

# Etymology in the Portuguese Academy of Sciences Dictionary

---

ANA SALGADO

NOVA CLUNL, FACULDADE DE CIÊNCIAS SOCIAIS E HUMANAS, UNIVERSIDADE NOVA DE LISBOA

ACADEMIA DAS CIÊNCIAS DE LISBOA, PORTUGAL

# Motivation

---

- DH Training workshop is a great opportunity to work with experts
- the conversion of the Portuguese Academy Dictionary to TEI Lex-0
- when the PDF document (print edition) was converted to TEI, not every detail of the etymology was properly annotated
- to improve my competences in TEI Dictionaries encoding
- to promote the discussion with the TEI Lex-0 community

# Portuguese Academy of Sciences Dictionary

- The *Dicionário da Língua Portuguesa Contemporânea* (DLPC) was published in 2001, in two-volume, 3809 pages.
- It's a general language contemporary dictionary with a descriptive nature and a normative concern.
- The word list of DLPC has a total of 69 426 entries with 167 556 senses.
- The digital version is being used as the basis for the creation of a new edition.

aprensão. 5. *Jur.* Retenção ilegal de pessoas privando-as da liberdade, com objectivos políticos, sociais ou para obtenção de um resgate. 6. Acto de se apoderar ilegalmente dos bens alheios. 7. O facto de estar afastado ou separado de um grupo, da sociedade. 8. *Med.* Fragmento de tecido ósseo que, ao longo da necrose, foi afastado do osso são. **sequioso, a** [sɐkiu̯z, -oz], *adj.* (De *seco* + *-i-* + sufix. *-oso*). 1. Que tem intenso desejo de beber; que tem sede. = **sedento**. 2. Que carece de água; que é muito seco. *A manada percorreu aquelas terras sequiosas e poeirentas, em busca de um sopro de orvalho.* 3. Que deseja ardentemente. = **ansioso**, **avido**, **faminto**, **sórego**. *É uma criança sequiosa de saber, sequiosa por aprender.* Pl. **sequiosos**, as [sɐkiu̯z, -wz]. **sequito** [sɐki̯tu] s. m. (Do it. *sequito*, do lat. *sequi* 'seguir'). 1. Conjunto de pessoas que acompanham uma individualidade por dever oficial ou por cortesia. = **acompanhamento**, **comitiva**, **cortejo**. *O sequito do Presidente segue-o sempre para onde quer que ele vá.* 2. Acompanhamento, companhia. 3. Amizade, estima; simpatia. **sequoia** [sɐkijɐ] s. f. (Do ingl. *sequoia*, de *Sequoiah*, antrop.). *Bot.* Árvore conífera, originária da América do Norte (*S. sempervirens*, Lamb., Endl., *S. wellingtonia*, Seem.), notável pelo seu tamanho gigantesco e pela sua longevidade. = **WILLINGTONIA**. **ser** [sɐr] v. (Do lat. *sedere* 'estar sentado'). I. Como verbo pleno significa: 1. Ter existência real. = **EXISTIR**. *Ser ou não ser: eis a questão. Se não fosse o meu tio, não estaria neste lugar. Esteve quase a desistir, não fossem os meus assistentes.* **FEITO** de **SER**. 2. Ter existência, ter acontecer, ocorrer, suceder. *Não se sabe com profundidade do navio. Se conseguirmos o lugar, (ele) será em vão.* 3. Situar-se, alguma coisa, em determinado lugar. = **ESTAR**, **ficar**, **LOCALIZAR-SE**. *A Torre de em Lisboa. A casa dele é na rua Palmeirim.* 4. Des-  
-se, de modo fixo, um evento, em determinado lugar, decorrer, passar-se. *A festa de São João é no Parque Espinho é em Paris.* 5. Ter lugar, um acontecimento, dado momento do tempo. = **OCORRER**, **REALIZAR-SE**. *ultimato foi em 1890. As festas dos santos populares Junho. O casamento é ao meio-dia.* 6. Chegar, a um determinado momento do tempo, uma estação, uma quadra, etc. *São dez horas. É meio-dia. Era Primavera. Até é Agosto. É Natal. É o século XX. É tempo de dar horas de trabalhar.* 7. Vir, alguém ou alguma coisa, terminado lugar. (Seguido da prep. *de*). = **PROVIR**. *os jogadores são do Brasil. Essa música é de Cabo Verde.* 8. Pertencer a alguém ou alguma coisa. (Seguido de *de* ou *poss.*). *As crianças são da vizinhança. Os livros são da biblioteca. Essas cadeiras são da sala de jantar. O pontão é nosso.* 9. Situar-se, alguém, entre; estar, colocado entre. (Seguido da prep. *de* e nome pl. colectivo). *Esses jogadores são dos bons. O capitão é português. Ela é da clique.* 10. Estar, alguma coisa, relacionada com; dizer, alguma coisa, respeito a. (Seguido de *com*). *O caso agora é com a polícia. Os incidentes são com os bombeiros.* 11. Ter, alguém, uma qualidade em algum grau. (Seguido da prep. *de*). *Ela é de uma abnegação à prova.* 12. Mostrar-se, uma qualidade, digna, por parte de alguém. = **CONSTR**. *Esse comportamento é de agradecer.* 13. Ter, alguém, capacidade, inclinação para, para ou *de*. *Este rapaz, não é para ser músico.* 14. Mostrar-se, alguém, favorável a, partidário. (Seguido da prep. *por*). = **TORCER**. *Muitos castelos são pela língua portuguesa. Esta clique é pela equipa.* 15. Custar, alguma coisa, determinado valor, por parte de alguém. *Quanto é o livro? Uma revista de papel é mil.* II. Como verbo copulativo ou de ligação entre o sujeito e o complemento predicativo, indica: 1. Identificar, eventualmente com o sujeito implícito. *Lisboa é a capital de Portugal. A capital de Portugal é Lisboa. Chegar*

*tro é a ambição de muitos políticos. São os nossos amigos. Quem é você?* 2. Classificação, inclusão. *Ela é professora. O castanheiro é uma árvore de grande porte. A democracia é respectivo pelos outros. Roubar é um delito.* 3. Atribuição, qualificação permanente. *A Terra é redonda. Essa criança é alegre. O muro é alto. Esta mesa é de madeira. É bom respirar a aragem da manhã, maneira de ser, modo de ser.* 4. Resultado de uma acção, quando, seguido de participio passado, funciona como auxiliar de construção passiva. *A praça foi ocupada pelos manifestantes. Mostrou-se apreensivo por não ser observado pelo médico. O avião foi abatido.* III. Como verbo modalizador indica: 1. Exortação, ordem, usado no presente do indicativo (3ª pers. sg.) e seguido de infinitivo. *Se tivermos de sair, é sair e acabar-se. Vem aí as férias, é saber aproveitar-as.* 2. Obrigatoriedade, quando está seguido de infinitivo regido pela prep. *de* ou *para*. = **DEVER**. *Essa decisão não é de surpreender. Se tivesse juízo, era para estar calado. Vamos, as mães são para comer.* 3. Reforço do valor dubitativo de frases interrogativas, nas formas *verá* e *seria*, seguidas de completiva. *Será que a tia acertou no total? Seria que eles tinham razão?* 4. *Fam.* Confirmação positiva, nas formas *é, foi, era...*, como resposta a determinadas perguntas afirmativas e em substituição do verbo destas. = **SIM**. *Você falou do desastre desta manhã? — É! (Sim, falou).* — *Uma loja fechou para o almoço? — Foi. (Sim, fechou).* 5. *Fam.* Confirmação positiva, usado em determinadas frases interrogativas, nas formas *não é, não foi, não era...*, em



# Why TEI Lex-0?

---

- The TEI dictionary chapter is the de facto encoding standard for dictionaries digitized from print sources.
- The community is discussing a new format for dictionaries – TEI Lex-0.
- Our goal is to have the dictionary in TEI Lex-0 for tests and to promote the discussion with the community, keeping the original version in our own interpretation of TEI.

---

## TEI Lex-0 — A baseline encoding for lexicographic data

---

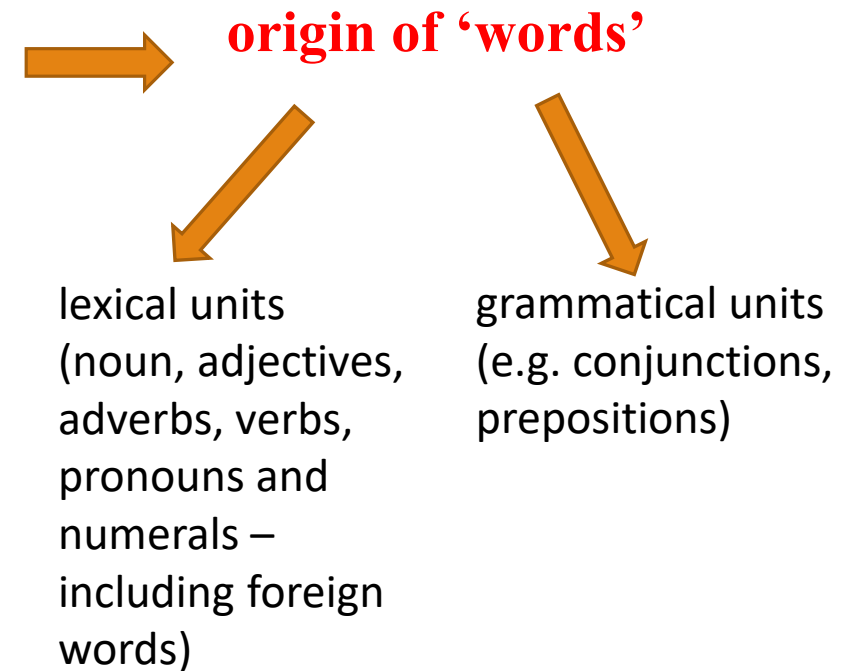
### Table of contents

1. [Introduction](#)
  - 1.1. [Context](#)
  - 1.2. [Dictionary model vs. XML serialization](#)
2. [Entries and entry-like objects](#)
  - 2.1. [Unified representation of entries](#)
  - 2.2. [Changes to the content model of entry](#)
  - 2.3. [Types of entries](#)
3. [Senses](#)
  - 3.1. [General remarks](#)
  - 3.2. [A limited number of possible contexts for def](#)
  - 3.3. [Encoding sense in entries with multiple parts of speech](#)
  - 3.4. [Case of bilingual dictionaries](#)
  - 3.5. [Deprecating hom](#)
4. [Written and Spoken Forms](#)
  - 4.1. [A note on inheritance](#)
  - 4.2. [Grammatical properties of lexical entries](#)
  - 4.3. [Representation of the lemma](#)
  - 4.4. [Representation of inflected forms](#)
  - 4.5. [Paradigms](#)
  - 4.6. [Representation of variants](#)
    - 4.6.1. [Orthographic variation](#)
    - 4.6.2. [Phonetic Variation](#)
    - 4.6.3. [Regional or dialectal Variation](#)
5. [Cross-references](#)
  - 5.1. [General remarks](#)
  - 5.2. [Using ref and xr to point to lexical entities](#)
    - 5.2.1. [Typology for xr/@type values](#)
    - 5.2.2. [Values of ref/@target](#)
    - 5.2.3. [How TEI Lex-0 differs from TEI Guidelines](#)
  - 5.3. [xr and ref: further examples](#)
    - 5.3.1. [Simple cross-reference to an entry](#)
    - 5.3.2. [More complex example including quotations](#)
    - 5.3.3. [Cross-references inside definitions](#)
6. [Usage information](#)
  - 6.1. [Label-like vs. narrative usage descriptions](#)
  - 6.2. [Types of usage information](#)
  - 6.3. [Restricting the scope of use](#)
7. [Etymology](#)

[https://dariah-eric.github.io/lexicalresources/pages/TEILex0/TEILex0.html#index.xml-body.1\\_div.2\\_div.2](https://dariah-eric.github.io/lexicalresources/pages/TEILex0/TEILex0.html#index.xml-body.1_div.2_div.2)

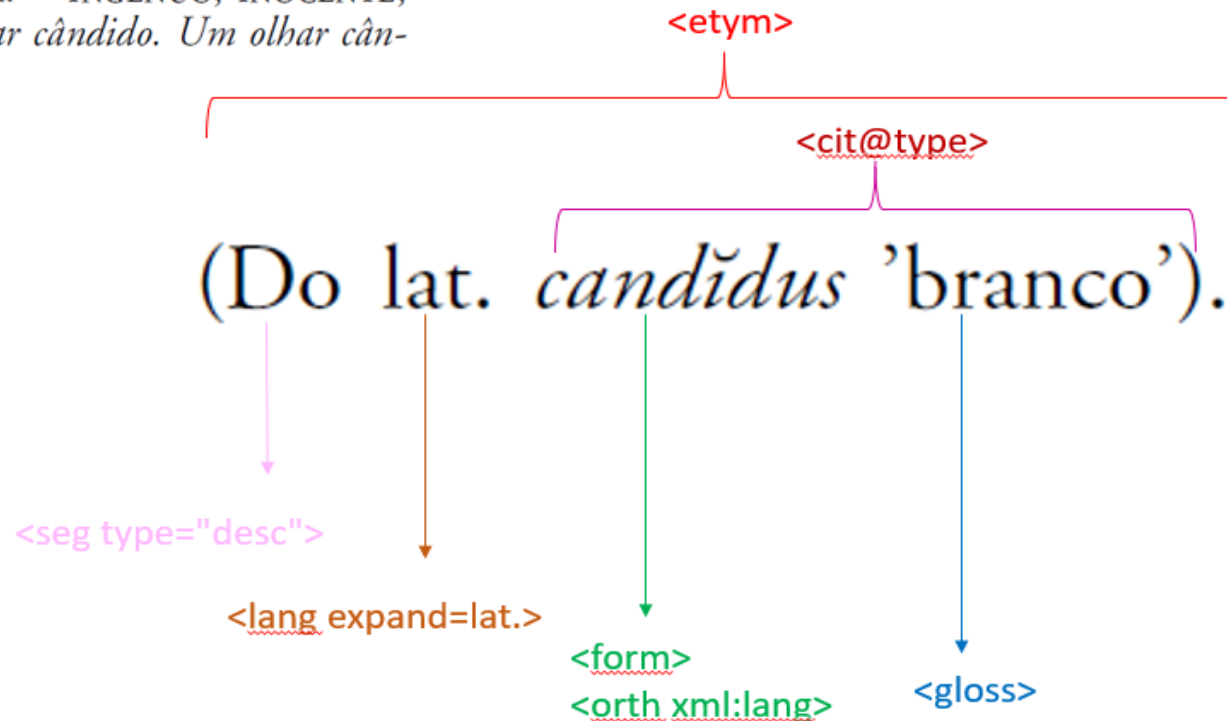
# DLPC etymology

**etimologia** [ɛtimuluʒiɐ]. *s. f.* (Do lat. *etymologia* < gr. ἔτυμολογία, de ἔτυμος 'verdadeiro' + λόγος 'palavra', 'razão'). **1.** *Ling.* Ciência que constitui uma parte da Gramática, cujo objectivo é o estudo da **origem das palavras** de uma língua e da evolução das mesmas até ao estado actual. «*a lexicologia é uma disciplina muito mais vasta que [...] inclui digamos de certo modo a própria etimologia portanto a origem das palavras e a sua história*» (OF, 4). **2.** Origem, génese, derivação de uma palavra. *Procurou a*

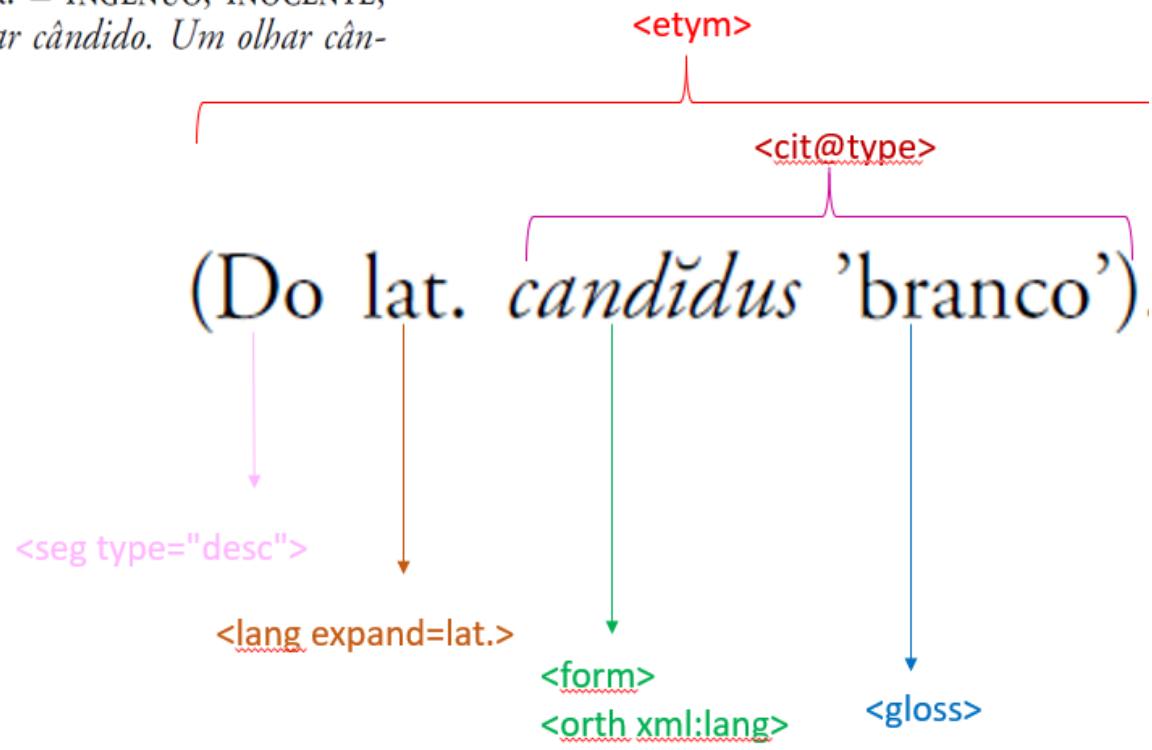


# DLPC etymology

**cândido, a** [kɛ́didu, -ɐ]. *adj.* (Do lat. *candīdus* 'branco').  
1. Que é branco; que tem candura. 2. Que revela pureza, inocência, simplicidade, candura. ≈ INGÉNUO, INOCENTE, PURO. *Uma alma cândida. Um ar cândido. Um olhar cândido.* Adv. candidamente.



**cândido, a** [kêdidu, -ɐ]. *adj.* (Do lat. *candīdus* 'branco').  
**1.** Que é branco; que tem candura. **2.** Que revela pureza, inocência, simplicidade, candura. ≈ INGÉNUO, INOCENTE, PURO. *Uma alma cândida. Um ar cândido. Um olhar cândido.* Adv. candidamente.



```

<entry xml:id="cândido" xml:lang="pt">
  <form type="lemma">
    <orth xml:lang="pt">cândido</orth>
  </form>
  <form type="inflected">
    <orth>cândido</orth>
    <pron>k'êdidu</pron>
    <gramGrp>
      <gram type="gen">m.</gram>
    </gramGrp>
  </form>
  <form type="inflected">
    <orth>cândida</orth>
    <gramGrp>
      <gram type="gen">f.</gram>
    </gramGrp>
    <pron>k'êdide</pron>
  </form>
  <!--etc.-->
  <etym type="inheritance">
    <pc></pc>
    <seg type="desc">Do</seg>
    <cit type="etymon">
      <lang expand="latim">lat.</lang>
      <form>
        <orth xml:lang="la">candīdus</orth>
        <pc>'</pc>
        <gloss xml:lang="pt">branco</gloss>
        <pc>'</pc>
      </form>
    </cit>
  </etym>
  <!--etc.-->
</entry>

```

# Methodology

---

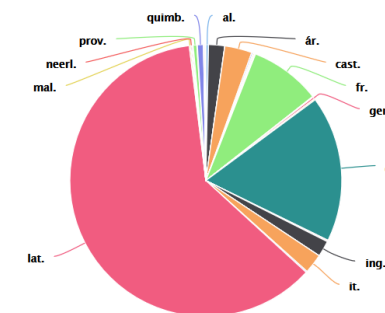
1. Analysis of the paper edition abbreviations list.
2. Gathering of abbreviations that occur in the field of etymology.
3. Gathering of examples illustrating the use of these abbreviations and different types of etymological phenomena.
4. Create a template (xlsx file) with abbreviations, expanded forms, informations about usage, examples and some observations.
5. Create xml files for each of the etymological examples.
6. Test typology of etymological processes in TEI Lex-0 (Bowers, Herold, Romary, 2018).
7. Create a set of regular expressions to detect clear portions of the etymology rules to an automatized encoding etymology components that cover all cases.



# Hands-on activities

| Abbreviations <etym> | Expanded          | Usage  | Entries                       | Observations                |
|----------------------|-------------------|--|-------------------------------|-----------------------------|
| a.                   | alto              | <i>always associated to a language</i>         | pífaro, represália            |                             |
| A.                   |                   | <i>initial of an anthroponym</i>               | salazarismo                   | IGNORE (+antrop.)           |
| abl.                 | ablativo          | <i>about ablative case in Latin</i>            | fidedigno                     |                             |
| abrev./Abrev.        | abreviatura       | <i>always associated to abbreviations</i>      | expo, LP                      |                             |
| absol.               | absoluto          |  | novíssimo                     | IGNORE??? +superl.???       |
| ac.                  |                   |  |                               | fixed; paper error; = acus. |
| acus.                | acusativo         | <i>about acusative case in Latin</i>           | requiem                       |                             |
| Adapt.               | adaptação         |  | ???                           | where???                    |
| adv.                 | advérbio          | <i>grammatical units</i>                       | daí                           |                             |
| al./Al.              | alemão            | <i>language</i>                                | bunker, búnquer               |                             |
| alt.                 |                   |  |                               | IGNORE = lat.               |
| ant.                 | antigo            | <i>always associated to a language</i>         | afã, mostarda                 |                             |
| antrop.              | antropónimo       | <i>units derived from the name of a person</i> | bechamel, masoquista          |                             |
| antropo.             | eliminado         |  |                               | IGNORE= antrop.             |
| ár./Ár.              |                   | <i>language</i>                                | Corão Do ár. quārān 'leitura' |                             |
| árab.                | eliminado         |  |                               | IGNORE = ár.                |
| arc.                 | arcaísmo, arcaico | <i>always associated to a written form</i>     | cova                          |                             |
| art.                 | artigo            | <i>grammatical units</i>                       | el                            |                             |

al. 102  
 ár. 629  
 beng. 1  
 cast. 1064  
 cat. 50  
 célt. 25  
 chin. 47  
 conc. 9  
 din. 2  
 drav. 1  
 esp. 16  
 finl. 3  
 fr. 2761  
 franç. 24  
 gal. 5  
 germ. 80  
 gót. 45  
 gr. 5683  
 hebr. 39  
 hindust.9  
 húng. 9  
 ing. 621  
 isl. 2  
 it. 761  
 jap. 43  
 lat. 20035  
 mal. 59  
 nauat. 1  
 neerl. 47  
 nor. 6  
 pers. 4  
 port. 35  
 prov. 143  
 quich. 17  
 quimb. 242  
 rom. 1  
 rus. 14  
 sâns. 2  
 sâns. 45  
 tib. 1  
 turc. 10  
 vasc. 12



xml:lang=« [language code]">[etymon]</orth>

| Abbreviations <etym> | Expanded | Usage                                  | Entries            |
|----------------------|----------|--|--------------------|
| a.                   | alto     | <i>always associated to a language</i> | pífaro, represália |
| ant.                 | antigo   | <i>always associated to a language</i> | afã, mostarda      |
| cláss.               | clássico | <i>always associated to a language</i> | açafrão            |
| tard.                | tardio   | <i>always associated to a language</i> | florir             |

# Data categories

## Borrowing (*Empréstimo*)

Donut is a loanword from english.

**donut** [dónut]. *s. m.* (Ingl.). *V. donute.*

**dónute** [dónuti]. *s. m.* (Do ingl. *donut*). Argola de massa frita enrolada em açúcar ou com diferentes coberturas, em forma de anel ou bola.

## Inheritance (*Herança*)

Portuguese “tingir” was inherited from Latin “tingere”.

**tingir** [tĩʒír]. *v.* (Do lat. *tingere*). **1.** Alterar a cor de determinado material através da aplicação de uma substância corante. *Mandei tingir os meus sapatos de preto.* **2.** Cobrir ou ficar coberta, uma parte de certa superfície, com determinada substância de cor diferente; pôr ou receber manchas de tinta ou de cor.  $\approx$  MANCHAR. *A criança tingiu o bibe com tinta. O vinho tingira-lhe a camisa. As mãos do mecânico tingiram-se de óleo.* **3.** Dar ou tomar determinada cor.  $\approx$  COLORIR. *O pôr-do-sol tingia toda a planície de um tom alaranjado. A sua face tingiu-se de rubor, ao ouvir aquelas palavras. «E ao acabar de dizer estas palavras tingiram-lhe as faces de uma vermelhidão intensa.»* (J. DINÍS, *Fidalgos*, p. 450). **tingir a cara de preto**, esconder-se ou tapar a cara de vergonha por ter cometido algum erro. **tingir as mãos/a espada no sangue de alguém**, ferir ou matar alguém.

## Derivation (*Derivação*)

**alquimista** [alkimíʃte]. *s. m. e f.* (De *alquimia* + suf. *-ista*). **1.** Pessoa que se dedicava à alquimia. **2.** O que transforma, transmuta; mágico. **3. Deprec.** Falsificador.

|

## Portmanteau (*Amálgama*)

**aparthotel** [ɐpartɔtél]. *s. m.* (De *apart<amento>* + *hotel*). Bloco de apartamentos com serviço de hotel. «Desde a década de 80, o 'aparthotel' é um êxito de investimento imobiliário em todo o mundo» (*Expresso*, 27.10.1990).

The word “aparthotel” is a portmanteau of “apartamento” + “hotel”.

SEE Etymological processes in TEI Lex-0  
(Bowers, Herold, Romary, 2018)

# Other lexical Portuguese resources

**opalina** [ɔpɐ'line]. *s. f.* (De *opala* + suf. *-ina*). **1.** Vidro translúcido, de aspecto semelhante ao da opala, que se usa no fabrico de objectos decorativos. *Jarra em opalina.* **2.** Objecto fabricado com essa espécie de vidro. *Colecção de opalinas.* **3.** Vidro espesso, de cor branca leitosa, utilizado no revestimento de paredes, tectos... **4. Zool.** Género (*Opalina*, Purk.) de protozoários ciliados, que compreende várias espécies.

“opalina” [opaline] (DLPC-ACL)

## ▶ opalina

*s.f.* (1899 cf. CF') **1** vidro de aspecto acetinado, irizado quando visto à contraluz, leitoso, ger. translúcido, us. no fabrico de objetos decorativos, jarras taças, lustres etc. **2 p.met.** o objeto fabricado com esse vidro *⟨tem uma coleção de opalinas francesas⟩* **3** vidro branco, leitoso, espesso, us. para forrar paredes, tetos etc. ◉ ETIM fr. *opaline* (1899) 'tipo de vaso vitrificado'; ver *opal(i)-*; a datação é para a acp., não registrada aqui, "gênero de infusórios que se encontra no ventre das rãs

“opalina” [opaline] (Houaiss)

## opalina

o.pa.li.na • ɔpɐ'line

nome feminino

1. vidro fosco translúcido usado para fabricar objetos decorativos
2. vidro branco e espesso usado para revestir paredes, tetos, etc.
3. MINERALOGIA ver [opala](#)



Do francês *opaline*, «tipo de vaso vitrificado»

“opalina” [opaline] (Infopédia): <https://www.infopedia.pt/dicionarios/lingua-portuguesa/opalina>

## o·pa·li·na

(francês *opaline*)

*substantivo feminino*

[Joalheria] Opala.

“opalina” [opaline] (Priberam): <https://dicionario.priberam.org/opalina>

# DLPC discussions topics

---

- Which elements can occur within <etym> or <cit> elements?
- How many data categories do we have for etymology description?
- Special cases: cognates, derivatives, variants, etc.
- Usage labelling vs. Values for describing etymology processes such metaphor, metonymy, etc.
- Can my methodology be used for automatized encoding?

# Thank you for your attention.

---

*I'm very excited for being here...*

*and I am really motivated to work hard*

*and to provide a comprehensive modeling and representation of etymology in DLPC.*