

Geração Automática de Abreviaturas para Sistema de Comunicação Alternativa do Português Europeu

Luís Garcia
LabSI/ESTIG-IPBeja
Rua Afonso III, 1, Beja
luisbgarcia@estig.ipbeja.pt

Luís Oliveira
INESC-ID/IST
R. Alves Redol, 9, Lisboa
lco@inesc-id.pt

Sumário

A expansão de abreviaturas é uma técnica que pode ser utilizada por pessoas com necessidades especiais para a aceleração da escrita. Neste estudo analisámos o desempenho de esquemas de abreviaturas para o Português Europeu gerados de forma totalmente automática. Na geração dos esquemas de abreviaturas foram utilizados corpus de origens distintas: (1) Textos de um aluno com necessidades especiais (CorpA); (2) Textos de manuais escolares da Porto Editora (CorpPE); (3) Textos do jornal “O Público” (CorpPub), e analisadas duas técnicas de abreviação - a contracção e a truncagem. Foram realizadas simulações de escrita com os vários esquemas de abreviaturas obtendo-se um máximo de 18% para a taxa de poupança de caracteres com um esquema de 500 abreviaturas. Com apenas 250 abreviatura esta taxa baixou para 16,2%. A introdução de bigramas e trigramas nos esquemas de abreviaturas não influenciou significativamente o desempenho. Os esquemas de abreviaturas gerados através da truncagem apresentaram regra geral os melhores resultados.

Palavras-chave

Expansão de Abreviaturas, Aceleração da Escrita, Pessoas com Necessidades Especiais, Comunicação Aumentativa e Alternativa, Processamento de Língua Natural, Interfaces com Língua Natural

1. INTRODUÇÃO

A American Speech-Language-Hearing Association (ASHA) define Comunicação Aumentativa e Alternativa (CAA) como uma área de prática clínica que tenta compensar, de forma temporária ou permanente, incapacidades de comunicação por parte de pessoas com dificuldades de expressão [Sevcik97].

Para o apoio a esta actividade têm vindo a ser desenvolvidos sistemas que apoiam a formulação, transmissão, recepção e interpretação de mensagens. Estes sistemas apresentam características distintas, como o nível de tecnologia, símbolos utilizados (e.g. imagens, pictogramas, letras) ou área de aplicação (e.g. conversação, escrita).

O sistema “Eugénio – O Génio das Palavras” (Figura 1) foi desenvolvido pelo Laboratório de Sistemas Interactivos da Escola Superior de Tecnologia e Gestão de Beja (LabSI/ESTIG), Instituto de Engenharia e Sistemas de Computadores/Instituto Superior Técnico (INESC/IST) e o Centro de Paralisia Cerebral de Beja (CPCB) para o apoio a pessoas com dificuldades de comunicação, quer ao nível da escrita, quer ao nível da fala. Este sistema possui uma página internet (<http://www.l2f.inesc-id.pt/~lco/eugenio>) a partir da qual pode ser obtido de forma gratuita.

O Eugénio é um agente de software que funciona no ambiente Microsoft Windows para sugerir palavras que completem o texto que está a ser editado. O Eugénio analisa a vizinhança do cursor e sugere um número configu-

rável de palavras que, na sua opinião, são mais relevantes no contexto. Caso a palavra pretendida se encontre entre as palavras sugeridas o utilizador pode completá-la automaticamente, evitando assim algum esforço adicional e possíveis erros de escrita. Para utilizadores com dificuldades de acesso ao teclado do computador o Eugénio dispõe de vários teclados de ecrã que permitem a escrita com recurso ao rato. O sistema também funciona em conjunto com o sintetizador de fala DIXI+, permitindo ao utilizador ouvir o texto que escreveu, ou transmitir uma mensagem a outra pessoa. Os vários componentes da

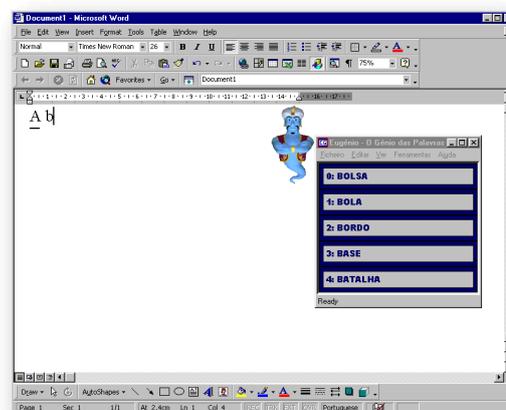


Figura 1 – O sistema “Eugénio – O Génio das Palavras”

interface do sistema podem ser adaptados às necessidades particulares de cada pessoa [Garcia03].

Estudos têm no entanto revelado que a carga sensorial e cognitiva imposta ao utilizador, na pesquisa da lista de palavras previstas, pode comprometer a aceleração da escrita. Num destes estudos [Venkatagiri94] utilizando tamanhos de 5, 10 e 15 palavras, para a lista de palavras previstas, verificou que aumentando o tamanho desta lista aumentava também a taxa de poupança de caracteres, pois as palavras eram propostas mais cedo ao utilizador, no entanto os valores obtidos para as taxas de comunicação mantiveram-se aproximadamente iguais nas três situações. Com a lista de 15 palavras foi obtida uma maior taxa de poupança de caracteres mas o tempo que o utilizador demorou a pesquisar as palavras pretendidas na lista contrabalançou o tempo poupado na escrita dos caracteres, não havendo por isso melhorias na taxa de comunicação. Estes resultados mostram o impacto que o tempo de pesquisa na lista de palavras previstas pode ter na performance do sistema. Estudos desenvolvidos por Koester e Levine (1994, 1996) revelam também este problema.

Para utilizadores mais rápidos, que não conseguem beneficiar da predição de palavras, existem técnicas de aceleração da escrita mais adequadas, que se baseiam na codificação de mensagens, como por exemplo a expansão de abreviaturas. Nos sistemas que disponibilizam esta técnica cada palavra ou frase possui uma abreviatura única. Após a escrita da abreviatura esta é expandida na palavra ou frase correspondente. A expansão de abreviaturas é apontada com uma das técnicas que potencialmente permite uma maior taxa de comunicação no entanto requer do utilizador uma grande capacidade de memorização.

No nosso trabalho de aconselhamento de tecnologias de apoio no Centro de Paralisia Cerebral de Beja temos também vindo a constatar que a predição de palavras não é adequada para utentes com uma razoável velocidade de escrita no computador. Desta forma decidimos analisar a utilização de abreviaturas na escrita do Português Europeu. Este estudo foi desenvolvido para um dos utentes que temos vindo a acompanhar, com o intuito de estender depois os resultados obtidos aos outros utentes em condições semelhantes. Trata-se de um aluno integrado no ensino regular que acompanhamos desde o 1º ano do 1º ciclo de escolaridade. É uma criança com paralisia cerebral, que apresenta alguns compromissos ao nível motor, e que frequentava no momento do estudo o 7º ano do ensino regular (actualmente frequenta o 8º ano). A nível cognitivo não revela quaisquer dificuldades manifestando por vezes algum cansaço no desenvolvimento de actividades intelectuais mais exigentes, como a realização de testes de avaliação. Para apoio à locomoção utiliza um andarilho e uma cadeira de rodas eléctrica. Ao nível da manipulação revela algumas dificuldades na motricidade fina que comprometem a realização de diversas actividades escolares. Por esse motivo utiliza um computador portátil como meio de apoio à realização destas actividades entre as quais a escrita.

2. EXPANSÃO DE ABREVIATURAS

Como referimos na expansão de abreviaturas cada palavra ou frase possui uma abreviatura única. Após a escrita da abreviatura esta é expandida na palavra ou frase correspondente, quer por iniciativa do sistema quer por ordem expressa do utilizador. Na primeira situação o sistema evita que a pessoa tenha de pressionar uma tecla adicional para a expansão da abreviatura, no entanto este método pode provocar a ocorrência de expansões indesejadas, quando o início da palavra que o utilizador se encontra a escrever coincide com uma abreviatura existente no sistema. Nestas ocasiões diz-se que houve uma colisão. Os sistemas que expandem as abreviaturas por iniciativa própria resolvem este problema disponibilizando ao utilizador uma tecla para desfazer as expansões indesejadas. Na segunda situação o utilizador necessita de pressionar uma tecla predefinida sempre que queira expandir uma abreviatura, não havendo assim, em relação à forma anterior, uma economia tão grande nas teclas pressionadas. No entanto apresenta a vantagem de colocar a decisão de expansão sob o controlo do utilizador.

As abreviaturas podem encontrar-se predefinidas no sistema ou então ser definidas pelo próprio utilizador. Para o Português Europeu não existem estudos conhecidos que analisem o desempenho de conjuntos de abreviaturas para utilizadores destes sistemas. O Eugénio, apesar de permitir a utilização desta técnica, também não incorpora qualquer conjunto de abreviaturas predefinido. É por isso um dos objectivos deste estudo a identificação de um conjunto ou conjuntos de abreviaturas que possa ser disponibilizado com o Eugénio.

2.1 Selecção do Vocabulário

A criação de um conjunto de abreviaturas requer que numa primeira fase sejam seleccionadas as palavras a abreviar. A selecção do conjunto de vocabulário importante para uma pessoa em particular não é tarefa fácil. Num estudo que analisa a capacidade dos técnicos em preverem qual o vocabulário necessário em duas actividades de lazer distintas, vela e café-internet, verificou-se que de todas as palavras previstas apenas 68% foram realmente utilizadas numa das actividades (Vela) e 47% na outra (Café-internet) [Leigha07]. Estes resultados estão em sintonia com outros estudos semelhantes [Balandin98].

Se as palavras mais importantes para o utilizador não forem correctamente identificadas o esquema de abreviaturas resultante não ficará adaptado ao utilizador e por isso prestará um apoio muito reduzido na aceleração da escrita.

Uma alternativa à selecção do vocabulário por parte dos técnicos consiste na identificação do conjunto de palavras mais importantes através do processamento automático de textos relacionados com o utilizador. Neste estudo pretendemos seguir esta abordagem e analisar o desempenho do esquema de abreviaturas gerado. Pretendemos ainda determinar as diferenças no desempenho resultantes da

utilização de diferentes esquemas de abreviação e diferentes corpora.

2.2 Atribuição de Abreviaturas

Após a selecção do conjunto de vocabulário é necessário proceder à atribuição de abreviaturas a cada palavra ou frase de modo a maximizar o desempenho do utilizador na escrita. Desta forma é conveniente a atribuição de abreviaturas mais pequenas às palavras ou frases com maior potencial de poupança de caracteres na escrita. Este potencial de poupança está relacionado com a taxa de utilização e com a quantidade de caracteres que compõem a palavra ou frase.

Existem várias formas através das quais podem ser criadas abreviaturas sendo a truncagem e a contracção as mais comuns [Bailey95]. Através da truncagem a abreviatura é formada com base nas primeiras letras da palavra, desta forma a abreviatura da palavra *com* poderia ser *c*. Através da contracção a abreviatura é formada retirando-se da palavra todas as vogais. Assim, para a mesma palavra a abreviatura seria *cm*. Para a abreviatura de frases podem ser utilizadas as iniciais das várias palavras que a compõem [Cook95], como por exemplo *BD* para *Bom Dia*.

3. GERAÇÃO AUTOMÁTICA DE ABREVIATURAS

Na realização deste estudo foi numa primeira fase seleccionada corpora adequada para em seguida se proceder à geração automática dos vários esquemas de abreviaturas. Por fim foi analisado o desempenho de cada um destes.

3.1 Selecção de Corpora

Para a realização deste estudo foram utilizados três corpus de textos: (1) Textos de um aluno com necessidades especiais (*CorpA*); (2) Textos de manuais escolares da Porto Editora (6º ano) (*CorpPE*); (3) Textos do jornal “O Público” (*CorpPub*). Do total dos corpus 90 % foi utilizado para treino (geração das abreviaturas) e 10% para testes (medição do desempenho das abreviaturas). Para uma caracterização destes corpus apresenta-se no Gráfico 1 os resultados da contagem das palavras (unigramas-UG), pares de palavras (bigramas-BG) e trios de palavras (trigramas-TG) existentes em cada um destes. É fácil verificar pelo gráfico a diferença de dimensão do *CorpA* em relação aos outros dois corpus.

3.2 Geração dos Esquemas de Abreviaturas

Para a geração dos vários esquemas de abreviaturas combinámos os diferentes valores das seguintes variáveis: (1) Corpus de treino (*CorpA*, *CorpPE*, *CorpPub*); (2) Técnicas de abreviação utilizadas (contracção, truncagem); (3) Estratégia de abreviação (apenas unigramas, unigramas+bigramas, unigramas+bigramas+trigramas). Resulta-

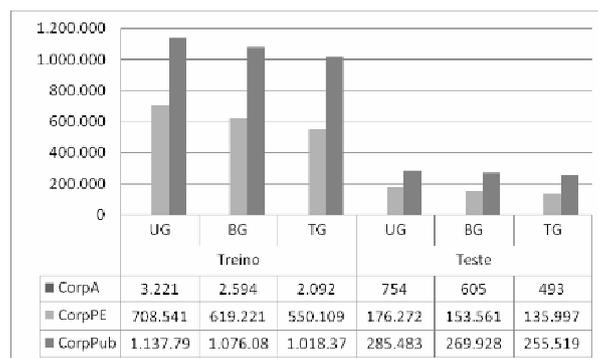


Gráfico 1 – Caracterização da Corpora

ram desta combinação 18 esquemas de abreviaturas diferentes.

Como referimos anteriormente é conveniente a atribuição de abreviaturas mais pequenas às palavras ou sequências de palavras com maior potencial de poupança de caracteres na escrita. Este potencial de poupança não depende unicamente da frequência de utilização de uma palavra ou sequência de palavras mas também da quantidade de caracteres poupados em cada utilização, que neste caso depende do tamanho da abreviatura atribuída.

Definimos como potencial de poupança de uma palavra a abreviar o produto da sua frequência de utilização num corpus pela quantidade máxima de caracteres poupados em cada utilização, ou seja, a diferença entre a quantidade de caracteres da palavra e a quantidade de caracteres da abreviatura mais pequena que lhe possa ser atribuída (Eq.1).

$$pp(pal) = f(pal) \times [comp(pal) - comp(ab_{min}(pal))] \quad (Eq.1)$$

Onde $f(pal)$ é a frequência da palavra num corpus, $comp(pal)$ o comprimento da palavra (quantidade de caracteres) e $ab_{min}(pal)$ a abreviatura mais pequena que possa ser atribuída àquela palavra.

No caso particular da abreviação por truncagem a Eq.1 apresenta a forma da Eq.2, dado que através desta técnica qualquer palavra pode obter uma abreviatura com apenas 1 caracter.

$$pp(pal) = f(pal) \times [comp(pal) - 1] \quad (Eq.2)$$

Definimos ainda como potencial de poupança de uma sequência de palavras a abreviar, $pal_1, pal_2, \dots, pal_n$, o somatório dos potenciais de poupança de cada uma das palavras, $pp(pal_1), pp(pal_2), \dots, pp(pal_n)$, acrescido de $n-1$, devido aos espaços poupados entre as palavras (Eq.3).

$$pp(pal_1, pal_2, \dots, pal_n) = \sum pp(pal_i) + n - 1 \quad (Eq.3)$$

Numa primeira fase do processo de geração das abreviaturas foi efectuado o levantamento dos unigramas, bigramas e trigramas existentes em cada corpus. Em seguida, para cada corpus, foram criados 3 conjuntos com estes elementos (apenas unigramas, unigramas + bigramas, unigramas + bigramas + trigramas) resultando 9 conjuntos diferentes. De seguida, utilizando a Eq.3, foi calculado o potencial de poupança de cada um dos elementos contidos nestes conjuntos, quer para a truncagem, quer para a contracção, donde resultaram 18 conjuntos diferentes. Finalmente procedeu-se à atribuição de abreviaturas de forma prioritária aos elementos com maior potencial de poupança. No caso da truncagem este procedimento assegurava as abreviatura mais pequenas para as palavras com maior potencial de poupança. Quando uma abreviatura já se encontrava ocupada por um elemento prioritário era analisada a possibilidade de atribuição da abreviatura com mais uma letra da palavra. No caso da contracção este procedimento garantia que as palavras com maior potencial de poupança teriam mais possibilidade de ser abreviadas do que outras com menor potencial. Quando uma abreviatura já se encontrava ocupada por um elemento prioritário não era possível a atribuição de outra abreviatura pois estas são geradas através da remoção das vogais da palavra, existindo por isso apenas uma possibilidade para cada palavra.

Após a criação dos esquemas de abreviaturas analisámos a sua constituição. Nos esquemas de abreviaturas associados ao corpus *CorpA* foi privilegiada, sempre que era possível, a abreviatura de sequências de palavras (bigramas e trigramas). Tal pode-se explicar por ser um corpus específico e por esse motivo certos bigramas ou trigramas surgirem com alguma frequência. Os esquemas de abreviaturas gerados com os corpus *CorpPE* e *CorpPub* apresentaram uma tendência diferente pois a abreviação de sequências maiores de palavras foi em muitos casos preterida em relação às sequências menores. Tal deve estar relacionado com a dimensão dos corpus que favorece uma taxa de repetição elevada dos unigramas.

3.3 Desempenho dos Esquemas de Abreviaturas

Para a análise do desempenho dos esquemas de abreviaturas foi desenvolvido um simulador para a escrita dos textos de teste com cada um destes esquemas de abreviaturas. Na escrita simulada dos textos assumiu-se que um utilizador tenta expandir por ordem de preferência trigramas, depois bigramas e finalmente unigramas. Se não fosse possível a expansão de uma abreviatura então este utilizador virtual tinha de escrever a palavra e o espaço necessário. No caso de expansão de uma abreviatura considerámos que o sistema introduzia automaticamente um espaço no final da palavra ou palavras. Não considerámos a escrita de espaços adicionais, sinais de pontuação, ou outras formatações do texto, pois o nosso objectivo consistia apenas em medir o desempenho das abreviaturas na escrita de texto. Também não foi considerado no cálculo do desempenho o facto de alguns caracteres requererem a

pressão de duas ou mais teclas pois alguns teclados para pessoas com necessidades especiais contêm por exemplo teclas com as letras acentuadas de modo a evitar a pressão de duas teclas. Desta forma os resultados são independentes do tipo de teclado.

4. RESULTADOS

A maior percentagem de caracteres poupados foi obtida com os esquemas de abreviaturas do corpus *CorpA* baseados na truncagem (Gráfico 2). O esquema do corpus *CorpA* que continha apenas abreviaturas de unigramas permitiu a poupança de 17,7% dos caracteres, o esquema que continha também abreviaturas de bigramas possibilitou a poupança de 18,0% e o esquema com abreviaturas de unigramas, bigramas e trigramas 17,0%. As taxas de poupança dos esquemas gerados com os corpus *CorpPE* e *CorpPub* foram menores e situaram-se entre os 10% e 11%. Acharmos que esta diferença de resultados se deve ao facto do corpus *CorpA* ser de menor dimensão e apresentar uma menor variabilidade de vocabulário, pois tratam-se de textos criados por um mesmo aluno em contexto escolar, o que permite um melhor comportamento dos esquemas de abreviaturas na escrita dos textos de teste. Verificámos no entanto que os esquemas que continham bigramas e trigramas não introduziram melhorias significativas na taxa de poupança de caracteres o que nos leva a questionar sobre a utilidade de esquemas de abreviaturas com estes elementos, que numa primeira impressão devem acarretar maiores custos de memorização quando comparados com esquemas de abreviaturas compostos apenas por unigramas.

Os esquemas de abreviaturas gerados através da contracção proporcionaram taxas de poupança de caracteres menores dos que os gerados com a truncagem (Gráfico 3). As taxas maiores continuaram a ser obtidas com os esquemas de abreviaturas gerados com o corpus *CorpA*, com um máximo de 10,8% obtido com o esquema que continha apenas unigramas. As taxas de poupança dos esquemas gerados com os corpus *CorpPE* e *CorpPub* situaram-se entre os 6,5% e 6,7%. Já indicamos anteriormente os factos que nos parecem explicar estas diferenças em relação ao corpus *CorpA*. Quanto à diferença na taxa de poupança entre os esquemas gerados com a truncagem

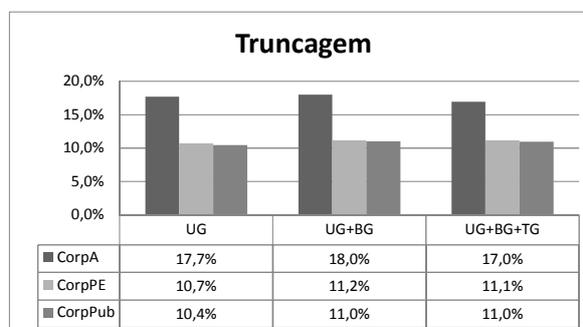


Gráfico 2 – Taxas de poupança de caracteres com esquemas de abreviaturas baseados na truncagem

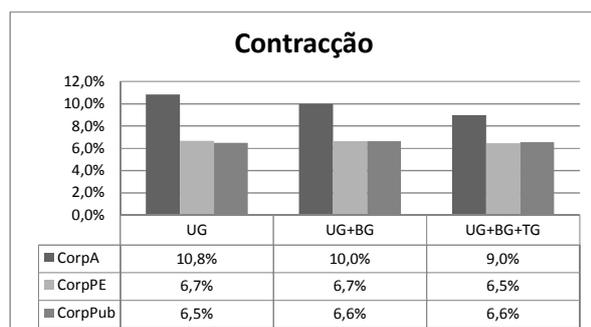


Gráfico 3 – Taxas de poupança de caracteres com esquemas de abreviaturas baseados na contração

e a contração a justificação é simples, a primeira técnica permite a geração de abreviaturas pequenas, em alguns casos de apenas 1 letra para as palavras com maior potencial de poupança, enquanto que na contração não existe esta possibilidade de optimização pois as abreviaturas são geradas através da remoção das vogais das palavras, não ficando as palavras com maior potencial necessariamente com as abreviaturas mais pequenas. Garante-se no entanto que em caso de colisão a palavra com maior potencial de poupança obtém a abreviatura.

Analisámos também as taxas de expansão proporcionadas pelos vários esquemas de abreviaturas na escrita dos textos (quantidade de expansões/quantidade de palavras do texto). A maior taxa de expansão foi obtida com o esquema de abreviaturas do corpus *CorpA* baseados em unigramas que foram 32,2% e 34,9%, para a truncagem e contração respectivamente. As taxas de expansão dos esquemas homólogos dos corpus *CorpPE* e *CorpPub* foram menores, e entre si muito semelhantes, situando-se entre 20,7% e 25,2%. Relativamente aos esquemas de abreviaturas que continham unigramas e bigramas verificámos que estes últimos foram muito pouco utilizados em relação aos unigramas. Nestes esquemas as taxas de expansão de unigramas variaram entre 18,9% e 20,5% para a truncagem e 12,1% e 13,7% para a contração. A taxa máxima de expansão de bigramas obtida foi 8,2%. De assinalar que mesmo nos esquemas de abreviaturas gerados com o corpus *CorpA*, onde os bigramas foram bastante valorizados em termos de potencial de poupança veio a verificar-se que com os textos teste acabaram por não ser muito úteis. Tal deve-se ao facto dos bigramas, trigramas apresentarem uma grande variabilidade. No que respeita aos esquemas de abreviaturas que também continham trigramas a tendência foi semelhante, estes foram muito pouco utilizados nas expansões, ainda menos do que os bigramas, com um máximo de taxa de expansão de 2,7%.

Decidimos ainda testar a adaptabilidade dos esquemas de abreviaturas produzidos. Para tal realizámos novas simulações mas agora os esquemas de abreviaturas eram utilizados na escrita dos textos teste de outros corpus. Estas medidas de desempenho forneceram-nos uma estimativa de como os esquemas de abreviaturas se comportariam com textos completamente diferentes daqueles com os quais

foram gerados. Este aspecto é importante se quisermos fornecer um conjunto de abreviaturas genérico para utilizadores de um sistema de comunicação alternativa. Os esquemas gerados com o corpus *CorpA* apresentaram os piores resultados a este nível, com um máximo de 4,4% e um mínimo de 1,4% nas taxas de poupança de caracteres quando testado com textos de outra corpora. Para os esquemas de abreviaturas gerados com o *CorpPub* e *CorpPE* estas taxas situaram-se respectivamente entre 8,0% - 3,7% e 8,3% - 3,7% mostrando uma maior versatilidade em relação aos esquemas do corpus *CorpA*.

Os testes realizados utilizaram esquemas constituídos por 500 abreviaturas que consideramos ser um número muito elevado para poder ser memorizado por crianças em idade escolar. Realizámos por isso novos testes de desempenho com esquemas de 250 abreviaturas. Constatámos que esta redução não influenciou proporcionalmente o desempenho dos esquemas de abreviaturas. De um máximo de taxa de poupança de 18,0% com um esquema de abreviaturas composto por unigramas e bigramas do corpus *CorpA* passámos para um máximo de 16,2% com o mesmo esquema. Em todos os esquemas as quebras de desempenho rondaram sempre os 2% quer para a truncagem quer para a contração. Estes resultados indicam-nos que é viável a redução do número de abreviaturas dos vários esquemas de modo a facilitar a sua memorização aos utilizadores.

5. DISCUSSÃO DOS RESULTADOS E IMPLICAÇÕES CLÍNICAS

Os resultados obtidos nas taxas de poupança de caracteres na escrita dos textos teste ficaram aquém das expectativas iniciais. Na literatura são referidos desempenhos de 31,0% nas taxas de poupança de um sistema de expansão de palavras para a língua inglesa [Higginbotham92] contra o máximo de 18,0% conseguido neste estudo. No esquema de abreviaturas do sistema analisado as 50 palavras mais frequentes obtiveram abreviaturas com 1 carácter, certas palavras raiz obtiveram abreviaturas com 2 caracteres, sendo necessário mais 1 carácter para a expansão de uma das palavras derivadas. Na construção das abreviaturas também foram utilizados algarismos o que aumenta a quantidade de abreviaturas com menores dimensões mas que pode dificultar a memorização por parte do utilizador. Estas optimizações juntamente com as diferenças da língua poderão explicar a diferença nos resultados obtidos. Assim será importante verificar para o Português Europeu se a criação de esquemas de abreviaturas recorrendo a algarismos e palavras raiz permitirão a obtenção de desempenhos semelhantes.

Os resultados obtidos não são totalmente satisfatórios e por isso deverão ser utilizados com cuidado na prática clínica. O esforço necessário para o domínio de um esquema de abreviaturas é considerável e por isso os técnicos envolvidos devem certificar-se que o investimento realizado compensará o utilizador. De assinalar também a perturbação que pode surgir em termos de memorização se após a adopção de um esquema de abreviaturas for

aconselhada uma nova alternativa, mais eficaz, ao utilizador. Por este motivo não aconselhamos ao utente referido no início do estudo qualquer um dos esquemas de abreviaturas gerados, pois pensamos proceder a novas optimizações que possibilitem maiores taxas de poupança de caracteres. Em todo o caso pensamos que os esquemas com 250 abreviaturas de unigramas por truncagem gerados com os corpus *CorpPub* ou *CorpPE* poderiam constituir uma solução. Apesar da taxa de poupança máxima ser inferior à obtida para os esquemas gerados com o corpus *CorpA* (15,3% contra 8,1% e 8,0%) os primeiros mostraram uma maior adaptabilidade a vários tipos de textos apresentando diminuições nesta taxa da ordem dos 2%, enquanto que os esquemas do *CorpA* apresentaram nestas situações de adaptabilidade taxas máximas de poupança de caracteres de 4,4%. Desta forma seria extremamente arriscado o seu aconselhamento para uma solução de comunicação alternativa. Provavelmente o esquema gerado com o corpus *CorpPE* seria a alternativa mais acertada por conter vocabulário de manuais escolares e por isso encontrar-se mais próximo do utente (criança em idade escolar). No entanto questionámos o utente sobre qual o conjunto de abreviaturas que preferiria de entre os que continham apenas abreviaturas de unigramas gerados através da contracção e truncagem a partir dos corpus *CorpPub* ou *CorpPE* (4 esquemas de abreviaturas) e numa primeira opinião mostrou preferência pelo esquema de abreviaturas de unigramas por truncagem do corpus *CorpPub* (a nossa segunda escolha).

6. CONCLUSÕES E TRABALHO FUTURO

Como referimos os resultados obtidos para a taxa de poupança de caracteres com os vários esquemas de abreviaturas ficaram aquém do esperado. Obteve-se um máximo de 18% de taxa de poupança de caracteres com um esquema de 500 abreviaturas gerado com um corpus de textos de um aluno com necessidades especiais (*CorpA*). Reduzindo o número de abreviaturas para 250 obteve-se uma taxa de poupança de caracteres de 16,2%. No entanto os esquemas de abreviaturas gerados com este corpus mostraram ter pouca adaptabilidade quando testados com textos de outros corpus. Na escrita de textos de outros corpus a taxa de poupança desceu dramaticamente para 4,4% mostrando a sua fragilidade em novas situações. A introdução de bigramas e trigramas nos esquemas de abreviaturas não teve influência significativa na taxa de poupança de caracteres pelo que pode não compensar o acréscimo de complexidade introduzido no esquema de abreviaturas. Consideramos que os esquemas com 250 abreviaturas de unigramas por truncagem gerados com os corpus *CorpPub* ou *CorpPE* poderão ser utilizados numa solução de comunicação aumentativa e alternativa desde que sejam consideradas as implicações clínicas referidas na secção anterior. Apesar da taxa de poupança máxima ser inferior à obtida para os esquemas gerados com o corpus *CorpA* (15,3% contra 8,1% e 8,0%) os primeiros mostraram uma maior adaptabilidade a vários tipos de textos apresentando diminuições nesta taxa da ordem dos

2%, enquanto que os esquemas do *CorpA* apresentaram nestas situações de adaptabilidade taxas máximas de poupança de caracteres de 4,4%. Os esquemas de abreviaturas gerados através da contracção apresentaram regra geral sempre os melhores resultados.

Como trabalho futuro pretendemos produzir optimizações que produzam ganhos nas taxas de poupança de caracteres. Também pretendemos analisar melhor a influência do número de abreviaturas do esquema no seu desempenho. Um outro ponto a estudar consiste na análise da facilidade de aprendizagem dos vários esquemas de abreviaturas propostos.

7. REFERÊNCIAS

- [Bailey95] Bailey, Robert W. Human Performance Engineering. 3ª ed. Prentice-Hall. New Jersey. 1995, pp. 504.
- [Balandin98] Balandin, Susan, Teresa Iacono. A Few Well-Chosen Words. *Augmentative and Alternative Communication*, Volume 14, Issue 3 September 1998, pages 147 – 161.
- [Cook95], Cook, Albert M. e Susan M. Hussey. *Assistive Technologies: Principles And Practice*. Mosby. Missouri. 1995, pp. 489.
- [Garcia03] Garcia, Luis. *Concepção, Implementação e Teste de um Sistema de Apoio à Comunicação Aumentativa e Alternativa para o Português Europeu*. Tese de Mestrado. Instituto Superior Técnico. 2003.
- [Higginbotham92] Higginbotham, D. Jeffery. *Evaluation of Keystroke Savings across Five Assistive Communication Technologies*. *AAC Augmentative Alternative Communication*. 1992.
- [Koester94] Koester, Heidi Horstmann, Levine, Simon P. *Learning and Performance of Able-Bodied Individuals Using Scanning Systems with and without Word Prediction*. *Assistive Technologies* 1994. RESNA. 1994.
- [Koester96] Koester, Heidi Horstmann, Levine, Simon P. *Effect of a Word Prediction Feature on User Performance*. *AAC Augmentative and Alternative Communication* 1996. ISAAC. Setembro, 1996.
- [Leigha07] Leigha, Dark, Susan Balandin, *Prediction and Selection of Vocabulary for Two Leisure Activities*. *Augmentative and Alternative Communication*, Volume 23, Issue 4 2007, pp. 288 – 299.
- [Sevcik97] Sevcik, Rose A. e MaryAnn Ronski. *ACC: More Than Three Decades of Growth and Development*. *American Speech-Language-Hearing Association (ASHA)*. 1997.
- [Venkatagiri94] Venkatagiri, Horabail S. *Effect of Window Size on Rate of Communication in a Lexical Prediction AAC System*. *AAC Augmentative Alternative Communication*. 1994.