original do projeto. Os passos relevantes reduzem-se à chamada de duas rotinas, a última das quais exhibe os resultados do exercício:

```

caminha_texto( );
mostra_desempenho( );

```

9.4. Tratamento dos dados; avaliação da hipótese

Para fins de avaliação da hipótese serão utilizados os dados fornecidos pelos testes de desempenho dos grupos. Esses dados, depois de devidamente levantados e computados, serão transcritos numa tabela que exibirá o número de caracteres corretos, o tempo utilizado e o escore para cada sujeito.

A configuração geral do experimento assumirá o seguinte esquema:

GA	T1 ----->	A ----->	T2
GQ	T3 ----->	Q ----->	T4

DGA = T2 - T1
DGR = T4 - T3

- Nesse esquema, GA = grupo ao qual é ministrado treinamento com o teclado ARDOM
- GQ = grupo ao qual é ministrado treinamento com o teclado QWERTY
- A = treinamento de noventa horas com o teclado ARDOM
- Q = treinamento de noventa horas com o teclado QWERTY
- > = tempo que passa
- T1 = pré-teste do GA
- T2 = pós-teste do GA
- T3 = pré-teste do GQ
- T4 = pós-teste do GQ
- D = Desempenho

A hipótese será avaliada com base na comparação dos desempenhos de cada grupo. O desempenho corresponde à diferença entre o pós e o pré-teste. A hipótese será confirmada na seguinte condição:

$$DGA > DGQ$$

O tratamento estatístico ainda analisará a covariância do escore total, estabelecerá o ajuste das médias e calculará o coeficiente t de Student para averiguar o nível de significância dos resultados.

9.5. Conclusão

Como tivemos oportunidade de verificar anteriormente, nenhum argumento convincente possuímos, além de plausibilidade decorrente do aspecto altamente técnico na elaboração do novo teclado, para postular uma maior eficiência dessa nova disposição. Menos ainda nos queremos iludir sobre a aceitação tanto pelas empresas fabricantes de teclados quanto pela comunidade consumidora em geral. A aceitação de alterações em padrões já consagrados pelo uso e pelo tempo demanda aspectos altamente convincentes sobre as vantagens dessas mudanças. Será somente através de dados empíri-

cos, fornecidos mediante um experimento, que provaremos, junto aos cientistas e engenheiros da computação, a eficácia do novo teclado e a conseqüente conveniência e até a necessidade de processar as modificações sugeridas.

REFERÊNCIAS BIBLIOGRÁFICAS

01. BISOL, Leda. Interferência de uma segunda língua na aprendizagem da escrita. In: Tasca & Poersch, *Suportes Lingüísticos para a alfabetização*. Porto Alegre, Sagra, 1986, p. 71-92.
02. CHERRY, Colin. A comunicação humana. São Paulo, Cultrix, 1971.
03. DEMASIO, Esther. Novidades fazem mercado crescer. *Info, a revista brasileira de informática*. Rio de Janeiro, Ed.IB, 7(81):40-51, out. 1989.
04. DVORAK, August. Is there a better typewriter keyboard? *National Business Education Quarterly*, 1943.
05. DVORAK, A., MERRICK, N. L., DEALEY, W. L., and FORD, G. C. *Typewriting behavior*. New York, American Book Company, 1936.
06. FRANÇA, Paulo Bianchi. *Introdução aos microcomputadores*. Rio de Janeiro, Livros Técnicos e Científicos, 1985.
07. GENTNER, D. R. and NORMAN, D. A. The typist touch. *Psychology Today*, March 1984, pages 66-72.
08. GRUPO EDUCAÇÃO E CULTURA. "O texto perfeito". *Software*, Rio de Janeiro, Rio Gráfica, 1984a.
09. GRUPO EDUCAÇÃO E CULTURA. "Problemas no teclado". *Chips & bytes* Rio de Janeiro, Rio Gráfica, 1984b.
10. GUIRAUD, PIERRE. *Problèmes et methodes de la statistique linguistique*. Dordrecht, D. Reibel Publishing Company, 1959.
11. HERDAN, GUSTAV. *The advanced theory of language as choice and chance*. Heidelberg, Springer-Verlag, 1966.
12. JONES, Christopher & FORTESCUE, Sue. *Using computers in the language classroom*. Londres, Longman, 1987.
13. LEMLE, Miriam. A tarefa da alfabetização: etapas e problemas do português. *Letras de Hoje*, Porto Alegre, PUCRS, (50):41-60, dez. 1982.
14. MALMBERG, Bertil. *As novas tendências da lingüística*. São Paulo Nacional, 1971.
15. MILLER, George. *Language and communication*. New York, McGraw-Hill Company, 1951.
16. OLSON, Donald W. and JASINSKI, Laurie E. Keyboard efficiency. *Byte*, Feb 1986, p. 241-244.
17. PARGHALY, Ali. A model for intelligent computer assisted language instruction (MICALI). *Computers and the Humanities*. Dordrecht, Kluwer Academic Publishers, 23 (3):235-250, jun. 1989.
18. POERSCH, José Marcelino. *Lingüística Computacional: Elaboração do DIPLOMATA para a língua portuguesa*. *Letras de Hoje*, Porto Alegre, PUCRS, 22 (1) 97-108, março de 1987.

19. _____. Redistribuição dos caracteres gráficos no teclado de microcomputadores com base na lingüística quantitativa. *Actas do IV Encontro da Associação Portuguesa de Lingüística*. Lisboa, Associação Portuguesa de Lingüística, 1988a. p. 249-80.
20. POERSCH, José Marcelino. Freqüência dos caracteres gráficos de microcomputadores com base na lingüística quantitativa. *Anais do XI Encontro Nacional de Lingüística*. Rio de Janeiro, PUCRJ 1988b. p. 145-160.
21. POERSCH, José Marcelino e BECKER, Bertilo F. ARDOM, uma nova configuração para teclados de microcomputadores. *Veritas*, 35(138):279-304, jun. 1990.
22. SILVA, Myriam Barbosa. *Letura, ortografia e fonologia*. São Paulo, Ática, 1981.
23. TREMBLAY, Jean-Paul & BUNT, Richard B. *Ciências dos Computadores; uma abordagem algorítmica*. São Paulo, McGraw-Hill do Brasil, 1983.
24. VOTRE, Sebastião Josué. *Um léxico para cartilha*. Rio de Janeiro, Universidade Gama Filho, MEC/INEP, 1983.
25. ZIPF, G. K. *Human behavior and the principle of least effort*. Cambridge (Mass.), Addison-Wesley Publishing Company, 1949.