

Regras Lingüísticas para Tratamento Computacional da Variação de Grafia e Abreviaturas do Corpus Tycho Brahe

RELATÓRIO IC FINAL ENTREGUE
A FUNDAÇÃO DE AMPARO À
PESQUISA DO ESTADO DE SÃO
PAULO - FAPESP

Proc 01/06529-6
Prof. Orientador: Helena Britto

UNICAMP
Agosto de 2002

[clique aqui](#)

ÍNDICE

1. INTRODUÇÃO	3
2. Regras Lingüísticas	3
2.1 Considerações gerais	3
2.2 Regras Geradas	4
2.1.1 Primeiro conjunto de regras para tratamento da variação de grafia encontrada	4
2.1.2 Refinamento 1: conjunto de regras acerca de consoantes dobradas	5
2.1.3 Refinamento 2: acentuação verbal especificamente	7
2.1.4 Refinamento 3: acentuação geral	8
3. ABREVIATURAS	11
4. CORPUS TYCHO BRAHE	12
5. CONCLUSÃO	13
6. BIBLIOGRAFIA	14
7. ANEXO I	16
7.1 Saudações	17
7.2 Despedidas	18
7.3 Formas de Tratamento:	20
7.4 Muito, Que, Quanto, Para, Por que, Por, Minha	22
7.5 Não encontrados no dicionário de abreviaturas, de M.H. Flexor	23
7.6 Outros	25

Regras Lingüísticas para Tratamento Computacional da Variação de Grafia e Abreviaturas do Corpus Tycho Brahe

1. INTRODUÇÃO

As diferenças entre grafias no século 17 mostram que nessa época não havia uma normatização na grafia dos textos portugueses. Pode-se dizer que essa normatização começou a tomar forma somente a partir do século 18 quando, no reinado de D. João V, foi reconhecida a conversão da norma regional do centro atlântico do reino em norma de prestígio, como afirma Rita Marquilha, em seu livro *Norma Gráfica Setecentista*.

Foi tomado como padrão socialmente prestigiado o português falado no centro político do reino, isto é, na corte, e este foi passado para o registro escrito. E assim, somente a partir de 1734, a Academia Real da História passou a adotar uma ortografia para a impressão das obras dos autores portugueses, crendo que a adoção a uma normatização da grafia seria bem aceita se fosse aplicada aos textos dos bons autores. Essa normatização foi necessária, especialmente, para ser usada pela tipografia.

Dos textos do Corpus Anotado do Português Histórico Tycho Brahe (daqui em diante CTB), usados como base de pesquisa para a variação gráfica contida neste trabalho, alguns foram impressos fiéis aos manuscritos, outros possivelmente passaram (ou não) por uma revisão tipográfica que normatizaria a escrita segundo as normas da tipografia em questão [\[1\]](#).

De qualquer modo, sendo o caso um ou outro, o fato é que muitas variações foram encontradas e apresentaram-se como um problema para a ferramenta automática de etiquetagem morfológica, como evidenciado em Menegatti (2002), apresentado a FAPESP em fevereiro de 2002. A ferramenta automática foi treinada computacionalmente a partir de textos do corpus cujas edições traziam os textos com grafia modernizada. Logo, por ter sido treinada a partir de um léxico do português moderno, tal ferramenta apresenta problemas para etiquetar palavras com grafias diferentes. Por meio de uma listagem das variações gráficas nos textos, é possível classificar essas variações, extrair regras a partir delas e finalmente treinar o etiquetador, para que ele passe a classificar corretamente as diferentes formas gráficas. No presente relatório apresentamos o conjunto de regras a que chegamos, lingüística e computacionalmente aceitáveis, a ser aplicado ao mais recente treinamento de uma nova versão do etiquetador.

Neste relatório apresentarei também o trabalho feito com as abreviaturas de alguns textos do CTB, que também se mostraram variadas e inconstantes, bem como falarei um pouco da participação no Projeto Temático [\[2\]](#).

2. Regras Lingüísticas

2.1 Considerações gerais

Uma possível solução para uma maior precisão na etiquetagem foi a de fazer o levantamento das palavras de grafia não modernizada, e depois classificar essas variações, visando-se a extrair regras de variação de grafia para cada um dos períodos históricos contemplados no corpus. Pela implementação computacional de tais regras, espera-se aumentar a performance do etiquetador, que poderá classificar corretamente mais de uma forma gráfica de determinada palavra.

Como apresentado no relatório anterior, os textos selecionados para o trabalho que não passaram por uma modernização na sua grafia são indicados a seguir:

[1] Textos com grafia não-modernizada do Corpus Tycho Brahe utilizados neste trabalho

1. D. JOAO III (1502-1557) Letters of John III - King of Portugal 1521-557 (The portuguese text edited with an introduction by J. D. M. Ford). Cambridge, Massachusetts. Harvard University Press, 1931. (número de palavras do texto: 56.604).
2. COUTO, Diogo do (1542-1606). Décadas (seleção, prefácio e notas de António Baião). Vol 1. Lisboa, Livraria Sá da Costa - Editora, 1947. (número de palavras do texto: 47.448).
3. COSTA, Manuel da (1601-1667). Arte de Furtar (seleção, introdução e notas de Roger Bismut). Lisboa, Imprensa Nacional Casa da Moeda, 1991. (número de palavras do texto: 52.867)
4. VIEIRA, António (1608-1697). Sermões (prefaciado e revisto pelo Rev. Padre Gonçalo Alves). Porto, Livraria Chardron - Lello & Irmão Editores, 1907. (número de palavras do texto: 53.855)
5. MELO, D. Francisco Manuel de (1608-1666). Cartas Familiares (seleção, prefácio e notas por M. Rodrigues Lapa). Lisboa, Livraria Sá da Costa, 1942. (número de palavras do texto: 58.070)
6. CÉU, Maria do (1658-1753). Rellação da Vida e Morte da Serva de Deos a Venerável Madre Elenna da Crus (transcrição do Códice 87 da Biblioteca Nacional precedida de um estudo histórico, por Filomena Belo). Quimera. Lisboa, 1993. (número de palavras do texto: 27.410)
7. COSTA, António da (1714-?). Cartas do Abade António da Costa (introdução e notas de Fernando Lopes Graça). Lisboa, Cadernos da Seara Nova, 1946. (número de palavras do texto: 27.096)
8. Marquês da Fronteira e d'Alorna (1802-1881). Memórias do Marquês da Fronteira e d'Alorna (revisadas e coordenadas por Ernesto de Campos de Andrada). Lisboa, Imprensa Nacional - Casa da Moeda, 1926. (número de palavras do texto: 54,588)

Cada palavra de um mesmo texto foi agrupada de acordo com sua característica: vogais dobradas, consoantes dobradas, oscilações (i.e. variações de grafia), acentuação, etc. A partir dessa listagem é que foram extraídas as possíveis regras para o treinamento do etiquetador.

2. 2 Regras Geradas

2.1.1 Primeiro conjunto de regras para tratamento da variação de grafia encontrada

Segue-se abaixo uma relação de regras geradas a partir de um estudo das listas de palavras levantadas, e que já foram aplicadas ao treinamento do etiquetador. Estas regras foram inicialmente apresentadas no relatório parcial deste projeto, e encontram-se aqui novamente evidenciadas, de maneira mais clara e organizada:

[2]

- q **Ditongos**
 - **aes è ais (no final da palavra)**
 - **eo, ea è eio, eia**
 - **Deos è Deus**

- q **Vogais**
 - **y è i**
 - **i è e**
 - **e è i**
 - **u è o**
 - **o è u**

- q **Queda do H**
 - **th è t**
 - **he è é**
 - **ph è f**

- q **Queda de consoantes**

- **mn è n**
- **mpt è nt**
- **m<consoante ¹p,b> è n**
- **<vowel>pt<vowel> è <vowel>t<vowel>**
- **cc è c**
- **ch è qu**
- **ch è x**
- **ch è c**
- **ct è t**
- **pt è t**
- **pc è c**
- **pç è ç**

¶ **Acentuação**

- **ò è ó**
- **<algo>à è á**
- **issim. è íssim.**
- **ee è é**

¶ **Nasais**

- **aã è ã**
- **oens è ões**
- **aõ è ão**

¶ **Consoantes modificadas**

- **u è v**
- **g è j**
- **j è g**
- **z è s**
- **s è ss,z**
- **c è sc**
- **ss è c**

¶ **Outros**

- **suff: ozo è oso, oza è osa**
- **acem è assem**
- **re è er**
- **er è re**

2.1.2 Refinamento 1: conjunto de regras acerca de consoantes dobradas

As consoantes dobradas ocuparam um grande espaço na listagem dos textos do CTB e mereceram especial atenção na geração de regras para modernização gráfica dos textos do corpus, que pudessem ser computacionalmente implementadas. Sobre elas foram feitas muitas considerações.

Do português moderno a suas épocas passadas pode-se observar o seguinte. O sistema ortográfico do português moderno prevê, em posição intervocálica, a ocorrência de dois únicos grafemas que podem ser dobrados: **r** e **s**, que representam na escrita os fonemas /r/ e /z/, ao lado de **rr** e **ss**, representantes graficamente dos fonemas /x/ e /s/. Em épocas passadas do português, entretanto, consoantes simples e suas correspondentes dobradas proliferavam na escrita da língua, sem, entretanto, terem obrigatoriamente o valor distintivo hoje observado. Inspiradas no sistema latino, no qual indicavam a vogal longa do vocábulo por virem posicionadas imediatamente após a esta, as consoantes dobradas observadas em textos históricos do português ora seguem o sistema latino (como em *anno*), ora acompanham, no onset silábico, a própria vogal tônica da palavra (como em *Tyrannia*).

Com base nas consoantes dobradas presentes em nosso corpus, o que observamos em Britto & Menegatti (em andamento) foi o fato de que dobrar consoantes não é uma propriedade que pode afetar aleatoriamente todo e qualquer membro do conjunto de consoantes da língua. Segundo Britto & Menegatti, a observação sistemática dos dados descritos em Menegatti (2002) indicou que

apenas as consoantes correspondentes às séries de oclusivas e fricativas latinas (tabela A) podem ser dobradas, não havendo um único caso de variação entre consoante simples versus consoante dupla para os grafemas que representam os fonemas fricativos sonoros /v/, /z/ e /ʒ/ característicos do português (tabela B)- nomeadamente **v** versus **vv**, **z** versus **zz** ou **j** versus **jj**.^[3]

Oclusivas	surdas	/p/	/t/	/k/
	sonoras	/b/	/d/	/g/
Fricativas	surdas	/f/	/s/	/•/
	sonoras	-	-	-

Tabela A
História da língua: série de oclusivas e fricativas latinas

Oclusivas	surdas	/p/	/t/	/k/
	sonoras	/b/	/d/	/g/
Fricativas	surdas	/f/	/s/	/•/
	sonoras	/v/	/z/	/ʒ/

Tabela B
História da língua: série de oclusivas e fricativas portuguesas

Nossa hipótese para a metódica não-ocorrência de contrastes como avisar vs avvisar, fazer vs fazzer ou tijela vs tijjela pode ser assim resumida:

- Consoantes dobradas consistem em uma marca característica do sistema latino;

- Consoantes **v**, **z** e **j** como representantes da historicamente recente série de fricativas sonoras /v/, /z/ e /ʒ/ são inovações do português frente ao latim;

- Em apoio ao que já apontado na literatura especializada, no sentido de a história da escrita do português ser caracterizada não por uma mudança, ou ruptura, radical frente à tradição escrita latina, mas por um continuum cumulativo desta para com a tradição romance (Emiliano, 1997; Viegas, 2001 apud Menegatti (2002)), na escrita do português são grafadas dobradamente, i.e. ao estilo latino, não toda e qualquer consoante, mas apenas aquelas conhecidas como tal pelo sistema latino.

Observa-se ainda, na escrita do português histórico observado, a ocorrência, para o inovador fonema fricativo palatal sonoro /ʒ/, da variação de grafia **g** vs **gg** - ambas grafias conhecidas pela tradição latina, embora nesta com valor fonêmico oclusivo palatal surdo /g/. Numa profunda fusão dos sistemas gráficos português e latino, alia-se a inovação /ʒ/ às tradições **g** vs **gg**.

Frente a esta descrição, chegou-se às seguintes regras:

[4]

Grafemas relativos a oclusivas

Fonemas

/p/ /t/ /k/

/b/ /d/ /g/

Grafemas correspondentes

p t c mantêm-se

b d g como

pp tt cc passam a

contexto

bb dd gg

Grafemas relativos a fricativas

Fonemas

/f/ /s/ /•/

/v/ /z/ /ʒ/

Grafemas correspondentes

f s|ss|c ch|x mantêm-se como tais

v s|z j|g

ff Æ^[4]|Æ^[5]|cc Æ^[6]|Æ^[7]

em quaisquer passam a simples

2.1.3 Refinamento 2: acentuação verbal especificamente

Alguns dos textos selecionados não apresentavam acentuação alguma, ou possuíam uma acentuação inconstante ou diferente da acentuação das palavras modernizadas do léxico do etiquetador.

Uma acentuação, presente em alguns autores, que chamou a atenção foi a dos verbos terminados em **-am** e **-ão**:

[5] MANUEL DA COSTA^[12].

Aceitaõ
Cursaõ
Custáraõ
Faráõ
Governáraõ
Excômungado

MARIA DO CÉU

Assentaõ
Deraõ
Examinaõ
Enamoraõ
Faziaõ
Julgaõ

Diante dessas ocorrências, é necessário explicitar regras ou equivalências que possam ser utilizadas para o treinamento do etiquetador, com por exemplo: em certos autores, os verbos no futuro recebem acento para diferenciá-los dos verbos no presente, já que em ambos, a terminação é **-ão**, ou **-aõ**. Para esses casos, extrai-se assim a seguinte regra:

acharãõ > achar am acharáõ > achar ãõ
Tabela C Regra para verbos terminados em ão ou aõ

2.1.4 Refinamento 3: acentuação geral

Como já dito anteriormente, o léxico do qual o etiquetador automático tycho brahe é dependente consta de palavras modernizadas do português, e ocorrências de acentuação como “Ávante”, “sómene”, “sôpas”, causam um estranhamento, levando, em geral, a atribuição, pelo etiquetador automático, de uma etiqueta qualquer, geralmente diferente daquela que seria dada a “Avante”, “somente”, “sopas”.

De modo a auxiliar o etiquetador, procuramos, de certa forma, identificar primeiramente a lógica que parece reger os sistemas de acentuação dos textos que compõem o CTB. Posteriormente, sugerimos certas regras. O processo pelo qual passamos para consolidar estas duas etapas, as quais também fazem parte de Britto & Menegatti (em andamento), pode ser assim resumido

Em primeiro lugar, é necessário observar que os textos do CTB que não possuem grafia modernizada, ou seja, aqueles a que nos dedicamos, são de dois tipos:

- [6]
- q aqueles inspirados em uma escrita etimológica, ou pseudo-etimológica (como defendem Emiliano, 1997 e Viegas, 2001), ou seja, francamente inspirada no latim;
 - q aqueles que seguem uma escrita de caráter fonético, no sentido de relevar as propriedades da nova língua, i.e. o português.

Nos dois casos, a escrita se distancia da escrita do português atual.

No que diz respeito ao português atual, Mattoso Camara Jr (1984), em seu dicionário de linguística, afirma que são do gênio da língua as paroxítonas terminadas em -a, -e, -o, e variantes (-as, -es, -os e -am, -em, -om), assim como as oxítonas terminadas em -i, -u e variantes. São do gênio da língua, porque 80% ou mais dos itens que compõem o léxico do português são paroxítonas terminadas em -a, -e, -o e variantes (caneta, caderno, cadeira, mesa, cabelo, calça, teto, janela, grade, teclado, telefone, quadro-negro, etc) ou oxítonas terminadas em -i, -u e variantes (saci, jaboti, (eu) subi, (eu) parti, etc). Ao comparar esta observação de Mattoso Camara com os gramáticos normativos (Celso Cunha, Rocha Lima, etc), conseguimos decifrar e explicar o porquê das regras do sistema de acentuação do português atual serem como são.

A lógica é assim:

Regras do gênio da língua		
Paroxítonas em -a, -e, -o		sem qualquer
acento		
Oxítonas em -i, -u		sem qualquer
acento		
Se invertemos o gênio da língua, i.e		
Paroxítonas em -i, -u		acento
Oxítonas em -a, -e, -o		acento
Tabela D		
Do gênio da língua às regras de acentuação do português atual		

Por isso, temos trazem vs armazém ou bate vs café, ou ainda saci vs júri. Além disso, temos também, no sistema atual que:

[7]

- q Todas as proparoxítonas são acentuadas (como na música do Chico Buarque, Construção);
- q As oxítonas terminadas em -ar, -er, -or não são acentuadas (amar, caber, compor).

Assim são (algumas d)as normas de acentuação do português moderno. Mas, como são os autores do CTB diante destas normas? Os autores do CTB apresentam, além de uma aparente falta de regra na acentuação das palavras, uma grande inconstância até mesmo ao acentuar uma mesma palavra: “Avós” vs “Avòs”^[13]. Alguns autores, como o Marquês da Fronteira e Alorna, quase não acentuam as palavras, com algumas raras excessões como advérbios, por exemplo (v. [8]).

Outros, como Francisco Manuel de Melo, mantém, algumas vezes, uma acentuação bem próxima a do atual português europeu: “Cômica”, “Cômico”, “Cómodo”. De qualquer forma o que se observa é que nenhum dos autores aplica integralmente o sistema de acentuação do português moderno. Assim sendo, como resolver esse problema de uma inconstância na acentuação, que parece não seguir regras?

Conjugando o fator consoantes dobradas (tratado na seção anterior) com o fator acentuação (tratado nesta seção), observamos que parecer haver uma certa sintonia entre eles. Os autores que dobram consoantes são aqueles que quase nunca acentuam, e os que não dobram, são os que acentuam, e fazem isto privilegiando a marcação da tônica - i.e. acentuam todas as tônicas. Isto está em sintonia com o dito em [6]: os que optam pela escrita etimológica dobram consoantes (como em latim) e não acentuam (como em latim); por outro lado, os que optam pela escrita fonética não dobram consoantes e acentuam. Assim parecem ser D. João ([+consoante dobrada], [- acentuação]), Diogo do Couto ([-consoante dobrada], [+acentuação]) e Marques da Fronteira [+consoante dobrada], [-acentuação].^[14]

Assim sendo, embora não tenhamos extraído regras precisas^[15], esta informação a que chegamos poderá ser útil para outros treinamentos do etiquetador automático, talvez sendo até interessante a adaptação de etiquetadores

diferentes para cada um dos tipos de escrita. Para os que seguem a escrita etimológica (ou pseudo-etimológica), devem ser inseridos acentos, para que as palavras fiquem com a grafia do português moderno, e assim, possam ser reconhecidas pelo etiquetador; para os que seguem a escrita fonética, os acentos devem ser trocados de lugar, ou retirados, quando se trata dos casos de paroxítona terminada em -a, -e, -o ou oxítonas terminadas em -i, -u, por exemplo (lembrar que os autores com escrita fonética acentuam primordialmente as tônicas, como Couto em êle ou fôsse, que são paroxítonas terminadas em -e)^[16].

[8]

Abstinencia8	cadaver8	episodio8	Lamentavel8
Acerrimo8	calculos8	escapulario8	Legionarios8
Adormeciamos8	calendario8	escrupulos8	Legitimos8
Africa8	cartorio8	especie8	Leguas8
Agradavel8	chapeu8	espectaculo8	Maniaco8
Agua8	Corôa8	espiritos8	Medicos8
Aguaia8	Coroneis8	formidavel8	Mediterraneo8
Alem8 ^[17]	Correspondencia8	fortissimo8	Mêdo8
Alguem8	Decadencia8	genio8	Melancolica8
Amavel8	Desculpavel8	gloria8	Memoravel8
America8	Detestavel8	habil8	Memorias8
Austriaco8	Diacono8	hereditarias8	Negocios8
Bebados8	diarias8	horrivel8	Ninguem8
Beneficencia8		hostia8	notaveis8
Breviario8		imaginavel8	numerosissima8
Brilhantissimo8		Imperio8	nupcias8
Britanicas8		Impertinencias8	
		Joias8	
		Juizo8	

3. ABREVIATURAS

As abreviaturas, comuns especialmente em textos epistolográficos (i.e. cartas), não foram mantidas no corpus tal como estavam grafadas no original. Afinal, como o etiquetador, ao se deparar com um ponto (.), o entendia como um ponto final, a classificação de uma frase com abreviatura ficava comprometida. Ela era “quebrada” na altura do ponto da abreviatura, que recebia a etiqueta de um ponto final.

[10]

Rogo a V. Sra. se sirva mandar^[18] > Rogo/VB-P a/P V/NPR ./.
> Sra/NPR/ ./.

A melhor solução encontrada para manter tanto a fidelidade ao original quanto a qualidade da etiquetagem, foi a de expandir, ou desenvolver, as abreviaturas. As palavras expandidas recebiam, como estabelecido no manual do corpus, a grafia modernizada, enquanto a abreviatura era mantida dentro do comando **<original>** :

[11]

Rogo a Vossa Senhoria <original> V Sra. </original> se sirva^[19] (...)

Para esse processo de expansão, foi utilizado o dicionário de abreviaturas de Maria Helena Flexor^[20], que serviu muito bem a nossos propósitos. Porém, algumas abreviaturas não foram encontradas, exigindo, então, um trabalho mais extenso de pesquisa sobre palavras que pudessem caber nas abreviaturas. Muitas foram encontradas, porém algumas ainda se encontram sem soluções – como por exemplo, abreviaturas de nomes próprios. Nesse caso, retiramos o ponto e colocamos a abreviatura entre os comandos <original> </original> .

[12]

hásde estimar que eu appele a ti para <original> p.^a </original> o Rod no <original> Rod.no </original> .[\[21\]](#)

As abreviaturas também mostraram certa inconstância já que, como dito antes, não havia na época uma normatização gráfica:

[13]

V. Exa. = Vossa Excelência.

V. Ex. = Vossa Excelência.

A princípio, cheguei a pensar que as abreviaturas pudessem ser modificadas de acordo com o século, mas as inconstâncias de grafia foram detectadas em um mesmo texto, no caso, Cartas de Garrett. (Almeida GARRETT, apresentação e edição por Segismundo Spina). São Paulo, Humanitas Publicações FFLCH/USP, 1997). Segue-se, abaixo, alguns exemplos de abreviaturas retiradas do texto acima citado, que puderam ser encontradas no livro de Maria Helena Flexor.

[14]

- Atento Venerador e Criado Obrigado <original> Att.o V.or C.o Obr.do </original>
- Criado Obrigadíssimo <original> C.do Obg.mo </original> e fiel captivo <original> capt.o </original>
- Criado muito venerador obrigado <original> C.do m.to v.dor obr.o </original>
- Criado e atento venerador <original> C.do e att.to V.or </original>

As abreviaturas que não foram encontradas em Flexor também foram listadas, e as suas expansões foram feitas a partir de pesquisas em dicionários, enciclopédias e textos da mesma época.[\[22\]](#)

[15]

- esta Secretaria d'Estado <original> d'Est.o </original>
- Recebi os despachos <original> desp.o </original> de Madrid e
- nem vai o Ministro <original> Min.o </original> Ingles
- Quarta-feira <original> 4f.ra </original> 28 Maio <nl>
- 5 d' Abril 10 Manhã <original> M. </original> (1852)

Os séculos em que foram usadas as abreviaturas de [7] já constavam no dicionário de Flexor, portanto não se mostrou necessário uma pesquisa sobre os séculos. Já as palavras que não foram encontradas no dicionário, que fazem parte de [8], são usadas no século em que o texto foi escrito.

As abreviaturas que não foram expandidas - pelos motivos já explicados - também foram listadas, e fazem parte do Anexo I que acompanha o presente relatório. Essas abreviaturas serão colocadas na página do Projeto, esperando sugestões para uma possível solução.

4. CORPUS TYCHO BRAHE

Como proposto no relatório anterior, eu ainda trabalharia diretamente na construção do CTB, sendo encarregada da chamada Segunda Revisão, bem como do treinamento dos novos bolsistas.

A Segunda Revisão é a checagem final do texto digitalizado, que precisa ser minuciosamente revisado antes de receber as etiquetas. Trata-se de um trabalho conjunto de dois bolsistas: um lê em voz alta o texto que está digitalizado, enquanto o outro acompanha a leitura pelo original, para que nenhum detalhe seja deixado de lado, e a fidelidade ao original seja sempre mantida. Minha tarefa foi a de efetuar a leitura oralizada, enquanto os bolsistas G. Menezes, R. Manduruca e P. Lourençatto seguiam essa leitura com os originais. Os textos revisados que contaram com minha participação foram:

q MANUEL PIRES DE ALMEIDA, “Poesia e Pintura”.

q MANUEL DE GALHEGOS, “Gazeta, em que se relatam as novas todas, que ouve nessa çorte, e que vieram de varias partes no mês de novembro de 1651” .

q ANDRÉ DE BARROS, “A Vida do Padre António Vieira”.

Também, juntamente com Patrícia Lourençatto, fiz o treinamento dos dois

novos bolsistas acima mencionados, que precisavam ser instruídos sobre a montagem do CTB: como escanear um texto, a parte da Primeira correção, a Segunda Revisão e todos os comandos que são necessários para se preparar um texto para receber as etiquetas. Segue-se um exemplo de um trecho do texto de Diogo do Couto [\[23\]](#) com alguns comandos:

[16]

```
<comment t. menegatti> digitalization and first revision by P. Abdo, T. Menegatti
and C. Namiut
<edition> COUTO, Diogo do. Décadas (seleção, prefácio e notas de António Baião). Vol
1. Lisboa, Livraria Sá da Costa - Editora, 1947. </edition>
<P_01>
<heading>
                QUINTA DÉCADA
                LIVRO OITAVO
                CAPÍTULO IX
        Do que aconteeo ao Governador Martim Affonso em Moçambique até partir
para a India: e de como a sua não se foi perder em Baçaim, e êle chegou a Goa;
e de como Dom <original> D. </original> Estevão da Gama lhe entregou a India
</heading>
```

Depois de digitalizados e revisados, os textos passarão pelo etiquetador automático, ficando pronto para ser disponibilizado na internet tanto nas versões ortograficamente transcritas quanto morfológicamente etiquetadas.

O texto “Gazeta, em que se relatam as novas todas, que ouve nessa corte, e que vieram de varias partes no mês de novembro de 1651”, de Manuel de Galhegos, precisou de muito tempo para sua Segunda Revisão, já que o texto impresso está em formato de microfilme, além de contar com uma grafia não modernizada. Para fazer essa revisão, Patrícia Lourençatto e eu precisamos usar o leitor de microfilme do Instituto de Estudos da Linguagem - IEL, para ser possível fazer a correção e inserção de comandos no texto digitalizado que já havia sido digitado. A Segunda Revisão não pôde ser concluída, por atuais problemas técnicos com o leitor de microfilme do Instituto, que, segundo consta, deverá estar resolvido em 30 dias.

5. CONCLUSÃO

Como afirma Marquilhas, em textos antigos, as sobreposições de grafia acontecem inevitavelmente, e os textos do CTB, cuja grafia não era modernizada, podem confirmar essa afirmação. A inconstância gráfica é muito comum, especialmente na acentuação, na dobra de vogais e consoantes, e em vários outros exemplos já evidenciados no primeiro relatório.

As variantes gráficas listadas em seis dos textos de grafia antiga não normatizada do CTB são muitas, e resolver essa variação para ser aplicada ao etiquetador é um trabalho longo. Algumas soluções já foram encontradas e aqui colocadas, facilitando assim o trabalho da ferramenta automática de etiquetagem, porém, o trabalho pode ser ainda bastante estendido.

As abreviaturas também abrem espaço para muitas pesquisas. Neste trabalho foram tratadas abreviaturas que constavam nos textos do CTB. Embora incomparável ao trabalho visto no dicionário de abreviaturas, de Maria Helena Flexor, que serviu de guia para o tratamento da grande maioria das abreviaturas aqui encontradas, o presente relatório traz abreviaturas não contempladas por Flexor. Sabendo, desde sua visita ao IEL - Unicamp, que a autora está a preparar uma nova edição de seu dicionário, as abreviaturas inéditas encontradas por nós serão enviadas à autora. Assim, poderemos participar na nova edição como colaboradora.

E a continuidade do trabalho direto com o Corpus possibilitou um maior enriquecimento cultural, já que os textos nele colocados são obras de grande valor literário e linguístico, pois pude conhecer autores importantes da literatura portuguesa e suas particularidades linguísticas.

6. BIBLIOGRAFIA

- Bergström, Magnus & Neves Reis (1999) *Prontuário Ortográfico e guia da língua portuguesa*. Lisboa, Notícias Editorial.
- Borges, Carla (1996) "As terminações em -õ, -ã, -ão". In: Mattos e Silva, R. V (org) *A Carta de Caminha - Testemunho Linguístico de 1500*. Salvador, Universidade Estadual da Bahia (UfBA).
- Câmara Jr., Joaquim Mattoso (1985) *História e Estrutura da Língua Portuguesa*. Rio de Janeiro, Padrão Livraria e Editora.
- Couto, Diogo do. *Décadas* (seleção, prefácio e notas de António Baião). Vol 1. Lisboa, Livraria Sá da Costa - Editora, 1947
- Cunha, Celso Ferreira (1979) *Gramática da Língua Portuguesa*. Ministério da Educação e Cultura - Fundação Nacional de Material Escolar, Rio de Janeiro.
- Flexor, M. Helena (1991) *Abreviaturas, Manuscritos do século XVI ao XIX*. Editora Unesp - secretariado Estado da Cultura - Arquivo do Estado de São Paulo,.
- Garrett, Almeida. *Cartas de Garrett*. (1997) (apresentação e edição por Segismundo Spina). São Paulo, Humanitas Publicações FFLCH/USP.
- Mattoso Camara Jr, Joaquim (1984) *Dicionário de Linguística e Gramática*. Petrópolis, Vozes.
- Marquilhas, Rita (1991) *Norma Escrita Setecentista - Do autógrafo ao Impresso*. Lisboa, Instituto Nacional de Investigação Científica, Centro Linguístico da Universidade de Lisboa.
- Teyssier, Paul (1997) *História da Língua Portuguesa*. Livraria Sá da Costa Editora, Lisboa.
- Viegas, Rui. (2002) "Da origem, formação e consolidação do português: breve história externa da língua portuguesa". *Calliope* 4 (2) (http://www.mediom.qc.ca/~estrudex/html_divers/parutions.html).
- Williams, Edwin B. (1975) *Do Latim ao Português*. Editora Tempo Brasileiro, Rio de Janeiro.

7. ANEXO I

ANEXO ABREVIATURAS

Durante o processo de seleção de abreviaturas, foram constatadas várias formas de ocorrência para uma mesma abreviatura. Depois de retiradas do texto, elas foram agrupadas em diferentes seções, que serão apresentadas a seguir. Os números que se encontram entre parênteses () indicam quantas vezes aquela abreviatura foi usada no texto.

7.1 Saudações

- α Ilustríssimo <original> Ill.mo </original>
- α Ilustríssimo Senhor <original> Ill.mo Snr. </original> (7)
- α Ilustríssimo Senhor <original> Ill.mo Sr. </original> (24)
- α Ilustríssimos Senhores <original> Ill.mos Srs. </original> (2)
- α Ilustríssimo Senhor Doutor <original> Ill.mo Sr. Dr. </original>
- α Ilustríssimo e Excelentíssimo Senhor <original> Ill.mo e Ex.mo Snr. </original> (12) <nl>
- α Ilustríssimo e Excelentíssimo Senhor <original> Ill.mo e Ex.mo Sr. </original> (7)
- α Ilustríssimo Excelentíssimo Senhor <original> Ill.mo Ex. mo Sr. </original> (1)
- α Ilustríssimo Excelentíssimo Senhor <original> Ill.mo Ex.mo Sr. </original> (6)

- q Ilustríssimo Excelentíssimo Senhor <original> Ill. mo Ex.mo S.r </original>
- q Ilustríssimo Excelentíssimo Senhor <original> Ill.mo Ex.mo Snr. </original> (3)
- q Excelentíssimo Senhor e Amigo <original> Exmo. Snr. e Am.o </original> <nl>
- q Excelentíssimo Senhor <original> Ex.mo Sr. </original> (6)
- q Excelentíssimo Senhor <original> Ex.mo S.r </original> (3)
- q Excelentíssimo Senhor <original> Exmo. Sr. </original> (4)
- q Excelentíssimo Senhor <original> Ex.mo Snr. </original> (2)
- q Senhor Doutor <original> Sr. Dr. </original>
- q Meu amigo e Senhor do Coração <original> am.o Sr. do C. </original>

7.2 Despedidas

- q Adeus <original> Ad.s </original> (8) até logo que fallaremos
- q Adeus Senhora <original> Ad.s Sra. </original> Bixa <nl>
- q Amigo <original> Am. o </original> certo <nl>
- q amigo certo grato <original> am.o certo gr.to </original>
- q Amigo certo, obrigadíssimo <original> Am..o certo, obgd.mo </original>
- q Amigo Criado <original> am.o Cr.o </original>
- q Amigo Criado <original> am.o c.do </original>
- q Amigo Criado e Venerador <original> C.do e V. or </original>
- q Amigo criado obrigadíssimo <original> Am.o C.o obg.mo </original> <nl>
- q Amigo Obrigado <original> Am.o obg.o </original> (2)
- q Atento venerador e obrigado <original> Att.o V.or e Obg.do </original>
- q Atento Venerador e Criado Obrigado <original> Att.o V.or C.o Obr.do </original>
- q Criado Obrigadíssimo <original> C.do Obg.mo </original> e fiel captivo <original> capt.o </original> <nl>
- q Criado muito venerador obrigado <original> C.do m.to v.dor obr.o </original>
- q Criado e atento venerador <original> C.do e att.to V.or </original>
- q Colega Obrigadíssimo amigo <original> Coll.a obg.mo am.o </original> certo
- q De Vosso <original> V. </original> <nl>

- q De Vossa Excelência <original> V Exa. </original> (2) <nl>
- q De Vossa Excelência <original> V.Exa. </original> (6) <nl>
- q De Vossas Senhorias <original> VV. SS. </original> <nl>
- q De Vossa Senhora amigo <original> V. Snr.a am.o </original> velho e Criado <original> C.do </original> <nl>
- q De Vossa Senhora Criado muito vosso <original> V. Snr. a C. do m.to v. </original>
- q Deus <original> D.s </original> guarde a Vossa Excelência <original> V Excia. </original> <nl>
- q Deus guarde a Vossa Excelência <original> g.de a V. Exa. </original>
- q Deus guarde a Vossa Excelência <original> g.e a V. Exa. </original>
- q João Baptista <original> J. B. </original> (48)
- q João Baptista <original> J. Bap.ta </original> <nl> (7)
- q João Baptista <original> J. B.ta </original> <nl>
- q Muito Atento Vosso Criado Amigo e Obrigado <original> Mto. Atto. Vo. Cº. Am.o e Obgdo </original> <nl>
- q Muito atento Venerador e Criado <original> M.to att.o V.or C.do </original> <nl>
- q Muito atento Venerador e Criado <original> M.to att.o V.or e C.o </original> (2)
- q Muito atento Venerador e Criado <original> M.o att. o V. or C. do </original>
- q Muito atento venerador e criado <original> M.to att.o v.or c.o </original>
- q Muito atento venerador e criado <original> m.to att.o v.r c.do </original> <nl>
- q Muito atento e criado <original> M.to att.o e C.o </original>
- q Sou de Vossa Senhora Amigo criado e colega <original> V. Snr.a Am.o c.do e coll.a </original>
- q Teu Criado <original> C. </original> <nl>

7.3 Formas de Tratamento:

- q Dom <original> D. </original> (9) Luiz de Noronha.
- q Vossa Excelência <original> V. Exa. </original> (123)
- q Recebi a carta de que Excelência <original> Exa. </original> me fez favor
- q Vossa Excelência e os Senhores <original> V.Exa. e os Srs. </original>
- q Vossa Excelência <original> V. Excia. </original> que na quarta-feira

- q como Vossa Excelência <original> V.E. </original> sabe
- q Confesso a Vossa Excelência <original> V. Exa </original> (2) que é mais a elle
- q desejaria muito que Vossa Excelência <original> m.to q. V. Exa </original> os visse hoje
- q parabens a Vossa Excelência <original> V. Ex.ia </original> e felicitações a Patria
- q rogo a Vossa Excelência <original> V Exa. </original> (2) o favor de me esperar
- q Rogo a Vossa Senhoria <original> V Sra. </original> se sirva mandar expedir
- q bem o sabe Vossa Senhoria <original> V. Snr.a </original> (13).
- q Vossa Senhoria <original> V. Sr.a </original> <nl>
- q Vossa Senhoria <original> V. Sra. </original> (7)
- q Vossa Senhoria <original> V. S.a </original> (33)
- q Vossa Senhoria <original> V. Sa. </original> (6)
- q e protesto a Vossa Senhoria <original> V. S. </original> (27)
- q Vossas Senhorias <original> V. Snr.as </original>
- q Vossas Senhorias <original> V. SSas. </original>
- q Vossas Senhorias <original> V.V. S.S. </original>
- q Vossas Senhorias <original> VV SS </original>
- q Vossas Senhorias <original> VV SS. </original> (1)
- q Visconde <original> V. </original>
- q Sua Majestade <original> S. M. </original> a Rainha
- q Senhor <original> Sr. </original> (30) Duque
- q O Senhor <original> Snr. </original> (4) Francisco Alves da Silva Taborda
- q Suas Majestades <original> SS.MM. </original> (2)
- q mandado por Sua <original> S. </original> (12) Majestade El Rei
- q com a Senhora <original> Sra. </original> (5) Condessa de Tancos
- q Minha Senhora <original> M.a Sra. </original> (3)
- q Princeza Dona <original> D. </original> (8) Amélia

7.4 Muito, Que, Quanto, Para, Por que, Por, Minha

- q muito que <original> m.to q. </original>

- q quanto <original> q.to </original> (6)
- q para <original> p. </original>
- q que <original> q. </original> (322)
- q muita <original> m.ta </original> (11)
- q Há muito que <original> m.to q. </original>
- q noto quanto <original> q.to </original> (6)
- q anda retardada a remessa de Boletins para <original> p. </original>
- q e me julga causa das demoras que <original> q. </original> (322) há.
- q ja vou com elles para <original> p.a </original> (85) fallar
- q muita <original> m.ta </original> (11) honra
- q honra de beijar por <original> p. r </original> mim as Mãos
- q Muito <original> M.to </original> (54) parabem
- q devo muitos <original> m.tos </original> (6) favores e distincões
- q por que <original> p. q. </original>
- q por que <original> pr. q. </original>
- q porque <original> prq. </original>
- q por que <original> p.r q. </original> (21) todo elle é uma violência
- q porque <original> pq. </original> (2) o plenipotenciario
- q agora o Autor para <original> A. p.a </original> lhe dizer
- q por <original> p.r </original> (63) causa
- q minhas <original> m.s </original> flores
- q ésta minha <original> m.a </original> (16) casa, onde os esperarei à hora indicada.
- q na minha <original> ma. </original> ignorancia
- q quando <original> q.do </original> (6) fiz o primeiro

7. 5 Não encontrados no dicionário de abreviaturas, de M.H. Flexor

- q secretário da Secretaria de Sua Majestade <original> S. de S.M. </original>
- q Secretário d'Estado dos Negócios <original> S. d'Est. dos Neg. </original> Estrangeiros <nl>
- q esta Secretaria d'Estado <original> d'Est.o </original>
- q Recebi os despachos <original> desp.o </original> de Madrid e

q nem vai o Ministro <original> Min.o </original> (2) Ingles
q Quarta-feira <original> 4f.ra </original> 28 Maio <nl>
q 5 d ' Abril 10 Manhã <original> M. </original> (1852) (2)
q Mando n'este momento a os interessantes Ofícios que <original> Off.os q.
</original> acabo de receber de Madrid.
q Quinta-feira <original> 5.a f.a </original> 1 d'abril <nl>
q não ir a despeito <original> desp.o </original> por estar bastante
<original> bast.e </original> incommodado e tomei ésta madrugada um
remedio.
q Pateo do Pimenta. Quinta-feira <original> Quinta. f.a </original> <nl>
q Quarta-feira <original> 4a. f.a </original> de manhan <nl>
q Pateo do Pimenta <original> Pim. </original> 23 - novembro <original>
9.bro </original> .
q convem preencher o nosso tribunal de Comércio <original> Com.cio
</original>
q dos Negócios Estrangeiros <original> NN. EE. </original>
q os negócios da Justiça <original> J.a </original> sôbre o direito
q estou actualmente incumbido <original> actualm.te incumb.o </original> .
q e não dão provavelmente <original> provalm.e </original> lugar
q Forte da Estrella. Junq <original> Junq.ra </original> 25 de julho de 1854
q aprovação com que a Câmara <original> q. a Cam.a </original> foi
dissolvida
q uma necessidade que realmente <original> necessid.e que realm.te
</original> era.
q Dia de 3 de Dezembro <original> D. de 3 de X.bro </original> com todas as
suas violências e ilegalidades <original> illegalid.es </original>
q com efeito sensivelmente <original> eff.to sensivelm.e </original> e sem
dúvida
q O orçamento <original> orçam.to </original> já é mais economico
q emprehender obras uteis e fazer duplicados <original> dupl.os (?)
</original> produtores e fomentadores
q A Senhora Infanta de Espanha <original> Sr.a Inf.a de Hesp.a </original>
q Lisboa, 7 de Setembro <original> 7.bro </original> de 1852
q suas cartas recebidas de 8 e 17 do presente <original> pp. </original>
q 24 de outubro <original> 8.bro </original>
q com Gomes que vilmente <original> q. vilm.te </original> me trahiu
q Tenha indulgência <original> indulg.a </original> com esta fraqueza
q acêrto e admirável prudência <original> prud.a </original>

- q dificuldades <original> difficuld.es </original>
- q O enviado Extremado Ministro Plenipotenciário <original> Extr. Ministro Plen. </original>
- q 24 de fevereiro <original> fv. </original>
- q verão ao meu querido <original> q. </original> Minho
- q gôsto de levar pessoalmente <original> pessoalm.te </original> a seus pés.
- q Vai o arrendamento <original> arrendam.to </original> do Bastos (n MH)
- q 13 de Dezembro <original> X.bro </original> de 1841
- q Sexta feira <original> f.ra </original> 26 do corrente <original> corr.e </original> ás 7 horas da noite
- q Sexta-feira de manhã <original> 6a. f. a de m. </original> <nl>

7.6 Outros

- q n ' este momento <original> mom.o </original>
- q Hoje verá aqui junta a próva d'este mau serviço. Peço-lhe que dê promptas e severas ordens a este respeito <original> resp. o </original> <nl>
- q E meu amigo <original> am.o </original> (55) muito
- q mais verdadeira <original> verd.a </original> (2) e alta estima e de todo o Coração <original> C. </original> (32) <nl>
- q Felizmente <original> Felizm.e </original>
- q número <original> n.o </original> 36
- q Mas é o mesmo <original> m.mo </original> (5)
- q Tomando o pretexto d'este último <original> ult.o </original>
- q da Rainha <original> R.a </original> de Portugal
- q apparece é minha inferioridade <original> m.a inferiorid.e </original> e a superioridade do franchinote
- q sabes perfeitamente <original> perfeitam.te </original>
- q da Rainha <original> R.a </original> de Portugal
- q apparece é minha inferioridade <original> m.a inferiorid.e </original> e a superioridade do franchinote
- q sabes perfeitamente <original> perfeitam.te </original>
- q Aqui o mais importante <original> import.e </original> é a geral approvação
- q chegou aqui bastante <original> bast.e </original> (2)doente,
- q chegou a dar cuidado <original> cuid.o </original> (3) serio a sua molestia

q Eu especialmente <original> especialm.te </original> lhe devo
 q satisfarei como quem <original> q.m </original> (5) sou
 q nem despedir-me de e de nossa correspondência <original> correspond.a.
 </original> official
 q Meu amigo colega <original> am.o Coll.a </original> <nl>
 q Duas palavras verdadeiras <original> verd.as </original> e singelas como
 eu.
 q felizmente <original> felizm.te </original> recebi
 q n'uma bestialidade <original> bestialid.e </original> sem nome e sem
 proveito.
 q da autoridade <original> auctorid.e </original> de um ministro
 constitucional
 q bons amigos <original> am.os </original>
 q Esta é a verdade <original> verd.e </original>
 q com particular <original> p.ar </original> estima
 q longa infirmitade <original> infirmid.e </original>
 q No primeiro momento <original> mom.to </original>
 q se não esqueça de algumas circunstâncias <original> circumst.as </original>
 q coisa porque insisto <original> insto. </original>
 q minha vontade <original> m.a vont.e </original>
 q Forte da Estrella Sexta-feira <original> Sexta-fa. </original>
 q Lisboa <original> Lx. </original> 14 de julho
 q quantia de 10000000 <original> 10.000\$000 </original>
 q 14 do corrente <original> corr.e </original>
 q Primo e amigo <original> Pr.o e am.o </original> velho

7.7 Palavras não encontradas

q Meu amigo M <original> am.o M.im (?) </original> <nl>
 q não quer dar ao A <original> A. </original> o incómodo
 q e o de ser o Rod min dos E E <original> Rod.ro min.o dos E. E. </original>
 q há de estimar que eu appele a Rod <original> Rod.no </original> isto é dos
 q A Licínio F. C. de Carvalho
 q Snakers do C <original> C. </original> do Lavradio e do Rodrigo

8. ANEXO II

ANEXO ABREVIATURAS

Listagem de acentuação retirada de T. Menegatti(2001).

DOM JOAO	DIOGO DO COUTO	MANOEL DA COSTA
acertará1	açucar2	abundancia3
africa1	alagôa2	accessorio3
agoa1	âmanhã2	ácolá3
agradavel1	aprasivel2	acômete-os3
allvara1	armazens2	acrédor3
almazem1	Arménia2	acrédores3
apontaveis1	Aureola2	admiraveis3
apos1	Babilónia2	admiravel3
Arabia1	Bisnagá2	adóce3
Arbitros1	Bofatás2	agradavel3
Aspero1	Camara2	agua3
Avida1	Caricias2	aguias3
Concordia1	Concluido2	ahi3
Cò1	Consul2	alcatêas3
Còta1	Crem2	álem-mar3
Dara1	despi-lo2	alguem3
Diminuissem1	dôce2	amendoa3
Dira1	dominio2	amigaô3
Dizies1	escandalo2	ancora3
Eficacea1	espadaúdo2	antidoto3
Estará1	espiritos2	Antonio3
Frances1	êsse2	Apices3
Has1	esteril2	Apostolica3
Juizo1		
Juizos1		
Merces1		

Misterio1
Notoreo1
Ordinaria1
Outrem1
Trara1

Farimá2
Fatima2
Fertil2
Fôlhas2
Fortissimamente2
Grãos2
Gravissima2
Hostia2
Incendio2
Infamia2
Inutil2
Juizo2
Louçãmente2
Oculos2
Ordinarias2
Papeis2
Parabens2
Paraiso2
Particula2
Prejuizo2
Quarteis2
Raizes2
Refens2
Regio2
Relampagos2
Reliquias2
Relogio2
Respondencia2
Ruinas2
Saido2
Santarem2
Santissima2
Serenissimo2
Soberbissimo2
Solido2
Sómente2
Subita2
Substancia2
Superfluo2
Taboas2
Tartaro2

Apréstaõ3
Arbitro3
Arêa3
Arganá3
Armazens3
Arruido3
Artifices3
Artificio3
Arvores3
Assistencia3
Atomos3
Attonito3
Atribuimos3
Audiencia3
Babilonia3
Bacalháo3
barbaria3
bençaõ3
bebados3
beneficio3
bótas3
boticario3
cábe3
cadêa3
cadimo3
cambio3
Camera3
Candeinhas3
Candêa3
Candido3
Capatáz3
Capêaõ3
carestia3
carnás3
Cartorio3
Catholico3
Cedulas3
Ceo3
Ceremonias3
Certissimo3
céva3
Christãa3
Christaõ3
Ciencia3
Circunstancia3
Clausula3
Clerigo3
Cõmercio3
Compor3
compró3
comparaçaõ3
composiçaõ3
conciencia3
concurrency3
conquistó3
cofórmes3
consequencia3
consideraveis3
constancia3
continuos3
contraditorio3
contrario3
controversia3
conveniencias3

Córando3
Córar3
Correspondencia3
Creditos3
Critico3
Crueis3
Dadivas3
Dadivas3
Decalogo3
Degráos3
Delirio3
Demonio3
Démo3
Depositos3
Desobediencia3
Destruído3
Detestavel3
Deuteronomio3
Diabolicas3
Dicordias3
Diétas3
Diminuíó3
Discipulo3
Discordias3
Dispoz3
Distraidas3
Dizima3
Dizimos3
Domesticos3
Domicilio3
Dominica3
Dominio3
Donatario3
Duvidas3
Duzia3
Ecclesiastico3
Elogíós3
Eminencia3
Emisferio3
Emprestimo3
Envoltorio3
Erario3
Escandalo3
Escritorio3
Escrupulo3
Especie3
Espirito3
Essencia3
Estancia3
Estavel3
Estimavel3
Estimulo3
Estipendio3
Estomago3
Evangelicos3
Evidencia3
Evora3
Excelencias3
Excluía3
Exercito3
Exercitos3
Extraordinaria3
Fabrica3
Facil3

Factive3
Fallencias3
Familias3
Fantastico3
femea3
Filosofo3
Furia3
Ganancia3
Gemeas3
Generos3
Grandissima3
Haereditario3
Harpêo3
Heroico3
Homicidios3
Horoscopos3
Hospedes3
Ignorancia3
Ignoraveis3
Iliada3
Iligitimo3
Illicito3
Immemoravel3
Impetos3
Importancia3
Impossivel3
Improvavel3
Incendio3
Incontrastaveis3
Industria3
Inexpugnavel3
Infalliveis3
Infatigavel3
Infieis3
Influencias3
Innocencia3
Innumeraveis3
Insensiveis3
Insolencias3
Instancias3
Instítuirã3
Inteligencia3
Intoleraveis3
Intrepidos3
Inuteis3
Invencivel3
Inviolavel3
Invisiveis3
Irremediaveis3
Jurisprudencia3
Legitima3
Lêm3
Léme3
Lepido3
Licitos3
Ligitimos3
Liquido3
Louvavel3
Magnificencia3
Malicia3
Mathematica3
Matricula3
Matrimonio3
Maximo3

Mecanicos3
Medicos3
Medio3
Memoria3
Meritos3
Milicia3
Minimos3
Ministerios3
Misericordias3
Modestia3
Molestia3
Mórtos3
Nádaõ3
Necesario3
Necessario3
Ninguem3
Notaveis3
Notavel3
Noticia3
Notoria3
Numero3
Obediencia3
Obediência3
Observancia3
Odio3
Oleo3
Opulencia3
Outrem3
Óvos3
Paciencia3
Pacifica3
Pàdar3
Pádeiras3
Palacio3
Papeis3
Parabola3
Paragrafo3
Paraiso3
Parentélla3
Partivel3
Patria3
Patrimonio3
Pensionario3
Perús3
Pezame3
Plenario3
Poderiamos3
Politico3
polvora3
Pompêo3
Pontífice3
Possuía3
Possuisse3
Potencias3
Preeminencias3
Prégador3
Prégar3
Premio3
Presidio3
Prestimo3
Primicias3
Primùm3
Principe3
Prióste3

Privilegio3
Prodigio3
Prohibitoria3
Propicio3
Propor3
Proposito3
Propria3
Proprietario3
Proprio3
prospero3
prosperos3
provavel3
proverbio3
providencia3
Provincia3
Proxima3
Publica3
Purgatorio3
Rectissimamente3
Regalia3
Reliquias3
Relogio3
Remedio3
Remedios3
Republicas3
Repugnancias3
Residencia3
Resistencia3
Restituia3
Retem3
Revéle3
révera3
reverencia3
Rhetorica3
Ridiculo3
Ruina3
Sabio3
Sacrilegos3
Sadio3
Sáquate3
Saguates3
Salario3
Saudavel3
Saude3
Sável3
Secretario3
Sedéla3
Serêas3
Serenissima3
Setima3
Silencio3
Simonia3
Simonias3
Sitio3
Sofriveis3
Subditos3
Subrepticia3
Subsidios3
Substancia3
Subtilissimos3
Tabaliôas3
Taboa3
Tacito3
Territorio3

		Terrivel3 Titulo3 Tocámos3 Tolére3 Tomò3 Trafegos3 Tragicos3 Tratavel3 Travéz3 Treguas3 Tres3 triennio3 tutelá3 tyrannicas3 ultima3 Ungría3 Unico3 Urgencias3 Util3 Valéra3 Valído3 Varias3 Vastissimo3 Vicio3 Victoria3 Vigilancia3 Vitoria3 Vituperios3 Viuva3 Voluntario3 Voluntarios3 Medio3
--	--	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

MELO	MARIA DO CEU	ANTONIO DA COSTA	MARQUÊS DA FRONTEIRA E ALORNA
Àvante5 Belem5 Bésta5 Cairam5 Canónico5 Ceremónia5 Cesar5 Ciume5 Cómica5 Cómico5 Cómodo5 Cónego5 Desdens5 Detem5 Discorressemos5 Frângãos5 Impios5 Inseparávelmente5 Jámais5 Officio5 Ordináriamente5 Papeis5 Parabens5 Prègação5 Prègador5 Reïteração5	à6 abstinencia6 abundancia6 adversario6 advertencia6 agua6 Agua6 Alem6 Ali6 Altissimo6 Ambar6 Amplissimo6 Ancia6 Animos6 Anuncios6 Aparencia6 Apostolo6 Às6 Ascendencia6 Aspera6 Aspero6 Assistencia6 Atomo6 Auzencia6 Cà6 Camara6	Àguardente7 bêsta7 cómodo7 compor7 compos-se7 corporeas7 crueis7 demónios7 dezóito7 estrêlas7 génio7 gondolas7 ideia7 leem7 malignas7 noutrem7 património7 prègações7 proibido7 roem7 saude7 segrêdo7 senti-las7 sôltas7 tempora7 vè-lo7	Abstinencia8 Acerrimo8 Adormeciamos8 Africa8 Agradavel8 Agua8 Agua8 Alem8 Alguem8 Amavel8 America8 Amigavel8 Andavamos8 Aneis8 Animos8 Aristocratica8 Assembleia8 Assiduos8 Audiencia8 Ausencia8 Austriaco8 Bebados8 Beneficencia8 Breviario8 Brilhantissimo8 Britanicas8

Saüdades5	Canà6	veras7	cadaver8
Saüdáveis5	Character6		calculos8
Tambem5	Catholico6		calendario8
Tránsito5	Ciumes6		cartorio8
Voces5	Clerigo6		chapeu8
Võe5	Colera6		circumstancia8
	Commodo6		cirio8
	Competencia6		claviculas8
	conciencia6		clerigos8
	conferencias6		coincidencia8
	consciencia6		colonia8
	consequencias6		comico8
	contrario6		concluia8
	constancia6		concorrenca8
	conveniencias6		condestaveis8
	cre6		condiscipulo8
	credito6		Conego8
	dà6		Consciencia8
	dè6		Consequencia8
	demonio6		Consideravel8
	designio6		Consortios8
	dilicias6		Contiguas8
	diligencias6		Continencia8
	dirà6		Continuos8
	discipula6		Conveniencia8
	divorsio6		Cór8
	domesticos6		Corôa8
	domicilio6		Coroneis8
	encastoarà6		Correspondencia8
	espírito6		Decadencia8
	estomago6		Dêdos8
	Excellentissimo6		Deleitavel8
	Excluido6		Desagradavel8
	Exercitos6		Desculpavel8
	Extraordinaria6		Detestavel8
	Familia6		Diacono8
	Fè6		diarias8
	Fee6		discipulos8
	Ficarà6		divergencia8
	Filosofos6		dôce8
	Graã6		domesticos8
	Haã6		dôr8
	Idolo6		economicas8
	Illuminatiua6		edificios8
	Imperio6		elastico8
	Importancia6		elegancia8
	Impossivel6		episodio8
	Inclemencias6		escapulario8
	Inclinadissima6		escrupulos8
	Incognitas6		especie8
	Incommodo6		espectaculo8
	Inconstancia6		espíritos8
	Inculpavel6		estereis8
	Indicatorio6		exercitos8
	Indicio6		exigencias8
	Indulgencias6		existencia8
	Inferirà6		exotica8
	Infortunio6		extraordinarias8
	Innocencia6		faceis8
	Insoportaveis6		familia8
	Instancia6		fanatico8
	Josè6		finissimas8
	Juizo6		formidavel8
	Lagrima6		fortissimo8

Licitos6
Magua6
Materia6
Memoria6
Merce6
Menhaã6
Misericordia6
Molítico6
Musico6
Nobilissimo6
Nòs6
Notavel6
Noticia6
Obedientissima6
Observancia6
Odio6
Officios6
Opéra6
Oraculo6
Outrem6
Ouviamos6
Paciencia6
Palacio6
Papeis6
Paraizo6
Particula6
Patria6
Patrocínio6
Pè6
Pendencia6
Penitencias6
Perfidia6
Perola6
Politica6
Ponderarà6
Pontifice6
Potencias6
Premio6
Presepio6
Primogenita6
Principios6
Proposito6
Proprio6
Prouidencias6
Proxima6
Prudencia6
Raizes6
Relogio6
Remedio6
Riquissimamente6
Ruido6
Sacrificio6
Sahira6
Santissima6
Santuário6
Saude6
Seculo6
Silencio6
Singularissima6
Sitio6
Sonolencia6
Soportaveis6
Subditos6
Subidissimo6

frescôr8
funebres8
gastronomo8
genio8
gloria8
habil8
hereditarias8
horriavel8
hostia8
identica8
ignorancia8
imaginarios8
imaginavel8
Imperio8
Impertinencias8
Importancia8
Impossivel8
Impoz8
Inconsolavel8
Inconveniencia8
Incrivel8
Independencia8
India8
Individuo8
Inexplicaveis8
Infancia8
Inqualificavel8
Instruida8
Intendencia8
Intimos8
Invalidos8
Inverosimil8
Irreparavel8
Joias8
Juizo8
Lamentavel8
Legionarios8
Legitimos8
Leguas8
Licôres8
Limitadissimo8
Lingua8
Luminarias8
Maçonicos8
Madreperola8
Magnificos8
Maiusculas8
Maniaco8
Medicos8
Mediterraneo8
Mêdo8
Melancolica8
Memoravel8
Memorias8
Ministerio8
Miseravel8
Missionario8
Miudo8
Modêlo8
Modestia8
Moidos8
Monotona8
Movei8
Negocios8

Substancia6
Tafetà6
Tambem6
Terà6
Timida6
Titulos6
Tres6
Triduo6
Tumulo6
Ultimo6
Và6
Vè6
Veneravel6
Vigilancia6
Vigilantissima6
Voo6

Ninguem8
notaveis8
numerosissima8
nupcias8
obsequios8
ocio8
oculos8
odio8
Ondê8
Opio8
Ordinaria8
Paciencias8
Paineis8
Paizes8
Palacio8
Panico8
Parabens8
Partidarios8
patibulo8
patria8
patricio8
patrioticas8
pecuniarios8
penitencias8
pequenissima8
pericia8
periodo8
perola8
pesadissimo8
pessimas8
pêtas8
pêzames8
pêzo8
plateia8
politica8
portatil8
possiveis8
poz8
preambulo8
prégador8
prejuizos8
premio8
presepio8
presidencia8
primogenito8
principios8
proposito8
proprietarios8
proprio8
provincia8
proximas8
quarteis8
rapida8
razoaveis8
referencia8
regencia8
reliquia8
remedio8
reminiscencias8
represalia8
repugnancia8
respeitavel8
reus8
revez8

		revolucionario8 Ridicula8 Riquissimas8 Risivel8 Romantico8 Sabio8 Sacratio8 Sacrificio8 Saude8 Seculo8 Seminario8 Semsabôres8 Semsaborissima8 Serios8 sitios8 socios8 solitarios8 sómente8 sôpas8 subsistencia8 substituidos8 suburbios8 Suecia8 Tambem8 Territorio8 terriveis8 timido8 tisico8 titulos8 tôla8 tôrto8 tragicas8 Tres8 Triplice8 Tristissimo8 Tumulo8 Ultimos8 Varios8 Védor8 Velhissimo8 Vesperas8 Vestuario8 Vigario8 Vigilancia8 Voluntarios8
--	--	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

-

-

[1] Dizemos *possivelmente*, porque, diferentemente de Rita Marquilha, que teve a oportunidade de confrontar manuscritos e suas respectivas edições, no caso do CTB a preocupação era, não a mesma de Marquilhas, mas apenas a de utilização de edições filologicamente confiáveis, no sentido de serem baseadas em manuscritos autógrafos ou acompanhadas pelo próprio

autor, de modo a garantir que a edição seria segura para estudos de caráter sintático. Algumas das obras utilizadas no *corpus*, como é o caso dos sermões do Pe. Antônio Vieira, nem têm mais sua versão manuscrita. De qualquer forma, como utilizamos uma edição baseada integralmente na edição feita pelo próprio autor, o critério filológico citado não foi violado nem neste caso, nem em qualquer outro. De todo modo, as variedades de edições encontradas no *corpus* são, no que diz respeito à grafia, de três tipos: umas são cópia fiel do manuscrito autógrafo em todos os aspectos lingüísticos; outras seguem tal e qual o manuscrito autógrafo no que diz respeito à sintaxe, mas não no que diz respeito à grafia; outras ainda foram feitas com base no manuscrito autógrafo, mas este já não existe (ou, se existe, não se sabe onde estará). Para este último caso, não sabemos dizer se a variação de grafia observada é do próprio autor, ou do tipógrafo/editor.

[2] Projeto Temático *Padrões Rítmicos, Fixação de Parâmetros e Mudança Lingüística*, coordenado por Charlotte Galves (IEL - UNICAMP) e subsidiado pela FAPESP (Proc. 98/3382-0).

[3] Lembrar oportunamente que, no sistema latino, *v* e *j* não têm valor consonântico, mas vocálico, uma vez que correspondem a *u* e *i*.

[4] Não há um único caso de grafia *ss* em contexto onde hoje esperaríamos *s* com valor fonêmico /s/.

[5] *ss* já é dobrado, não se esperando, portanto, como confirmam os dados, ocorrências de *ssss*.

[6] Não foi observada nenhuma ocorrência de *ch* dobrado.

[7] Não foi observada nenhuma ocorrência de *x* dobrado.

[8] Como já observado anteriormente, *v* dobrado tem uso categoricamente bloqueado.

[9] Não há um único caso de grafia *ss* em contexto onde hoje esperaríamos *s* com valor fonêmico /z/.

[10] Como já observado anteriormente, *z* dobrado tem uso categoricamente bloqueado.

[11] Também como já observado anteriormente, *j* dobrado tem uso categoricamente bloqueado.

[12] COSTA, Manuel da (1601-1667). *Arte de Furtar* (seleção, introdução e notas de Roger Bismut). Lisboa, Imprensa Nacional Casa da Moeda, 1991. (número de palavras do texto: 52.867)

CÉU, Maria do (1658-1753). *Relação da Vida e Morte da Serva de Deos a Venerável Madre Elenna da Crus* (transcrição do Códice 87 da Biblioteca Nacional precedida de um estudo histórico, por Filomena Belo). Quimera. Lisboa, 1993. (número de palavras do texto: 27.410)

[13] CÉU, Maria do (1658-1753). *Relação da Vida e Morte da Serva de Deos a Venerável Madre Elenna da Crus* (transcrição do Códice 87 da Biblioteca Nacional precedida de um estudo histórico, por Filomena Belo). Quimera. Lisboa, 1993. (número de palavras do texto: 27.410)

[14] Não houve tempo para observar isto nos outros autores.

[15] Também não houve tempo suficiente para observar isto.

[16] Este relatório, assim como Britto & Menegatti (em preparação), será enviado ao Prof. Marcelo Finger, que é membro do projeto temático Fapesp ao qual o CTB está vinculado e é o responsável pelo etiquetador automático. Conforme o professor nos disse no último encontro do projeto temático, de 5 a 16 de agosto de 2002, o novo etiquetador está sendo treinado com base em um algoritmo utilizado em genética. O algoritmo funciona assim: tendo uma cadeia de gens A e outras duas, B e C, calcula-se quais modificações devem ser feitas em B e C para que fiquem como A. O problema da variação de grafia é o mesmo: temos a escrita *ele* e outras duas, *elle* e *ële*. Dever-se-á calcular quais modificações devem ser feitas em *elle* e *ële* para chegarmos a *ele*.

[17] O número ao lado da palavra indica o autor de cujo texto a palavra foi retirada, como dito no relatório parcial deste projeto.

[18] GARRETT, Almeida. *Cartas de Garrett*. (apresentação e edição por Segismundo Spina). São Paulo, Humanitas Publicações FFLCH/USP, 1997.

[19] GARRETT, Almeida. *Cartas de Garrett*. (apresentação e edição por Segismundo Spina). São Paulo, Humanitas Publicações FFLCH/USP, 1997.

[20] FLEXOR, Maria H. *Abreviaturas, Manuscritos do século XVI ao XIX*. Editora Unesp – secretaria do Estado da Cultura – Arquivo do Estado de São Paulo, 1991.

[21] GARRETT, Almeida. *Cartas de Garrett*. (apresentação e edição por Segismundo Spina). São Paulo, Humanitas Publicações FFLCH/USP, 1997.

[22] A lista completa de abreviaturas encontradas no CTB, e não presentes em Flexor, encontra-se anexa ao trabalho.

[23] COUTO, Diogo do. *Décadas* (seleção, prefácio e notas de António Baião). Vol 1. Lisboa, Livraria Sá da Costa - Editora, 1947.