

**Projecto para a Análise e Classificação das Ocupações**  
**NORMAS de TRANSCRIÇÃO de DOCUMENTOS em COMPUTADOR**  
versão 2.0  
Junho/Outubro de 1998

**PACO é o acrónimo do Projecto para a Análise e Classificação das Ocupações. Este é um projecto multidisciplinar no domínio das ciências humanas (antropologia, economia, história, linguística e sociologia), financiado pelo Programa Praxis XXI e pelo Ministério do Emprego e Formação Profissional, que teve início em Março de 1998 e terá a duração de dois anos.**

**O desenvolvimento do projecto está a cargo de três equipas de investigadores: uma no ISCTE, que é também a instituição coordenadora, outra na FLUP e outra ainda no ILTEC.**

**Este projecto pretende desenvolver um estudo sobre a evolução da divisão social do trabalho em Portugal e sobre os sistemas de classificação das ocupações, com base em documentação arquivística e em fontes impressas, produzidas entre o séc. XIV e 1940. O levantamento destes documentos permitirá criar fundos informativos de suporte à investigação, armazenados em bases de dados textuais e relacionais. Por outro lado, o projecto procurará elaborar uma proposta para a classificação das ocupações, em sectores relacionados com a actividade agrícola e a indústria, nomeadamente, a agricultura, as pescas e os têxteis.**

**Neste artigo apresentam-se dois instrumentos de trabalho relacionados com o tratamento dos documentos arquivísticos que integram o corpus textual do projecto.**

**Para a documentação circular livremente em vários ambientes e para ela ser reconhecida por diversos utilizadores há que partir de um conjunto de escolhas comuns. No entanto, a uniformização de regras e procedimentos não pode ser tão inflexível que redunde em perda de informação e comprometa as diferenças próprias do trabalho interpretativo. Seleccionar poucas regras e prever soluções opcionais para casos de variabilidade foi a estratégia seguida para assegurar maior versatilidade.**

**Metodologicamente, consideram-se dois planos distintos: o primeiro diz respeito à informação normalizada em campos fixos, cuja finalidade é a de funcionar como um conjunto de apontadores que facilitam o rastreio rápido dos dados. Data, local, referência bibliográfica ou arquivística e identificação prosopográfica dos indivíduos são exemplos de itens onde se pode conseguir grande eficácia através da manipulação de campos fixos com informação estandardizada tanto ortograficamente como no próprio formato. O segundo plano refere-se à transcrição de texto - em formato mais ou menos livre - e visa apreender os detalhes linguísticos do documento escrito.**

## **1. Campos pré-definidos**

**É sabido que os documentos-fonte produzidos por instituições tendem a apresentar formas mais fixas de dados, nos quais se reproduz uma estrutura ordenada ou, pelo menos, se repetem certas categorias de informação. Mas mesmo nas fontes dispersas e com representações mais variáveis é possível e recomendável isolar certas componentes em campos pré-definidos de modo a gerir**

eficazmente a informação. Por isso se aconselha que seja qual for o modelo de bases de dados (no projecto PACO são bases de dados relacionais e bases de dados textuais em formato livre) se reservem os primeiros campos de informação para armazenar referências espaço-temporais e arquivísticas, utilizando um vocabulário normalizado e um formato fixo. Estes campos não pretendem substituir a transcrição integral de documentos ou de partes de documentos, com todas as suas variantes e especificidades, mas sim criar duplicados capazes de facilitarem as tarefas de rastreio, seriação, escolha, relacionamento e classificação de dados.

### 1.1. Referências cronológicas

É frequente encontrar datações rigorosas dos documentos, susceptíveis de serem representadas num campo de tipo data com dez caracteres: DD/MM/AAAA. Trata-se de um formato actualmente disponível na maior parte dos produtos de *software* e pode ser facilmente reconvertido para outras formas de referência temporal. A grande desvantagem advém do facto de as facilidades fornecidas pela maioria dos sistemas de Gestão de Bases de Dados, no que diz respeito aos calendários internos - operações aritméticas, identificação de anos bisextos e validação do numero de dias por mês, apenas funcionarem para as referências cronológicas do século XX. Um outro problema é o da existência, ou coexistência, de datações rigorosas com datações sistematicamente vagas e lacunares, devido a: (i) ausências parcelares de informação sobre determinado dia e/ou mês; (ii) indicações só do conjunto mês-ano , ou apenas de um determinado ciclo do ano (na Quaresma, pelo S. João, no Inverno, no primeiro semestre); (iii) ausência completa de referências temporais no documento , que podem ser ultrapassadas mediante crítica interna e externa do texto e da atribuição de um intervalo de tempo aproximado (i.e. segunda metade do século XVI, entre 1743 e 1749 , etc.).

Neste quadro, a solução que assegura maior versatilidade, respeitando as características próprias de cada fonte, passa por considerar três campos distintos: um para o dia , outro para o mês, e outro para o ano (com 2, 2, e 4 dígitos, num campo de tipo numérico, respectivamente). Esta separação vai permitir a recuperação, reconversão e fusão dos elementos separados para um formato próximo do nível de discriminação dos documentos. É aconselhável a sua implementação sempre que se trabalhe com bases de dados relacionais. ou com folhas de cálculo.

Na eventualidade de ser impossível descobrir qualquer elemento de datação, o campo relativo ao ano será preenchido como o valor NULL. Em termos de tecnologia de bases de dados o NULL (invisível na listagem de ecrãs), representa a ausência de valor e pode portanto ser equiparado à tradicional referência bibliográfica de "S/D" - "sem data", sendo manipulado por primitivas próprias.

Para o caso especial das Bases de Dados de pesquisa textual e análise de conteúdo, que não autorizam a representação de informação em formatos diferentes do alfanumérico, aconselha-se a utilização do seguinte texto:

"ano"+ (espaço em branco) + ":" + (espaço em branco)+ quatro dígitos do ano + (Enter).

Ex: ano : 1678

Esta referência sintética será acompanhada, depois, da transcrição integral da data tal como ocorre no texto do documento.

Finalmente nas situações em que as balizas cronológicas são fixadas criticamente pelo próprio

investigador, dentro de um intervalo razoável de tempo, sugere-se o recurso a campos de tipo alfanumérico.

## **1.2. Referências geográficas**

**1.2.1. Para efeitos de referência espacial , consideram-se oito níveis de descrição da informação geográfica:**

- (i) Bispado**
- (ii) Distrito**
- (iii) Comarca**
- (iv) Concelho**
- (v) Couto (bases de dados medievais e início do período moderno)**
- (vi) Localidade**
- (vii) Freguesia ou paróquia**
- (viii) Residência**

Estes níveis correspondem obviamente a localizações mais precisas ou mais amplas das acções referidas nos documentos e não é obrigatório realizar investigações complementares para preencher todos os campos pré-definidos de informação geográfica, particularmente no momento de recolha de fundos arquivísticos. No entanto, sempre que esta informação for directamente acessível a partir das fontes, aconselha-se a planificação da base de dados de modo a obter o maior grau de precisão e discriminação geográfica possível. Por forma a não sobrecarregar os ecrãs de recolha de dados, apenas a freguesia irá constar do ecrã principal das Bases de Dados Relacionais, sendo necessário chamar uma janela (através de um botão) para ter acesso aos restantes atributos espaciais.

Nos casos em que o horizonte temporal da investigação abarca diferentes períodos de divisão administrativa do território e a relação entre níveis de descrição geográfica não é fixa mas depende do valor "data", será imprescindível dispor de tabelas de conversão com a divisão do território de cada reforma administrativa.

Quanto à transcrição do local de residência num campo pré-definido (i.e implicando a duplicação da transcrição conservadora e contextual), esta apenas parece justificável quando o projecto de investigação envolve fins prosopográficos, cruzamento nominativo de indivíduos, empresas ou instituições entre fontes diferentes, ou, ainda, reconstituição de redes de relações espaciais. O grau de precisão depende naturalmente do objecto de estudo. Por residência entender-se-á, simultaneamente, a indicação do LUGAR que é um elemento constitutivo das freguesias , e também a indicação do topónimo relativo à residência e/ou sítio de trabalho.

No projecto PACO prevê-se desde já a possibilidade de articulação, numa fase de posterior à conclusão do projecto, das bases de dados criadas por diferentes investigadores. Serão utilizadas como etiquetas geográficas de referência às Nomenclaturas das Unidades Territoriais,

correspondentes a um nível intermédio entre os Distritos e os Concelhos e conhecidas por NUT III. Para possibilitar a reconversão de toda e qualquer informação sobre o território nas nomenclaturas de NUT III, privilegiar-se-á o nível de descrição geográfica do Concelho. Sempre que possível, será esse um denominador comum de registo de informação, tendo em vista utilizações futuras e a percepção das bases de dados como entidades que têm um ciclo de vida que não se esgota no projecto de investigação em curso.

1.2.2. Na descrição da Localidade adopta-se o princípio de normalizar todas as citações relativas a "termo", "súrbios", "arredores", "vizinhanças", uniformizando-os pela bitola comum de "termo".

Ex: "termo da Vila da Feyra" > termo Vila da Feira

"arredores de Çintra" > termo Sintra

Na uniformização suprimem-se os determinantes "de" e "da" e observa-se a regra de transcrever a palavra "termo" (com minúscula) + 1 espaço em branco + o nome standardizado da Localidade.

1.2.3. Na descrição do Local de Residência adopta-se o princípio de transformar as referências espaciais relativas em referências absolutas.

Ex: "abaixo da Sé" > Sé

"junto ao Pelourinho" > Pelourinho

Suprimem-se logicamente os determinantes, as contracções e as conjunções "de", "da", "à", "ao", etc.

Este procedimento, bem como o descrito em 1.2.2., visa naturalmente melhorar a eficácia para efeitos de cruzamento entre registos de informação. No entanto, para que não existam diferenças subtis mas importantes entre o que está no computador e o que está no original, é imprescindível duplicar a informação normalizada dos campos Localidade e Residência, com a transcrição conservadora destes mesmos dados num outro local da base de dados.

### 1.3. Referências arquivísticas

A referência arquivística deve ser feita num único campo de tipo alfanumérico, através da descrição de uma sequência que poderá englobar quatro a seis itens, separados pela vírgula e por um espaço em branco:

Ex: Arquivo, Fundo documental, Série, Documento.

Ex: Arquivo, Fundo documental, Secção, Subsecção, Série, Documento.

Os primeiros itens tomam a forma ou de siglas ou de abreviaturas, enquanto o último é descrito por extenso.

Ex: ANTT, Min.Rei., C.Faz., Liv. 414 fl. 21 v.

Ex: C.D.E.D.P., Ferr.Dias., Min.Econ.Ind., Organ.Ofic., Doc. 32.

Fundos que ainda não foram objecto de tratamento arquivístico são referenciados através do

sistema de classificação utilizado pela instituição de origem ou, alternativamente, pelo descritor dos guias de remessas.

Fontes arquivísticas já publicadas, comportam uma dupla referência bibliográfica: a primeira para o livro publicado, a segunda para a referência arquivística da fonte de origem.

#### 1.4. Nomes dos indivíduos

Apenas se justifica o recurso a campos pré-definidos com o nome próprio dos indivíduos quando se antevê a possibilidade de a informação de tipo nominativo ser cruzada entre si ou com a informação de uma outra fonte, como sucede, por exemplo, na ligação entre registos paroquiais e róis de confessados. O computador terá primeiro de reconhecer e depois de confirmar que dois nomes registados em fontes diferentes correspondem à mesma pessoa. O primeiro passo para tornar esta operação possível é modernizar os nomes próprios e os apelidos. De notar que esta normalização em campos pré definidos não deve eliminar a transcrição conservadora do nome num outro campo da base de dados, pois a duplicação é fundamental para assegurar a reversibilidade das interpretações e ligações .

O reconhecimento da identidade dos nomes é apenas o primeiro passo dum processo de certificação no qual influem outros testes de congruência lógica: até o computador validar a conclusão final de que o indivíduo "Abc" na ocorrência documental K e o indivíduo "Abc" na ocorrência documental Y são uma e a mesma mesma pessoa, ele vai ter de perguntar (por exemplo) se as datas / idades de ambas as fontes são reciprocamente compatíveis, se o cônjuge é o mesmo, na ausência de segundos casamentos, se a naturalidade é igual, etc.

1.4.1. A transcrição modernizada e normalizada de nomes históricos é feita num único campo de formato fixo. No entanto, importa desde já considerar situações tipo que tornam as comparações literais e as inferências automáticas potencialmente equívocas, exigindo técnicas mais sofisticadas de análise.

1.4.1.2. Um primeiro problema, bastante comum em algumas fontes históricas, é o da variação do número de palavras utilizadas para transcrever um nome: "Jorge Manuel de Sousa Rodrigues", tanto pode aparecer sob a forma de "Jorge Rodrigues", de "Jorge Manuel Rodrigues" ou de "Jorge de Sousa Rodrigues". O computador obviamente não reconhece como idênticas as variantes mais abreviadas ou mais extensas . Assim sendo, devemos não só perguntar se dois nomes são idênticos em ocorrências diferentes , mas também perguntar se, não sendo eles idênticos, possuem alguma verosimilhança.

1.4.1.3. O segundo aspecto é o das transformações no nome provocadas pelo ciclo de vida das pessoas, isto é as mudanças que remetem para a biografia e as trajectórias individuais. A sua incidência faz-se sentir na alteração da quantidade de palavras que um nome tem e na construção social dos sobrenomes. O exemplo mais óbvio é o da assimilação do(s) apelido(s) do marido pela mulher depois do casamento, embora outras situações sejam igualmente frequentes, caso das alcunhas que com o correr do tempo se fixam em apelidos, ou dos falsos patronímicos, geralmente indicando locais de origem ou designações profissionais que igualmente se fixam no nome. A concepção da informação deve ser bastante flexível para reconhecer que "José Carlos" e "José Carlos Bexiga" podem referir-se a uma só e mesma pessoa. Assim sendo, utilizar-se-á como procedimento de base, ajustado obviamente às grandes variações na qualidade da informação de fonte para fonte e de local para local, a seguinte regra: sempre que o primeiro nome de "Abc" na fonte K for igual a "Abc" na fonte Y, e um dos outros nomes de Abc(K) estiver contido em Abc(Y),

o computador selecciona essas duas ocorrências para um teste de validação (ver acima 1.4.), apresenta os resultados ao investigador e pede a confirmação / anulação da identidade.

1.4.2. Sempre que for necessário definir de forma conceptual (i.e. não histórica como em 1.4.1.3.) o que é um "apelido", considerar-se-á apenas o último nome do indivíduo, com a excepção os casos em que o nome de família tem uma composição verdadeiramente lexicalizada (Espírito Santo, Castelo Branco) incluindo os nomes que contêm preposição (Cabeça de Vaca). Já volta a registar-se só o último nome quando uma conjunção liga dois nomes de origem geográfica (de Vasconcelos e Sá>Sá, de Sousa e Brito>Brito).

1.4.3. Se nos nomes nacionais se adopta o princípio de os normalizar e estandardizar ortograficamente num campo pré-definido, para os nomes estrangeiros mantém-se a transcrição tal como aparece na fonte.

## 1.5. Transcrição de números

1.5.1. A transcrição de números é feita de forma compacta , sem cifrão ou qualquer outro sinal, e sem separação de casas entre milhares, milhões, etc. A vírgula assinala sempre a presença de números decimais.

1.5.2. Qualquer valor representado pelo recurso a fracções ( por exemplo as fórmulas utilizadas nos câmbios  $10 \frac{2}{5}$  ) é sempre convertido no número decimal correspondente.

1.5.3. Sempre que os números se referem a uma determinada unidade de grandeza empregue para a medida de sólidos, líquidos, comprimentos, unidades monetárias ou outras, esta unidade é transcrita sob forma abreviada. A abreviatura surgirá, na base de dados de texto livre, à direita do número, após um espaço branco. Na base de dados relacional ou na folha de cálculo ocupará um atributo/coluna separado, com a dimensão de quatro caracteres. As abreviaturas mais frequentemente utilizadas constam do *Vocabulário Ortográfico da Língua Portuguesa*, da Academia das Ciências de Lisboa, à excepção das abreviaturas assinaladas com \* que são de nossa responsabilidade. Estas abreviaturas são as seguintes:

Palavra		Abreviatura
Almude	=	alm.
Alqueire	=	alq.
Arratel	=	art*.
Arroba	=	arr.
Centavos	=	ctvs
Conto de réis	=	c.
Coroa (moeda)	=	cor.
Côvado	=	côv
Cruzado		crz*
Dinheiros(moeda)		dnh*

Dúzia	=	dz
Escudo	=	esc
Grama	=	g
Grade (medida)	=	grd
Grão(peso)	=	gr
Grosa	=	grz
Hectare	=	ha
Hectolitro	=	hl
Hectovátios	=	hw
Quarteirão	=	qtão*
Quilograma	=	kg
Légua	=	Lég
Libra(peso)	=	lb
Libra(moeda)		lbm*
<b>Palavra</b>		<b>Abreviatura</b>
Litro	=	l
Maravedi	=	marv*
Milha(s)	=	mil
Milheiro		mlhe*
Moio	=	m
Onça(peso)	=	onç
Palmo(s)	=	ps
Pé(medida)	=	p
Polegada	=	pol
Quartilho	=	quart
Quintal(peso)	=	q
Reais	=	rl*
Reais brancos	=	rlbr*
Reais pretos	=	rlpr*
Réis	=	rs
Sacos	=	sc
Soldo(moeda)	=	sld*
Teiga	=	tei
Tonel	=	ton
Tonelada	=	t
Vara(medida)	=	vr*
Unidades	=	und*
Quilómetro	=	km

**Quadro 1****Síntese com as dimensões mínimas aconselhadas para os campos pré-definidos**

Nome do Campo	Tipo de campo	Nº caracteres	Informação
Dia	Numérico	2	Indicação do Dia
Mês	Numérico	2	Indicação do Mês
Ano	Numérico	4	Indicação do Ano
Datatribuida	Alfanumérico	40	Período aproximado do documento
NutConcelho	numérico	9	Inf.Geográfica actualizada
Distrito, Bispado, Comarca, Concelho, Couto	Alfanumérico	25	Inf. Geográfica
Localidade	Alfanumérico	25	Inf. Geográfica
Freguesia	Alfanumérico	25	Inf. Geográfica
Residência	Alfanumérico	40	Inf. Geográfica
Nomefonte*	Alfanumérico	60	Transcrição conservadora do nome e do apelido
Nomenorm*	Alfanumérico	60	Nomes estandardizados
Fonte	Alfanumérico	60	Referência arquivística
Unidmed	Alfanumérico	4	Abreviatura das medidas de sólidos, líquidos , etc.

Note-se que a designação campo relativo ao nome e do apelido (cf. \*) pode obviamente tomar muitas formas consoante o quadro concreto da investigação (nomeproprietário, apelidoproprietário, nomerendeiro, nomepai, etc).

## 2. Normas de Transcrição - orientações gerais

O desenvolvimento do PACO inclui uma fase inicial de transcrição de texto orientada para a constituição de um *corpus* que alimentará investigações ulteriores de ordem historiográfica, linguística, antropológica, sociológica e económica. Torna-se absolutamente necessário que os critérios seguidos na transcrição das fontes (manuscritas e impressas, antigas e modernas) sejam ao mesmo tempo uniformes, exequíveis e úteis. Esta primeira proposta de sistematização de normas de transcrição procura atingir esses três objectivos.

A história da normalização gráfica dos textos, sobretudo dos medievais, tem gerado dezenas de procedimentos diversos dos quais resultaram, para além de inesperados antagonismos, duas ideias claras: textos dirigidos a um público com preocupações filológicas e linguísticas são preferentemente conservados nas suas características gráficas originais, enquanto textos destinados a uma divulgação didáctica, literária ou historiográfica são transcritos à custa de uma actualização ortográfica mais ou menos moderada. Conservação e modernização são, pois, as opções extremas que se podem tomar no



início de uma campanha de transcrição, sendo depois possível adoptar um dos sistemas já formulados por paleógrafos, linguistas ou editores literários. É claro que factores como as edições existentes, uma singular originalidade gráfica ou a complexidade da utilização prevista para um texto podem motivar a formulação de novas normas. É o que tem de acontecer com as fontes do PACO, projecto interdisciplinar com diversas perspectivas concentradas no mesmo objecto - o aspecto gráfico das fontes transcritas tem de ser satisfatório perante diferentes exigências. Resulta daqui que o PACO adoptará, inicialmente, normas de transcrição conservadoras que darão lugar, em fase posterior, a normas de publicação modernizadoras, subdivididas estas em normas para textos dos séculos XIV, XV e XVI, por um lado e dos séculos XVII, XVIII e XIX por outro.

A ideia é a de permitir o tratamento linguístico para determinação da etimologia dos nomes das ocupações e das respectivas propriedades gramaticais. Isto vem tornar indispensável o respeito pela originalidade gráfica de todos os textos escritos até 1911. Tal procedimento reduz a ambiguidade da matéria gráfica aos olhos dos linguistas mas carrega o texto de excessivas variações quando passam só a interessar as informações extra-linguísticas nele contidas. O resultado do registo conservador terá por conseguinte um perfil tal, que permita um tratamento modernizador para efeitos de edição. Antes de entrarmos nestas matérias convirá porém abordar alguns aspectos que facilitam o reconhecimento, identificação e seriação dos textos no computador.

Se bem que haja a preocupação em não camuflar a originalidade gráfica dos textos, pode proceder-se logo de início àquelas normalizações que não têm efeitos sobre a interpretação linguística das formas. A primeira consequência desta opção é eliminar toda a alografia, quer nos casos dos alógrafos tradicionalmente ignorados (*s* longo e *s* redondo, por exemplo), quer no de *i* e *j*, *u* e *v*.

O desenvolvimento de abreviaturas seguirá, nesta primeira fase, uma orientação desequilibrada: algumas abreviaturas não se desenvolvem (o til sobre vogal não deverá, em caso algum, abandonar a primeira das letras vocálicas que modifica), outras desenvolvem-se sem indicação (as abreviaturas sistemáticas, excepto o referido til), outras ainda, se bem que desenvolvidas, têm de ser assinaladas (as abreviaturas vocabulares). Não se justifica, contudo, o trabalho lento da sinalização individual das letras repostas pelo desenvolvimento: toda a palavra transcrita à custa da resolução de abreviatura é globalmente precedida e seguida de sinais convencionais.

Os erros não se emendam, mas são assinalados. Conserva-se a pontuação, sendo apenas regularizada a morfologia dos sinais e conserva-se a distribuição de maiúsculas e minúsculas.

### **3. Transcrição conservadora - normas**

#### **3.1. Transcrição das vogais**

3.1.1. Transcrição de *v* e *j*, respectivamente, como *u* e *i* quando tiverem valor vocálico ou semi-vocálico:

Ex. Vrraca>Urraca, Lovrenzo>Lourenzo, Jdania>Idania, Lejrea>Leirea.

3.1.2. Conservação de *y* quer tenha valor de vogal, quer de semi-vogal:

Ex. my>*my*, mays>*mays*

3.1.3. Transcrição das vogais nasais tal qual ocorrem no documento. Como o recurso ao diacrítico til sobre determinadas vogais é dificultado por muitos processadores, convencionou-se que o trema o substituirá na transcrição de "u", "e" e "i". Estas vogais transcrevem-se respectivamente como "ü", "ë", "ï".)

Ex. com til (sobre a primeira se houver geminação - *freo*>*frëo*}, seguidas de *m* ou *n* (*poemte*, *senpre*), ou simplesmente não marcadas (*testemoyo*)

3.1.4. Conservação das vogais geminadas quer seja etimológico ou não o seu agrupamento:

Ex. *peolho*, *maao*

## 3.2. Transcrição das consoantes

3.2.1. Transcrição de *u* e *i*, respectivamente, como *v* e *j* ou *g* quando tiverem valor consonântico.

Ex. *aumentado*>*aventurado*, *Iunio*>*Junio*, *ieolio*>*geolio*

3.2.2. Conservação do *h*, mesmo que não tenha justificação etimológica:

Ex. *hüa*

3.2.3. Conservação de *c*, *ç*, *z*, *s*, *ss*, *g*, *j* e *y* de valor consonântico sem qualquer regularização, nem mesmo a da cedilha:

Ex. *noca*, por "nossa", *Ção*, por "São", *pazo*, por "paço", *Lourenzo*, por "Lourenço", *pressa*, por "presa", *gusto*, por "justo", *jente*, por "gente", *yeira*, por "jeira")

3.2.4. Desenvolvimento do til em *m* ou *n* quando for abreviatura de consoante heterossilábica:

Ex. *ströento*>*stromento*, *ãno*>*anno*

3.2.5. Conservação de todas as consoantes geminadas:

Ex. *rreyno*, *donno*

## 3.3. Abreviaturas

3.3.1. Transcrição do sinal tironiano com valor de copulativa como *e*:

Ex. *meu corpo 7 ma alma*>*meu corpo e ma alma*

3.3.2. Desenvolvimento das abreviaturas sistemáticas sem indicação especial:

Ex. *per*, *pro*, *que*, *de*, *-eiro*, *-mente*, *-er-*, *-re-*?

3.3.3. Desenvolvimento das abreviaturas de formas do léxico comum e onomástico orientadas, preferencialmente, pelos hábitos do escriba; senão, pelos dos documentos aparentados ou pela grafia da época. Mas no caso de a ortografia portuguesa contemporânea ter soluções de desenvolvimento que já

existiam na época do documento, opta-se por essa grafia totalmente moderna. Estes desenvolvimentos são assinalados rodeando a forma transcrita de dois sinais convencionais ( " \" no início e " { " no final).

### 3.4. Erros e lacunas

Os erros não são emendados e as lacunas não são preenchidas: assinalam-se os primeiros com (*sic*) e as segundas com (?).

### 3.5. Maiúsculas

Mantém-se a distribuição de maiúsculas e minúsculas, se bem que se converta em maiúscula a forma intermédia da semimaiúscula.

### 3.6. Ligação e separação de palavras

Moderniza-se a separação de palavras excepto nos casos da contracção e do pronome clítico - mantém-se aqui a grafia original, aglutinada ou separada, sem recurso a sinais de ligação como o apóstrofo ou o hífen.

Ex: mencomendasse e não \*"m'encomendasse", cospialhe e não \*"cospialhe".

### 3.7. Pontuação e acentuação

Mantêm-se ambas mas reduzem-se, esquematizando-os, os sinais de pontuação aos quatro hoje disponíveis: , ; : . As plicas também se transformam em ' '.

## Referências

### AAVV

1. Normas de transcrição para textos medievais portugueses

*Boletim de Filologia* XXXII (417-425)

### Academia das Ciências de Lisboa

1940 *Vocabulário Ortográfico da Língua Portuguesa*

Lisboa: Imprensa Nacional

Ivo **Castro**

1983. Vidas de santos de um manuscrito alcobacense

*Revista Lusitana*, Nova Série, 4 (separata)

1985. Vidas de santos de um manuscrito alcobacense

*Revista Lusitana*, Nova Série, 5 (separata)

Avelino de Jesus da **COSTA**

1. *Normas Gerais de Transcrição e Publicação de Documentos e Textos Medievais e Modernos*

**1993** 3ª edição muito melhorada

Coimbra: Faculdade de Letras da Universidade de Coimbra

Daniel **GREENSTEIN**

**1991** *Modelling Historical Data*

Gottingen: Max-Planck Institut für Geschichte

Clarinda de Azevedo **MAIA**

1. Edição de textos antigos galego-portugueses

*História do Galego-Português. Estado linguístico da Galiza e do Noroeste de Portugal desde o século XIII ao século XVI (19-40)*

Coimbra: INIC

Ana Maria **MARTINS**

1. *Clíticos na História do Português. Apêndice documental: documentos notariais dos séculos XIII a XVI do Arquivo Nacional da Torre do Tombo (xvii-xxvii)*

Dissertação de Doutoramento

Faculdade de Letras da Universidade de Lisboa:

Eduardo Borges **NUNES**

Introdução

**1969** *Álbum de Paleografia Portuguesa*

Dissertação de Doutoramento

Faculdade de Letras da Universidade de Lisboa

C. M. **SPERBERG-MCQUEEN** e Lou **BURNARD**

*Guidelines for the encoding and interchange of machine-readable texts*

**1991** 2.<sup>a</sup> edição

Chicago/Oxford, T.E.I.