

# Working with lexical data: Experiences with Uralic languages

Course: NLP for Endangered Languages of the Amazon.  
From a Uralic perspective. Lecture 3.

Jack Rueter, Niko Partanen,  
Mika Hämmäläinen & Khalid Alnajjar

# Dictionaries

Arranging a lexicon is a central task in language documentation

Dictionaries also have a large role in language planning and teaching

Two distinct, but connected, work phases:

- Dictionary creation (i.e. for endangered languages)
- Dictionary maintenance (adding new terms, editing meanings)

Especially in multilingual dictionaries the relationships are important

- Derivational relations, variants, synonyms, antonyms, etymologies etc.

# Dictionary editing

There are several commonly used tools and databases:

- [FLEx](#)
  - [TLex](#)
  - [Lexique Pro](#)
  - [Lexonomy](#)
- 
- In rule based natural language processing we often use so called LEXC files or various XML standards
  - None of these solutions address very well the connections between multilingual dictionaries, and the need for multiple uses for the same data

# Ve'rdd – an open-source dictionary editing platform

The goal has been to combine various dictionaries so that entries can be linked

Not tied to a specific format:

- Uses an internal database, but imports and exports different structures
- Looking for a way to keep ahead of XML and JSON

Write a word or annotation ONCE

Still in experimental phase, but used with several research groups in Finland

Back

# Verdd

gsocrueter

Logout

## Filter

Lexeme:

Exact

Language:

POS:

Inflex Type:

Source:

Range from:

Range to:

Processed:

Contlex:

Order by:

Search

Download

Bulk Approve

## Filter

Lexeme:

laranja

Exact

Language:

-----

POS:

-----

Inflex Type:

-----

Source:

Range from:

-----

Range to:

-----

Processed:

-----

Contlex:

Order by:

-----

Search

Download ▾

Bulk Approve

ID	Lexeme	POS	Contlex	Inflex Type	Language	Notes	Actions
532069	laranja				por		<ul style="list-style-type: none"><li><a href="#">view</a></li><li><a href="#">edit</a></li><li><a href="#">delete</a></li></ul>

Total: 1

English (en)

suomi (fi)

Русский (ru)

эрзякс (myv)

español (es)

Português (pt)

العربية (ar)

- vie
- edi
- del

English ▼

## Filtrar

Lexeme:

laranja

É idêntico

Língua:

-----

Classe gramatical:

-----

Tipo de inflexão:

-----

Fonte:

Desde:

-----

Até:

-----

Processado:

-----

Léxico continuado:

Ordem:

-----

Pesquisar

Baixar ▾

Aprovar tudo

ID	Lexeme	Classe gramatical	Léxico continuado	Tipo de inflexão	Língua	Notas	Ações
532069	laranja				por		<ul style="list-style-type: none"><li><a href="#">mostrar</a></li><li><a href="#">editar</a></li><li><a href="#">excluir</a></li></ul>

Total: 1



Lexemes  
anteriores

# Lexeme: laranja ([mostrar](#)/[editar](#))

ID: 532069

Kieli (ISO 639-2): por

Classe gramatical:

ID do homónimo: 0

Léxico continuado:

Tipo:

ID de inflexão:

Especificação:

Tipo de inflexão:

ID do lema:

Afiliações ([adicionar](#)):

Processado: Não

Alterado recentemente: 6 de Agosto de 2020 às 22:02

Notas:

Metadados ([adicionar](#)):

Exemplos ([adicionar](#)):

Raízes ([adicionar](#)):

- 0 - laranja (abadia\_\_n) ([editar](#)/[excluir](#))
- 0 - laranja (anti\_\_adj) ([editar](#)/[excluir](#))

Paradigmas curtos ([adicionar](#)):

Relações ([adicionar](#)):

Aprovar

Lexemes  
seguintes

Mostrar paradigmas curtos

# Relações (adicionar):

Aprovar

Retornar



# Ve'rdd

## Adicionar relação:

Desde: laranja

Até

Type\*

Tradução

Notes

Notas

Salvar

Retornar



# Adicionar relação:

Desde: laranja

Até

Type\*

- ✓Tradução
- Etimologia
- Palavra composta
- Derivação
- Variante
- Fraseologia
- Hyponymy
- Hyperonymy
- Holonymy
- Meronymy
- Co-hyponym
- Synonym
- Antonym
- Outro

Salvar

# Adicionar relação:

Desde: laranja

Até

appelsiini

**appelsiini** (N) - 0 krl

**appelsiini** (N) - 0 fin

**appelsiini-** (A) - 0 fin

**appelsiinijuoma** (N) - 0 fin

**appelsiinilehto** (N) - 0 fin

**appelsiinilikööri** (N) - 0 fin

**appelsiinilimonadi** (N) - 0 fin

**appelsiini**marmeladi (N) - 0 fin

**appelsiini**mehu (N) - 0 fin

**appelsiini**mehutiiviste (N) - 0 fin

**appelsiini**n (A) - 0 fin

**appelsiini**inkeltainen (A) - 0 fin

**appelsiini**inkokoinen (A) - 0 fin

Retornar



# Adicionar relação:

Desde: laranja

Até

appelsiini (N) - 0 fin

Type\*

Tradução

Notes

Notas

Salvar



Desde

Lexeme: laranja  
([mostrar](#)/[editar](#))

ID: 532069

Kieli (ISO 639-2): por

Classe gramatical:

ID do homónimo: 0

Léxico continuado:

Tipo:

ID de inflexão:

Especificação:

Tipo de inflexão:

ID do lema:

Afiliações ([adicionar](#)):

Processado: Não

Alterado recentemente: 6 de Agosto de 2020 às 22:02

Notas:

Até

Lexeme: appelsiini  
([mostrar](#)/[editar](#))

ID: 30542

Kieli (ISO 639-2): fin

Classe gramatical: N

ID do homónimo: 0

Léxico continuado:

Tipo:

ID de inflexão:

Especificação:

Tipo de inflexão:

ID do lema:

Afiliações ([adicionar](#)):

- [Akusanat: Fin:appelsiini](#) ([editar](#)/[excluir](#))

Processado: Não

Alterado recentemente: 6 de Agosto de 2020 às 18:16

Metadados ( <a href="#">adicionar</a> ):	Notas:
	Metadados

## Relação:

**Kieli (ISO 639-2):**

**Tipo:** Tradução

**Processado:** Não

**Notas:**

**Alterado recentemente:** 29 de Junho de 2021 às 14:03

## Fontes

- [adicionar](#)

## Exemplos

- [adicionar](#)

## Metadados

- [adicionar](#)

Aprovar



Lexemes anteriores

[lampeiro](#)

[lamuriador](#)

[lanceador](#)

[lanceolado](#)

[lancetado](#)

[lancetável](#)

[lanchado](#)

[lancinante](#)

[landês](#)

[landeiro](#)

[langoroso](#)

[lanoso](#)

[lanuginoso](#)

[lançado](#)

[lançador](#)

[lançável](#)

[laparotomizável](#)

[lapidado](#)

[lapidificável](#)

[lapidoso](#)

[lapidário](#)

[lapidável](#)

[lapidífico](#)

[laqueado](#)

[laqueável](#)

# Lexeme: laranja ([mostrar](#)/[editar](#))

ID: 532069

Kieli (ISO 639-2): por

Classe gramatical:

ID do homónimo: 0

Léxico continuado:

Tipo:

ID de inflexão:

Especificação:

Tipo de inflexão:

ID do lema:

Afiliações ([adicionar](#)):

Processado: Não

Alterado recentemente: 6 de Agosto de 2020 às 22:02

Notas:

Metadados ([adicionar](#)):

## Exemplos ([adicionar](#)):

## Raízes ([adicionar](#)):

- 0 - laranja (abadia\_\_n) ([editar](#)/[excluir](#))
- 0 - laranja (anti\_\_adj) ([editar](#)/[excluir](#))

## Paradigmas curtos ([adicionar](#)):

Mostrar paradigmas curtos

## Relações ([adicionar](#)):

ID	Desde	Até	Tipo	Fontes	Exemplos	Metadados	Notas	Ações
1139182	(por) <a href="#">laranja</a>	(fin) appelsiini	Tradução	• <a href="#">adicionar</a>	• <a href="#">adicionar</a>	• <a href="#">adicionar</a>		• <a href="#">mostrar</a> • <a href="#">editar</a> • <a href="#">excluir</a>

[Retornar](#)

Ve'rdd

## Adicionar relação:

Desde: laranja

Até

fruta

des**frutado** () - 0 por

des**frutador** () - 0 por

des**frutar** () - 0 por

des**frutação** () - 0 por

**fruta** (N) - 0 por

**frutado** () - 0 por

[Salvar](#)

Retornar



# Adicionar relação:

Desde: laranja

Até

Type\*

Hyperonymy

Notes

Notas

Retornar



## Adicionar relação:

Desde: laranja

Até

fruta (N) - 0 por

Type\*

Hyperonymy

Notes

Notas

Salvar

• 0 = fruta (abadia\_11) ([editar](#)/[excluir](#))

[fronteira](#)

[frota](#)

[frouxidade](#)

[frugalidade](#)

[frugalismo](#)

[fruição](#)

[frumentação](#)

[frumento](#)

[frustração](#)

## Paradigmas curtos ([adicionar](#)):

## Relações ([adicionar](#)):

ID	Desde	Até	Tipo	Fontes	E
1094577	(apu) txipukury	<u>(por) fruta</u>	Tradução	<ul style="list-style-type: none"><li><a href="#">adicionar</a></li></ul>	
1139192	(por) laranja	<u>(por) fruta</u>	Hyperonymy	<ul style="list-style-type: none"><li><a href="#">adicionar</a></li></ul>	

Aprovar

ID	Desde	Até	Tipo	Fontes	Exemplos	Metadados	Notas	Ações
1094577	(apu) txipukury	<u>(por) fruta</u>	Tradução	<ul style="list-style-type: none"><li>• <a href="#">adicionar</a></li></ul>	<ul style="list-style-type: none"><li>• <a href="#">adicionar</a></li></ul>	<ul style="list-style-type: none"><li>• <a href="#">adicionar</a></li></ul>		<ul style="list-style-type: none"><li>• <a href="#">mostrar</a></li><li>• <a href="#">editar</a></li><li>• <a href="#">excluir</a></li></ul>

Desde	Até
<p>Lexeme: txipukury (<a href="#">mostrar</a>/<a href="#">editar</a>)</p> <p>ID: 674676</p> <p>Kieli (ISO 639-2): apu</p> <p>Classe gramatical: N</p> <p>ID do homónimo: 0</p> <p>Léxico continuado: N_SUFFIXES_01_Msc_re2/ry</p> <p>Tipo:</p> <p>ID de inflexão:</p> <p>Especificação:</p> <p>Tipo de inflexão: X</p> <p>ID do lema:</p> <p>Afiliações (<a href="#">adicionar</a>):</p> <p>Processado: Não</p> <p>Alterado recentemente: 6 de Agosto de 2020 às 23:03</p> <p>Notas:</p> <p>Metadados (<a href="#">adicionar</a>):</p>	<p>Lexeme: fruta (<a href="#">mostrar</a>/<a href="#">editar</a>)</p> <p>ID: 554277</p> <p>Kieli (ISO 639-2): por</p> <p>Classe gramatical: N</p> <p>ID do homónimo: 0</p> <p>Léxico continuado:</p> <p>Tipo:</p> <p>ID de inflexão:</p> <p>Especificação:</p> <p>Tipo de inflexão:</p> <p>ID do lema:</p> <p>Afiliações (<a href="#">adicionar</a>):</p> <p>Processado: Não</p> <p>Alterado recentemente: 4 de Maio de 2021 às 19:20</p> <p>Notas:</p> <p>Metadados (<a href="#">adicionar</a>):</p>

Relação:

[Retornar](#)

## Filtrar

Lexeme:

Língua:

apu

Classe gramatical:

N

Tipo de inflexão:

-----

Desde:

-----

Até:

-----

Processado:

-----

Pesquisar

ID

Lexeme

Classe gramatical

Léxico



[Retornar](#)

## Filtrar

Lexeme:

É idêntico

Língua:

Classe gramatical:

Tipo de inflexão:

Fonte:

Desde:

Até:

Processado:

Léxico continuado:

Ordem:

[Pesquisar](#)[Baixar](#)[Aprovar tudo](#)

ID	Lexeme	Classe gramatical	Léxico continuado	Tipo de inflexão	Língua	Notas	Ações
674622	hăkiti	N	N_SUFFIXES_01_Msc	X	apu		<ul style="list-style-type: none"><li><a href="#">mostrar</a></li><li><a href="#">editar</a></li><li><a href="#">excluir</a></li></ul>
674623	herĕkatxi	N	N_SUFFIXES_01_Msc_0/txi	X	apu	herĕka >> herĕkatxi	<ul style="list-style-type: none"><li><a href="#">mostrar</a></li><li><a href="#">editar</a></li><li><a href="#">excluir</a></li></ul>

[Retornar](#)

# Ve'rdd

## Filtrar

Lexeme:

Língua:

Classe gramatical:

Tipo de inflexão:

Fonte:

Desde:

Até:

Processado:

☒ -----  
Sim  
Não

Léxico continu

Pesquisar

Filtrar

Lexeme:

É idêntico

Língua:

apu

Classe gramatical:

N

Tipo de inflexão:

-----

Fonte:

Desde:

-----

Até:

-----

Processado:

-----

Léxico continuado:

Pesquisar

ID	Lexeme	Classe gramatical	Léxico continuado	Tipo de inflexão	Língua	Notas
674622	hãkiti	N	N_SUFFIXES_01_Msc	X	apu	
674623	herẽkatxi	N	N_SUFFIXES_01_Msc_0/txi	X	apu	herẽkatxi

- ✓ -----
- Lexeme
- Lexeme (Decrescente)
- Classe gramatical
- Classe gramatical (Decrescente)
- Léxico continuado
- Léxico continuado (Decrescente)
- Tipo de inflexão
- Tipo de inflexão (Decrescente)
- Consonância
- Consonância (Decrescente)
- Consonância (inverso)
- Consonância (inverso) (Decrescente)
- Assonância
- Assonância (Decrescente)
- Assonância (inverso)
- Assonância (inverso) (Decrescente)

Filtrar

Lexeme:

É idêntico

Língua:

apu

Classe gramatical:

N

Tipo de inflexão:

-----

Fonte:

Desde:

-----

Até:

-----

Processado:

-----

Léxico continuado:

Ordem:

-----

Pesquisar

Baixar ▾

Apri

ID	Lexeme	Classe gramatical	Léxico continuado	Tipo de inflexão	Língua	Notas
674622	hãkiti	N	N_SUFFIXES_01_Msc	X	apu	

- CSV
- LEXC

hãkiti+v1+N:hãkiti N\_SUFFIXES\_01\_Msc "" ;  
hãkiti+v2+N:ãkiti N\_SUFFIXES\_01\_Msc "" ;  
herẽkatxi+v1+N:herẽka N\_SUFFIXES\_01\_Msc\_0/txi "" ;  
herẽkatxi+v2+N:erẽka N\_SUFFIXES\_01\_Msc\_0/txi "" ;  
herẽkatxi+v3+N:arẽka N\_SUFFIXES\_01\_Msc\_0/txi "" ;  
amaru+N:amaru N\_SUFFIXES\_01\_Fem-ru\_te/0 "" ;  
amary+N:amary N\_SUFFIXES\_01\_Msc-ry\_te/0 "" ;  
kanawa+N:kanawa N\_SUFFIXES\_01\_Msc\_te/0 "" ;  
iuwata+N:iuwata N\_SUFFIXES\_01\_Fem\_ne/0 "" ;  
namatxi+N:nama N\_SUFFIXES\_01\_Msc\_0/txi "" ;  
ukypitẽtxi+N:ukypite N\_SUFFIXES\_01\_Msc\_0/Ntxi "" ;  
kiritãtxi+N:kirita N\_SUFFIXES\_01\_Msc\_0/Ntxi "" ;  
xinhĩtxi+N:xinhi N\_SUFFIXES\_01\_Msc\_0/Ntxi "" ;  
wakũtxi+N:waku N\_SUFFIXES\_01\_Msc\_0/Ntxi "" ;  
pitxĩtxi+N:pitxi N\_SUFFIXES\_01\_Msc\_0/Ntxi "" ;  
sukũtxi+Hom1+N:suku N\_SUFFIXES\_01\_Msc\_0/Ntxi "" ;

ID, Language, Lexeme, POS, Contlex, Inflex Type

674622, apu, hãkiti, N, N\_SUFFIXES\_01\_Msc, X

674623, apu, herẽkatxi, N, N\_SUFFIXES\_01\_Msc\_0/txi, X

creenshot

674624, apu, amaru, N, N\_SUFFIXES\_01\_Fem-ru\_te/0, X

674625, apu, amary, N, N\_SUFFIXES\_01\_Msc-ry\_te/0, X

674626, apu, kanawa, N, N\_SUFFIXES\_01\_Msc\_te/0, X

674627, apu, iuwata, N, N\_SUFFIXES\_01\_Fem\_ne/0, X

674628, apu, namatxi, N, N\_SUFFIXES\_01\_Msc\_0/txi, X

674629, apu, ukypitẽtxi, N, N\_SUFFIXES\_01\_Msc\_0/Ntxi, X

674630, apu, kiritãtxi, N, N\_SUFFIXES\_01\_Msc\_0/Ntxi, X

674631, apu, xinhĩtxi, N, N\_SUFFIXES\_01\_Msc\_0/Ntxi, X

674632, apu, wakũtxi, N, N\_SUFFIXES\_01\_Msc\_0/Ntxi, X

674633, apu, pitxĩtxi, N, N\_SUFFIXES\_01\_Msc\_0/Ntxi, X

674634, apu, sukũtxi, N, N\_SUFFIXES\_01\_Msc\_0/Ntxi, X



Source

Rich Text



Recompile



```

5858 \begin{dictionaryentry}
5859 \entry{vihreä}{A}{}
5860 \rawtext{(\textit{aineesta,
esineestä}) ruänn,}
5861 \rawtext{(\textit{kasveista})
ruônâs (ruôn'nâz, ruôn'nsa),}
5862 \rawtext{\textit{\MakeLowercase{näk
yä vihreänä}} ruân'n'jed (ij
ruôn'nu).}
5863 \end{dictionaryentry}
5864
5865 \begin{dictionaryentry}
5866 \entry{vihta}{N}{}
5867 \rawtext{västt (väästast, vasttu).}
5868 \end{dictionaryentry}
5869
5870 \begin{dictionaryentry}
5871 \entry{vihtoa}{V}{}
5872 \rawtext{västtad (väästam,
vasttum),}
5873 \rawtext{(\textit{itseään})
väästõõttâd (väästõõdam,
väästâätt).}
5874 \end{dictionaryentry}
5875

```

**vihreä** (*aineesta, esineestä*) ruänn, (*kasveista*) ruônâs (ruôn'nâz, ruôn'nsa), *näkyä vihreänä* ruân'n'jed (ij ruôn'nu).

**vihta** västt (väästast, vasttu).

**vihtoa** västtad (väästam, vasttum), (*itseään*) väästõõttâd (väästõõdam, väästâätt).

**vihuri** rëvvai (rëvjest, rëvja).

**vihurirokko** vihuriruäkkôs.

**viidakko** ju'ngel (ju'ngel, ju'ngla).

**viidakkoeläin** ju'ngelskoott (-skoottâst, -skoo'tte).

**viidakkomies** ju'ngeloomaž.

**viidennes** viidôs (viidõõzzâst, viidõ'sse).

**viides** viidad (viidee'st, viidää), *viides osa* viidôs (viidõõzzâst, viidõ'sse).

**viidesosa** viidôs (viidõõzzâst, viidõ'sse).

**viidesti** vi'ttešt.

**viidestoista** vittmlääggad.

48

---

viihde

**viihde** ääi'jääjtôs.

Lexemes  
anteriores

# Lexeme: hãkiti ([mostrar](#)/[editar](#))

ID: 674622

Kieli (ISO 639-2): apu

Classe gramatical: N

ID do homónimo: 0

Léxico continuado: N\_SUFFIXES\_01\_Msc

Tipo:

ID de inflexão:

Especificação:

Tipo de inflexão: X

ID do lema:

Afiliações ([adicionar](#)):

Processado: Não

Alterado recentemente: 6 de Agosto de 2020 às 23:03

Notas:

Metadados ([adicionar](#)):

## Exemplos ([adicionar](#)):

## Raízes ([adicionar](#)):

- 0 - hãkiti (N\_SUFFIXES\_01\_Msc) ([editar](#)/[excluir](#))
- 1 - ãkiti (N\_SUFFIXES\_01\_Msc) ([editar](#)/[excluir](#))

## Paradigmas curtos ([adicionar](#)):

[Mostrar paradigmas curtos](#)

## Relações ([adicionar](#)):

ID	Desde	Até	Tipo	Fontes	Exemplos	Metadados	Notas	Ações
1094892	( <a href="#">apu</a> ) <a href="#">hãkiti</a>	(eng) cougar	Tradução	<ul style="list-style-type: none"><li>• <a href="#">adicionar</a></li></ul>	<ul style="list-style-type: none"><li>• <a href="#">adicionar</a></li></ul>	<ul style="list-style-type: none"><li>• <a href="#">adicionar</a></li></ul>		<ul style="list-style-type: none"><li>• <a href="#">mostrar</a></li><li>• <a href="#">editar</a></li><li>• <a href="#">excluir</a></li></ul>



Notas:

Metadados ([adicionar](#)):

Exemplos ([adicionar](#)):

Raízes ([adicionar](#)):

- 0 - hãkiti (N\_SUFFIXES\_01\_Msc) ([editar/excluir](#))
- 1 - ãkiti (N\_SUFFIXES\_01\_Msc) ([editar/excluir](#))

Paradigmas curtos ([adicionar](#)):

Relações ([adicionar](#)):

ID	Desde	Até	Tipo
1094892	<u>(apu) hãkiti</u>	(eng) cougar	Tradução

# Paradigmas curtos (adicionar):

ID	Descrição morfossintática	Forma da palavra
-	N+Pl+Gen	appelsiinien
-	N+Pl+Nom	appelsiinit
-	N+Sg+Ill	appelsiiniin

Mostrar paradigmas curtos

# Ve'rdd & FLEx

FLEx has different export formats

- For example, LIFT XML file could be imported

It's possible to imagine a workflow where FLEx lexicon is exported and synchronized regularly into Ve'rdd, where more complex relations are stored

We have done this once with dictionaries stored in MediaWiki

We would be happy to experiment with a new language that has FLEx corpus!