

# Indicalex: áreas do saber a serviço do estudo da língua

Fabiane Cristina Altino<sup>1</sup>, Celciane Alves Vasconcelos<sup>1</sup>,  
Cinthyan Renata Sachs C. de Barbosa<sup>2</sup>, Edio Roberto Manfio<sup>3</sup>,  
Fábio Carlos Moreno<sup>4</sup>, Maria Renata de Mira Gobbo<sup>2</sup>, Marcelo Ozawa de Sousa<sup>2</sup>

<sup>1</sup>Departamento de Letras Vernáculas e Clássicas – Universidade Estadual de Londrina (UEL) – 86057-970 – Londrina – PR – Brazil

<sup>2</sup>Departamento de Computação – Universidade Estadual de Londrina (UEL) 86057-970 – Londrina – PR – Brazil

<sup>3</sup>Faculdade de Tecnologia de Garça Deputado Júlio Julinho Marcondes de Moura (FATEC) – 17400-000 – Garça – SP – Brazil

<sup>4</sup>Centro de Ciências Tecnológicas – Universidade Estadual do Norte do Paraná (UENP) 86300-000 – Cornélio Procópio – PR – Brazil

fabiane@uel.br, celcianevas@hotmail.com, cinthyan@uel.br,  
edio.manfio@fatec.sp.gov.br, fbio\_moreno@yahoo.com.br,  
mr.gobbol@gmail.com, hakozaawa@gmail.com

**Abstract.** *This work describes the Interdisciplinary Project for Dialectology, Computing, Cartography and Lexicography – INDICCALEX. It brings together researchers and professionals from various theoretical and methodological perspectives to meet the wishes of undergraduate and graduate students of the literature course, as well as of researches in the areas involved in this knowledge field from various Higher Education Institutions and it is intended to maintain cooperative agreements with these institutions for the development of projects related to these subareas. This paper will demonstrate a bit about the history of the INDICCALEX, its steps and some of the results achieved.*

**Resumo.** *Este trabalho descreve o Projeto Interdisciplinar de Dialetoologia, Computação, Cartografia e Lexicografia - INDICCALEX. Esse reúne pesquisadores e profissionais de diversas perspectivas teóricas e metodológicas para atender aos anseios dos alunos de graduação e pós-graduação do curso de letras, bem como de pesquisas nas áreas envolvidas neste campo do conhecimento de diversas Instituições de Ensino Superior e tem por objetivo manter acordos de cooperação com essas instituições para o desenvolvimento de projetos relacionados a essas subáreas. Este artigo irá demonstrar um pouco sobre a história do INDICCALEX, seus passos e alguns dos resultados alcançados.*

## 1. Introdução

Há algumas décadas, o Departamento de Letras Vernáculas e Clássicas da Universidade Estadual de Londrina (UEL) vem se dedicando ao desenvolvimento de pesquisas

voltadas para o estudo diacrônico da Língua Portuguesa, tais como: Atlas Linguístico do Paraná – ALPR [Toniolo 2008] [Almeida-Baronas 2013], Atlas Linguístico do Brasil – ALiB [Cardoso 2014a] [Cardoso 2014b] [Moreno, Manfio e Barbosa 2017] [Moreno et al. 2015], Tesouro Patrimonial do Léxico Galego e Português [Toniolo 2008] [Almeida Baronas 2013], TEDIPOR, Léxico Histórico do Paraná - LHisPAR [Almeida Baronas 2006] [Almeida Baronas e Almeida 2009] [Company y Melis 2002] e Léxico Histórico do Português Brasileiro - LHisPB, os quais deram origem a um projeto Interdisciplinar de Dialectologia, Computação, Cartografia e Lexicografia – INDICCALEX, que liga-se ao projeto nacional e interinstitucional Para a História do Português Brasileiro – PHPB

Atualmente o Projeto INDICCALEX é desenvolvido na UEL e seu principal objetivo é apresentar, sob tratamento lexicográfico, o conteúdo lexical dos documentos manuscritos datados dos séculos XVII a XIX. Com o avanço da tecnologia, muitos pesquisadores da área da Lexicografia, principalmente aqueles que se dedicam à cartografia de dados linguísticos e toponímicos e à elaboração de glossários e vocabulários precisaram recorrer ao auxílio de novas ferramentas computacionais para a busca, recuperação e indexação de dados linguísticos.

Dessa forma, a proposta do estudo interdisciplinar se justifica pela necessidade da composição de um léxico histórico que possa ser viabilizado por meio de mapas e *web*, oferecendo um material para estudos sincrônicos e diacrônicos, sob tratamento lexicográfico, apresentando os itens lexicais e suas variantes gráficas vigentes em épocas pretéritas.

O referido trabalho, por ser interdisciplinar, demanda estudos de várias áreas do saber: da Linguística (Dialectologia, Geolinguística, Sociolinguística, Lexicologia e Lexicografia), da Geografia (Cartografia) e da Computação (Processamento de Linguagem Natural, Estrutura de Dados, Processamento de Imagens, Banco de Dados e Engenharia de Software). Para atender aos anseios de cada área de conhecimento, o INDICCALEX

- promove a integração de pesquisadores, docentes e discentes das subáreas supracitadas;
- oferece um espaço para pesquisadores e docentes de várias IES, discute a possibilidade de desenvolvimento de projetos interdisciplinares, interdepartamentais e interinstitucionais com vistas ao avanço do conhecimento científico e tecnológico;
- oferta, sob a forma de banco de dados informatizado, um vasto material lexicográfico para estudos sincrônicos e diacrônicos do léxico registrado em documentos manuscritos durante os três séculos mencionados;
- proporciona um instrumento de consulta rápida do uso de vocábulos vigentes na época e em várias regiões do Brasil Colônia e Brasil Império;
- apresenta, com a frequência de uso, os itens lexicais e suas variantes gráficas registrados em cerca de 2500 documentos emanados das então províncias e vilas sediadas nos atuais estados da Bahia, Rio de Janeiro, São Paulo, Paraná, Minas Gerais, Paraíba, Ceará, Pernambuco e Rio Grande do Sul.

Assim, este artigo mostra na seção 2 a seleção de textos e composição dos *corpora* para a realização deste trabalho interdisciplinar. A seção 3 apresenta um pouco

do tratamento lexicográfico e fala sobre a ferramenta computacional. A seção 4 aborda a inserção do site institucional e o analisador léxico desenvolvido. A seção 5 aponta as conclusões e trabalhos futuros.

## **2. Seleção de Textos e Composição dos Corpora**

Os 1.394 documentos manuscritos que constituem a base para este trabalho fazem parte do acervo documental maior pertencente às equipes do projeto Léxico Histórico do Português do Brasil, doravante LHiPB, a saber: Léxico Histórico da Bahia – LHisBA (28), Léxico Histórico de Pernambuco – LHisPE (51), Léxico Histórico de São Paulo – LHisSP (196), Léxico Histórico da Paraíba – LHisPB (203), Léxico Histórico do Paraná – LHisPAR (730) e Léxico Histórico do Rio de Janeiro – LHisRJ (141).

Para a composição dos *corpora* foi necessário fazer um recorte inevitável, por não ser possível abarcar tudo o que foi produzido entre os séculos XVII e XIX. Em vista disso, foram selecionados, para a constituição do LHisPB (Léxico Histórico do Português do Brasil), textos de natureza oficial e texto de caráter coloquial (familiar, pessoal). Esta escolha se justifica pelo fato de, observando o *corpus* constituído de um dos projetos regionais que faz parte do LHisPB, *Para a história do português paranaense: estudos diacrônicos em manuscritos dos séculos XVII a XIX*, desenvolvido na UEL, a leitura e estudo desses manuscritos têm indicado que, apesar de oficiais (documentos notariais), a ortografia e a gramática ali aplicadas estão longe de ser consideradas a padrão ou de estar rigorosamente dentro da gramática normativa.

O fato acima tem sido atestado por [Toniolo 2008] [Almeida Baronas 2006] [Almeida Baronas e Almeida 2009] em seus estudos, confirmando que, pelo contrário, a forma expressa em muitos dos manuscritos está muito próxima de uma modalidade de língua provavelmente falada por boa parte dos brasileiros no tempo da Colônia e do Império.

Diante da diversidade de gêneros coletados pelas seis equipes (São Paulo, Bahia, Rio de Janeiro, Minas Gerais, Paraná e Paraíba), do grande volume de material já disponível e, tentando compor um *corpus* geral menos heterogêneo, optou-se por selecionar dois tipos de documentos manuscritos: os oficiais (correspondências eclesiásticas, notariais, cartas ao governo, lista de nomes, relatórios, testamentos, inventários, por exemplo) e os de caráter coloquial (familiar, pessoal, petições, entre outros).

## **3. Tratamento lexicográfico: ferramenta computacional**

Finalizada a transcrição e revisão dos 1.394 documentos manuscritos (730 manuscritos paranaenses; 141 fluminenses; 28 baianos, 45 cearenses, 196 paulistas, 51 pernambucanos; 203 paraibanos), a próxima etapa consiste no o tratamento dado aos documentos para a constituição do corpus a ser submetido à ferramenta computacional Lexico3, desenvolvido pela equipe SYLED-CLA2T [Lamalle et al. 2003] [Lexico3, 2019], da Université de La Sorbonne Nouvelle - Paris3. Trata-se de um programa de aplicação lexicométrica que permite:

- a) balizar livremente o texto a ser analisado, determinando e distinguindo as partes do texto previamente digitalizado;
- b) determinar o tamanho do contexto de um segmento a ser pesquisado;

- c) fazer o levantamento das ocorrências do segmento, indicando a distribuição das palavras dentro do texto;
- d) exibir as concordâncias; e
- e) apresentar, por meio de gráficos, as frequências relativa e absoluta de uma palavra.

### **3.1 Transcrição 1: Tratamento para utilização na ferramenta LEXICO3**

O tratamento para utilização à ferramenta LEXICO3 resume-se em dois procedimentos:

- a) Junção e desmembramento de palavras de acordo com a ortografia atual:
  - d omesd e Maiod od itto annonestaVi lla>do mes de Maio do ditto anno nestta Villa;
  - pasoaprezente>paso a presente;
  - ese lhes Con staSeja dema conduta >e se lhes Consta Seja de ma conduta
- b) Eliminações: da identificação, assinaturas, quebras de linhas, espaços, algarismos, sinais de pontuação.

### **3.2 Transcrição 2: tratamento para aplicação à ferramenta LEXICO3**

O tratamento para aplicação à ferramenta LEXICO3 envolve:

#### **Junção de todos os fólhos por estado em um só arquivo;**

- a) Distinção das homógrafas: a<sup>1</sup> = artigo, a<sup>2</sup> = pronome demonstrativo, a<sup>3</sup> = pronome oblíquo, a<sup>4</sup> = preposição; a<sup>5</sup> = contração e a<sup>6</sup> = verbo haver sem o h;
- b) Passagem para caixa alta, no modelo sem formatação;
- c) Submissão ao Programa Lexico3.

Segue abaixo a Figura 1, uma pequena amostragem da ferramenta Léxico 3 que serviu de base para o tratamento do *corpus* escrito:

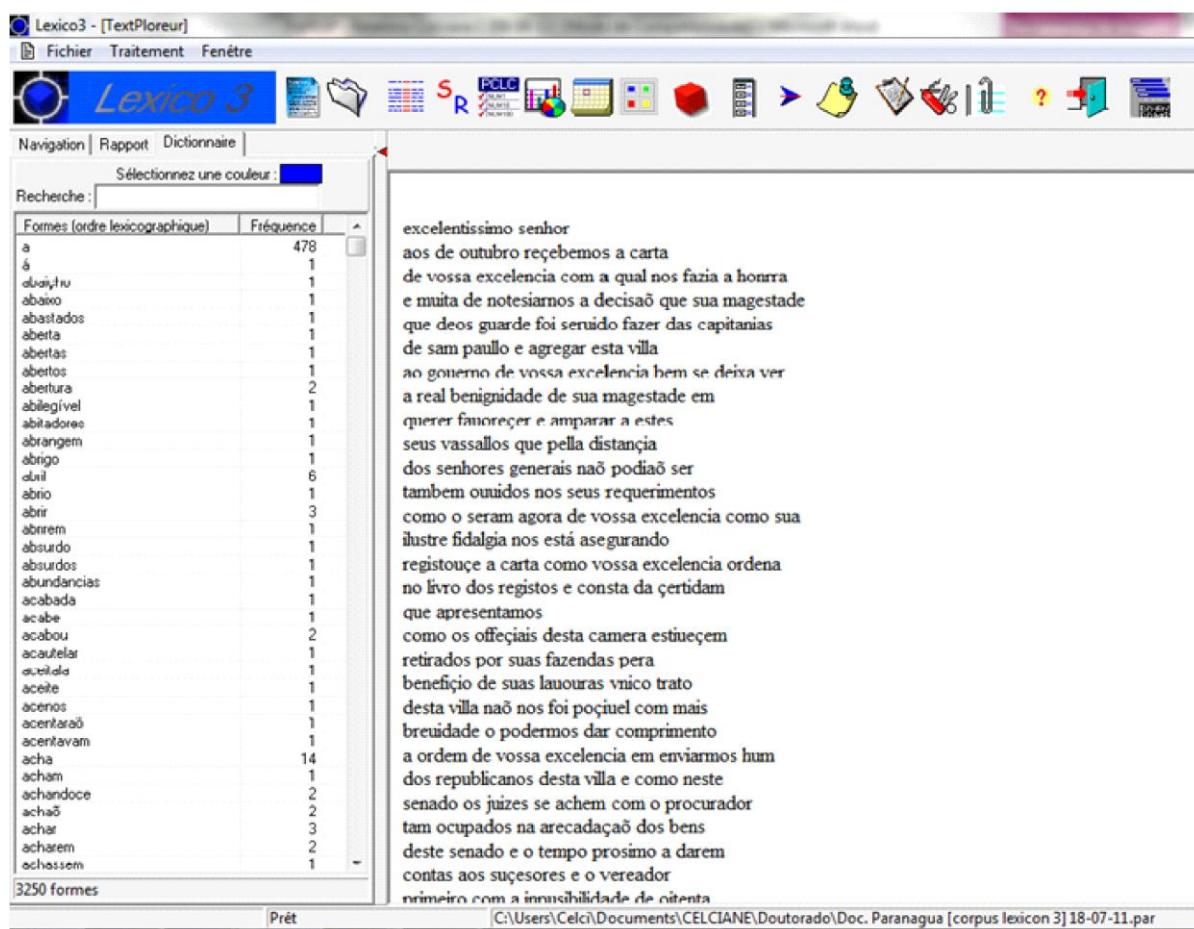


Figura 1. Exemplo de tratamento do *corpus* escrito: listagem dos vocábulos em ordem alfabética, número de frequência e texto transcrito conforme normas exigidas do programa Lexico3.

### 3.3 Construção do verbete

Os verbetes foram construídos seguindo os moldes da obra de Company y Melis (2002). Ao final do processo, pode-se oferecer um léxico lematizado que corresponde a um dicionário de construção e uso em que as definições semânticas dos lemas não serão incluídas, mas depreendidas pelos contextos das abonações. Assim, o verbete é constituído por:

- a) entrada com a norma ortográfica atual;
- b) significado(s) no documento;
- c) abonação com o local e
- d) número do documento, variantes e frequência global.

No LHisPB estão contidos registros de alguns dos usos da língua portuguesa com exemplificação dos diversos contextos gramaticais e semânticos extraídos dos manuscritos, datados dos séculos XVII, XVIII e XIX, editados pelas equipes estaduais do PHPB (São Paulo, Rio de Janeiro, Bahia, Minas Gerais, Paraná,

Paraíba, além das equipes mais recentes Alagoas, Ceará, Pernambuco e Rio Grande do Sul).

### 3.4 Exemplo de verbetes

Quanto ao modo de apresentar suas características, os verbetes são assim exibidos:

abrimento: Recebemos a Carta que Vossa Excelencia nos derigio datada de 6 de Junho do prezente anno na qual nos facultava o **abrimento** do Caminho desta Villa para a Freguezia de Saõ Jozé de Curitiba (GRT, 1805, 198, 03).

FORMAS DOCUMENTADAS: abrimento (1).

FREQUÊNCIA GLOBAL: 1.

adivinhar. quando he sertto que eles camaristas na ã podiam **adivinhar** o foturo que tal ruina havia de acomteser (CAS, 1805, 736,12).

FORMAS DOCUMENTADAS: adivinhar (1).

FREQUÊNCIA GLOBAL: 1.

mãe. He muito certo que ficaõ desamparadas as **Maens** destes Orfaons mensionados os quaes por menores naõ se utilizaõ a si nem á suas **Maens** (CTB, 1799, 81,20); pela esta forma ficaõ dispidas as **Mays** para vestir os estranhos ficando aquellas em huã total necessidade e desamparo (ANT, 1798, 415, 13).

FORMAS DOCUMENTADAS: maens (3), mays (1).

FREQUÊNCIA GLOBAL: 4.

## 4. Analisador Morfológico

Uma das propostas do INDICCALEX é a inserção dos resultados na página oficial da UEL. Para tanto, este projeto envolve alunos e pesquisadores dos cursos de Letras Vernáculas e Clássicas, Ciência da Computação, Design, além de colaboradores externos. A estrutura do site para o LHisPB está sendo construída conforme a Figura 2.

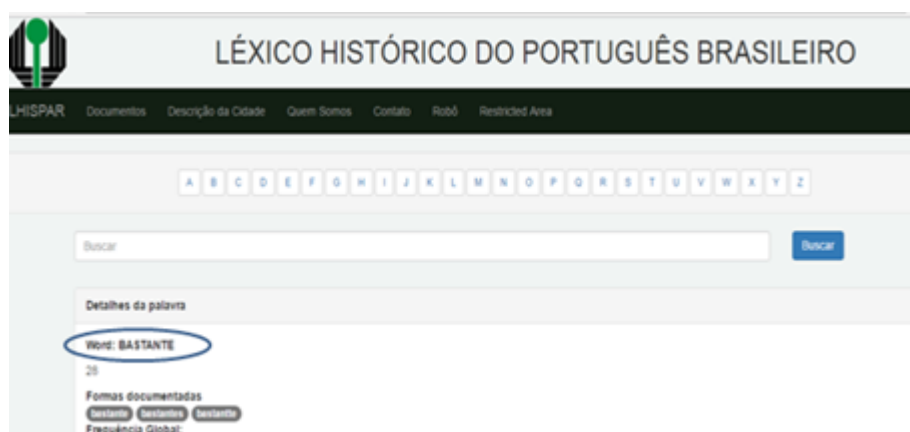


Figura 2. Protótipo do site para o LHisPB

Na página principal há as seguintes informações: (i) logomarca que dá acesso à página principal, (ii) o texto de apresentação, (iii) todos os botões relacionados na Tabela 01, (iv) o campo para digitação de palavras, (v) o campo para exibição dos

resultados, e (vi) o campo em que figuram as imagens relacionadas e os links para os sites relacionados (UEL, Projeto ALiB, ALPR, parceiros, órgãos de fomento, etc.).

Atualmente a equipe é composta por pesquisadores das seguintes universidades: UEL, FATEC, UNIVEM e UENP. Analisadores Morfológicos foram implementados por discentes na disciplina de *Processamento da Linguagem Natural* do curso de Ciência da Computação da UEL com o intuito desses alunos fazerem análise morfológica de manuscritos, como nas Figuras 3 e 4.

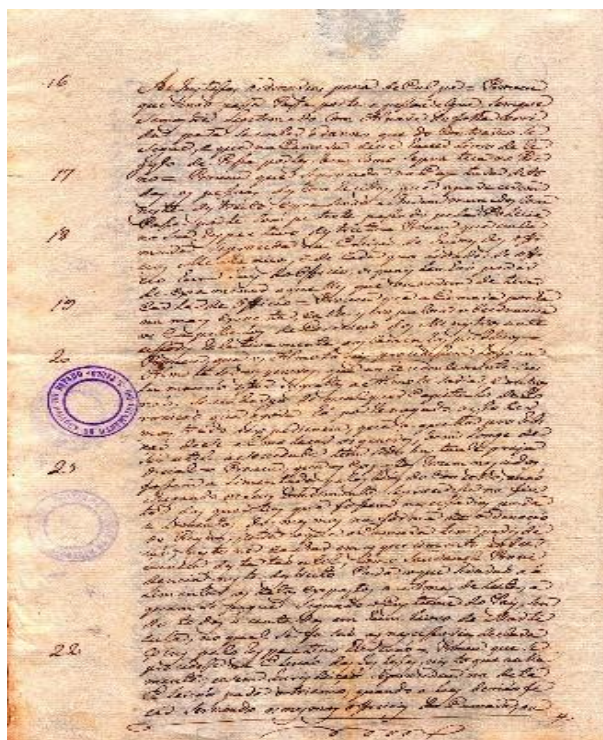


Figura 3. Manuscrito texto 500

### Texto do documento 500

*As justissas ordinarias pena de culpa Proveu que senao passe passaporte a pessoa alguma sem que se mostre Legitimado com Alvara de folha corrida para se cortar o danno que do contrario se segue e que na Camara deve haver Livro de registo de Passaportes bem como se pratica no Reino Proveu que se proceda na Captura de todas as pessoas desconhecidas que apparecerem neste dstricto e que senaõ acharem munidos com Passaporte competente passado pela Policia no seu respectivo dstricto Proveu que em Camara se proceda na Eleiçaõ de Juizes dos officios Mechanicos e de cada qualidade de officio hum Juis do officio os quais tenhaõ poder de examinar aquelles que ouverem de tirar carta de officio Proveu que a Camara ponha na mais exacta e athe escrupulosa obcervancia os capitulos de Correiaçõ dos Menistros antecessores relativamente aos caminhos publicos Proveu que os Almotaceis procedessem logo em Almotasar os generos que antecedentemente costumavaõ estar sujeitos a Almotasaria e outros naõ se embargo de qualquer capitulo de Correiaçõ que agora há por revogado visto ter mostrado a experiencia que aquella prohibiçaõ de se almotazar os generos bem longe de ser util a sociedade tem sido bastante prejudicial Proveu que os expostos recém nascidos fossem alimentados pelos bens do concelho e naõ chegando os seus rendimentos se procedesse na finta das quantias que fossem necessarias para alimentos dos mesmos na forma da ordenaçãõ do Reyno para o que a Camara tem jurisdicaõ visto ver e achar em esquecimento a obcervancia desta taõ util como saudavel Providencia neste dstricto Para o que se daraõ a alimentar os ditos expostos a Amas de Leite a quem se pague segundo o costume do Pais sendo todos acentados em hum livro de Marticula no qual se fasaõ as necessarias declaraçoins pelo respectivo Escrivaõ Proveu que se procedesse na Eleiçaõ das justissas visto que nullamente e sem jurisdicaõ se procedeu na dita Eleiçaõ para o trienio quando alias deviaõ ficar servindo os mesmos officiaes da Camara ou*



Figura 4. Manuscrito texto 501

### Texto do documento 501

*Ou proceder somente na Eleicão para hum anno e inteiramente portanto elle corregedor revalida a eLeicão taõ somente para este anno de mil oito centos e quatro e anulla para os dous annos annos seguintes com declaracão que prevista a eleicão do Juiz dos orfaõs pelo bem que tem servido e que elle ditto Menistro tem obcvorado na revista do Cartório Proveu que se proceda na factura ou reforma do Curral do Conselho aonde sejaõ recolhidos os animais de qualquer qualidade que forem achados a fazer Danno ou ainda mesmo aquelles que senaõ achar possuidor certo de que elle ditto Menistro espera ver aprompta execucao para utilidade publica Proveu que se proceda pelos officiais da Camara a fazerem aRematar o Provimento das carnes do Asougue desta villa pondo-õ logo em praca e arematandoõ a quem por menos der ao Povo cada libra por ver elle Menistro que a ampla facultade que neste Pais tem qualquer demandar vender carne à enxerga hé prohibido pela Ley Novissima asim como taõ bem se acha prohibido vendella sem Licenca delle ditto Menistro mas por ninhum modo fora do Lugar destinado pela Camara para abastecimento do povo e publica utilidade Proveu que em qualquer parte aonde houvessem citios ou Lugares pantanozos e charcos dagua em que as agua se achem encharcadas e sem darem Lugar à penetrar os ardores do sol cauzando huns vapores pestiferos que inficionaõ o Ar cauzando grandno denno e prejuizo a saude dos povos e ate Epidemias e os seus tristes e funestos efeitos querendo elle ditto Menistro como Magistrado da Policia promover a publica seguranca e Conservacão dos habitante por deste districto que se proceda na abertura destas paragens para dar expedicaõ as ditas Aguas convocando a todos os habitantes por hua Escala de baixo da Inspecão da Camara na aubzencia delle ditto Menistro ao qual o fariãõ saber achando se neste districto e villa para melhor expedicaõ e prompta execucao deste Provimento que como tem em vista a*

Importante lembrar que embora não tenha sido projetado para operar exclusivamente com o Lexico3, o analisador morfológico, desenvolvido na linguagem de programação Python (versão 2.7.12), tem a função de selecionar os verbetes já ordenados por aquele software. A comunicação em tempo real do analisador com o website ainda não foi considerada neste trabalho, pois consiste em outra etapa da pesquisa a ser apresentada em momento oportuno. A seguir serão mostradas execuções do parser para várias categorias morfológicas desses textos que serão exibidas nas Figuras de 5 a 10.



```
-----  
Digite a palavra desejada: justissas  
  
justissas  
Correcao: justiças  
Classe: substantivo  
Genero: feminino  
Numero: plural  
  
Contagem: 5  
Texto 500: T  
Texto 501: F  
Texto 502: T  
-----
```

Figura 5. Substantivo justiças

```
-----  
Digite a palavra desejada: pantanozos  
  
pantanozos  
Correcao: pantanosos  
Classe: adjetivo  
Genero: masculino  
Numero: plural  
pantanosos  
Contagem: 1  
Texto 500: F  
Texto 501: T  
Texto 502: F  
-----
```

Figura 6. Adjetivo pantanosos

```
-----  
Digite a palavra desejada: na  
  
na  
Correcao:  
Classe: contração  
Genero: feminino  
Numero: singular  
em + a  
Contagem: 19  
Texto 500: T  
Texto 501: T  
Texto 502: T  
-----
```

Figura 7. Contração na

```
-----  
Digite a palavra desejada: proveu  
  
proveu  
Correcao:  
Classe: verbo  
Genero:  
Numero: singular  
prover  
Contagem: 15  
Texto 500: T  
Texto 501: T  
Texto 502: T  
-----
```

Figura 8. Verbo prover

```

-----
Digite a palavra desejada: menos
menos
Correcao:
Classe: advérbio
Genero:
Numero:

Contagem: 1
Texto 500: F
Texto 501: T
Texto 502: F
-----

```

Fig 9. Advérbio menos

```

-----
Digite a palavra desejada: elle
elle
Correcao: ele
Classe: pronome
Genero: masculino
Numero: singular

Contagem: 6
Texto 500: F
Texto 501: T
Texto 502: T
-----

```

Figura 10. Pronome ele

Note-se que o analisador pode não apenas discriminar as diferentes classes de palavras em português como também extrair delas características gramaticais próprias de sua natureza. Outra vantagem é que não depende da ortografia padrão como ocorre em analisadores convencionais, ou seja, verbetes como ‘elle’ (grafado com dois ‘eles’) ou pantanozo (grafado com ‘z’) não geram inconsistência.

## 5. Considerações Finais

Das etapas realizadas até o momento inclui-se a revisão e confecção de verbetes do Paraná, Rio de Janeiro, Bahia, Paraíba, São Paulo e Pernambuco e os testes com o analisador morfológico. Dentre as metas a serem alcançadas podemos listar: **(i)** inclusão dos verbetes do Paraná no site LHisPAR, **(ii)** inclusão dos verbetes dos demais corpora, **(iii)** incorporação dos verbetes em um que unifique as mesmas entradas. Ex.: officio para os diversos léxicos organizados em uma única entrada (ou agrupados), **(iv)** melhoria do funcionamento da ferramenta, **(v)** inclusão de dados nos campos, **(vi)** inclusão dos fac-símiles no site, **(vii)** construção do tutorial para a navegação no site, e **(viii)** transformação do site do LHisPAR em LHisPB.

Acreditamos que o trabalho interdisciplinar traz uma grande contribuição aos estudiosos da área da linguagem que se dedicam aos estudos diacrônicos. A conclusão do Atlas Linguístico do Brasil [Cardoso 2014a] [Cardoso 2014b], é um exemplo disso, pois partes dele já digitalizadas [Moreno, Manfio e Barbosa 2017] [Moreno et al. 2015] constituem material virtual passível de ser consultado por meios puramente digitais e isso tem se estendido a dicionários, glossários e léxicos históricos do português, bem como à elaboração de obras dessa natureza lexicográfica.

## Referências

- Almeida-Baronas, J. E. (2013). Para a História do Português Paranaense. *Revista Brasileira de Linguística*. Niterói, 2: pages 207-217.
- Almeida-Baronas, J. E. (2006). Manuscritos paranaenses: fontes para um estudo diacrônico. In: *V Selisigno e VI Simpósio de Leitura da Universidade Estadual de Londrina*. Londrina-PR.
- Almeida-Baronas, J. E. e Almeida, P. L. C. de (2009). A escrita dos séculos XVIII e XIX: em busca de dados diacrônicos. In: *XIX Seminário do CELLIP – Centro de Estudos Linguísticos e Literários do Paraná*. UNIOESTE – Universidade Estadual do Norte do Paraná, Cascavel-PR.
- Cardoso, S. A. M. S. et al. (2014a). *Atlas Linguístico do Brasil: Introdução*, 1. Londrina: Eduel.
- Cardoso, S. A. M. S. et al. (2014b). *Atlas Linguístico do Brasil: Cartas Linguísticas I*, 2. Londrina: Eduel.
- Concepción Company e Melis, C. (2002). *Léxico histórico del español de México*. México: Universidad Autónoma de México.
- Lamalle, C., Martinez, W., Fleury, S. and Salem, A. (2003). *Lexico 3*. <http://lexi-co.com/ressources/manuel-3.41.pdf>
- Lexico3 - SYLED - CLA2T Université de la Sorbonne nouvelle - Paris 3. <http://lexi-co.com/ressources/manuel-3.41.pdf>.
- Moreno, F. C., Manfio, E. R. e Barbosa, C. R. S. C. de (2017). Tical: um chatbot que versa sobre assuntos linguísticos. In: *14º Encontro de Linguística de Corpus*, 51.
- Moreno, F. C., Manfio, E. R., Barbosa, C. R. S. C. de. e Brancher, J. D. (2015). Tical: Chatbot sobre o Atlas Linguístico do Brasil no WhatsApp. In: *26º Simpósio Brasileiro de Informática na Educação*, Maceió. 1: pages 279-288.
- Manfio, E. R., Moreno, F. C. e Barbosa, C. R. S. C. de (2014). Professor Tical e ALiB: Interação Humano Computador em Diferente Campo. In: *19º Conferência Internacional sobre Informática na Educação*, Fortaleza, pages 782-787.
- Manfio, E. R., Barbosa, C. R. S. C. e Moreno, F. C. (2016). Um Robô Linguista que “Ouve” e “Fala”: Geolinguística, PLN e Tabelas Hash em Concurso. *Revista Signum*. 19, pages 415-439.
- Toniolo, E. (2008). Possíveis marcas da oralidade em manuscritos paranaenses. In: *Filologia e Linguística Portuguesa*. FLP 10/11. <http://www.fflch.usp.br/dlcv/lport/flp/images/arquivos/FLP10-11/Toniolo.pdf>