



XXXI Encontro Nacional da Associação Portuguesa de Linguística
28-30 Outubro 2015, Universidade do Minho

EFFE-ON: NOVA BASE DE DADOS ONLINE

Correspondência de escrita e fala de 1º Ciclo do EB

Celeste Rodrigues (FLUL/CLUL)

Maria do Carmo Lourenço-Gomes (CLUL / ILCH-UMinho)

Isabel Alves (FLUL/CLUL)

Apresentação do *corpus*

EFFE-On: *Escreves como falas – Falas como escreves?*



Equipa

Celeste Rodrigues (Universidade de Lisboa, FLUL-CLUL)

Maria do Carmo Lourenço-Gomes (Universidade do Minho, CEHUM)

Isabel Alves (Universidade de Lisboa, FLUL/CLUL)

Maarten Janssen (Universidade de Lisboa, CLUL)

Colaboradores

Patrícia Pardal da Costa (FLUL)

Silvana Martinez (FLUL)

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

Projeto UID/LIN/00214/2013



Universidade do Minho



Apresentação do *Corpus*

- O EFFE-On é um *corpus* de dados de escrita e fala de crianças nos primeiros anos de escolaridade. Foi elaborado com base no projecto “EFFE - Escreves como Falas - Falas como Escreves?”, que começou a ser desenvolvido em 2012 no Centro de Linguística da Universidade de Lisboa (CLUL).
- Os dados são recolhidos de acordo com procedimentos metodológicos sistemáticos e consistem em descrições de imagens com estímulos controlados e adequados às faixas etárias inquiridas.

Apresentação do *corpus*

- O *corpus* foi criado no sistema ©TEITOK (Janssen 2014) para ser disponibilizado online.
- O EFFE-On foi pensado para ser sucessivamente ampliado e tem como principal objetivo fornecer dados a professores de Português língua materna, investigadores e terapeutas, além de possibilitar a fundamentação de um leque amplo de estudos em diferentes áreas, incluindo a Linguística, a Educação e a Terapia da Fala.

Aplicações

Inicialmente, o *corpus* irá permitir, entre outras aplicações:

- Identificar formas ortográficas não-convencionais (FN-Cs), em crianças falantes do português de diferentes áreas dialetais
- Comparar as FN-Cs e produções orais espontâneas das mesmas crianças e entre crianças
- Comparar dados recolhidos num momento definido (coleta transversal) e ao longo do tempo (coleta longitudinal)

Aplicações

- Auxiliar o desenvolvimento de materiais pedagógicos e a adequação de estratégias de ensino para diferentes regiões dialetais
- Examinar aspectos fonológicos, sintáticos, semânticos e pragmáticos na escrita infantil
- Pesquisar frequências de palavras e de estruturas sintáticas

Constituição do *corpus*

- Composições descritivas produzidas a partir de imagens-cenário para elicitación de palavras (nas modalidades oral e escrita), balanceadas com base em critérios fonéticos, fonológicos, lexicais e de frequência.
- Composições narrativas na modalidade escrita, construídas a partir de imagens em série (histórias sem texto).

Constituição do *corpus*

Lisboa

2º Ano (*códigos 2A_CM ou 2B_CM*)

- 48 crianças; 2º ano; escola particular de Lisboa
- Idade: 7 anos
- Distribuição por sexo: 23 masculino e 25 feminino
- 96 textos
- 58 entrevistas gravadas

4º Ano (*códigos 4A_CM ou 4B_CM*)

- 54 crianças; 4º ano; mesma escola particular de Lisboa
- Idade: 9 anos
- Distribuição por sexo: 31 masculino e 23 feminino
- 108 textos

Lisboa: 204 textos

Constituição do *corpus*

Porto

2º Ano (*códigos P_2_OsC ou P_2_PA*)

- 44 crianças; 2º ano; escola pública do Porto
- Idade: 7 anos
- Distribuição por sexo: 23 masculino e 21 feminino
- 88 textos
- 41 entrevistas gravadas

Porto: 88 textos

Procedimentos de recolha dos dados

Textos narrativos a partir de banda desenhada sem texto

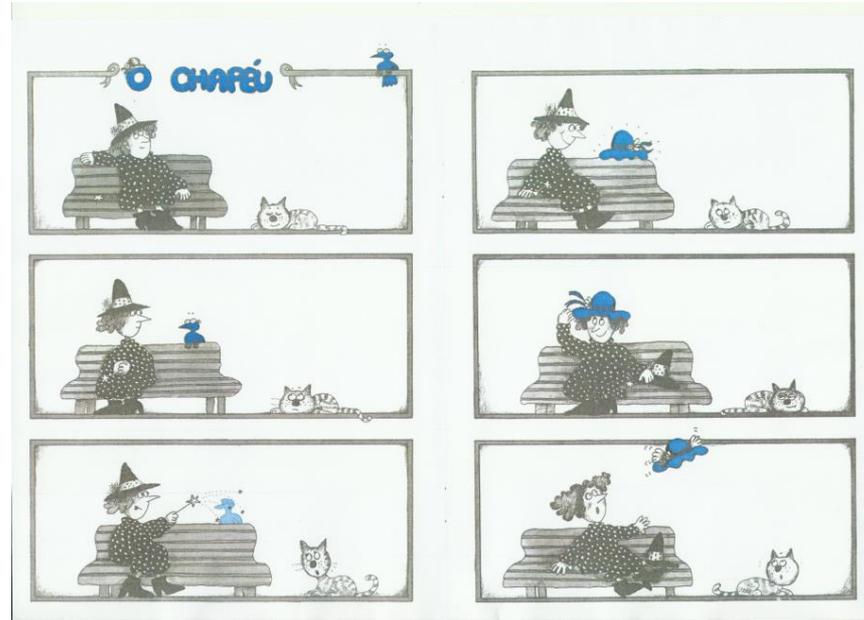
- História da Bruxinha (Furnari, 1991)
 - “O Chapéu” no 2º ano
 - “O Telefone” no 4º ano

Textos descritivos de imagens temáticas

- Imagens (Guerreiro, 2007)
 - Casa de Banho (CB)
 - Cidade (CI)
 - Cozinha (CZ)
 - Floresta (FL)
 - Sala (SA)

Exemplos de Materiais

Composição narrativa a partir de banda desenhada sem texto
História da Bruxinha – “O Chapéu” – 2º ano

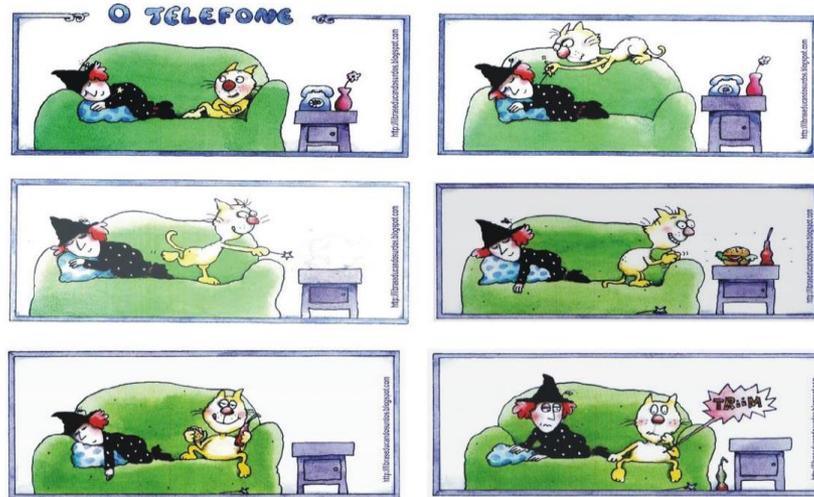


Era uma vez uma Bruza cama-da Ana e um dia ela estava a olhar para o céu e o gato estava á dormir - e de repente o gato acordou e estava la um passaro e a bruza estava a olhar para ele. E a bruza lansso um feitico sobre o passarinho. e o gato estava espantado. e tranformole nom chapéu. e a Ana pose o chapéu. e o chapeu ganhou asas e a Ana ficou toda despentiada.

12_BR_2B_CM

Exemplos de Materiais

Composição narrativa a partir de banda desenhada sem texto História da Bruxinha – “O Telefone” – 4º ano



A magia de um gato

Numa bela manhã a bruxa Mimi estava a dormir no seu **sófa** verde claro, com o seu gato Amarelinho, ao lado do **sófa** estava uma mesa cor de rosa com um telefone e um jarro com uma **flôr**.

Depois o gato fica a pensar que não tem ali a sua amiga bruxa por isso vai fazer um feitiço. Rouba-lhe a varinha **magica** sorrateiramente para não acordar a bruxa Mimi.

De repente transforma o telefone e o jarro num **hamburger** e num sumo de morango. (...)

Exemplos de Materiais

Composições descritivo-narrativas de imagens temáticas

CASA DE BANHO



Éra uma vez uma menina que se chamava Catarina ela estava a tomar banho e estava a gastar muita água a água estava a transbordar o seu irmão Francisco estava afelitinho para fasere chi-chi o ceu outro irmão Jozé ia agora lavar os dentes mas eu não sei porqueque ele não fasia chi-chi estava mesmo opé da sanita! todos estavãose a rir dele a Catarina envez de rirse do Francisco podia desliguar a água e sair da banhaira vitoria vitori acabouse esta istoria.

31_CB_2A_CM

Exemplos de Materiais

Composições descritivo-narrativas de imagens temáticas CIDADE



Ontem de manhã abrimos a janela e vi um menino que **chotou** uma bola pufe e partiu o vidro. Ei quem **partiu** o vidro **derrepente** vi um carro que quase **atroplava** um menino **depos** vi um comboio eu acho que leva muitas **pesouas** **derrepente** **pasou** um trator carregado de palha. Vi um menino a **nada** na **pichina** e estava muito divertido e depois **aparecerão** um menino e uma menina que estavam a brincar. A menina estava a brincar com uma boneca e o menino com um carro.

44_CI_2A_CM

Exemplos de Materiais

Composições descritivo-narrativas de imagens temáticas COZINHA



A festa

A menina **segou** a casa e **viu** a cozinha e eu gostava de ter **ésta** festa, porque **a** um bolo de anos e **a** dois ratos a dançar. E a menina tentou **xegar** ao prenda e ela **coprou** as **zelas**. E o **otro** ratinho a comer queijo e a menina **camache** Laura e ela foi **a** janela e disse Olá lua e estrelas e foi ao **ferígorífico** e foi para a cama. E um café de leite e leite.

12_CZ_2B_CM

Exemplos de Materiais

Composições descritivo-narrativas de imagens temáticas FLORESTA



De manhazinha fui passar na floresta, vi muitos animais vi um caracol, vi uma zebra, vi uma galinha com dois pintainhos e dois ovos, vi uma borboleta, vi um trige dois passaros uma serpente um paixe a mergulhar no lago e um dragão a cospir fogo, tambam aviam muitas árvores, de repente comessou a chuver depois apareceram luses mas depois o sol apareceu, a fluresta comessou a tramer era um elefante, a zebra fico espantada o elefante foi o por num monte lá muito ao fundo.

38_FL_2A_CM

Exemplos de Materiais

Composições descritivo-narrativas de imagens temáticas

SALA

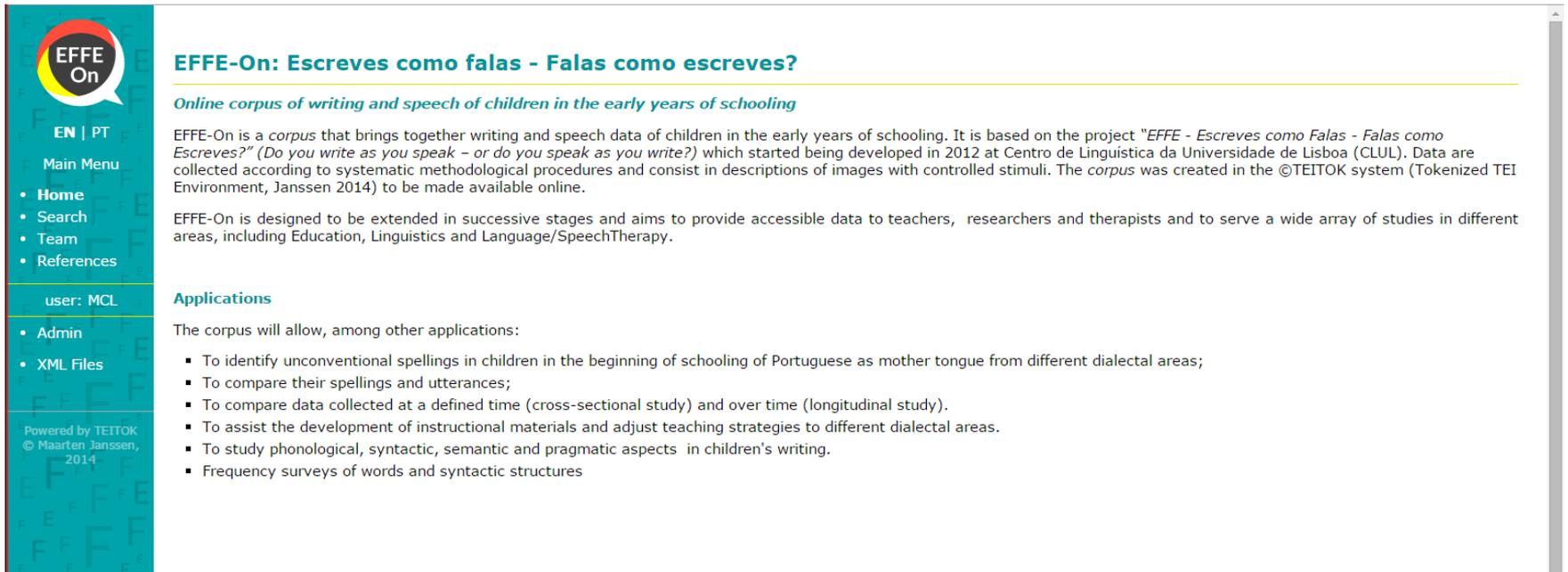


O meu pai quando eu chegei a casa o meu pai estava a cantar e o meu pai estava olado de uma mesa que tem um jornal três lapis uma tesora dinheiro e um desanho dois livros no chão um paiaso de brincar um martelo um jogo de tableir atras da janela esta uma brucha a voar na soa vassora uma igraisia um gato a brincar com uma bola um cofre de dinheiro uma televisão um radio doas caichas um vaso com plantas um sofa almofadas um coadro com a imaxai dom por do sol um gelobo um saco de dinheiro.

[10 SA 2B CM](#)

EFFE-On Homepage

<http://alfclul.clul.ul.pt/teitok/effe/en/index.php?action=home>



The screenshot shows the homepage of EFFE-On. On the left is a teal sidebar with the EFFE-On logo (a speech bubble with 'EFFE On' inside), language options 'EN | PT', a 'Main Menu' section with links for Home, Search, Team, and References, a 'user: MCL' section with links for Admin and XML Files, and a footer with 'Powered by TEITOK © Maarten Janssen, 2014'. The main content area has a teal header with the title 'EFFE-On: Escreves como falas - Falas como escreves?' and a subtitle 'Online corpus of writing and speech of children in the early years of schooling'. Below this is a paragraph describing the corpus, followed by an 'Applications' section with a list of uses.

EFFE-On: Escreves como falas - Falas como escreves?

Online corpus of writing and speech of children in the early years of schooling

EFFE-On is a *corpus* that brings together writing and speech data of children in the early years of schooling. It is based on the project "EFFE - Escreves como Falas - Falas como Escreves?" (*Do you write as you speak - or do you speak as you write?*) which started being developed in 2012 at Centro de Linguística da Universidade de Lisboa (CLUL). Data are collected according to systematic methodological procedures and consist in descriptions of images with controlled stimuli. The *corpus* was created in the @TEITOK system (Tokenized TEI Environment, Janssen 2014) to be made available online.

EFFE-On is designed to be extended in successive stages and aims to provide accessible data to teachers, researchers and therapists and to serve a wide array of studies in different areas, including Education, Linguistics and Language/SpeechTherapy.

Applications

The corpus will allow, among other applications:

- To identify unconventional spellings in children in the beginning of schooling of Portuguese as mother tongue from different dialectal areas;
- To compare their spellings and utterances;
- To compare data collected at a defined time (cross-sectional study) and over time (longitudinal study).
- To assist the development of instructional materials and adjust teaching strategies to different dialectal areas.
- To study phonological, syntactic, semantic and pragmatic aspects in children's writing.
- Frequency surveys of words and syntactic structures



EFFE-On

- A base de dados estará disponível em português e em inglês
- Estarão disponibilizados materiais e referências sobre o seu funcionamento
- O cabeçalho de cada ficheiro contém as seguintes informações:

The screenshot shows a web browser window displaying the EFFE-On interface. The page title is '6_CB_2B_CM.xml' and the main heading is 'CASA DE BANHO'. The template used is 'teiHeader-edit.tpl'. The form contains the following fields:

Title	
Title	CASA DE BANHO

Student	
ID	6
Gender	fem
Age	7
Birthdate	10/08/05
Nationality	portuguesa/espanhola

Text	
Type	Describe image
Task	Casa de banho
ID	CB

School	
Name	CM
Class	2
City	Lisboa

Parents	
Education father	Licenciatura
Education mother	Licenciatura

Audio	
Filename (URL)	Audio/6_CB_2B_CM.wav
File type (audio/wav)	audio/wav
Audio description	This sound file only contains a selection of words of the text

At the bottom of the form is a 'Save' button. The left sidebar contains navigation links for 'EN | PT', 'Main Menu', 'Home', 'Search', 'Team', 'References', 'user: MCL', 'Admin', and 'XML Files'. The footer indicates the application is powered by TEITOK, © Maarten Janssen, 2014.

EFFE-On

(Opções de visualização)

Original / Forma escrita



EN | PT

Menu principal

- Início
- Pesquisar
- Team
- References

user: MCL

- Admin
- XML Files

Powered by TEITOK
© Maarten Janssen, 2014

10_BR_4B_CM.xml

História da Bruxinha

Idade	9
Gênero	masculino
Turma	4B
Tarefa	História da Bruxinha - Telefone

- mais informação
- edit teilHeader

Opções de visualização

Texto: [Original](#) | [Forma escrita](#) | [Forma normalizada](#) - Mostrar: [Cores](#) | [Imagens](#)

Edit the information about each word of this file by clicking on the word in the text below, or click [here](#) to edit the raw XML

A magia que corre mal

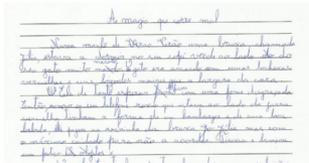
Numa manhã de **Verão** Verão uma bruxa chamada Júlia, estava a dormir no seu sofá verde ao lado de do seu gato muito **maroto** maroto. O gato era amarelo com umas **bochexas** vermelhas e uns bigodes maiores que a largura da cara.

Ele de tanto **ficom**f ficou com uma fome desgraçada. Então, como o seu telefone roxo que estava ao lado da jarra vermelha **tinham** a forma de um **hamburgar** e de uma boa bebida, **ele** pegou na varinha da bruxa **Jo** Júlia mas com o máximo cuidado para não a acordar. **Tiroua** e lançou um feitiço. **Os-objetos**

Os objetos tinham-se transformado numa refeição ótima para ele. Comeu, comeu e comeu, ficou muito cheio.

Mas a magia não correu bem. porque ele não era um bruxo era um gato. isso fez

10_BR_4B_CM



Pesquisar

- Team
- References

user: MCL

- Admin
- XML Files

Powered by TEITOK
© Maarten Janssen, 2014

Opções de visualização

Texto: [Original](#) | [Forma escrita](#) | [Forma normalizada](#) - Mostrar: [Cores](#) | [Imagens](#)

Edit the information about each word of this file by clicking on the word in the text below, or click [here](#) to edit the raw XML

A magia que corre mal

Numa manhã de Verão uma bruxa chamada Júlia, estava a dormir no seu sofá verde ao lado do seu gato muito maroto. O gato era amarelo com umas **bochexas** vermelhas e uns bigodes maiores que a largura da cara.

Ele de tanto ficou com uma fome desgraçada. Então, como o seu telefone roxo que estava ao lado da jarra vermelha **tinham** a forma de um **hamburgar** e de uma boa bebida, pegou na varinha da bruxa Júlia mas com o máximo cuidado para não a acordar. **Tiroua** e lançou um feitiço.

Os objetos tinham-se transformado numa refeição ótima para ele. Comeu, comeu e comeu, ficou muito cheio.

Mas a magia não correu bem, porque ele não era um bruxo era um gato, isso fez

EFFE-On

(Opções de visualização)

- Forma normalizada / Imagem - facsímile

The screenshot shows the EFFE-On web application interface. The browser address bar displays the URL: `alfclul.dul.ul.pt/teitok/effe/pt/index.php?action=edit&id=10_BR_4B_CM.xml`. The page title is **10_BR_4B_CM.xml** and the main heading is **História da Bruxinha**.

Metadata fields are displayed:

- Idade: 9
- Género: masculino
- Turma: 4B
- Tarefa: História da Bruxinha - Telefone

Navigation options include: [mais informação](#) and [edit teiHeader](#).

Opções de visualização

Texto: Original Forma escrita Forma normalizada - Mostrar: Cores Imagens

Edit the information about each word of this file by clicking on the word in the text below, or click [here](#) to edit the raw XML

A magia que corre mal

Numa manhã de Verão uma bruxa chamada Júlia, estava a dormir no seu sofá verde ao lado do seu gato muito maroto. O gato era amarelo com umas **bochechas** vermelhas e uns bigodes maiores que a largura da cara.

Ele de tanto ficou com uma fome desgraçada. Então, como o seu telefone roxo que estava ao lado da jarra vermelha **tinha** a forma de um **hambúrguer** e de uma boa bebida, pegou na varinha da bruxa Júlia mas com o m: **hamburger** acordar. **Tirou-a** e lançou um feitiço.

Os objetos tinham-se transformado numa refeição óptima para ele. Comeu, comeu e comeu, ficou muito cheio.

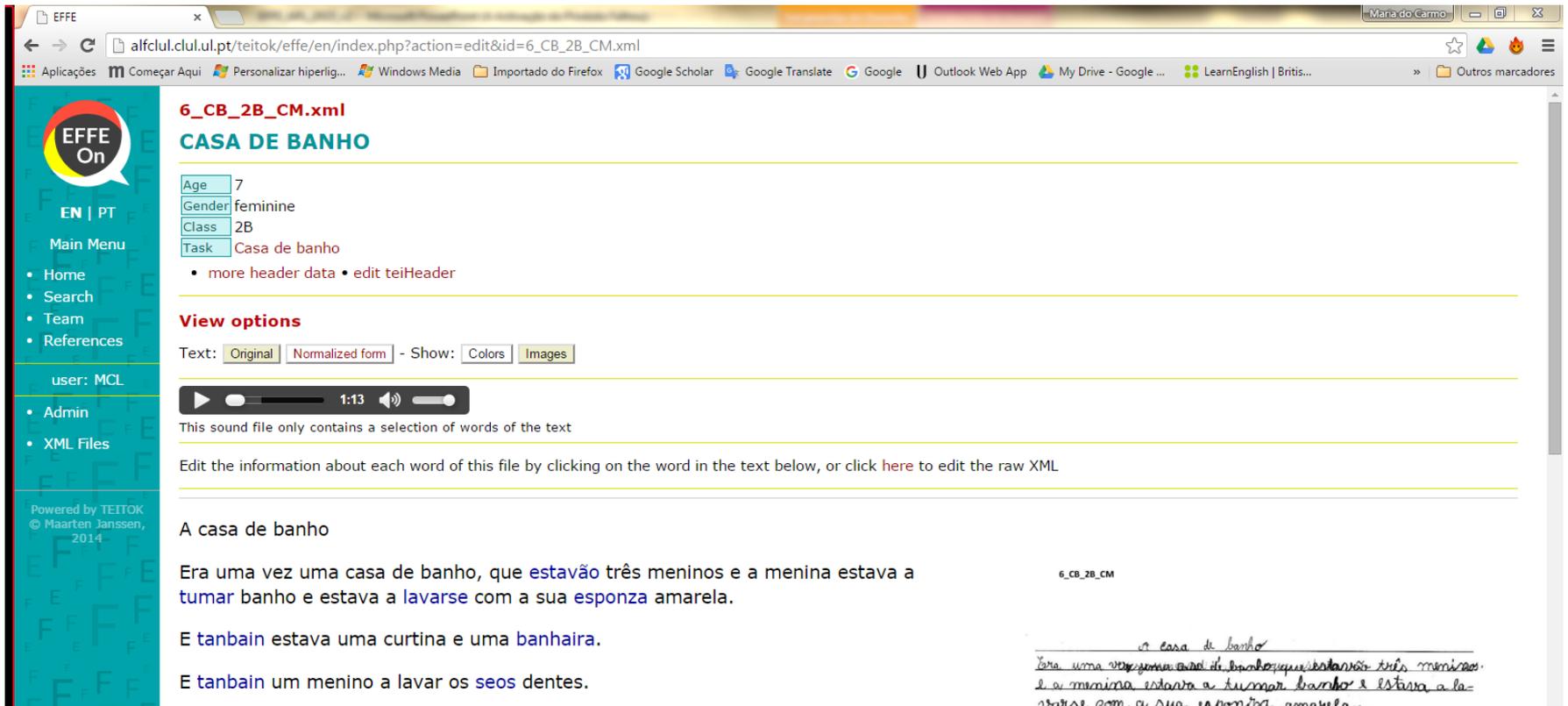
Mas a magia não correu bem, porque ele não era um bruxo era um gato, isso fez **com**

10_BR_4B_CM

EFFE-On

(Opções de visualização)

Áudio



The screenshot shows a web browser window with the URL `alfclul.dul.ul.pt/teitok/effe/en/index.php?action=edit&iid=6_CB_2B_CM.xml`. The page title is **6_CB_2B_CM.xml** and the main heading is **CASA DE BANHO**. On the left, there is a sidebar with the EFFE-On logo, language options (EN | PT), a Main Menu (Home, Search, Team, References), user information (user: MCL), and Admin/ XML Files links. The main content area includes a metadata table:

Age	7
Gender	feminine
Class	2B
Task	Casa de banho

Below the table, there are links for `more header data` and `edit teiHeader`. A **View options** section contains a `Text:` dropdown menu with `Original` and `Normalized form` options, and a `Show:` dropdown menu with `Colors` and `Images` options. An audio player is visible with a progress bar at 1:13. Below the player, it states: "This sound file only contains a selection of words of the text". There is also a link to "Edit the information about each word of this file by clicking on the word in the text below, or click here to edit the raw XML".

The text content is:

A casa de banho

Era uma vez uma casa de banho, que **estavão** três meninos e a menina estava a **tumar** banho e estava a **lavarse** com a sua **esponza** amarela.

E **tanbain** estava uma cortina e uma **banhaira**.

E **tanbain** um menino a lavar os **seos** dentes.

On the right side of the page, there is a small image of a handwritten note titled "a casa de banho" with the text: "Era uma vez uma casa de banho que estavam três meninos e a menina estava a tomar banho e estava a lavar-se com a sua esponja amarela."



Pesquisas na EFFE-On

Corpus Search

CQP Query: Search [advanced](#)

Type in a search query in the CQP (Corpus Query Protocol) format in the text box above to search in the corpus. Which fields can be searched on depends on the corpus, but they typically include:

- word: the written version of the word
- nform: the normalized/regularized version of the word
- lemma: the lemma
- pos: the Part-of-Speech tag

The [advanced search](#) page includes several of the corpus-specific search options

The CQP Query syntax uses an intuitive system of defining properties of words you are looking for, as in for instance:

```
[lemma="casa"] [pos="A.*"]
```

for form of the word *casa* followed by a adjective.

Within each query, it is possible to use full regular expression searches, to find words that have a diphthongue "ue" in it, you can use:

```
[word="[^q]ue.*"]
```

Powered by TEITOK
© Maarten Janssen,
2014



Pesquisas na EFFE-On

The screenshot displays the EFFE-On Corpus Search interface. The browser address bar shows the URL: `alfclul.clul.ul.pt/teitok/effe/en/index.php?action=cqp&act=advanced`. The page title is "Corpus Search".

Text Search

Search method: CQP Word Search

Written form: matches [input type="text"]

Normalized form: matches [input type="text"]

POS tag: matches [input type="text"]

Lemma: matches [input type="text"]

Display method: KWIC Attribute list

Form to show: Written form [input type="text"]

Context size: 5 [input type="text"] words

Sort on: Word [input type="text"]

Matching strategy: Longest match [input type="text"]

Document search

Student ID: [select]

Age: [select]

Gender: [select]

School year: [select]

School city: [select]

Education father: [select]

Education mother: [select]

Task: [select]

Left Sidebar:

- EN | PT
- Main Menu
 - Home
 - Search
 - Team
 - References
- user: MCL
- Admin
- XML Files
- Powered by TEITOK
© Maarten Janssen, 2014



Pesquisas na EFFE-On (Corpus Query Protocol, CQP)

Corpus Search

CQP Query: Search advanced

Type in a search query in the CQP (Corpus Query Protocol) format in the text box above to search in the corpus. Which fields can be searched on depends on the corpus, but they typically include:

- word: the written version of the word
- nform: the normalized/regularized version of the word
- lemma: the lemma
- pos: the Part-of-Speech tag

The advanced search page includes several of the corpus-specific search options

user: MCL

The CQP Query syntax uses an intuitive system of defining properties of words you are looking for, as in for instance:

```
[lemma="casa"] [pos="A,*"]
```

for form of the word casa followed by an adjective.

Within each query, it is possible to use full regular expression searches, to find words that have a diphthongue "ue" in it, you can use:

```
[word="[ue].*"]
```

Corpus Search

CQP Query: Search advanced

178 results • Showing 0 - 100 (next)

/4_BR_2B_CM	banco asegir a broxa lansou	faitso	contra o pasaro . De
/38_FL_2A_CM	dois passaros uma serpente um	paixe	a mergulhar no lago e
/39_FL_2A_CM	mostra-se ai ja congeuias ter	mais	amigos agora tenta ser bom
/38_BR_2A_CM	do passaro e então nunca	mais	voltou a fazer aquilo .
/37_BR_2A_CM	perferia estar a dormir em	baixo	. A Mariana reparou num
/31_BR_4A_CM	e disse uma coisa muito	baixinho	: - Uuuuaaa-catrapás o telemovel
/17_FL_2B_CM	o dragao e vi um	paixe	e uma cobra e vi
/56_SA_2A_CM	zomal os livros o tiaginho	vai	arromar rapido se não arramars
/23_BR_2B_CM	pasaro e depois o pasaro	vaio	para ca a baicho para
/23_BR_2B_CM	pasaro vaio para ca a	baicho	para o baco e depois
/23_BR_2B_CM	depois a brocha fes um	faitso	ao pasaro e o pasaro
/10_BR_4B_CM	bochexas vermelhas e uns bigodes	maiores	que a largura da cara
/10_BR_4B_CM	a mão por goela a	baixo	como se não houve-se amanhã
/10_BR_4B_CM	, não era nada de	mais	, por isso ficou todo
/28_FL_2B_CM	ovos O tempo estava a	daitar	chova o caracol etava nas
/28_FL_2B_CM	e a elva etane a	daitar	chova e o terra etava
/2_BR_2B_CM	a voar . O chapéu	mais	feito estava sempre no banco
/20_CZ_2B_CM	tinha ratos era bom soque	sain	ningani não tinha piada era
/20_CZ_2B_CM	a danssare com o parabanis	bain	doque eu vaio a lua
/20_CZ_2B_CM	o parabanis bain doque eu	vaio	a lua e as estrelas
/44_BR_2A_CM	ele vuou até cá a	baicho	e plim transformo-se nun chapueu
/52_SA_2A_CM	Eu ontem vi o meu	pai	a falar na televisão e
/52_SA_2A_CM	coadro bonito e o meu	pai	disse na televisão . -
/30_CB_2B_CM	o fim . A menina	vai	se lavare e á agua
/30_CB_2B_CM	á agua e o menino	vai	fasere xixi e o otero
/56_BR_2A_CM	um chapueu que era mointo	mais	lindo e comeso a vorar
/41_BR_2A_CM	chapéu voo a bruxa nunca	mais	quis ver o chapéu a



Pesquisas na EFFE-On (Corpus Query Protocol, CQP)

Corpus Search

CQP Query: Search advanced

Type in a search query in the CQP (Corpus Query Protocol) format in the text box above to search in the corpus. Which fields can be searched on depends on the corpus, but they typically include:

- word: the written version of the word
- nform: the normalized/regularized version of the word
- lemma: the lemma
- pos: the Part-of-Speech tag

The **advanced search** page includes several of the corpus-specific search options

The CQP Query syntax uses an intuitive system of defining properties of words you are looking for, as in for instance:

```
[lemma="casa"] [pos="A.*"]
```

for form of the word *casa* followed by a adjective.

Within each query, it is possible to use full regular expression searches, to find words that have a diphthong "ue" in it, you can use:

```
[word="[ ^q]ue.*"]
```

Powered by TEITOK
© Maarten Janssen, 2014

Corpus Search

CQP Query: Search advanced

3 results

/20_CZ_2B_CM	rato a danssare com o parabanis bain doque eu vajo a
/26_CZ_4B_CM	minha irmã . Cantamos os parabéns e depois fomos para a
/25_CZ_4B_CM	à minha festa cantámos os parabéns depois dessa festa fui dormir

Download results as TXT

Frequency Options

Use the query above to calculate:

- Frequency by POS
- Frequency by lemma
- Frequency by POS+lemma

Or run an additional custom CQP command on the results above (Matches):

CQP Query: Apply

Powered by TEITOK
© Maarten Janssen, 2014



Pesquisas na EFFE-On (word search)

The screenshot displays the EFFE-On Corpus Search interface. On the left is a navigation sidebar with the EFFE-On logo, language options (EN | PT), a main menu (Home, Search, Team, References), user information (user: MCL), and administrative links (Admin, XML Files). The main content area is titled "Corpus Search" and is divided into "Text Search" and "Document search" sections. The "Text Search" section includes search method options (CQP, Word Search), input fields for "Written form", "Normalized form" (containing "afitinho"), "POS tag", and "Lemma", and options for "Display method" (KWIC, Attribute list), "Form to show" (Written form), "Context size" (5 words), "Sort on" (Word), and "Matching strategy" (Longest match). The "Document search" section features a list of filters (Student ID, Age, Gender, School year, School city, Education father, Education mother, Task) and a search bar with the query "[nform='afitinho']". Below the search bar, it shows the search results: "Matching strategy: longest", "Normalized form = afitinho", and "3 results". The results list three entries with their normalized forms and original text: "/17_CB_4B_CM e com as calsas vermelhas **afelitinho** para fazer xixi . A", "/31_CB_2A_CM o seu irmão Francisco estava **afelitinho** para fasere chi-chi o ceu", and "/6_CB_2B_CM E um menino que estava **afelitinho** para fazer xixi . A". A "Download results as TXT" button is present. A "Frequency Options" section offers calculations for "Frequency by POS", "Frequency by lemma", and "Frequency by POS+lemma", with an "Apply" button. The interface is powered by TEITOK and Maarten Janssen, 2014.



Considerações Finais

The screenshot displays a web application interface for XML data. On the left, a sidebar features the EFFE On logo and a navigation menu with options like 'Início', 'Pesquisar', 'Team', and 'References'. The main content area is divided into two sections, each representing an XML file.

10_SA_2B_CM.xml
SALA
Idade: 7
Gênero: masculino
Turma: 2B
Tarefa: Sala
• mais informação • edit teiHeader

Opções de visualização
Texto: Original | Forma escrita | Spoken | Forma normalizada - Mostrar: Cores | Formatação | <lb> | Imagens | Audio

1:55

This sound file only contains a selection of words of the text

Edit the information about each word of this file by clicking on the word in the text below, or click [here](#) to edit the raw XML

O meu pai quando eu **chegei** a casa o meu pai estava a cantar e o meu pai estava **olado** de uma mesa que tem um jornal três **lapis** uma **tesora** dinheiro e um **desenho** dois livros no chão um paiaso de brincar um martelo um jogo de **taboleir** **atras** da janela **esta** uma **brucha** a voar na **soa** **vassora** uma **igraisa** um gato a brincar com uma bola um cofre de **dinhairo** uma televisão um radio **doas** **caichas** um vaso com plantas um **sofa** almofadas um **coadro** com a **imaxai** **dom** **por** do sol um **gelobo** um saco de dinheiro.

10_SA_4B_CM.xml
SALA
Idade: 9
Gênero: masculino
Turma: 4B
Tarefa: Sala
• mais informação • edit teiHeader

Opções de visualização
Texto: Original | Forma escrita | Forma normalizada - Mostrar: Cores | Imagens

Edit the information about each word of this file by clicking on the word in the text below, or click [here](#) to edit the raw XML

A bruxa e o pai

Eu cheguei a casa, da escola e estava tudo **dezarromado** com o meu a cantar, achei estranho porque é raro ele cantar a mas as coisas **dezarromadas**, dois livros no chão um violeta e outro roxo, o meu jogo favorito espalhado no chão laranja a caixa deitada , uma caixa azul e a sua cor preferida, dinheiro fora do cofre cinzento em cima da mesa, o nosso globo estava a girar e **um** o palhaço estava no chão.

Mas **o** por **detráz** disto estava uma bruxa que estava à janela quase a chorar de rir. Ela tinha feito um feitiço para o meu cantar e o meu pai quando canta **dalhe** vontade de dançar, mas ele não sabe dançar por isso e que estava todo **dezarromado**, ele **esta** estava **cem** a sua camisa amachucada a sua gravata torta, as calças **azuis** estavam bem.

Fla **actra** estava ao né da igreja e ainda não me tinha visto. Por isso subi até ao

Considerações Finais

- Base de dados em construção
- Em andamento:
 - - novos critérios de busca
 - - recolha e edição de novos materiais
 - - tratamento dos ficheiros-áudio

Fim!

Pesquisa no corpus

Padrão CQP:

Search

4 resultados

/44_BR_2A_CM feitiço esta mal feito . **vitoria** vitoria acabou-se a estoria .
/44_BR_2A_CM esta mal feito . vitoria **vitoria** acabou-se a estoria . O
/31_CB_2A_CM ágoa e sair da banhaira **vitoria** vitori acabose esta istoria .
/20_BR_2B_CM a ! o meno chapeo|vitoria **vitoria** acabosse a estoria . Era

Download results as TXT

Referências

- Guerreiro, H. W. Z. M. R. (2007) *Processos fonológicos na fala da criança de cinco anos*. Tese (Mestrado em Ciência da Fala). Instituto de Ciências da Saúde da Universidade Católica Portuguesa-Escola Superior de Saúde do Alcoitão da Santa Casa da Misericórdia de Lisboa.
- Lourenço-Gomes, M. C; Rodrigues, C.; Alves, I. (to appear) Escreves como Falas - Falas como escreves? *Revue Romane*, John Benjamins.
- Yavas, M.; Hernandorena, C. L. M & Lamprecht, R. R. (1991) *Avaliação fonológica da criança*. Porto Alegre: Artes Médicas.



Obrigada!

EFFE – ON

Contactos

Celeste Rodrigues

celesterodrigues@campus.ul.pt

Maria do Carmo Lourenço-Gomes

mclgomes@ilch.uminho.pt

