

# Procorph: um Sistema de Apoio à Criação de Dicionários Históricos

Arnaldo Candido Junior  
Sandra Maria Aluísio



# Roteiro

- O projeto DHPB
- Motivação do trabalho
- O corpus do projeto
- O sistema Procorph
- Conclusão

# O projeto DHPB

- Dicionário Histórico de Português do Brasil
  - O primeiro do gênero
  - Português brasileiro começou a divergir do Português europeu (cultura, fauna, flora)
  - Período pré-imprensa do Brasil (1500-1808)
- Baseado em um grande corpus de textos históricos
  - Autores brasileiros ou portugueses que viveram no Brasil por um longo período

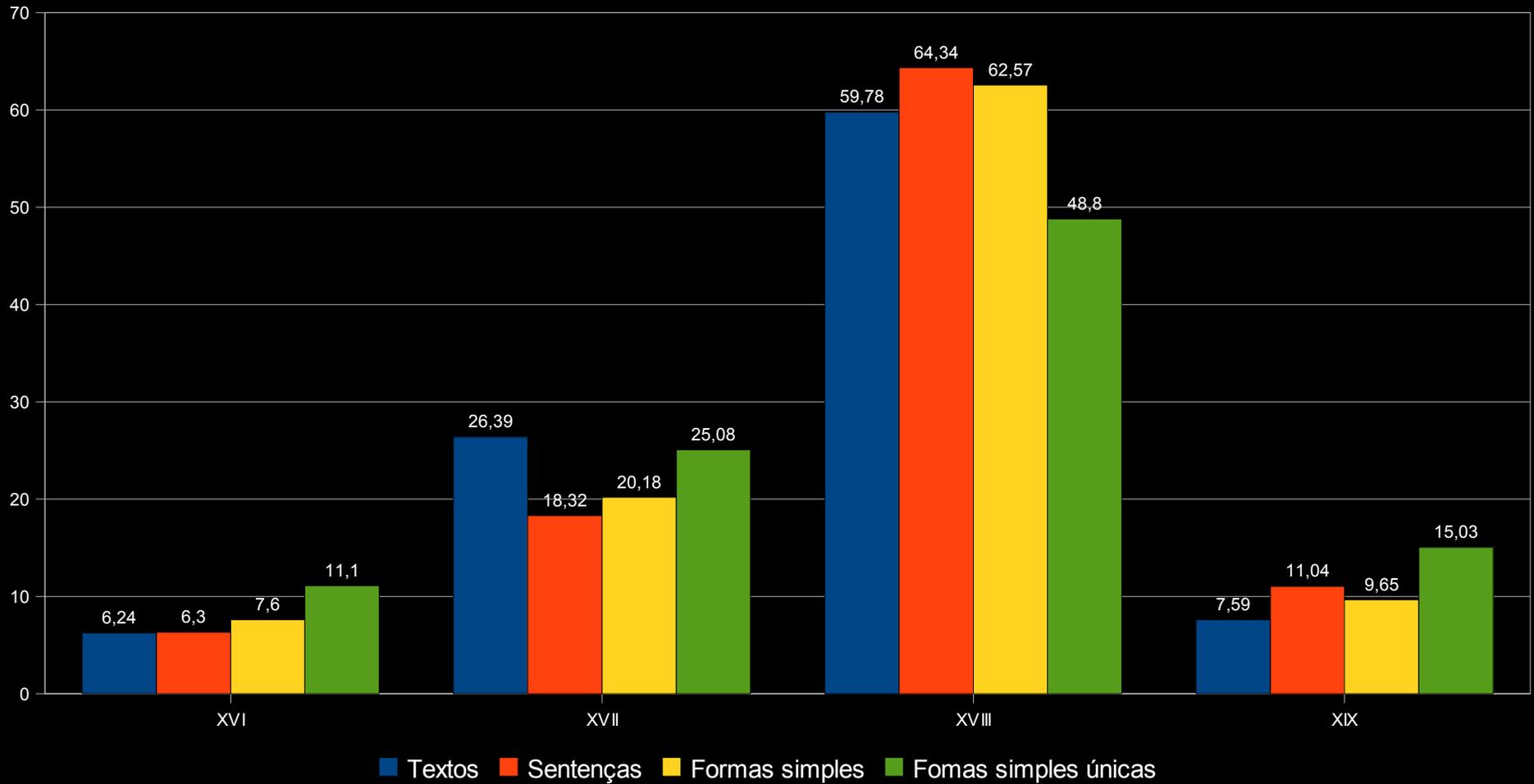
# Motivação

- Necessidade de edição colaborativa
- Formatação de verbetes se mostrou uma tarefa demorada
- Ferramentas existentes (Haddad, 1999) não satisfazem a necessidade de dicionários históricos
  - Tratamento de variantes de grafia e de símbolos especiais
  - Controle de datação

# Dados do C3rpus

- 2.5 mil textos
- 7.5 milh3es de palavras
- 369 mil palavras 3nicas
- 82 MB (UTF-16LE)

# Textos por século (%)



# Carta do padre Manuel da Nóbrega ao padre Simão Rodrigues (1549)

7. - BAÍA 9 DE AGOSTO DE 1549

127

convertidos, onde estaremos Vicente Rodriguez e eu, e hum soldado<sup>19</sup> que se meteo comnosco para nos servir, e está agora em Exercicios, de que eu estou muy contente. Faremos nossa igreja, onde insinemos os nossos novos christãos, e aos domingos e festas visitarey a Cidade e pregarey. <sup>205</sup>

O Padre Antonio Pirez e o P.<sup>e</sup> Navarro estaram em outras Aldeas longe, onde já lhes fazem casas. E portanto hé necessario V. R. mandar officiaes, e am-de vir já com a paga, porque cá diz ho Governador que, ainda que venha Alvará de S. A. para nos dar o necessario, que nom o averá <sup>210</sup> hi para isto. Os officiaes que cá estão tem muito que fazer, e que o nom tenham, estão com grande saudade do Reyno, porque deixão lá suas molheres e filhos, e nom aceitaram a nossa obra depois que cumprirem com S. A., e tambem ho trabalho que tem com as viandas e o mais os tira disso. <sup>215</sup> Portanto me parece que avião de vir de lá, e, se possivel

# Características de textos históricos

- Rydberg-Cox (2003)
  - Ausência de hifenização
  - Símbolos tipográficos incomuns
  - Junções de palavras como “éamor”
  - Abreviaturas
  - Variações de grafia

# Abreviaturas

- Expansões de B<sup>o</sup>: bairro, Bartolomeu, bastardo, beco, bento, Bernardo, etc.
- Abreviaturas de janeiro: jan., jan.<sup>ro</sup>, jan<sup>ro</sup>, janr.<sup>o</sup>, jan.<sup>o</sup>, etc.
- Solução: criação de um glossário de abreviaturas extraído de Flexor (1991)

# Variantes de grafia

- Inexistência de um sistema ortográfico unificado em textos históricos
- Exemplo: villa, vyla, vjlla, vylla, vjla
- Abordagem 1: regras de transformação (Giusti et al., 2007)
- Abordagem 2: Distância de edição (Philologic + Agrep)

# Regras de transformação

- Exemplo: (e[ao] e ei)
  - "e[ao]" cobre "aldeia", "meo", "cheas", etc
  - "e" subcadeia a ser substituída (aldeia, meo, cheas, etc)
  - "ei" cadeia de substituição (resultado em aldeia, meio, cheias, etc)
- 43 regras de transformação, 18.082 agrupamentos e 41.170 variantes

# O sistema Procorph

- Verbetes
  - Edição colaborativa de verbetes e sub-entradas
  - Formatação automática
  - Referências automáticas (Autor. Título, ano)
  - Conversão de formato

# O sistema Procorph (cont)

- Recursos auxiliares
  - Concordâncias (para abonações)
  - Glossário de variantes de grafia
  - Inserção simplificada de símbolos especiais
- Atualmente: testado por 4 dos 21 redatores

# Níveis de acesso

- **Consulente:** visualiza verbetes, concordâncias e glossários
- **Lexicógrafo:** cria glossários
- **Revisor:** altera verbetes de outros usuários
- **Administrador:** adiciona usuários na base

# Tela inicial

## ProCorPH

PROcessador de CÓRpus do Português Histórico

[Ajuda](#) [Contato](#) [Sair](#)

### Procorph

[Início](#)

[Preferências](#)

[Símbolos especiais](#)

### Dicionário

[Verbetes](#)

### Córpus

[Textos](#)

[Concordâncias](#)

[Variantes de grafia](#)

### Administração

[Usuários](#)

## Bem-vindo(a) Arnaldo Candido Junior

Último acesso: 25/10/2008 às 00:50:02.

Acesso rápido:

- [Verbetes](#)
- [Textos](#)
- [Formatação de texto](#)

## Sobre

O Processador de Córpus de Português Histórico (Procorph) é uma sistema voltado para o gerenciamento de verbetes históricos e processamentos de córpus históricos, desenvolvida para o projeto do Dicionário Histórico do Português do Brasil (DHPB). Esta ferramenta é software livre, e está sujeita aos termos da licença [GNU/GPL](#).

# Verbetes

- Dados gerais
- Variantes

## Alterar verbete

Verbetes:

Classe e atributo:

Situação:

Redator:

Data de criação: 2008-01-24

### Variantes de grafia

|  |  |
|--|--|
| Variante: <input type="text" value="prejuizo"/>  | <a href="#">Acima</a> / <a href="#">Abaixo</a> / <a href="#">Remover</a> |
| Variante: <input type="text" value="preiuizo"/>  | <a href="#">Acima</a> / <a href="#">Abaixo</a> / <a href="#">Remover</a> |
| Variante: <input type="text" value="preioizo"/>  | <a href="#">Acima</a> / <a href="#">Abaixo</a> / <a href="#">Remover</a> |
| Variante: <input type="text" value="preijuiso"/> | <a href="#">Acima</a> / <a href="#">Abaixo</a> / <a href="#">Remover</a> |
| Variante: <input type="text" value="preyuizo"/>  | <a href="#">Acima</a> / <a href="#">Abaixo</a> / <a href="#">Remover</a> |
| Variante: <input type="text" value="preyoizo"/>  | <a href="#">Acima</a> / <a href="#">Abaixo</a> / <a href="#">Remover</a> |
| Variante: <input type="text" value="prejoizo"/>  | <a href="#">Acima</a> / <a href="#">Abaixo</a> / <a href="#">Remover</a> |
| Variante: <input type="text" value="prejuiso"/>  | <a href="#">Acima</a> / <a href="#">Abaixo</a> / <a href="#">Remover</a> |
| Variante: <input type="text" value="perjuizo"/>  | <a href="#">Acima</a> / <a href="#">Abaixo</a> / <a href="#">Remover</a> |
| Variante: <input type="text" value="prejuifo"/>  | <a href="#">Acima</a> / <a href="#">Abaixo</a> / <a href="#">Remover</a> |

# Verbetes

- Abonações
- Acepções
- Referências

## Acepções

Acepção: Dano ou perda de qualquer natureza, como a honra, a saúde. Detrimento.

Atributos:

Abonação: Chamei o Padre Antão Gonçalves e examinando bem o caso, achei que tudo fôra uma equivocação, que causara uma má intelligencia das palavras no animo do capitão-mór. Comtudo, como já se não haviam de dar bem ambos, e podia de lá seguir algum \*prejuizo\* á missão, tirei o padre de Tapeacorú, pondo em seu lugar o Padre Manoel Rabello e o Padre João [p.531] de Avellar, o qual por. mais velho, na religião, ficou por missionario principal daquela residencia, sendo por suas muitas virtudes ambos

Texto: A00\_0548 padre. joão felippe betendorf [1699]. *capitulo 15 - acudo á aldea dos guajajaras no marcú e remedeio a um desgosto do capitão-mor do tapeacorú, no maranhão*

Página: 530

[Acima](#) / [Abaixo](#) / [Remover](#)

Acepção: Dano financeiro, perda de dinheiro.

Atributos:

Em 21 do passado me recolhi a esta sua caza de volta da minha jornada, e em aliqua couza aproveittou respeito a cobrar aliqua couza do que se deve a esta caza, e mais o havia de ser se o não tivesse empedido a falta de sulimão naquella caza de moeda, que ja passa de dous mezes, q. não

# Verbetes

- Relacionados
- Observações
- Primeira datação

## Verbetes relacionados

[Adicionar verbete relacionado](#)

## Observações

Observação:

A forma 'prejuizo' é a mais frequente.

[Acima](#) / [Abaixo](#) / [Remover](#)

[Adicionar observação](#)

## Primeira datação

Primeira datação:

E todos tem bem o que hão mister, e a necesidade lhes não fará \*prejuizo\* algum. Estão espantados de ver a magestade com que entramos e estamos, e temem-nos muito, o que tambem ajuda.]

Texto:

A00\_0694 padre manuel da nobrega . [1549]. *carta que o padre manonel da nobrega, preposito provincial da companhia de jesus, em o brasil, escreveu ao padre mestre simão o anno de 1549. (ms. copiado da livraria publica)*

Página:

460

# Sub-entradas

**mulher ama** (substantivo feminino): redigido por *Aparecida Negrí Inquarto* em 17/03/2008. Situação atual: Redigindo.  
[ver](#) | [alterar](#) | [apagar](#)

**mulher branca** (substantivo feminino): redigido por *Aparecida Negrí Inquarto* em 17/03/2008. Situação atual: Redigindo.  
[ver](#) | [alterar](#) | [apagar](#)

**mulher do Reino** (substantivo feminino): redigido por *Aparecida Negrí Inquarto* em 17/03/2008. Situação atual: Redigindo.  
[ver](#) | [alterar](#) | [apagar](#)

**mulher meretriz** (substantivo feminino): redigido por *Aparecida Negrí Inquarto* em 17/03/2008. Situação atual: Redigindo.  
[ver](#) | [alterar](#) | [apagar](#)

**mulher moça** (substantivo masculino): redigido por *Arnaldo Cândido Junior* em 17/03/2008. Situação atual: Redigindo.  
[ver](#) | [alterar](#) | [apagar](#)

**mulher parda** (substantivo feminino): redigido por *Aparecida Negrí Inquarto* em 17/03/2008. Situação atual: Redigindo.  
[ver](#) | [alterar](#) | [apagar](#)

**mulher pública** (substantivo feminino): redigido por *Aparecida Negrí Inquarto* em 17/03/2008. Situação atual: Redigindo.  
[ver](#) | [alterar](#) | [apagar](#)

**mulher preta** (substantivo feminino): redigido por *Aparecida Negrí Inquarto* em 17/03/2008. Situação atual: Redigindo.  
[ver](#) | [alterar](#) | [apagar](#)

**mulheres índias** (substantivo feminino): redigido por *Arnaldo Cândido Junior* em 17/03/2008. Situação atual: Redigindo.  
[ver](#) | [alterar](#) | [apagar](#)

**mulheres da terra** (substantivo masculino): redigido por *Aparecida Negrí Inquarto* em 17/03/2008. Situação atual: Redigindo.  
[ver](#) | [alterar](#) | [apagar](#)

# Visualização

**açúcar:** substantivo feminino.

Variantes: assúcar, asúcar

1. Substância doce fabricada industrialmente, extraída, em geral, da cana de açúcar.

Porque quem vos mostrara, há hoje trezentos anos, uma cana de que se faz o **açúcar** e vos dissera que daquela cana se havia de formar com a indústria humana um pão de **açúcar** tão formoso como hoje o vemos, tê-lo-íeis por coisa ridícula; e, pelo conseguinte, se vos fôsse mostrado um pedaço de pano velho de linho e vos afirmassem que daquele pano se havia de fazer o papel em que escrevemos, quem duvida que o teríeis por zombaria? **ambrósio fernandes brandão [1618]. diálogo primeiro, p. 1 .**

**Primeira datação:** A renda, que El-Rei cá tem nesta Baía, é esta, scilicet: as miunças que rendem cento e vinte mil réis em que andam arrendadas; o peixe e mandioca e algodão andam em cento e trinta mil réis; pagos em ordenado, que é um terço menos, pode valer em dinheiro oitenta mil réis; o **açúcar** do Engenho anda em cento e cinquenta cruzados. **p. manuel da nóbrega [1558]. carta do p. manuel da nóbrega ao p. miguel de torres, baía 8 de maio 1558, p. 26.**

# Conversão de formato

**comarca:** substantivo feminino.

Variantes: comarqua, comarcão, comarquã, comarça

1. Cada uma das circunscrições judiciárias em que se divide o território de um Estado, ou seja, divisão judicial, que fica sob a alçada de um juiz de direito.

Tirando e extinguindo de todo a Casa da Relação da Bahia, podia em seu lugar criar no Estado três corregedores com título da **comarca**, da maneira que os há no Reino e com a mesma alçada; e quando se lhes acrescentassem mais alguma quantidade, não o teria por desacertado. **ambrósio fernandes brandão [1618]. diálogo primeiro, p. 30 .**

2. Território situado entre os limites político-administrativos de duas áreas vizinhas e limítrofes.

Tambe mandou algue doze irmãos pera que estudassem grammatica e juntamente servissem de interpretes p<sup>a</sup> os Indios e assi se começou o estudo da grammatica de proposito e a conversão do Brasil porque na quella aldea se ajuntarão muitos indios daquela **comarca** e tinham douctrina ordinaria pola manhaa e á tarde e missa aos dias sanctos e a pr.<sup>a</sup> se disse dia da conversão de S. Paulo do mesmo anno e se começarão a batisar e casar e viver como xpãos, o qual ate aquelle tempo não se tinha feito nem na Baya nem em algua outra parte da costa. **desconhecido [1584]. enformação do brasil, e de suas capitancias, p. 426 .**

**Primeira datação:** E assi deseião os daquella **comarca** Padres em sua terra como se

Pág 1   Seção 1   1/2   Em 2,6 cm   Lin 1   Col 1   GRA   ALT   EST   SE   Português (

# Referências ao corpus

## Lista de textos

1 a 50 de 2458 [Avançar >](#) [Fim >>](#)

| Texto    | Ano       | Detalhes  |
|----------|-----------|---|
| A00_0335 | 1500      | pero vaz de caminha (1964). <i>carta de pero vaz de caminha.</i>  |
| A00_0078 | 1530-1532 | pêro lopes de sousa (1968). <i>diário da navegação de pêro lopes de sousa.</i>  |
| A00_1588 | 1548      | luís de góis (1965). <i>carta de luís de góis a d. joão iii, dando conta dos perigos que corria o brasil com a presença dos corsários franceses , que já chegavam à baía do rio de janeiro; vila de santos, 12 de maio de 1548 (4).</i>                     |
| A00_0001 | 1549      | p. manuel da nóbrega (1956). <i>carta do p. manuel da nóbrega ao p. simão rodrigues, baía [10 ? de abril] 1549.</i>   |
| A00_0002 | 1549      | p. manuel da nóbrega (1956). <i>carta do p. manuel da nóbrega ao p. simão rodrigues, baía 9 de agosto 1549.</i>   |
| A00_0003 | 1549      | dr. pero borges (1956). <i>carta do dr. pero borges a d. joão iii rei de portugal, porto seguro 7 de fevereiro 1549.</i>  |
| A00_0004 | 1549      | p. manuel da nóbrega (1956). <i>carta do p. manuel da nóbrega a d. joão iii rei de portugal, olinda [pernambuco] 14 de setembro 1549.</i>   |
| A00_0694 | 1549      | padre manuel da nobrega . (1885). <i>carta que o padre manoel da nobrega, preposito provincial da companhia de jesus, em o brasil, escreveu ao padre mestre simão o anno de 1549. (ms. copiado da livraria publica).</i>                                    |
| A00_0695 | 1549      | padre manuel da nobrega . (1885). <i>carta do padre nobrega para o padre mestre simão, do anno de 1549.</i>   |
| A00_0696 | 1549      | padre manuel da nobrega . (1885). <i>carta que o padre manoel da nobrega, companhia de jesus em as terras do brasil, escreveu ao padre mestre simão, preposito provincial da dita companhia em portugal no anno de 1549. (copia do ms. da bibliotheca).</i> |

# Concordâncias

## Busca por concordâncias

Buscar por:

[<< Início](#) [< Voltar](#) [201 a 300 de 378](#) [Avançar >](#)

das quais houve intensidade Tem êste rio bastante **cabedal** de águas cristalinas e saborosas Um sucesso dêste  
êste lugar da cachoeira do Ribeirão um rio assaz **cabedal** de boa água com uma laje de mais de passos  
nteza grande o que concorre para representar mais **cabedal** do que o próprio Guaporé nesta foz Também nos falt  
rdade a huma multidaõ de Indios ã fazia o grande **Cabedal** dos Paulistas e com ã de necessidade em hum dia  
ras visto ã os Índios as não lavraraõ nem tinhaõ **Cabedal** p ^a isso E ã p ^a o dito effeito  
efeitorio o qual serve denfermaria estando o dito **cabedal** enterrado em panellas de cobre e por cima pedra mi  
de oiro Enquanto os Taberneiros ajuntavam Imenso **cabedal** em poucos anos CARTAS CHILENAS Sem terem nas Taber  
to Tiraste de gastar em frias festas CARTA Imenso **cabedal** que o bom Senado Devia consumir em cousas santas S  
De muito mais abono e a quem devia Um grosso **cabedal** o Régio Erário Mal acaba Marquésio o seu triênio O  
fe Mostrar um grande zelo nas cobranças Do imenso **cabedal** que todo o povo Aos Cofres do Monarca está devendo  
Exclama o bom Ribério que não pode Pois todo o **cabedal** que tem cobrado Ou está nas demandas consumido Ou  
ella importante e larga fronteira como pelo maior **cabedal** de agoas destes grandes rios que facilitam o chega  
a do Pará se faz para VillaBella tanto pelo maior **cabedal** de agoas dos Amazonas Madeira Mamoré e Guaporé em  
que os primeiros descobridores vendo o seu maior **cabedal** de aguas tomaram pelo principal rio conflue com a  
oré pela mesma margem de norte rio de não pequeno **cabedal** daguas Do Cautarios são legoas de navegação a rumo  
sua margem de sul O rio Baures de extensão e **cabedal** daguas igual ao Guaporé de que é o maior confluent  
sendo o Mamoré rio de grande largura e de maior **cabedal** daguas elle traz as suas origens da latitude de gr  
p ^a ajuda das expensas desta expediçaõ fiz do meo **Cabedal** dezejezo de que exforçandose os demais Vassallos de  
metendose nos mesmos donde nascerão e com todo o **cabedal** de agoas que abraçarão se dirigem as ditas barras  
ados braços que tomão diversos nomes e com todo o **cabedal** de agoas desembocão em suas Barras deixando de cam  
soldados Ama a gente assisada a honra a vida o **cabedal** tão pouco que ponha uma acção destas nas mãos dum p  
os meus amores Eu ó cega não tenho um grosso **cabedal** dos pais herdado não o recebi no emprego nem tenho  
o seacha Excrito choice sic decabedaL sic corr de **cabedal** corr choice pera poder acudir ao sus tento da infa  
CÂMARA esta Cidade athe o tempo prezente enam tem **cabedal** para poder alugar Cazas e pagar pelas razoens asima  
the aqui tem Sucedido havendo pessoas demais poço **Cabedal** que há doze annoz selhetem tomado suas Cazas choic  
onforme os mesmoz Capitaens ordenarem etiverem de **Cabedal** choice sic deque sic corr de que corr choice daram

# Variantes de grafia

Busca por variantes de grafia

Digite uma palavra

variantes encontradas: villa, vila, vyla, vjlla, vylla, vjla.

# Símbolos especiais

## Inserir símbolos especiais

Clique nos símbolos da tabela abaixo que deseja inserir.

No caso de acentos, digite antes a letra que será acentuada.

**Dica:** os atalhos abaixo podem ser utilizados ao digitar nomes de verbetes.

Texto:

| Símbolo | Atalho    | Exemplo  |
|---------|-----------|----------|
| Æ       | {AE}      | gratiæ   |
| æ       | {ae}      | gratiæ   |
| f       | {f}       | feito    |
| œ       | {oe}      | cœteris  |
| ſ       | {s}       | difcurfo |
| ƒ       | {F-inv}   | ƒixit    |
| à       | {a-inv}   | passade  |
| ^       | {circ}    | quarÿ    |
| ~       | {til}     | coñãnde  |
| ̄       | {macron}  | cacaõ    |
| ¨       | {trema}   | muÿ      |
| ˆ       | {gancho}  | sõmente  |
| ˙       | {virgula} | tinhaó   |
| °       | {anel}    | Åfonso   |
| ´       | {agudo}   | quaeés   |
| ˘       | {breve}   | apanhë   |

# Conclusões

- O sistema apresentado pode ser adaptado a outros projetos semelhantes
- Recursos extras também estão disponíveis
  - Concordanceador
  - Glossários (junções e variantes)
  - Ferramentas de pré-processamento
  - [nilc.icmc.usp.br/nilc/projects/procorph/](http://nilc.icmc.usp.br/nilc/projects/procorph/)