



# The place of lexicography in (computer) science

Laurent Romary

## ► To cite this version:

Laurent Romary. The place of lexicography in (computer) science. The Future of Academic Lexicography: Linguistic Knowledge Codification in the Era of Big Data and AI, Frieda Steurs; Dirk Geeraerts; Niels Schiller; Marian Klammer; Iztok Kosem, Nov 2019, Leiden, Netherlands. hal-02358218

**HAL Id: hal-02358218**

**<https://inria.hal.science/hal-02358218>**

Submitted on 11 Nov 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# The place of lexicography in (computer) science

Laurent Romary, Inria, team ALMAAnaCH

# Overview

Understanding the role of lexicographic work in other scholarly fields:

- Dictionaries as primary sources in the humanities
- The CS perspective: from data modelling to data mining
  - Lexicography as a rich playground for data modelling
    - Current developments in international standardisation
  - Automatic analysis of legacy print dictionaries
    - The GROBID-dictionary experience
- Perspectives – automatic data enrichments

# Lexicographic works as a primary source

- Dictionaries integrate a wealth of linguistic information, but also represent a mirror of their times
  - The objective of comprehensiveness makes them essential primary sources of further humanities studies
- A wide range of possible reuse possibilities, illustrated through 2 examples:
  - Digitising the *Dictionnaire Universel* from Trévoux-Basnage (ANR project BASNUM)
  - The *Vocabulario en lengva misteca* at the service a language documentation project

# An encyclopaedic witness – digitising the 1701 *Dictionnaire Universel*

- The *Dictionnaire Universel* (DU), the first truly encyclopaedic dictionary
  - Covers general language, but above all terms from arts, crafts and sciences
  - Highly influential throughout Europe both directly and indirectly
- Antoine Furetière (1619-1688)
  - Ex-member of the *Académie française*, at loggerheads over his personal universal dictionary. The DU was published posthumously in 1690 in the Netherlands
  - Created the *Dictionnaire universel* as an encyclopaedic dictionary including all words used in France of his day
- Henri Basnage de Beauval (1657-1710)
  - A Protestant lawyer, son of a leading member of the Parliament of Normandy
  - Forced into exile after the Revocation of the Edict of Nantes that outlawed protestants
  - Succeeds the protestant philosopher Pierre Bayle as editor of a literary and philosophical journal - *Histoire des ouvrages des savants*
  - Engaged by Leers, publisher of the DU, to compile a revised and enlarged version, published 1701
  - Uses experts to write scientific entries

# The complex editorial history of the *Dictionnaire universel*

1690: first DU version, written and published by Antoine Furetière

- normative approach

*1701: second DU version, entirely revised by Basnage de Beauval and much augmented (1/3)*

- *descriptive approach*

1702: reprint of Basnage's version in 2 volumes

1708: second reprint (3 vols)

1725: new version (4 vols) revised by Brutel de La Rivière

**CARME.** f. m. Ordre de Religieux, qui est l'un des quatre Mendians, qui pretend tirer son nom du Mont Carmel en Syrie, qu'on dit avoir été habité par Elie. Ils ont été amenez en France par le Roi Louis IX. Il est celebre par la devotion du Scapulaire, & par la vision de Simon Stock, à qui il fut donné par la Sainte Vierge. Sur quoy de Launoy a écrit une curieuse Dissertation. La Vierge attacha ce privilege au Scapulaire, & à l'habit des *Carmes*, que ceux qui meurent le Samedi chargez de ces pieuses depouilles, sont exempts des flâmes du Purgatoire. Les *Carmes* se disent oncles de J. CHRIST, & freres de la Vierge. On dit les *Carmes du grand Couvent*; les *Carmes Mitigez*, qu'on nomme à Paris *Billetes*; & les *Carmes Dechaussez*, qui ont été reformez des autres. Dans des Theses soutenues à Beziers mentionnées dans le Journal de Hollande, on dit qu'il est fort probable que Pythagore étoit *Carme*, & que les Druides des Gaulois avoient aussi les observances regulieres des *Carmes*.

**MONT CARMEL**, est un Ordre Militaire de Chevaliers Hospitaliers, fondé par le Roi Henri IV. sous le titre, l'habit & la Regle de Nôtre Dame du *Mont Carmel*; & en consequence des Bulles du 16. Fevrier 1607. il a été uni à l'Ordre des Chevaliers de St. Lazare de Jerusalem, par acte du dernier Octobre 1608. avec toutes les Commenderies, Prieurez & autres biens pour sa dotation.

**CARME**, est aussi une espece d'acier. Voyez **ACIER**.

**CARME**, est aussi un vieux mot qui signifioit un vers. Il vient du Latin *carmen*; & en ce sens il est tout-à-fait hors d'usage.

<entry xml:lang="fre" xml:id="Carme">

<form><orth rendition="#uc">carme</orth>

<gramGrp><pos expand="Substantif">s.m.</pos></gramGrp></form>

<sense n="1">

<def> Ordre de Religieux, qui est l'un des quatre Mendians, qui pretend tirer son nom du Mont Carmel en Syrie, qu'on dit avoir été habité par Elie. </def>

<note> Ils ont été amenez en France par le Roi Louis IX. Il est celebre par la devotion du Scapulaire, & par la vision de Simon Stock, à qui il fut donné par la Sainte Vierge. Sur quoy de Launoy a écrit une curieuse Dissertation. La Vierge attacha ce privilege au Scapulaire, & à l'habit des <hi rendition="#i">Carmes</hi>, que ceux qui meurent le Samedi chargez de ces pieuses depouilles, sont exempts des flâmes du Purgatoire. Les <hi rendition="#i">Carmes</hi> se disent oncles de <name ref="#Jesus\_Christ\_ISN0000000120370699">

>J.Christ</name>, & freres de la Vierge. On dit les <seg

type="terme\_variant" rendition="#i">Carmes du grand

Couvent</seg> ; les <seg type="terme\_variant" rendition="#i">

>Carmes Mitigez</seg>, qu'on nomme à Paris <hi

rendition="#i">Billetes</hi> ; & les <seg

type="terme\_variant" rendition="#i">Carmes Dechaussez</seg>,

qui ont été reformez des autres. Dans des Theses soutenues à Beziers mentionnées dans le Journal de Hollande, on dit qu'il est fort probable que Pythagore étoit <hi rendition="#i">

>Carme</hi>, & que les Druides des Gaulois avoient aussi

les observances regulieres des <hi rendition="#i">

>Carmes</hi>.</note></sense>

<sense n="2"><form><orth rendition="#sc">mont carmel</orth>,</form>

<def> est un Ordre Militaire de Chevaliers Hospitaliers, fondé par le Roi Henri IV, sous le titre, l'habit & la Regle de Nôtre Dame du Mont Carmel ; & en consequence des Bulles du 16. Fevrier 1607. </def>

<note>Il a été uni à l'Ordre des Chevaliers de St.Lazare de Jerusalem, par acte du dernier octobre 1608. avec toutes ses

# Understanding the dictionary: some humanities research questions

- What source texts? Dictionaries and other texts?
- What links between the DU and *Histoire des ouvrages des savants*
- What lexicographical model between prescription and description?
  - the place of 'good' usage and the 'best ' authors
  - the dictionary as a language teaching tool – what users? what means?
  - The characterisation of terms
  - The role of contemporary scientific and literary networks
- To what degree the 1701 DU was theologically a "protestant dictionary"
- What changes were brought in between 1690 and 1701, between 1701 and 1725/27, and between 1701 and the 1704 Trévoux
- Who authored which entries in 1701? - Basnage and his specialist informers.

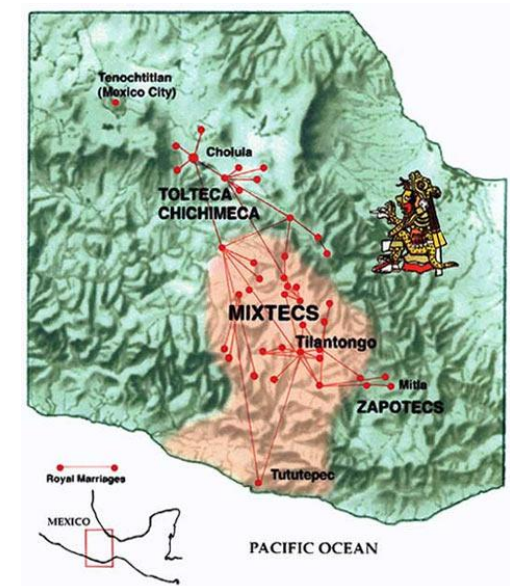


# Linguistic description of Mixtepec-Mixtec

- Sa'an Savi “*rain language*”
- ISO 639-3 code: ‘mix’
- San Juan de Mixtepec - Juxtlahuaca district (Oaxaca, MEX)
- “Vigorous” status but highly under-resourced
- Oto-Manguean, Mixtecan, Mixtec-Cuicatec, Mixtepec-Mixtec
- Tonal
- Spoken data mostly collected in sessions working with speakers from a small village called Yucunani in the San Juan Mixtepec municipality
- Estimated (+-9,000 -10,000 speakers)

*Source: (INEGI, 2010)*

- Phonology has been studied by Pike and Ibach (1978); Paster and Beam de Azcona (2004-2007);
- Beckman and Nieves-SIL (2005-current) published booklets and are working on developing orthography



# The research project

- Language documentation: (Jack Bowers' *PhD*)
- Primary sources of language data:
  - Speaker consultations (recordings, new written material..)
  - +- 40 Children's Booklets (SIL)
  - Public sources (YouTube, *other*)
  - Examples from academic papers
- Goals:
  - TEI Corpus
  - Linguistic descriptions
  - TEI Dictionary (*actually 2 dictionaries, 1 general, 1 inflectional*)
  - (*Etymology*) would like to create data contents and structure that can be copied and integrated into treatment of related languages

# Overview of the Source & Output

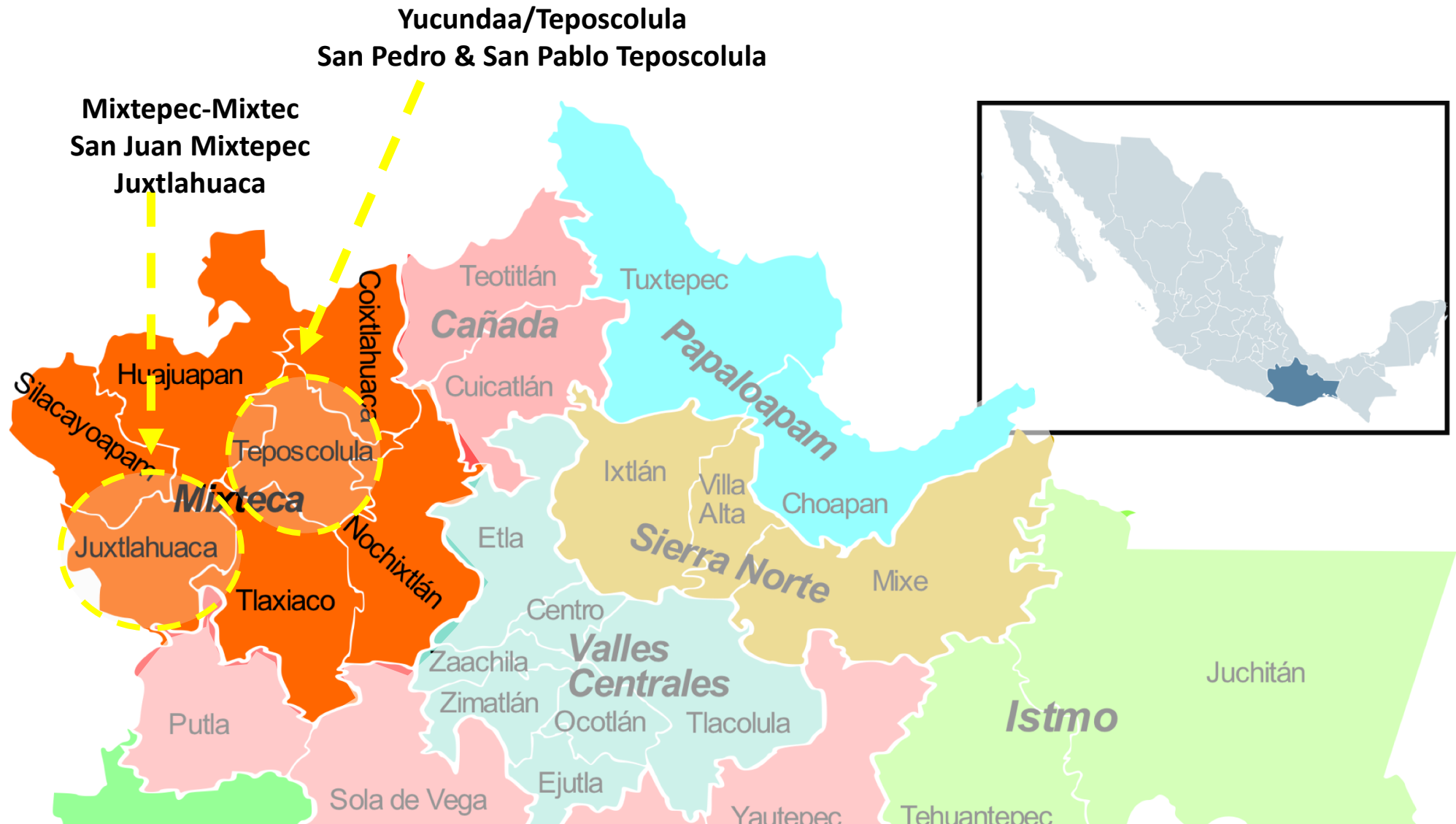
- ‘Vocabulario en lengua misteca’ published by the Dominican Francisco de Alvarado (1593)
- Variety from Teposcolula Mexico (Mixteca Alta)
  - Classical Mixtec/Colonial Mixtec/Yucu Ndaa
- Entries based on three earlier dictionary sources:
  - Castilian-Nahuatl (Valley of Mexico, 1571)
  - Castilian-Zapotec (Valley of Oaxaca, 1578)
  - Castilian-Latin (1492)
- PDF re-organized, modernized version ‘Voces de Dzaha Dzahui’ (Jansen & Pérez Jiménez, 2009)
- TEI dictionary produced contains roughly 26,600 entries and related entries.

# Versions of the resource

- Original (Printed: 1593) > (facsimile edition 1965)
- Mesolore (Bakewell & Hamman, 2001)
  - Digitized from scanned copy
- Jansen and Pérez Jimenez 2009

A. ANTE B.					
<div><div><div>A</div></div><div>PRIMERA letra del a. b. c. A. preposición por cerca, apud. nuu como nuuyahui. A. preposición por hazia. nuu quaha dzúhua, A. adverbio de llamar. dzi. A. adverbio para llamar cuándo se me ha olvidado algo. dzi. vt. na hadzi. nahacadzi. A. interiection. del que se queixa de alguna enfermedad. Hmij. A. interiection desiderantis. ha. A. interiection ridentis. ha ha ha. A. interiection admirantis. a a. a. hi. hi. hi. A. del que halla a otro en maleficio. a. a. A. alguna parte o en alguna parte. adverbio. huadza, huadzaca. Abad. Prelado. dzutu. yodadzi siñañu, dzutu ñohofina ñuhu, dzutu yocuvui dzini, yocuvui nuu. Abad. fer. yondadzi siñañu, f. cond. ñohondi siñañu. f. coho. yocuvuidzini. yocuvui nuundi. Abadia dignidad. f. indadzi siñañu hu, f. siñohu siñañu, f. sasicuvui nuu. Abadesa, prelada de monjas. ñaha yondacañaha yuq dzehe. ñaha-yocuvuidzini. Abadejo efcaraajo. tenuyuq. Abahar algo con el huelgo. yochidzo yoco yuhundi yotaa yoco yuhundi.</div></div>	<div><div>Abahar así cosa honda. yotniño yoco yuhundi. Abahada cosa así. f. así cohondaa yoco. f. sinucoho yoco yuhu sanisa codzo yoco yuhu. Abahar algo poniendolo al baho q fale dela olla. yodzaqnindi. yodzacuidzindi yoco. Abahar algo dentro de alguna olla para que se cōserue con el baho y no se pudra. yodzacaanuoyocondi. Abahar fopas. yodzachitundi yoco yotniño yocondi. yodzacutu yocondi. Abahadas fopas. dzita ninucoho yoco. Abahar por hechar el huelgo de la boca. yodzacaindi yoco. yodzacañadi yoco. yodzazqhūdi yoco. Abahar aposento. yodzacuidzindi huahi. yotniño yocondi. yondadzacuidzindi. yondadzañindi. Abahado aposento. huahi ninucuiñe huidzi. huahi ninucoho yoco. huahi ninacuidzi. huahi ninduvui ini. Abalançarfe, hechándose por los suelos. yocoo cavuan dayedi, yosacavuandayendi. f. qcavua. Abalançarfe de arriba a baxo. yon diyoninondi. Abalançarfe metiendose entre otros. yofanudzavuadi. f. qnanu. yofanandzavuandi. f. qnanu, yofivui cuinomañundi. yocai nānu dzavuandi. f. qivui. yosivui utnahandi. Aballar mouer cō difficultad. duz</div></div>	<div><div>folio 1r column 1</div><div>PRIMERA letra del a. b. c. A. preposicion por cerca, apud. nuu como nuuyahui. A. preposicion por hazia. nuu quaha dzuhua. A. adverbio de llamar. dzi. A. adverbio para llamar quando se me ha olvidado algo. dzi. vt. nahadzi. nahacadzi A. interiection. del que se queixa de alguna enfermedad. Hmii. A. interiection ridentis. ha ha ha. A. interiection admirantis. a a. a. hi. hi. hi. A. del que halla a otro en maleficio. a. a. A. alguna parte o en alguna parte. adverbio. huadza, huadzaca. Abad. Prelado. dzutu. yodadzi siñañu, dzutu ñoho siñañu, dzutu yocuvui dzini, yocuvui nuu. Abad. ser. yondadzi siñañu, f. cond. ñohondi siñañu. f. coho. yocuvuidzini. yocuvui nuundi. Abadia dignidad. sasindadzi siñañu, f. sasiño siñañu, f. sasicuvui nuu. Abadesa, prelada de monjas. ñaha yondacañaha yuq dzehe. ñaha yocuvuidzini. Abadejo escaraajo. tenuyuq. Abahar algo con el huelgo. yochidzo yoco yuhundi yotaa yoco yuhundi.</div></div>	<div><div>folio 1r column 2</div><div>Abahar así cosa honda. yotniño yoco yuhundi. Abahada cosa así. sasi cohondaa yoco. sasinucoho yoco yuhu sanisa codzo yoco yuhu. Abahar algo poniendolo al baho que sale de la olla. yodzaqnindi. yodzacuidzindi yoco. Abahar algo dentro de alguna olla para que se cōserue con el baho y no se pudra. yodzacaanuoyocondi. Abahar fopas. yodzachitundi yoco yotniño yocondi. yodzacutu yocondi. Abahadas fopas. dzita ninucoho yoco. Abahar por hechar el huelgo de la boca. yodzacaindi yoco. yodzacañadi yoco. yodzazqhūdi yoco. Abahar aposento. yodzacuidzindi huahi. yotniño yocondi. yondadzacuidzindi. yondadzañindi. Abahado aposento. huahi ninucuiñe huidzi. huahi ninucoho yoco. huahi ninacuidzi. huahi ninduvui ini. Abalançarfe, hechándose por los suelos. yocoo cavuandayendi, yosacavuandayendi. f. qcavua. Abalançarfe de arriba abaxo. yon diyoninondi. Abalançarfe metiendose entre otros. yosanudzavuandi. f. qnanu. yosanudzavuandi. f. qnanu, yosivui cuinomañundi. yocai nānu dzavuandi. f. qivui. yosivui utnahandi. Aballar mouer con difficultad. duz</div></div>	<div><div>A</div><div>a a: a (del que halla a otro en maleficio) a a a: a (interiection admirantis) a dzuchica añandaa: poco más o menos a dzuchica caa cuvui: poco más o menos a dzuchica coo cuvui: poco más o menos a hua dzevui: o no a hua dzevui dzavua: pues no a huilica añandaa: poco más o menos a huitnani: ahora poco a na ndehe cuvui ndatu nicay: o bienaventurado, o dichoso a ñaha: o no; pues no a sa dzevui: pues no a yoo: por ventura alguno a yoo ee ñahando: por ventura alguno de vosotros aa: de manera que aa: ya, acordándoseme lo que se me había olvidado aa dzuhua huii: así, así (sonriéndose) aa ndica huii: ya (acordándoseme lo que se me había olvidado) adzi: o (disyuntiva); por ventura; quizás adzi: suave cosa adzi cuvui: o (disyuntiva); por ventura; quizás adzi q cuvui: por ventura adzi yoo: por ventura alguno adzi yoo ee ñahando: por ventura alguno de vosotros ahua: ay, quejándose la mujer ama: así</div></div>	<div><div>ama: bien está (otorgando); si amana: ¿cuándo? (adverbio interrogativo), ¿en qué tiempo? amana cuiya: en algún tiempo amana cuvui huatu inindo: cuandoquiera que quisieres amana na ndita ñumana nuundo: ¿cuándo has de despertar? amana na ndotondo: ¿cuándo has de despertar? amana na tahui inindo: ¿cuándo has de volver en ti?; ¿cuándo has de despertar? amana quevui: en algún tiempo amana quevui sa cuvui inindo: cuandoquiera que quisieres amanaca: ¿cuándo, en qué tiempo? amani: de tarde en tarde o raras veces andaya: infierno, lugar de dañados andevui: cielo andevui isi ndaa tiño: cielo estrellado angel nicoo coo ndaa ndita ña: ángel de mi guarda angel yondaca ñaha: ángel de mi guarda aniñe: palacio anuhu: abismo; centro de la tierra; infierno, lugar de dañados anuhu: profundo anuhu maa: profundo anuhu naa: infierno, lugar de dañados anuhu ndahui: infierno, lugar de dañados atana: ojalá atu: amarga cosa; áspero al gusto aya: amarga cosa; áspero al gusto</div></div>

# La Mixteca (Mixtec Region)



# Utility/Purpose of Endeavour

- Increase coverage of relevant lexical material in Mixtepec-Mixtec documentation (ISO 639-3 [mix])
  - Link and cross-reference in Mixtepec TEI dictionary
- Machine searchable data set for:
  - Study of the Yucu Ndaa variety
  - Historiographical and philological research
- Create a more cohesive body of pan-Mixtecan resources
  - Vocabulary for cross Mixtecan comparison; (81 Varieties of Mixtec)
- TEI format can easily be exported into other formats for non-TEI users



# Integration into Mixtepec-Mixtec Project:

## TEI Structure of Output

- Goal to match the structure used in the Mixtepec-Mixtec TEI dictionary (Bowers & Romary 2018)

```
<entry xml:id="fruit-plantain">
  <form type="lemma">
    <orth xml:lang="mix">nchika</orth>
    <pron xml:lang="mix" notation="ipa">nɔ̌jiká</pron>
  </form>
  <gramGrp>
    <pos>noun</pos>
  </gramGrp>
  <sense corresp="http://dbpedia.org/resource/Plantain">
    <usg type="domain">Fruit</usg>
    <cit type="translation">
      <form>
        <orth xml:lang="en">plantain</orth>
      </form>
    </cit>
    <cit type="translation">
      <form>
        <orth xml:lang="es">plátano</orth>
      </form>
    </cit>
  </sense>
</entry>
```

...

Mixtepec-Mixtec

**nchika** [nɔ̌jiká] (*noun*)  
[FRUIT] plantain, plátano

```
<entry xml:id="plátano">
  <form type="lemma">
    <orth>chita</orth>
  </form>
  <gramGrp>
    <pos>noun</pos>
  </gramGrp>
  <sense corresp="http://dbpedia.org/resource/Plantain">
    <usg type="domain">Fruit</usg>
    <def xml:lang="es">plátano</def>
    <def xml:lang="en">plantain</def>
  </sense>
</entry>
```

**chita** (*noun*)  
[FRUIT] plantain, plátano

Classical Mixtec

Going further: modelling and  
standardising lexical resources



# Lexical resources in their varieties

- A variety of contexts and forms
  - Legacy dictionaries, dialectological studies, NLP lexica
  - Full form, etymology, corpus based research
  - Word document, database, shoebox, proprietary XML...
  - Lexical vs. Editorial views
  - Onomasiological vs. semasiological structures

# Lexicography or terminology

- Lexicography
  - Generic view on “words”
  - Attempt to provide a large coverage of a language
  - Semasiological view
    - Word > meaning(s)
- Terminology
  - *Term*: form associated to a specific concept within a given domain
  - Onomasiological view
    - Concept > various possible linguistic forms
- Depends on available data, objectives and user scenarios

# Comparing approaches

## **Semasiological approach**

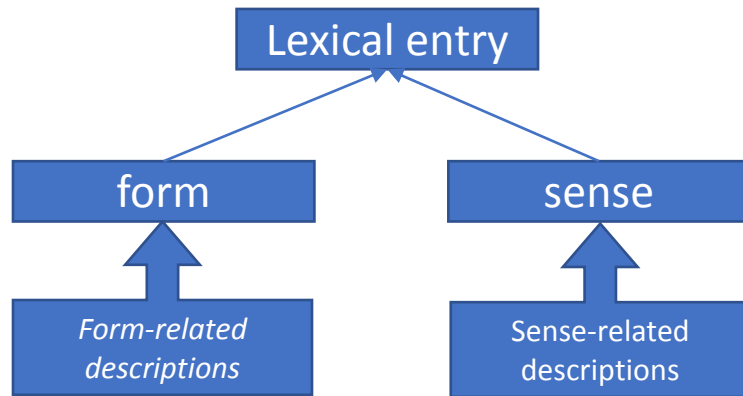
- Large coverage
- All parts of speech
- Build-in polysemy
  - Multiple senses for the same entry
- Referential synonymy

## **Onomasiological approach**

- Domain oriented
- Essentially nouns
  - Extension to verbs, adjectives
- No polysemy (needs to be reconstructed)
- Build-in synonymy
  - Multiple terms for the same concept

# Basic modelling of lexical components

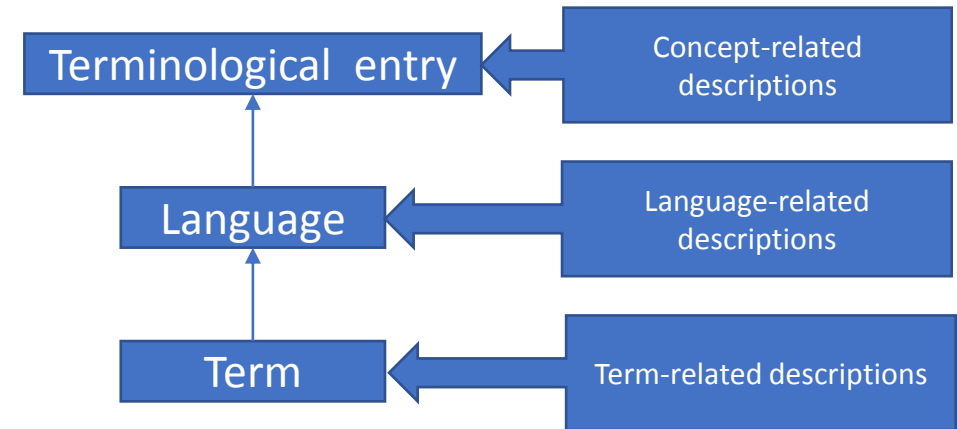
## Semasiological models



### Issues

- Various levels/sensibilities in entry groupings: homonyms, families (e.g. roots)
- Providing a neat way of representing lexical dependencies: from “see also” to multi-word expressions

## Onomasiological models



### Issues

- Representing conceptual relations between entries
- Providing fine-grained semantic information at term level (e.g. usage, translation equivalents)

### Issues

- Making the appropriate choice of model
- Integrating information between the two types of models

# Why standardizing all this?

- Defining methods, models and format to facilitate
  - Exchange of lexical data
  - Pooling heterogeneous lexical data
  - Interoperability between software components
    - Search engines, layout, extraction of linguistic properties
  - Comparability of results
    - E.g. Linguistic coverage of lexical databases
  - Exchange of ideas within a community with a common background

# Standardization initiatives for lexical/terminological resources

- TEI
  - P5 edition of the guidelines
    - Cf. specification platform (ODD)
    - Dictionary chapter
    - Former terminology chapter (ancestor of TBX)
- ISO
  - ISO/TC 37: Language and terminology
    - ISO/TC 37/SC 3: ISO 16642 (TMF), ISO 30042 (TBX)
    - ISO/TC 37/SC 4: ISO 24613 (LMF)
- W3C
  - SKOS, Ontolex

# In the beginning



Text archives  
Humanities  
Standards  
SGML

*Not intended  
(immediately)  
for individual  
scholars*

*1. Novembre 1987: Vassar  
College, Poughkeepsie*

# A quick historical overview

- 1960's — GML (Generalized Markup Language) by IBM
- 1970's & 1980's — ANSI initiates project to develop a Standard text-description language based on GML
- 1983 — SGML becomes an industry standard
- 1986 — SGML (Standard Generalized Markup Language) becomes an ISO standard: ISO 8879:1986
- **1987 — TEI (Text Encoding Initiative)**
- 1990 — HTML 1.0 (HyperText Markup Language)
- 1992 — TEI edition P3 (Michael Sperberg-McQueen and Lou Burnard, eds)
- 1997/1998 — XML 1.0 (eXtensible Markup Language) (Tim Bray, Jean Paoli and Michael Sperberg-McQueen, eds)



# The TEI Dictionary chapter

- Initially designed within a working group lead by N. Ide and J. Veronis
- Accounts for both presentational and database views
  - Cf. <entry>, <entryFree>, ... and <dictScrap>
- Based on a hierarchical abstract model (crystals)
  - <form>: for characterizing the orthographic or phonetic form of the word
    - <orth>, <pron>, etc.
  - <gramGrp>: grammatical features
    - May characterize an entry, a specific form or a specific sense
    - <pos>, <gen>, generic <gram> feature
  - <sense>: iterative and recursive
    - May contains definitions, examples, etymological information, translations, etc.

# IRL: Petit Larousse illustré (1906)

```
<entry xml:id="pléthore" n="1906-001_unknown">
  <form type="lemma"><orth>PLÉTHORE</orth></form>
  <gramGrp><pos expand="nom">n.</pos>
    <gen expand="féminin">f.</gen></gramGrp>
  <etym><pc>(</pc>du <lang expand="grec">gr.</lang>
    <mentioned>plêthorê</mentioned><pc>,</pc>
    <gloss>plénitude</gloss><pc>)</pc><pc>.</pc></etym>
  <sense><def>Surabondance de sang, d'humeurs</def><pc>.</pc></sense>
  <sense><usg type="style" rend="italic" expand="figuré">Fig.</usg>
    <def>Surabondance quelconque amenant un état fâcheux</def>
    <pc>:</pc>
    <cit type="example">
      <quote>la pléthore des capitaux cause la diminution du taux de l'intérêt</quote>
    </cit><pc>.</pc></sense>
</entry>
```

# Advantages of being in the TEI framework

- Benefitting from the TEI environment
  - TEI Modelling Language: ODD
    - Customizing the guidelines within a project (e.g. restraining possible values)
  - Availability of a wealth of additional elements (~600)
    - E.g. annotating textual content, reflecting the specificities of the source etc.
- Standardisation reactivity
  - Issuing GitHub tickets for resolving bugs or introducing new features
- A community of experts
  - Support through the mailing list
- Main characteristic (drawback?): +very+ flexible

# Going ISO to provide a stable background

- Advantages of going ISO
  - International approval of ISO members
    - And international expert participation by construction
  - Stable background that is easy to reference (and known by third parties, non linguistic geeks, our institutions etc.)
- From a lexical point of view
  - Providing a generic model, independently of any specific implementation/serialisation
    - Stabilizing concepts, constraints and vocabulary
  - With the on-going LMF revision: Introducing the TEI as one possible model

# LMF

ISO 24613:2008 Language resource management — Lexical markup framework (LMF)

- Developed within ISO TC37/SC4/WG4
  - TC 37: Language and terminology
  - TC37/SC4/WG4: Lexical resources
- Shortcomings
  - Bulky: a single document with annexes
    - Major hindrance to revision
  - Complex modelling
    - Complex relationships between classes, redundant mapping mechanisms
  - Complex and ad hoc serialisation
  - Does not cover prominent information: Etymology and Diachrony



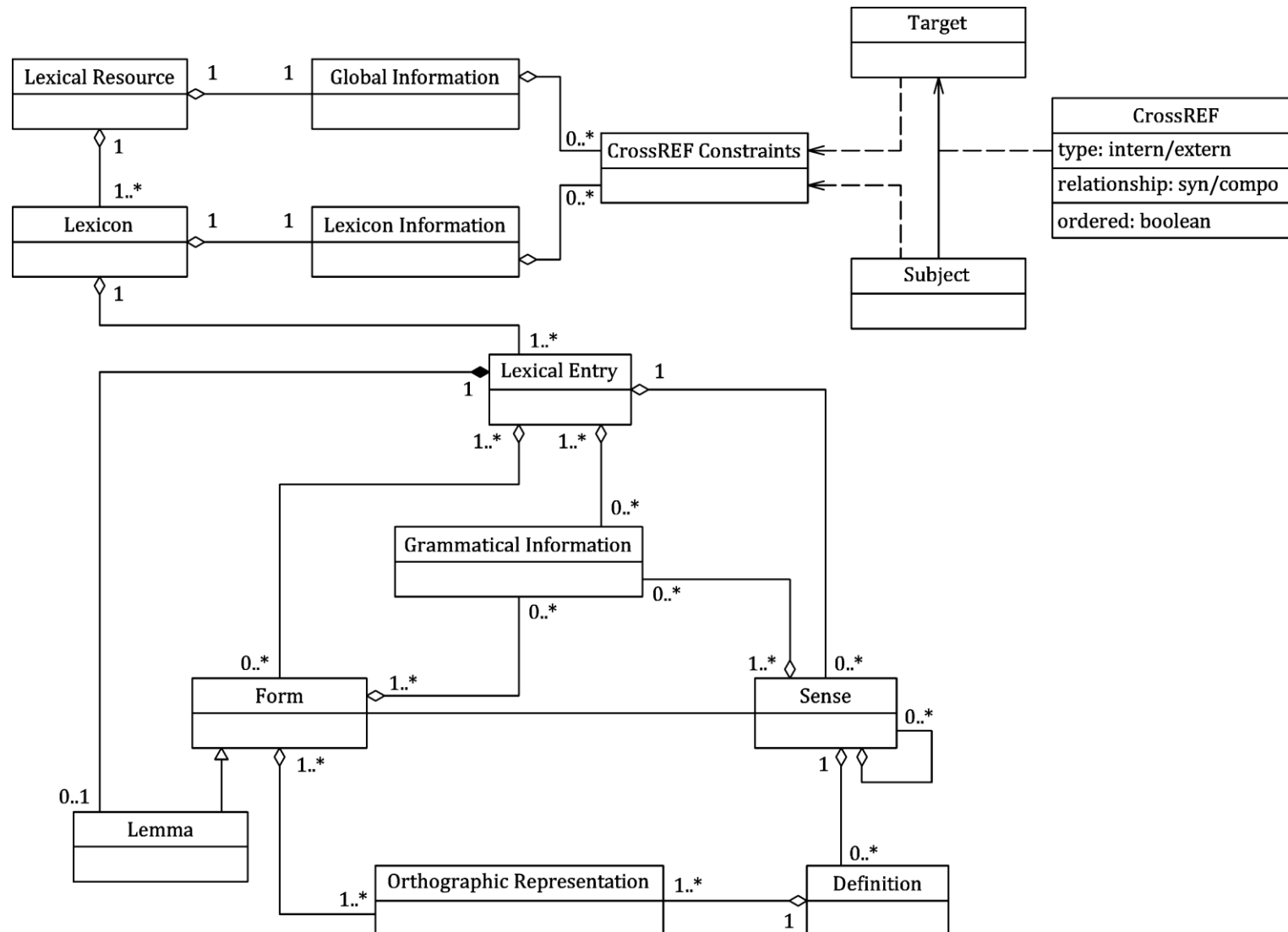
Laurent Romary, Mohamed Khemakhem, Fahad Khan, Jack Bowers, Nicoletta Calzolari, et al.. LMF Reloaded. *AsiaLex 2019: Past, Present and Future*, Jun 2019, Istanbul, Turkey. [hal-02118319](#)

# LMF Reloaded: Abstract Modelling

Restructuring: Multi-part standard

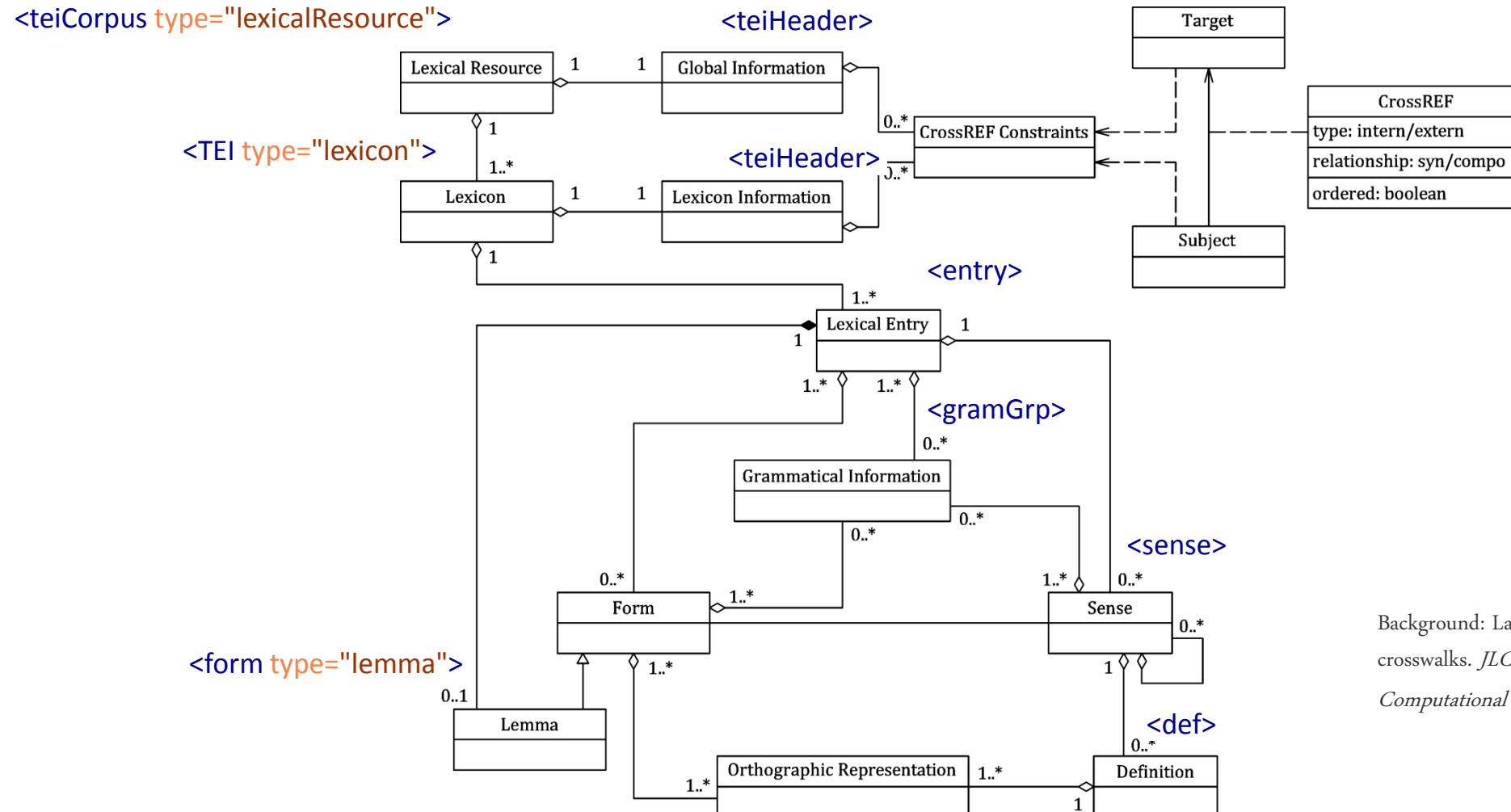
- ISO 24613-1 - Core model (published in June 2019)
- ISO 24613-2 - Machine Readable Dictionaries (MRD) model
- ISO 24613-3 - Diachrony-Etymology
- ISO 24613-4 - TEI serialisation
- ISO 24613-5 - LBX serialisation
- ISO 24613-6 - Syntax and Semantics
- ISO 24613-7 - Morphology

# Language resource management — Lexical markup framework (LMF) — Part 1: Core model



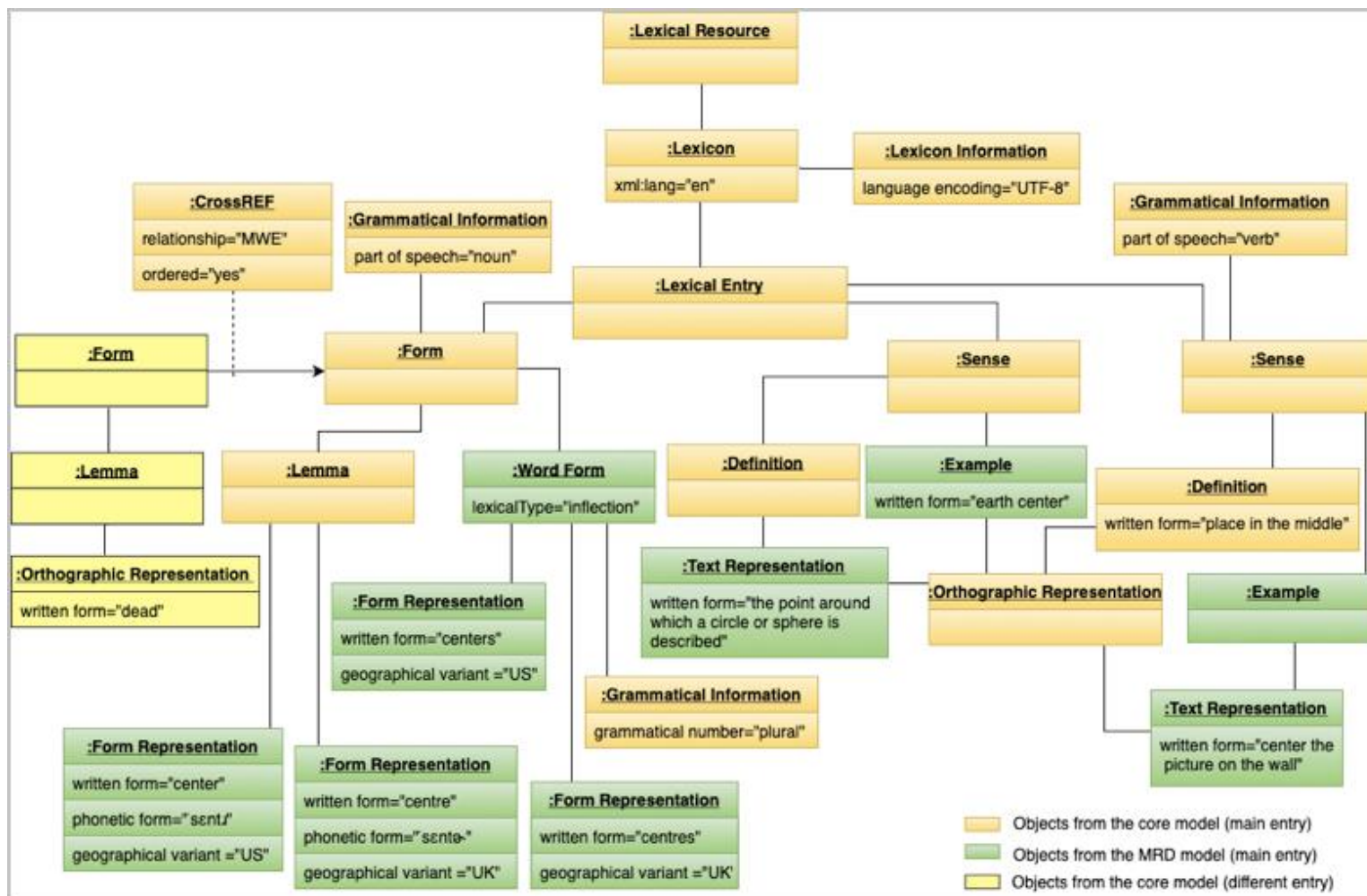


## From LMF to TEI – serialising one with the other (Part 4 – TEI serialisation)



Background: Laurent Romary. TEI and LMF crosswalks. *JLCL - Journal for Language Technology and Computational Linguistics*, 2015, 30 (1). [hal-00762664v4](#)

# Example: inflectional (full-form) lexicon



```

<entry>
  <form type="lemma" xml:id="center_form">
    <orth>center</orth>
    <pron>'sentʃ</pron>
    <gramGrp>
      <pos>noun</pos>
    </gramGrp>
    <usg type="geo">U.S</usg>
    <form type="variant">
      <orth>centre</orth>
      <usg type="geo">U.K</usg>
      <pron>'sentə</pron>
    </form>
  </form>
  <form type="inflected">
    <orth>centers</orth>
    <usg type="geo">U.S</usg>
    <gramGrp>
      <number>plural</number>
    </gramGrp>
  </form>
  <form type="inflected">
    <orth>centres</orth>
    <usg type="geo">U.K</usg>
    <gram type="number">plural</gram>
  </form>
  <sense>
    <def>the point around which a circle or sphere is described</def>
    <cit type="example">
      <quote>earth center</quote>
    </cit>
  </sense>
  <sense>
    <gramGrp>
      <pos>verb</pos>
    </gramGrp>
    <def>place in the middle</def>
    <cit type="example">
      <quote>center the picture on the wall</quote>
    </cit>
  </sense>
  <re type="multiWordExpression">
    <form>
      <seg corresp="#dead_form" n="1">dead</seg>
      <seg corresp="#center_form" n="2">center</seg>
    </form>
  </re>
</entry>
  
```

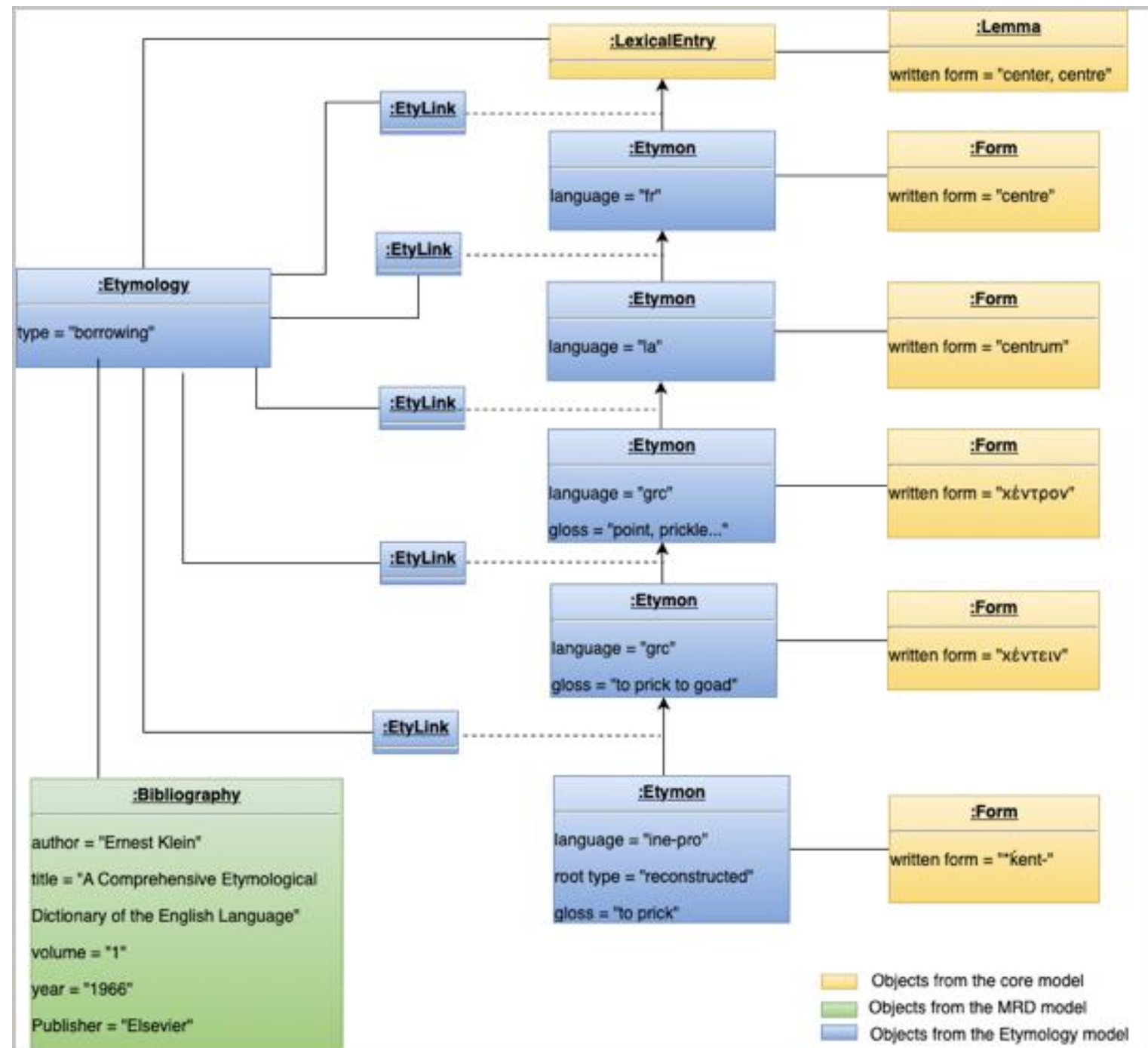
# LMF reloaded - Etymology

ISO 24613-3, prepared for DIS ballot (as of November 2019)

new classes: Etymology, Etymon, Cognate and EtymLink

**center, centre, n.** — F. *centre*, fr. L. *centrum*, fr. Gk. κέντρον, ‘point, prickle, spike, ox goad, point round which a circle is described’, from the stem of κέντειν, ‘to prick, goad’, whence also κέντωρ, ‘a goader, driver’, κεστός (for \*κεντ-τός), ‘embroidered’, κέστρον, ‘pickaxe’, κοντός, ‘pole’, fr. I.-E. base \**kent-*, ‘to prick’, whence also Bret. *kentr*, OIr. *cinteir*, ‘a spur’, OHG. *hantag*, ‘sharp, pointed’, Lett. *sīts*, ‘hunter’s spear’, *situ*, *sist*, ‘to strike’, W. *cethr*, ‘nail’. Cp. **centrifugal**, **centripetal**, **concentrate**, **eccentric**, **Dicentra**, **paracentesis**. Cp. also **cestrum**, **cestus**, ‘girdle’, **kent**, ‘a pole’, **quant**, ‘a pole’.  
**Derivatives:** *center*, *centre*, intr. and tr. v., *center-ing*, *centr-ing*, *centre-ing*, n.

Source: Klein’s *Comprehensive dictionary of the English Language*



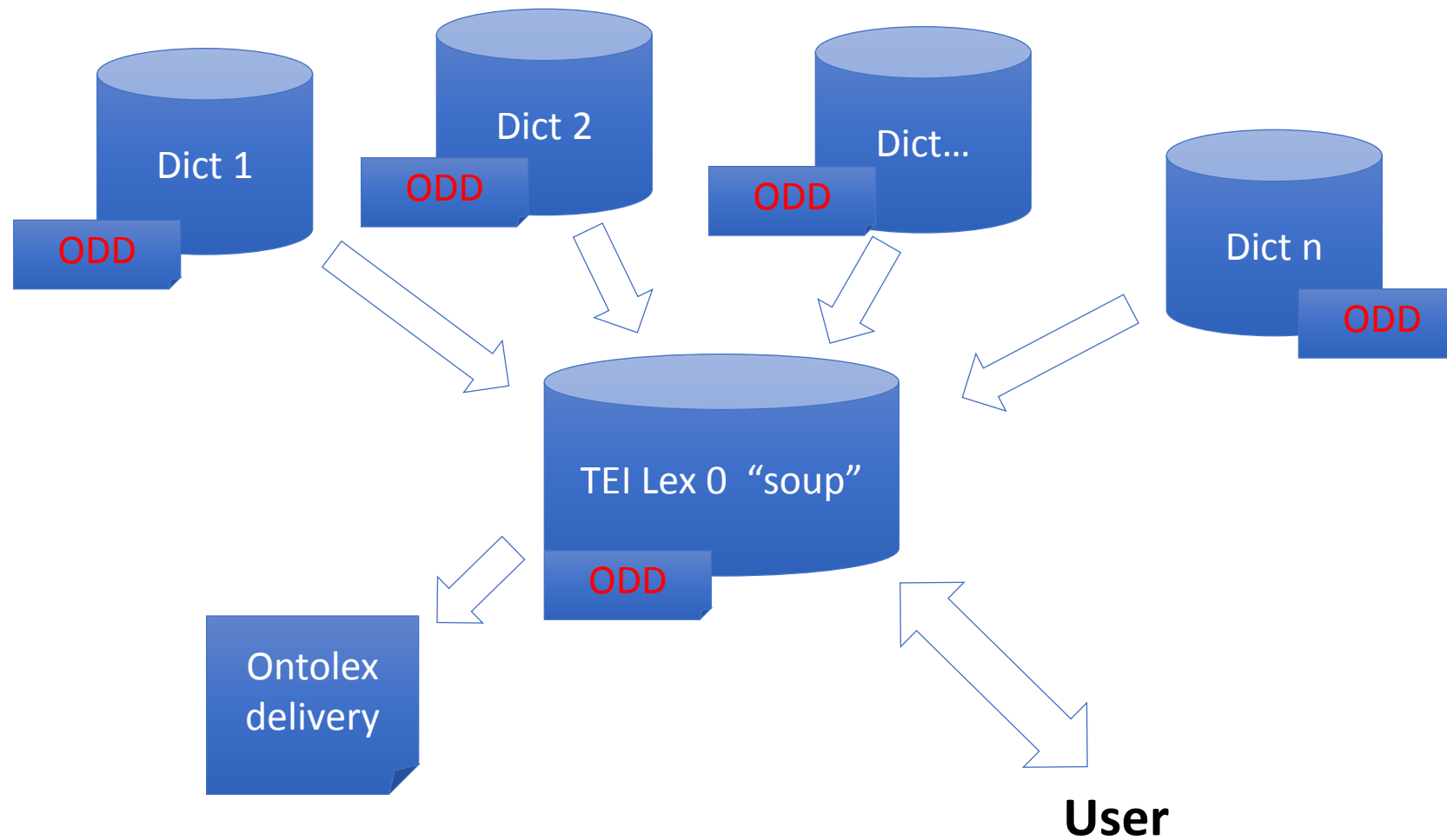
# TEI lex 0: tightening the guidelines

- Initiative set up in the context of the DARIAH working group on lexical resources, supported by the EU project Elexis
- Objectives
  - Designing a target format for heterogeneous lexical data integration
- Trade-off
  - Compliance with the standard
  - Fine-tuning to the needs of a specific context/scenario
- Back to what a standard is: a common reference for a transaction
  - Perfect to say: “I am compliant to standard X except for...”
  - The TEI guidelines provides the means to carry out such customizations

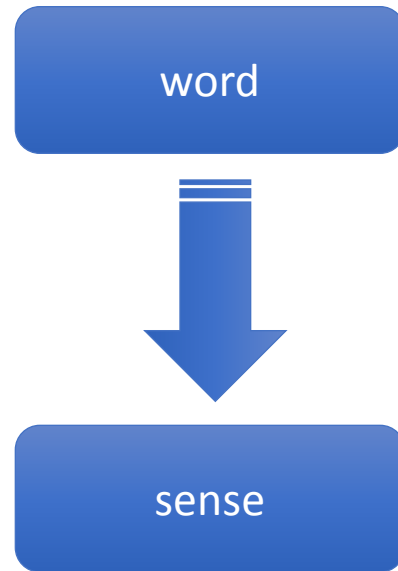
# Application – the ELEXIS project

- European Lexicographic Infrastructure
  - 2018-02-01 – 2022-01-31
  - “integrate, extend and harmonise national and regional efforts in the field of lexicography”
    - Focuses on “efficient access”
  - Cooperation with CLARIN and DARIAH for long-term sustainability
- Lexical formats and standards in ELEXIS
  - Double sided approach TEI – Ontolex
  - TEI Lex 0 specification at the core of the data hub

# The ELEXIS centralized hub



# Enforcing the semasiological model



```
<entry>  
  <form type="lemma">  
    ...  
  </form>  
  <sense>  
    <def>...</def>  
  </sense>  
</entry>
```

# Simplifying the dictionary micro-structure

- Current situation
  - Containing vs. contained entries
    - <superEntry> – <entry> – <re>
  - Structured vs. unstructured entries
    - <entry> – <entryFree>
- The TEI Lex-0 vision
  - Representing all entry-like objects as <entry>
    - Making <entry> recursive
    - Making more use of <dictScrap>



# Recursive entry

## - example

```
<entry type="wordFamily">
  <form type="base">
    <orth>Haus-</orth>
  </form>
<pc>,</pc>
  <form type="base">
    <orth>haus-</orth>
  </form>
<pc>:</pc>

<!-- possibly some shared usg information -->
<entry type="wordForm">
  <form type="lemma">
    <orth expand="Hausaltar">-altar</orth>
    <pc>,</pc>
    <gramGrp>
      <gen value="masculine">der</gen>
    </gramGrp>
  </form>
  <sense>...</sense>
</entry>
<entry type="wordForm">
  <form type="lemma">
    <orth expand="Hausandacht">-andacht</orth>
    <pc>,</pc>
  </form>

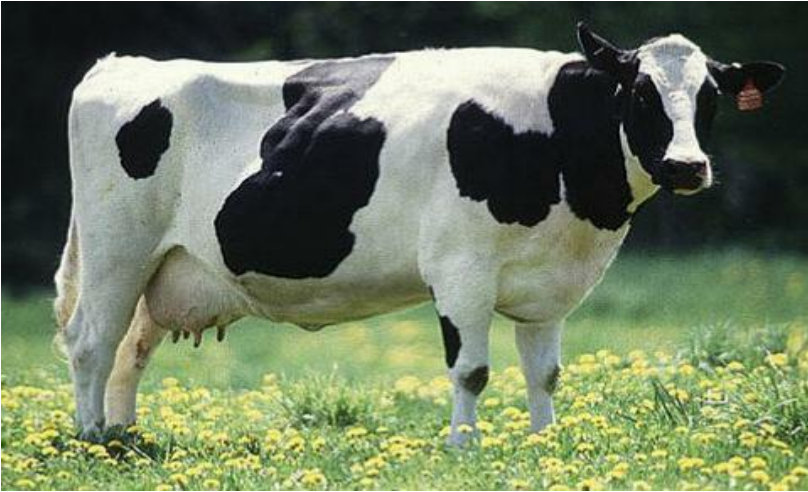
  <!-- ... -->
</entry>
<!-- ... -->
</entry>
```

Do we really have to encode all  
this manually?

# Considering machine learning techniques

- Strong layout regularities within a given dictionary
- Similarities within a family of dictionaries
- Supervised – unsupervised  $\Leftrightarrow$  Less data – more data
- Features to be considered:
  - Layout
  - Lexical
- Sequencing task: well adapted for so-called graphical models

# Why GROBID?



**Cow (structured data)**



**Hamburger (unstructured data)**

“Converting PDF to XML is a bit like converting hamburgers into cows. You may be best off printing it and then scanning the result through a decent OCR package.”

Michael Kay (<http://lists.xml.org/archives/xml-dev/200607/msg00509.html>)

Inspired from: Duncan Hull

# GROBID-Dictionaries

- Automatic extraction of TEI structures from digitised dictionaries (Khemakhem et al. 2017)
  - Input: PDF (soon ALTO)
  - Output: TEI compliant lexical resource
- Spin-off from GROBID (Romary and Lopez 2015)
  - Initiated in 2007
  - Automatic extraction of structural data from scholarly papers
    - Metadata (author, title, affiliations, keywords, abstract), bibliography, ... full text
  - And open source...
- Uses Conditional Random Fields (CRF) (Lavergne et al. 2010)
  - Probabilistic models for sequence labelling tasks

# State Of the Art

- Rule based approaches dominate
  - (*Khemakhem et al. 2009, Mykowiecka et al. 2012, Fayed et al. 2014*)
- Few machine learning attempts
  - Promote CRF for sequence labelling in dictionaries (*Crist 2011*)
  - Reduce the annotation time for labels (*Bago et al. 2015*)



# Cascading CRF models

Eugène IV à Ferrare en 1438-1439, puis à Florence de 1439 à 1442), de Latran (1512-1517), de Trente (1545-1563) [où fut décidée la réforme générale de l'Eglise catholique en face de la Réforme protestante], de Vatican I (1870) [où fut défini le dogme de l'infailibilité pontificale], de Vatican II (1962-1965) [où fut définie l'attitude de l'Eglise romaine à l'égard du monde moderne].  
**conciliable** adj. Qui peut se concilier avec une autre chose.  
**conciliabule** [kɔ̃siljabyl] n. m. (lat. conciliabulum) Réunion secrète entre un groupe de personnes.

**concordant, e** adj. Qui s'accorde : témoignages concordants.  
**concordat** [kɔ̃kɔrda] n. m. (lat. concordatum). Traité entre le pape et un gouvernement sur les affaires religieuses. || Dr. Accord entre le commerçant qui, ayant déposé son bilan, a été admis par le tribunal de commerce au règlement judiciaire et ses créanciers.  
 — Les plus anciens concordats sont le concordat de Worms (1122), entre Calixte II et Henri V; le concordat de 1516, entre Léon X et François I<sup>er</sup>. Le concordat entre

**concupiscence** n. f. (du lat. concupiscere, désirer). Penchant à jouir des biens terrestres, particulièrement des plaisirs sensuels.  
**concupiscent** [kɔ̃kypisɑ̃] e l-at adj. Qui exprime la concupiscence : regards concupiscent. Attaché aux plaisirs sensuels.  
**concurrer** [kɔ̃kyrɑ̃] v. t. Par concurrence. Par un concours mutuel, de concert : agir concurremment avec quelqu'un.  
**concurrence** n. f. Rivalité entre plusieurs personnes qui visent un même but : entre

**condenser** [kɔ̃dɑ̃se] v. t. (lat. condensare, rendre épais). Rendre plus dense, réduire à un moindre volume. || Liquéfier un gaz par refroidissement ou compression : le froid condense la vapeur d'eau. || Fig. Exprimer d'une manière concise, en peu de mots :

**conclaviste** n. m. Personne qui s'enferme au conclave avec un cardinal, pour le servir.  
**concluant, e** adj. Qui prouve bien ce qu'on a avancé : argument concluant.  
**conclure** [kɔ̃klyʁ] v. t. (lat. concludere) [con]. 62. Achever, terminer : conclure une affaire. || Tirer une conséquence : conclure une chose d'une autre. || — V. l. Donner son avis, ses conclusions : se prononcer : on vous demande de conclure. || Etre probant : les témoignages concluent contre lui.  
**conclusion** n. f. (lat. conclusio). Arrangement définitif : conclusion d'un traité. || Fin, résultat final : la conclusion d'un discours. || Conséquence d'un argument : la conclusion d'un syllogisme ne doit pas dépasser les prémisses. || — Pl. Présentations respectives de chacune des parties dans un procès. || Ecrit exposant ces présentations. || Réquisition du ministère public. || — En conclusion loc. adv. En conséquence, pour conclure.  
**concocter** v. t. Fam. Elaborer avec soin : concocter une lettre de réclamation.  
**concombre** [kɔ̃kɔbr] n. m. (anc. provenç. cocombre). Plante potagère de la famille des cucurbitacées, cultivée pour ses fruits allongés que l'on consomme comme légume ou en salade. || Ce fruit.  
**concomitamment** adv. De façon concomitante.  
**concomitance** [kɔ̃kɔmitɑ̃s] n. f. Coexistence, simultanéité de deux ou de plusieurs faits.  
**concomitant, e** adj. (lat. concomitans). Qui accompagne, qui se produit en même temps : des faits concomitants. • Variations concomitantes, variations simultanées et proportionnelles de certains phénomènes.  
**concordance** n. f. Conformité, accord : concordance de témoignages. || Géol. Disposition parallèle des couches sédimentaires. • Concordance de phases (Phys.), état de plusieurs vibrations sinusoïdales de même nature et de même période, dont la différence de phases est nulle. || Concordance des temps, règles de syntaxe d'après lesquelles le temps du verbe d'une subordonnée varie selon celui du verbe de la principale.

concours. || Lutte sportive : concours hippique. • Concours général, concours annuel entre les premiers élèves des classes supérieures des lycées, collèges et écoles normales.  
**concret** [kɔ̃kʁɛ] e l-et adj. (lat. concretus). Epais, condensé : huile concrète (vieilles). || Qui exprime quelque chose de réel, de positif : obtenir des avantages concrets. || Qui a le sens des réalités précises : esprit concret. || Gramm. Se dit d'un terme qui désigne un être ou un objet pouvant être perçu par les sens. • Musique concrète, technique de composition qui utilise les bruits produits par divers objets sonores enregistrés sur bande magnétique et susceptibles de transformation.  
**concret** n. m. Qualité de ce qui est concret.  
**concrètement** adv. De façon concrète.  
**concréter** v. t. (cong. 3). Rendre concret, solide.  
**concrétion** [kɔ̃kʁɛsjɔ̃] n. f. (de concret). Action de s'épaissir : la concrétion de l'huile, du sang. || Réunion de parties en un corps solide : concrétion saline. || Agrégation solide dans les tissus vivants : concrétions biliaires.  
**concrétiser** v. t. Rendre concret ce qui est abstrait : concrétiser une idée, un avantage.  
**concubin, e** adj. (lat. concubina). Relatif au concubinage. || — N. Personne qui vit en concubinage.  
**concubinage** [kɔ̃kybinɑ̃ʒ] n. m. Etat d'un homme et d'une femme qui vivent ensemble sans être mariés. (On dit aussi UNION LIBRE.)

**concussion**.  
**condamnabile** adj. Qui mérite d'être condamné : acte condamnable.  
**condamnation** [kɔ̃dɑ̃nasjɔ̃] n. f. (lat. condemnatio). Décision d'un tribunal imposant à l'un des plaideurs de s'incliner au moins partiellement devant les prétentions de son adversaire. || Décision d'une juridiction prononçant une peine contre l'auteur d'un crime, d'un délit ou d'une contravention. (En cour d'assises, le jury juge la culpabilité de l'accusé, et la cour prononce la condamnation.) || La peine infligée : une condamnation à la réclusion criminelle. || Blâme, désapprobation : la condamnation des abus.  
**condamnatoire** adj. Qui porte condamnation.  
**condamné, e** n. Personne qui a subi une condamnation. || — Adj. Qui ne peut échapper à un sort prévu : malade condamné.  
**condamner** [kɔ̃dɑ̃ne] v. t. (lat. condemnare). Prononcer un jugement contre un plaideur ou un inculpé : condamner un criminel. || Astreindre, réduire à : condamner au silence, à l'immobilité. || Désapprouver, blâmer : condamner une opinion, un usage. || Interdire : la loi condamne la bigamie. || Déclarer perdu, incurable : les médecins l'ont condamné. || Barter, muter : condamner une porte.  
**condensable** adj. Qui peut être condensé, réduit à un moindre volume.  
**condensateur** n. m. Phys. Appareil servant à emmagasiner une charge électrique : la bouteille de Leyde est un condensateur électrique. || Lentille servant à éclairer un objet dont on veut former une image.  
**condensation** n. f. Action de condenser ou effet qui en résulte. || Liquéfaction d'un gaz. || Soudure de plusieurs molécules chimiques, avec élimination d'eau.  
**condensé** n. m. Résumé d'une œuvre littéraire.  
**condenser** [kɔ̃dɑ̃se] v. t. (lat. condensare, rendre épais). Rendre plus dense, réduire à un moindre volume. || Liquéfier un gaz par refroidissement ou compression : le froid condense la vapeur d'eau. || Fig. Exprimer d'une manière concise, en peu de mots :



# Cascading CRF models

a, A 1. Türk alfabesinin ilk sırasında yer alan ve A adı verilen bu harf, ses bilimi bakımından kalın ünlülerin düz ve geniş olanını gösterir. 2. *miz*. Nota işaretlerini harflerle gösterme yönteminde *la* sesini bildirir. a'dan z'ye (kadar) baştan aşağı, tamamen, tamamıyla, bütünüyle: *Evini a'dan z'ye değiştirdi*.

a ünl. (a) Şaşma, hatırlama, sevinme, acıma, üzülme, kızma vb. duyguların anlatımına güç kazandıran söz: *A, ne güzel! A, sen burada mıydın?*

a / e ünl. Dilek kipinin ikinci teklik ve çokluk şahıslarının çekiminden sonra gelecek anlamı pekiştiren ve güçlendiren bir söz: *"Azıcık dursana oğlum, dedi."* -A. Kabaklı. "O mu bana getirirsen!" -N. Hikmet. "Başka gazetelere baksanız! Onlar da yazıyor." -N. F. Kısakürek. "Şimdi de başka çıkmazdayız deseniz!" -N. Uygur.

ab a. (a:b) Far. *ab* esk. Su. → *abihayat*, *abikevser*, *abuhava*

aba (I) a. *hik*. 1. Aba. 2. Anne.

aba (II) a. *Ar*. 'abâ 1. Yünün dövlmesiyle yapılan kalın ve kaba kumaş. 2. Bu kumaştan yapılmış yakasız ve uzun üstlük. 3. *sf*. Bu kumaştan yapılan. 4. *esk*. Bu kumaştan yapılan ve dervişlerce giyilen hırka. 5. Kepek (I). *aba altında er yatar* "bir insanın değeri giymiyle kumaşıyla ölçülemez" anlamında kullanılan bir söz. *aba altından sopa* (veya değnek) göstermek birini imalı bir biçimde tehdit etmek. *aba gibi kaba* ve kalın (kumaş). *aba vakti yaba, yaba vakti aba* "gereksinimler vaktinden önce ve ucuz olduğu zaman karşılanmalıdır" anlamında kullanılan bir söz. *abanın kadri yağmurdadır* bilinir "bir şeyin gerçek değeri ona gereksinim duyulduğunda anlaşılır" anlamında kullanılan bir söz. (bir yer) abayı sermek 1) istenilmediği hâlde teklifsizce yerleşmek, 2) uzun süre yerleşip kalmak. *abayı yapmak* *öz*. birine aşırı bir biçimde gönül vermek, tutulmak, aşık olmak: *"Sen mi verince ona gönül yoksa o mu yakıt sana daha önce*

abayı?" -O. C. → *aba güreşi*

abacı a. 1. Abi Abadan giyice *sf. mec*. Asala beci (ara yeti lendirmeyen mında kullanı

abacılık, -ğı a. .

abadi a. (a:ba: man renginde kalınca bir ya

aba güreşi a. *şy* bağlanarak ya

abajur a. *Fr*. a lamak, doğru önlemek için den veya renl peri. 2. Genel sı veya ayakk telefon, ortası *bajur*." -N. F.

abajurcu a. Ab

abajurculuk, -ı

abajurlu *sf*. Al *abajurlu*, par Safa.

abajursuz *sf*. *i abajursuz*, çe *okuyordu*." -S

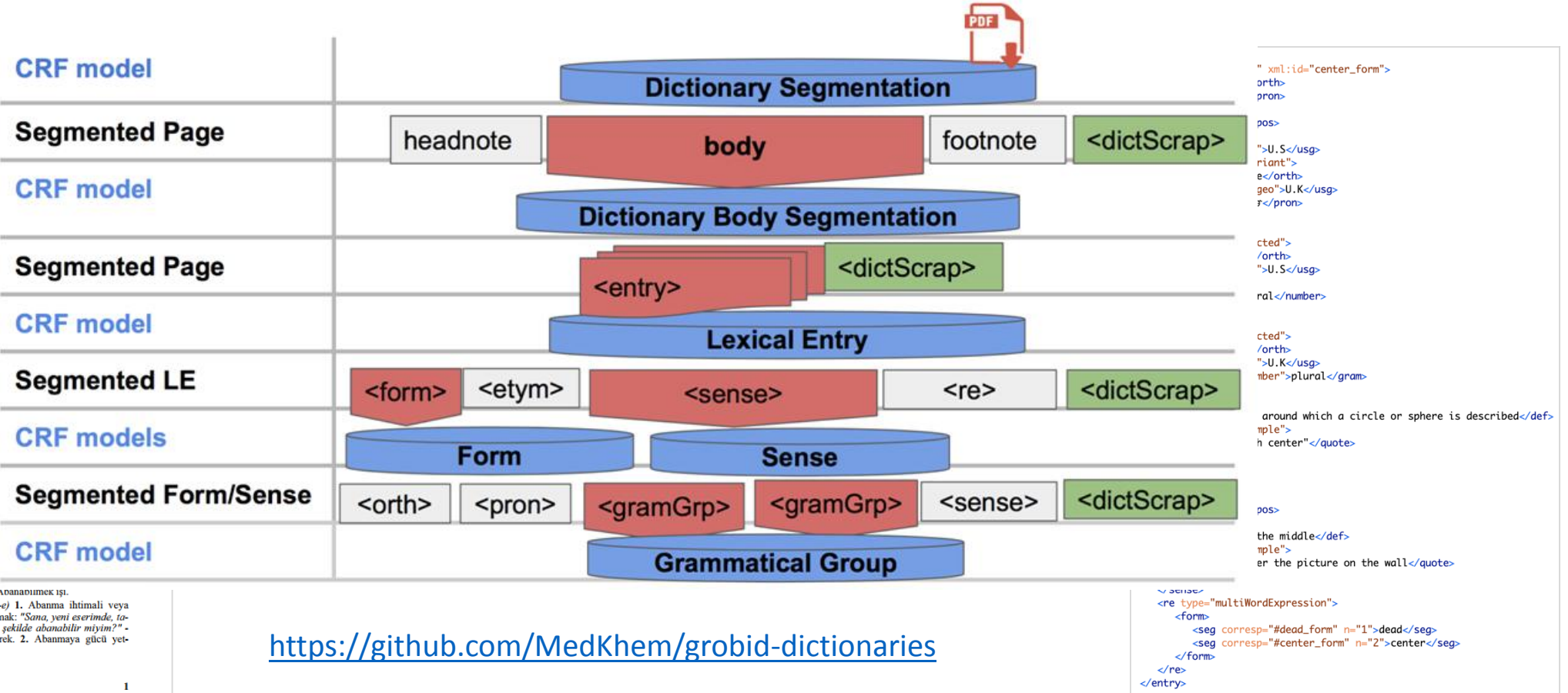
abaküs a. *Fr*. a *mim*. Sütün t konan ve kei taş blok.

abalı *sf*. Aba gi

Abana *öz*. a. (c ilçelerden biri

abanabileme a. *Λonanarımek ışı*.

abanabilemek (-e) 1. Abanma ihtimali veya imkânı bulunmak: *"Sana, yeni eserimde, ta-kat getirilmez şekilde abanabilir miyim?"* -N. F. Kısakürek. 2. Abanmaya gücü yetmek.



<https://github.com/MedKhem/grobid-dictionaries>



# First experiments

## B

**b** /bi:/, **B** *noun* the second letter of the alphabet, between A and C

**baby** /'beɪbi/ *noun* 1. a very young child  
○ Most babies start to walk when they are about a year old. ○ I've known him since he was a baby. 2. a very young animal ○ a baby rabbit (NOTE: The plural is **babies**. If you do not know if a baby is a boy or a girl, you can refer to it as **it**: The baby was sucking its thumb.)

**back** /bæk/ *noun* 1. the part of the body which is behind you, between the neck and top of the legs ○ She went to sleep lying on her back. ○ He carried his son on his back. ○ Don't lift that heavy box, you may hurt your back. 2. the opposite part to the front of something ○ He wrote his address on the back of the envelope. ○ She sat in the back of the bus and went to sleep. ○ The dining room is at the back of the house. ■ **adjective** 1. on the opposite side to the front ○ He knocked at the back door of the house. ○ The back tyre of my bicycle is flat. 2. (of money) owed from an earlier date ○ back pay ■ **adverb** 1. towards the back of something ○ She looked back and waved at me as she left. 2. in the past ○ back in the 1950s 3. in the state that something was previously ○ Put the telephone back on the table. ○ She watched him drive away and then went back into the house. ○ She gave me back the money she had borrowed. ○ I'll phone you when I am back in the office. (NOTE: **Back** is often used after verbs: to give back, to go back, to pay back, etc.) ■ **verb** 1. to go backwards, or make something go backwards ○ He backed or backed his car out of the garage. 2. to encourage and support a person, organisation, opinion or activity, sometimes by giving money ○ Her colleagues were willing to back the proposal. ○ to put someone's back up to annoy someone

**back up** *phrasal verb* 1. to help or support someone ○ Nobody would back her up when she complained about the service. ○ Will you back me up in the vote? 2. to make a car go backwards ○ Can you back up, please – I want to get out of the parking space.

**background** /'bækgraʊnd/ *noun* 1. the part of a picture or view which is behind all the other things that can be seen ○ The photograph is of a house with mountains in the background. ○ His white shirt stands out against the dark background. Compare **foreground** ○ in the background while other more obvious or important things are happening 2. the experiences, including education and family life, which someone has had ○ He comes from a working class background. ○ Her background is in the restaurant business. 3. information about a situation ○ What is the background to the complaint?

**backward** /'bækwəd/ *adverb* US same as **backwards**

**backwards** /'bækwəd/ *adverb* from the front towards the back ○ Don't step backwards. ○ 'Tab' is 'bat' spelt backwards. ○ backwards and forwards in one direction, then in the opposite direction ○ The policeman was walking backwards and forwards in front of the bank.

**bacon** /'beɪkən/ *noun* meat from a pig which has been treated with salt or smoke, usually cut into thin pieces

**bacteria** /'bæktɪəriə/ *plural noun* very small living things, some of which can cause disease (NOTE: The singular is **bacterium**.)

**bacterial** /'bæktɪəriəl/ *adjective* caused by bacteria ○ a bacterial infection

**bad** /bæd/ *adjective* 1. causing problems, or likely to cause problems ○ Eating too much fat is bad for your health. ○ We

badge

23

ban

were shocked at their bad behaviour. 2. of poor quality or skill ○ He's a bad driver. ○ She's good at singing but bad at playing the piano. 3. unpleasant ○ He's got a bad cold. ○ She's in a bad temper. ○ I've got some bad news for you. ○ The weather was bad when we were on holiday in August. 4. serious ○ He had a bad accident on the motorway. (NOTE: **worse** /wɜːs/ – **worst** /wɜːst/)

**badge** /bædʒ/ *noun* a small sign attached to someone's clothes to show something such as who someone is or what company they belong to

**badly** /'bædli/ *adverb* 1. not well or successfully ○ She did badly in her driving test. 2. seriously ○ He was badly injured in the motorway accident. 3. very much ○ His hair badly needs cutting. (NOTE: **badly** – **worse** /wɜːs/ – **worst** /wɜːst/)

**bag** /bæɡ/ *noun* 1. a soft container made of plastic, cloth or paper and used for carrying things ○ a bag of sweets ○ He put the apples in a paper bag. 2. same as **handbag** ○ My keys are in my bag. 3. a suitcase or other container used for clothes and other possessions when travelling ○ Have you packed your bags yet?

**baggage** /'bæɡɪdʒ/ *noun* cases and bags which you take with you when travelling

**bake** /beɪk/ *verb* to cook food such as bread or cakes in an oven ○ Mum's baking a cake for my birthday. ○ Bake the pizza for 35 minutes.

**baker** /'beɪkə/ *noun* a person whose job is to make bread and cakes ○ the baker's a shop that sells bread and cakes ○ Can you go to the baker's and get a loaf of brown bread?

**balance** /'bæləns/ *noun* 1. the quality of staying steady ○ The cat needs a good sense of balance to walk along the top of a fence. 2. to keep your balance not to fall over ○ As he was crossing the river on the tightrope he lost his balance and fell. 2. an amount of money remaining in an account ○ I have a balance of £25 in my bank account. 3. an amount of money still to be paid from a larger sum owed ○ You can pay £100 now and the

balance in three instalments. ○ The balance outstanding is now £5000. ■ **verb** 1. to stay or stand in position without falling ○ The cat balanced on the top of the fence. 2. to make something stay in position without falling ○ The waiter balanced a pile of dirty plates on his arm.

**balcony** /'bælkəni/ *noun* 1. a small flat area that sticks out from an upper level of a building protected by a low wall or by posts ○ The flat has a balcony overlooking the harbour. ○ Breakfast is served on the balcony. 2. the upper rows of seats in a theatre or cinema ○ We booked seats at the front of the balcony. (NOTE: The plural is **balconies**.)

**bald** /bɔːld/ *adjective* having no hair where there used to be hair, especially on the head ○ His grandfather is quite bald. ○ He is beginning to go bald.

**ball** /bɔːl/ *noun* 1. a round object used in playing games, for throwing, kicking or hitting ○ They played in the garden with an old tennis ball. ○ He kicked the ball into the goal. 2. any round object ○ a ball of wool ○ He crumpled the paper up into a ball. 3. a formal dance ○ We've got tickets for the summer ball. ○ to start the ball rolling to start something happening ○ I'll start the ball rolling by introducing the visitors, then you can introduce yourselves. ○ to play ball to work well with someone to achieve something ○ I asked them for a little more time but they won't play ball. ○ to have a ball to enjoy yourself a lot ○ You can see from the photos we were having a ball.

**baller** /'bæleɪ/ *noun* 1. a type of dance, given as a public entertainment, where dancers perform a story to music 2. a performance of this type of dance ○ We went to the baller last night.

**balloon** /bə'luːn/ *noun* 1. a large ball which is blown up with air or gas 2. a very large balloon which rises as the air inside it is heated, sometimes with a container attached for people to travel in ■ **verb** to increase quickly in size or amount

**ban** /bæn/ *noun* an official statement which says that people must not do

AYO

— 58 —

AZI

nuit à tel endroit. Bia bukh ayôa va, nous couchons ici. Syn.: azakh (dêk).

**AYÔE** (h) n.4, pl. meyoê (vb yôe h). Ayôe mezim, action de faire chauffer de l'eau. Syn.: awôghê mezim, gyôgha mezim.

**AYÔL** (b) n.4, ss pl. (vb yôl b). Amer, mauvais, amertume physique ou morale. Ebmuma ayôl, fruit amer. Mezô m'ayôl, paroles amères. Contr.: anêkh, nêghba. Ayôl est aussi une odeur, l'odeur de quelque chose qui est amer (voir nyumayôle).

**AYÔM** (m) n.4, ss pl. (vb yômê b). Vieillesse. Ayôm e to nye e nyôl, il se fait vieux. Syn.: ayômbe.

**AYÔM** (h) n.4, pl. meyôm. De même tribu. Bie-nye bi ne ayôm, lui et moi sommes de la même tribu. Un homme du clan des Esamelôkh et un homme du clan des Esisom sont tous les deux de la tribu des Esamelôkh. Si l'un va chez l'autre, il dit: Ma be ayôm dam.

**AYÔMBE** (bm) n.4, pl. meyômbe (vb yômbe b). Vieillesse. Syn.: ayôm.

**AYÔMLE** (h) n.4, pl. meyômle (vb yômle h). Bénédiction fétiche, parole qui porte bonheur. Voir segehe ntô e ntô.

**AYÔN** (h) n.4, ss pl. (vb yôh h). 1. Chaud, chaleur. Mezim me ne ayôh, l'eau est chaude. Mais on dit irrégulièrement: mezim meyôh, eau chaude (et mezim mewê, eau froide). Voir meyôh. — 2. Zèle, force, vivant, tempérament bouillant. Zal e ne ayôh, le village est plein d'animation. Voir alugha (b). Ntêgha ayôh, soyons zélés. Contr.: awêd. Abôkh di e ne ayôh, cette danse est très entraînante.

**AYÏ** (m) n.4, pl. meyÏ. Arbre à bois très dur qui sert à faire des bûches en bois pour creuser des trous (évan), ou des manches de haches. Autre bois pour les mêmes usages: ébam.

**AYÏA** (bm) n.4, pl. meyÏa (vb yÏa b). Mécontentement, fait d'être fâché. Syn.: éyÏ.

**AZA** (h) n.4, pl. meza (vb za h). Destruction. Syn.: azôh.

**AZAKH** (h) n.4, pl. mezakh. 1. Endroit arrangé par le chimpanzé ou le gorille pour y dormir, ce qui lui tient lieu de maison. C'est assez près du sol. Azakh e wagha, azakh e âgh. — 2. Campement d'homme, étape pour la nuit (dêk). Syn.: ayôa.

**AZAMÉ** (h) n.4, pl. mezamé (vb zamé h). Action de pardonner, de laisser. Syn.: biamé.

**AZAN** ! (h) n.4, et interj. (Atô) (vb zañ h). Imprécation pour le serment. Azôh bôr ! É bô be âga man-e-zôh me, tous les miens qui sont morts, je jure par eux. Autre phrase analogue: ma bête, me to mimim, je l'affirme, je vois les morts.

**AZANÉ** (h) n.4, pl. mezané (vb zañ h). Destruction, fait d'être détruit, de mourir en grand nombre. Syn.: aza.

**AZAP** (b) n.4, pl. mezap. Nom d'arbre. Syn.: aza.

**AZEBE** (bm) n.4, pl. mezebe (vb zebe b). Enterrement, funérailles.

**AZÉE** (bm) n.4, pl. mezé (vb zé b). Action d'ensevelir. Action d'écarter les bûches du foyer pour éteindre le feu. Azé môm. Azé biakh, zé.

**AZEGHA** (bm) n.4, pl. mezegeha (vb zegha b). Dernier soupir, fait d'expirer. Syn.: ayie.

**AZEM** (h) n.4, pl. mezem. 1. Paquet de feuilles de manioc pilées avec sel, plante, viande ou poisson. — 2. Petite piment au bord des ruisseaux.

**AZI** (h) n.4, ss pl. (vb zî h). Aliment, chose qui se mange. Syn.: bié, azie.

**AZIE** (h) n.1, pl. bazie. Boue que les femmes mettent sur les barrages de rivières (myekh) pour les rendre étanches.

**AZIE** (bm) n.4, pl. mezé (vb zî b). Action d'enfoncer une pointe. Azie akoñ.

**AZIE** (h) n.4, pl. mezé (vb zî h). Action de manger. Be vagha zi azie avôd, ils ont mangé une fois. Mezia meôh, ils ont mangé deux fois.

**AZIGHA** (b) n.4, pl. mezigha (vb zigha b). Azigha mam, inventaire. Action de compter des choses.

**AZIGHE** (h) n.4, pl. mezighe (vb zighe h). Action de brûler quelque chose. Azighe tsé, action de brûler un débroussaement pour plantation.

**AZIGHÉ** (h) n.4, pl. mezighe (vb zikh h). Incendie, fait de brûler soi-même.

**AZIKH** (h) n.4, pl. mezikh (vb zigha h). Flots de paroles dans une palabre pour en finir vite.

**AZIMÉ** (h) n.4, pl. mezimé. 1. Faute, tort, fait de se tromper ou de se perdre (vb zimé h). — 2. Azimé âgon, pl. mezimé me âgon, fin de lunaison, nouvelle lune (voir atêl âgon), ou encore coucher de lune. Azimé zô, coucher du soleil (vb zimé h).

**AZIN** (b) n.4, pl. mezin (vbs zin h

AZI

— 59 —

BA

et ziâda h). 1. Variété de palmier rotin épineux et grimpant qu'on voit surtout dans l'Abanga et dans la Lolo. Il ressemble au âkan, et sa base s'appelle aussi ako. On emploie la base qui est grosse comme le bras comme râpe en enlevant seulement les pointes des épines. Azin servait autrefois à faire des flûtes appelées azin. — 2. Nom de toute espèce de râpe. Voir akasigha.

**AZIR** (m) n.4, ss pl. Louré, poids, pesanteur (vb zir b). Akokh azir, pierre lourde. Nlô azir, tête dure. Ôson wa ce nlô azir, la honte alourdit la tête. Syn.: aném.

**AZO** (b) n.4, pl. mezo. 1. Un des plus beaux arbres de la forêt équatoriale qui peut atteindre 40 mètres. Arbre à beurre. Nom commercial: moubi. (Mimusops djavé). Le tronc est très droit et les branches horizontales. Son bois est très apprécié pour la belle menuiserie. Il fruit (ébôna) est gros et comestible, et les noyaux contiennent une bonne huile (ézo). Une légende ancienne veut que toutes les tribus des Fañ dans leurs migrations du nord au sud aient passé par une certaine cavité pratiquée entre un azo et un ébom qui se touchaient. Il n'y avait pas d'autre issue possible, et l'ouverture que l'on fit entre ces deux arbres s'appelle azo mbôgha. Fañ bese be âga lôr azo mbôgha. Mbôgha veut dire entaillé (vb bôkh b). Syn. de azo: azap. Voir éndôh azo, byézo (faux azo). — 2. Azo éwé (bm) n.4, pl. mezo m'éwé. Variété d'igname. Syn.: abule.

**AZOKH** (h) n.4, ss pl. (vb zôkh h). Fait de nager, nage, natation. A vagha ke ôvêl ayar azokh, il a traversé la rivière à la nage. Syn.: nêghba (h).

**AZOM** (m) n.4, pl. mezom. Amome, espèce de roseau à grande palme, très résineux, qui pousse dans les anciens

débroussaements (Aframomum Melegueta). Les tiges sèches (kômé) servent de torches dans les pays où manque la résine d'okoumé (âghuma); elles servent aussi à mettre le feu à un nouveau débroussaement. Le fruit rouge (zôn) se trouve presque au ras du sol, au bas de la tige, et il peut se manger. Les feuilles servent à couvrir les abris de forêt, mais sans préserver de la pluie. Les terrains où poussent les mezom sont fertiles. Voir ézom. Il y a trois variétés d'azom: 1. Le grand azom qui est le plus commun. — 2. Obalé-azom (voir ce mot). — 3. Ndôh (h) (voir ce mot).

**AZÔ** (b) n.4, pl. mezô (peu usité). 1. Ecume de l'eau produite par les poisssons groupés près du rivage. — 2. Beaucoup. Azô e ko, écume faite par les poisssons, ou: il y a beaucoup de poisssons. Ko za ko azô, les poisssons font de l'écume dans l'eau. Mais cela signifie aussi qu'il y a beaucoup de poisssons. Azô e bôr, beaucoup d'hommes. Azô e byôm, beaucoup de marchandises. Azô bêtair, beaucoup de bêtes. Voir avul, eêgha.

**AZÔ** (h) n.4, pl. mezô (vb zô h). Parole, affaire, différend. Mezô me anôm (m), paroles des vases.

**AZÔBA** (bm) n.4, pl. mezôba (vb zôba b). Repentance, regret. Voir bié.

**AZÔÉ** (h) n.4, pl. mezé (vb zô h). Fait d'être bien accordé (instrument de musique). Azô meér.

**AZÏ** (b) n.4, pl. mezi. Banane douce, grosse, à peau rouge.

**AZÏ** (h) n.4, pl. mezi (vb zî h). Action de régner, de commander, règne, commandement. Syn.: zêh.

**AZÏVÉ** (bm) n.4, pl. mezié (vb zîv b). Action d'ouvrir, ouverture. Azîé mbi, action d'ouvrir la porte.

## B

qu'un un croc-en-jambe (syn.: tu éba, fer éba).

**BA** (b) (lg) adj. poss. Nos. Béa ba, nos pères.

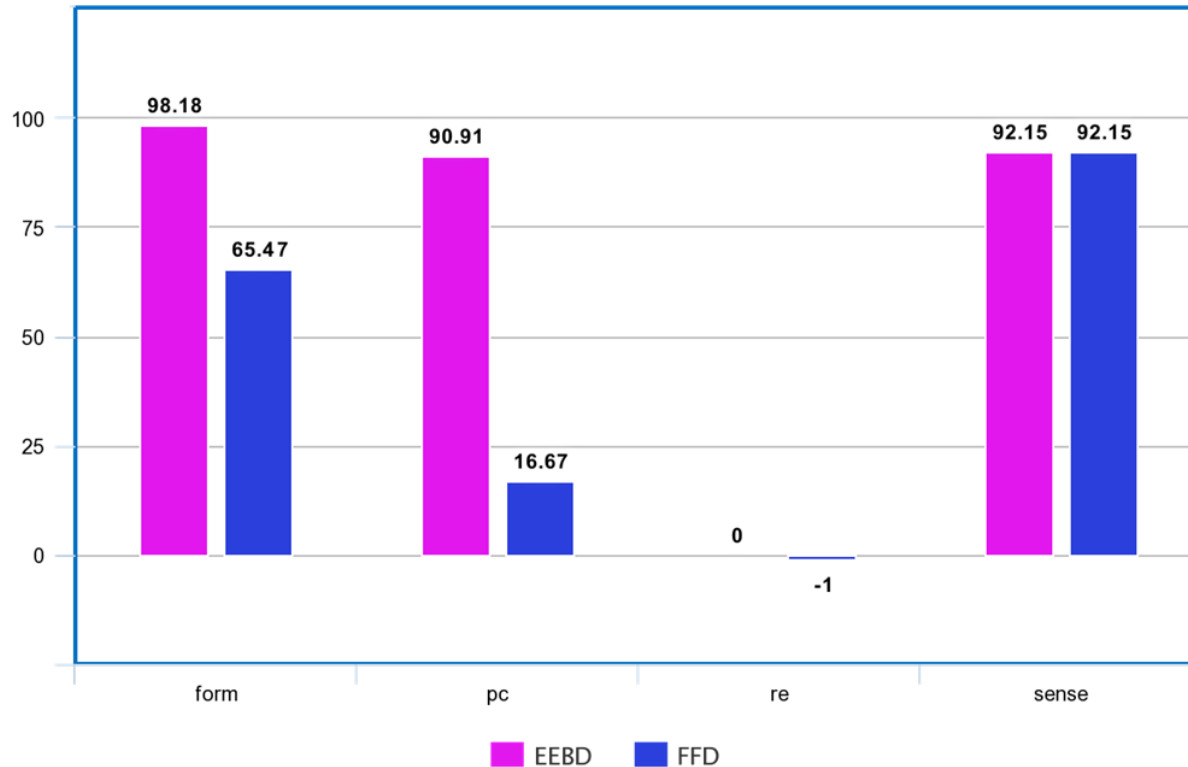
**BA** (b) (lg) interj. 1. Oui. — 2. interj. marquant l'impatience et le refus.

**BA** (b) (lg) n.1, pl. beba. Papa, mon père. Se dit surtout dans le haut. Ba a

# Evaluation: Token Level - F1 Score

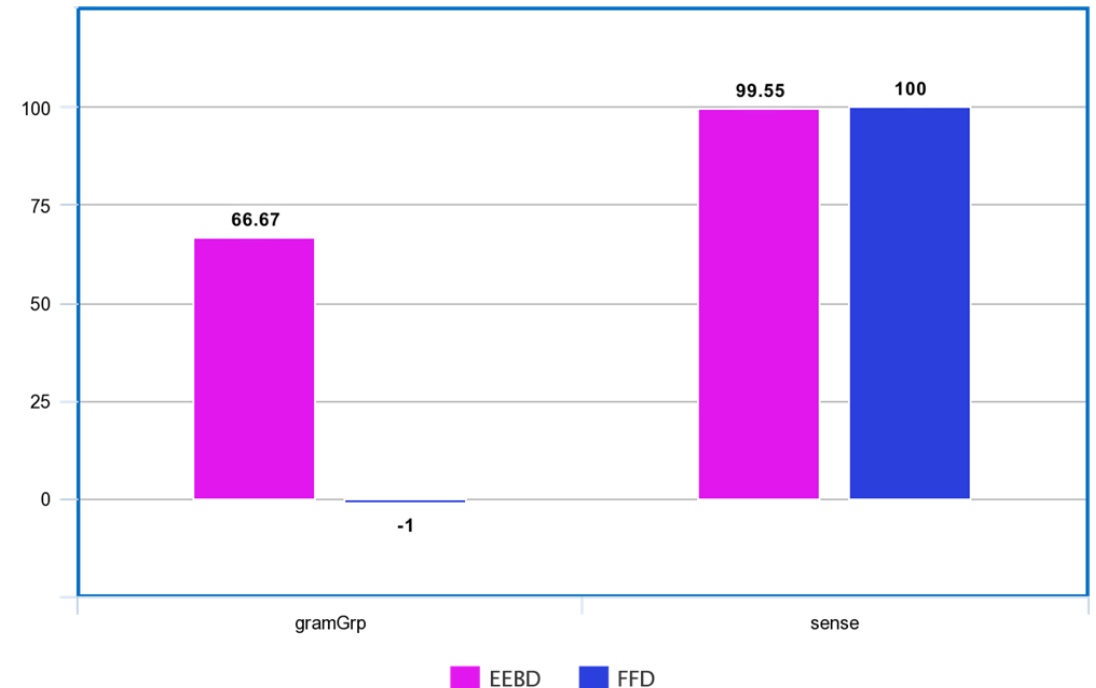
## Lexical Entry

- EEBD: 100 LE (8 pages): 76 training, 24 evaluation
- FFD: 71 LE (3 pages): 47 training, 24 evaluation



## Sense

- EEBD: 30 blocks (6 pages): 15 training, 15 evaluation
- FFD: 90 LE (4 pages): 71 training, 19 evaluation





## Further complexities

**C A C.**

**CACHÉ-NEZ.** f. m. Vieux mot qui signifioit autrefois un *masque*.

**CACHER.** v. æt. Mettre quelque chose en un lieu secret, où il ne puisse être vu ni trouvé par d'autres qu'avec beaucoup de difficulté. Les païsans *cachent* leur argent dans la terre, afin que les soldats ne le puissent trouver. Menage après Guyot derive ce mot de *caçier*, qui signifie *chasser, pourvoir*. On dit en ce sens, que la nature nous a *caché* ses trésors, ses plus merveilleuses opérations.

CHACHER, signifie aussi, Voiler, déguiser, ne paroître pas à la vue. Cette fille est si modeste, qu'elle se cache le visage de ses cheveux, et son visage. Cet homme m'a *caché* son nom, son fond. Il *cache* du bois; pour dire, Il fait *cache* à mon aise. Dans cette cécipité, *Lune cache* la moitié du disque du Soleil. Les arbres en se couchant *se cachent* dans l'onde. Voilà un bois qui nous cache la vue de ce chateau. On demandait à quelqu'un qui portoit quelque chose *caché* sous son manteau, ce qu'il portoit; je le *cache*, dit-il, afin qu'on ne le sache point. *MENE.*

*MENEUX* qui *laisse* de *bonne* fortune.

On dit *cacher son jeu*, & cette expression a trois divers sens. Elle signifie I. empêcher que quelqu'un ne voie son jeu. II. Dissimuler son adresse en faisant semblant de ne sçavoir pas bien jouer. III. *Cacher* ses desseins, en sorte que personne ne les puisse découvrir. Il est tout-à-fait figuré en ce dernier sens.

a-t-*il* *hugué* en se *récruser* dans  
 ces *choses* *morales*. C'est  
 un hypocrite qui *cahier* *cahier* *la* *turpitude*. Ce font  
 de bons *amis* qui ne *cahier* *rien* à l'autre. Il est  
 avantageux *fouvent* de *cahier* de *cahier*, de diffi-  
 muler *la* *colère*, *son* *amour*. Les *Payens* *cahier*  
 beaucoup de *fecres* de la *nature* *fois* *le* *voile* de *leurs*  
*fautes*. Dica a *cahier* *cahier* *cahier* *cahier* *cahier*  
 & les *autres* *amis* *simples*. Brutus *cahier* *une*  
*prudence* *fois* *une* *apparence* de *folie*. Il ne *vous*  
*fait* *peris* de *cahier* *plus* *long* *temps* *vos* *peis*  
 de *diffinitions*. O*z*, M. La *bacille* de *ce* *homme*  
*paroit* *aut* *plus* *qu'on* *ne* *cahier*. On *peut*  
*cahier* *ce* *contingens* *des* *fautes* *ingénieuses*. A *dieu* *cahier*  
*cahier* *cahier* *cahier* *cahier* *cahier* *cahier* *cahier*  
*se* *impuise* & *deracinables*. P*ort*-R. Il a de  
 l'*adrecie* à *bien* *cahier* *la* *passion*. A*ut*-R. *cahier* *la*  
*haine* *fois* *cahier* *la* *passion*. R*ac*. On s'*endie* *plus*  
*dans* *le* *monde* à *cahier* *cahier*, qu'à *acquies* *la*  
*veru*. W*c*. Rien n'*est* *plus* *aimable* que *la* *timidité*  
*du* *jeu* *Bergere*, qui n'*est* *plus* *se* *montrer*, ni *cahier*  
*cahier* *cahier* *cahier* *cahier* *cahier* *cahier* *cahier*  
*meux* que *l'indolence* de *ceux* qui ne *font* *rien*  
*la* *peine* de *dequies* *leurs* *deuix*. B*el* L.

On dit absolument, *Se cacher*; pour dire, Vivre en retraite, ou se mettre en lieu de sûreté pour n'être pas pris ni découvert. Les Saints *se cachent* aux yeux des hommes, pour se donner tout à Dieu. Cet homme craint la prison, il *se cache*, il ne va que la nuit, il se retire & *se cache* dans les maisons des Princes, en des asyles. Après avoir reçu un tel affront, il se faut *cacher*, & ne plus paroître en public.

On dit proverbiallement, *Cacher* à vie : c'est un des préceptes d'Epicure, dont Plutarque a fait un beau Traité; pour dire, qu'il ne faut pas faire connaître à tous les hommes ce que l'on fait. Le péché que l'on *cache* est demi pardonné, dit Regnier.

**CACHÉ**, ÉE, part. & adj. Cet homme est un trefor *caché*. On appelle un *caché*, Celui qui ne veut pas se faire connaître dans le monde, fort qu'il soit modeste, fort qu'il soit fantasque. Un esprit *caché*, un homme profond, & dissimulé. Une science *cachée*, celle qui est abstruse, ou connue de peu de personnes,

C A C.

comme l'Algebre, la Cabale, la Steganographie, l'Ecriture dit qu'il n'y a rien de si *cacher* qui ne le revele, qui ne paroisse quelque jour. Dieu nous tient les secrets *cacher*, afin que nous ne cessions pas de prier.

M. D E M.

*Quand la vertu gemit sous le pouvoir du vice,  
Adieu du Seigneur soit le joyeuxment cacher.*

*Adore du Seigneur les jugemens cachez.*  
L'AB. TETU.  
**CACHET.** *f. m.* Petite ſceau qui porte une gravure particulière de quelques Armes ou chiffres qu'on imprime ſur de la cire, ou du pain à chanter, pour empêcher qu'on n'ouvre un paquet fermé & marqué de cette empreinte. Les Anciens n'avoient point d'autres *cachets* que leurs anneaux, qui portoient des pierres gravées. Ce mot vient de *cachet*, à cauſe qu'il ſert à cacher l'écriture. MEN.

CACHET, se dit aussi de la figure, de la marque imprimée sur la cire. Le *cachet* est entier, il n'a point été rompu.

On appelle *Cachet volant*, la marque du *cachet* imprimée sur un papier, avec lequel on pourra fermer quand on voudra une lettre qu'on donne ouverte.

**LETTRE DE CACHET**, est une lettre cachetée du *cachet* du Roi, & signée d'un Secrétaire d'Etat, qui contient quelque ordre, commandement, avis, ou autre chose qu'on envoie de la part du Roi.

**CACHETER**, v. act. Appliquer un cachet sur quelque chose qu'on veut envoyer fermée. *Cacheter* un paquet, une boîte, une bouteille.

**CACHETÉ, ÉE.** part. & adj. Il m'a rendu vos lettres *cachetées*.

**CACHETTE.** f. f. Petite cache. Il y a bien des *cachettes* dans ce bois.

EN CACHETTE. adv. D'une manière cachée, secrète. Les livres défendus ne se vendent qu'en *cachette* & sous le manteau. Quand on fait les choses en *cachette*, il y a du péché & de la honte ordinairement. Il a fait cela en *cachette* de moi, c'est-à-dire, il n'a pas voulu que je le sçusse. Le jugement ne fut donné qu'en *cachette*. PATRU. On ne doit pas user de duel si on peut tuer son homme en *cachette*. PASC.

**CACHOS.** *L. m.* Plante qu'on ne trouve que dans les montagnes du Pérou. Elle croît comme un arbrisseau, & c'est d'un fort beau vert. Sa feuille est ronde & mince. Son fruit est plat d'un côté, rond de l'autre finissant en pointe, de couleur cendrée, d'un goût agréable & sans acrimonie, contenant une semence fort menue. Les Indiens font beaucoup d'us de cette plante à cause de ses rares qualités : car elle fait uriner & chasser le sable & la pierre hors des reins ; & ce qui est plus admirable, c'est qu'on tienne que par son usage elle biffe la pierre dans la vessie, si elle est encore tendre, & qu'elle se puisse rompre par quelque médicament. En Latin *cachos*, qu'on appelle *quiquira* [sic] *retunda tenui*.

C. ACHOT. f. m. Prison noire & obscure, qui est au dessous du rez de chauffée, & où on ne gîte que sur la paille. On met dans les *cachots* les criminels condamnés, ou accusés de grands crimes, ou qui font des rebellions dans la prison. Vous decrirai-je ces *cachots*, ou plutôt ces sepulchres funèbres, où l'on enterre des hommes vivans, pour qui il semble que le soleil ait cessé de luire, & que la nuit ait pris la place du jour ? Ft.

CACHOT, se dit aussi d'une sorte de petite loge qui est fermée à clef, & qui n'a qu'une petite ouverture à la porte, par laquelle on donne à boire & à manger au fou qui est dedans.

**ACHOU.** f. m. Petit grain qui se fait d'une composition de musc & d'ambre, qui sert à parfumer l'halaine. Sa bafe est une gomme qui se tire d'une decoction epaisse d'un certain arbre qui croit aux Indes. Cet arbre que les Auteurs appellent *kauru*, & qu'au Bresil on nomme

**ETYMOLOGISCHES  
WÖRTERBUCH  
DER  
PREUSSISCHEN SPRACHE.**

[278]

**A.**

**abbai** 'beide' **abbans** acc.: lit. **abū** le. **abi** abg. **oba**: gr. **ἄμφω** lat. **ambō**.

**abasus** V. 294. Gr. **abbas** 'Wagen': entlehnt aus poln. **obóz**; lit. **ābazas** ebenfalls entlehnt 'Heerlager, Heer'.

**aber** Cat. I: das deutsche Wort.

**aboros** V. 228 'Raufe': entlehnt aus p. **obora** 'Viehhof'; lit. **abarà**

'Hofraum' desgl.

**abse** 'Esche' V. 606: le. **apse** abg. **osina** ahd. **aspa** lit. **apuszis**.

**ader**, **adder**: aus dem deutschen; wie im ostpreuss. Dialekt für 'oder' und 'aber'.

**addle** V. 596 'Tanne': lit. **ėglė** aus \***edlė**, poln. **jodła**.

**aglo** V. 470 'Regen' für **aglu**: gr. **ἀχλὺς** 'Nebel, Wolke': **akh-** neben

**ak-**? in **āklas** 'blind', lat. **aquilo** 'Nordwind'.

**ayculo** V. 470 'Nadel': gr. **αἰχλὸι** **αἱ γωνίαι** τοῦ βέλους Hesych.

abg. **igla** wäre nur zu vgl. wenn es für **ayculo** verschrieben.

**ains** 'ein', aot. **ainan**, gen. **ainassel**, dat. **ainesmu** fem. **ainā**: lit. **vėnas** abn. **inb** lat. **oinos** gr. **οἶνός** 'Eins auf dem Würfel' aot. **ains**.

[279]

ainonts 'Jemand': acc. ainontin, gen. ainontsi, dat. ainontsmu: lit. vēnuntas 'einzig', vēnunta 'jemals'. niainonts acc. niainonton 'niemand'.  
erains dat. erainesmu 'Jeder'. er 'bis': le. ikwīns 'jeder', ik 'bis'.  
niains, niainan; gen. niainessa; niainā f. 'kein'.  
ainat 'allezeit': lit. vēnāt 'einzig, nur'. terains 'allein'.  
aysmis V. 363 'Spieß': lit. ėszmas le. ėsmis; 'hölzerner Bratspieß', gr. αἰχμή.  
aytegenis V. 363 'Kleinspecht': Pierson vermutet 'Spitzenspecht' (mhd. kleine 'Spitze' und vergleicht lit. ėtis 'Spitzel' (so wäre dann Kurschat's jėtis zu lesen). Möglich auch zu lett. aita 'Schaf': ai. etaša 'bunt, Antilope', ved. etā 'schnelles Tier'.  
aketes V. 255 'Eggen': lit. akėtės, akėczios, ođiva Hes; lat. occa camb. ocet ahd. eqida.

## D U R.

parceque sa chair est plus ferme que celle des autres pèches.

**DURANDAL**, f. m. est le nom de l'épée de Roland Chevalier Heros de l'Arioste. On s'en sert en cette phrase proverbiale : pour expliquer qu'une viande est fort dure, on dit que c'est *durandal*, l'épée de Roland.

**DURANT**, Preposition. Pendant, tandis qu'une chose subsistera. *Durant* qu'on est dans l'emploi il faut faire sa fortune. Il faut faire les provisions *durant* l'été. N'ai-je pas su que *durant* votre voyage, vous avez été de la plus belle humeur du monde? **LE T. PORTUG.** Si jamais la voye du Chretien est étroite, c'est *durant* les persecutions. **FL.** Ce mot se met quelquefois après le nom qu'il regit. J'y été malade six ans *durant*. On lui a assigné une pension fa vie *durant*.

**DURCIR**, v. act. & n. Rendre dur. On *durcit* le fer à force de le battre. Le soleil *durcit* l'ambre, *durcit* les perles. L'air *durcit* le corail. Un œuf trop cuit se *durcit*. La viande *durcit* pendant la gelée.

**DURCIR**, se dit aussi figurément de l'esprit, & signifie, Rendre ferme. Cela *durcit* l'esprit.

**DURCI**, se. part.

**DURE**, f. f. On ne le dit qu'en cette phrase : Coucher sur la *dure*; c'est-à-dire, sur la terre, ou sans matelas.

**DURE**, f. f. Perseverance des choses dans leur être; temps mesuré par la subsistance de quelque chose. Le temps est défini par les Philosophes, La *durée* d'un mouvement. Dieu a promis à ses élus une gloire d'éternelle *durée*. Cette fougue est trop violente, elle ne sera pas de *durée*. Nous ne jouissons de la vie qu'à mesure que nous la perdons : chaque moment en abrège la *durée*. **MORALE DE P.** On juge de la *durée* du temps selon la disposition où l'on se trouve : celui qui est accablé de tristesse s'ennuie de la *durée* du temps, parcequ'elle lui est pénible, & qu'il y fait plus d'attention. **MALB.** La *durée* des heures, au regard de l'ennui, & du chagrin, se fait plus sentir que celle des années. **BOU.** Les Dieux ne sont immortels que par la *durée* de leurs plaisirs. **DAC.** Les passions veulent être conduites avec art pour en étendre la *durée*, afin qu'elles ne s'épuisent pas trop tôt. **LE CH. D'H.** La *durée* de nos passions ne dépend pas plus de nous, que la *durée* de notre vie. **LA ROCH.** Je ne mesure pas ma vie par la *durée* du temps; mais par la *durée* de la gloire.

**BOU.** Les Dames pour l'ordinaire trouvent leurs maris de longue *durée*. **LE CH. DE M.** Cette femme s'est mis dans l'esprit d'égaliser la *durée* de son deuil à celle de sa vie, & a choisi cette triste, & fatigante voye pour acquiescer de la reputation. **M. ESP.** Il n'est rien ici bas d'éternelle *durée*. **MALB.** Cette tendre amitié par tant de fois jurée, Qui devoit surpasser les siècles en *durée*, A la fin s'est éteinte. **VOI.**

**DUREMENT**, adv. D'une manière dure. Il a été traité *durement* par son Maître, &c. Ces Religieux sont couchés bien *durement*. Luther s'est exprimé *durement* en parlant de la predestination. **CL.** Il ne faut pas dire *durement* les choses dures. **NIC.**

**DURE-MERE**, f. f. Terme d'Anatomie. C'est la membrane qui enveloppe le cerveau. Membrane du cerveau grossière & dure, qui est attachée à l'os du crâne.

**DURER**, v. n. Subsister pendant quelque espace de temps. Une femme se défait de son galant quand elle veut; mais il faut qu'elle garde son mari tant qu'il *dure*. **LE CH. DE M.** L'absence, pour peu qu'elle *dure*, nuit à l'amitié aussi bien qu'à l'amour. **ID.** Rien n'approche de l'ennui que donne une passion qui *dure* trop.

## D U R.

**ST. EV.** Un engagement qui doit *durer* jusqu'à la mort, ne se doit jamais faire qu'avec de grandes précautions. **MOI.** L'amour *durait* un monde au bon vieux temps. **MAR.** Le monde a déjà *duré* cinq à six mille ans. Ce que Malherbe écrit *dure* éternellement. **MALB.**

**DURER**, se dit aussi de ce qui est solide; qui subsiste long temps; qui est fort; qui s'use difficilement. Le drap d'Espagne est d'un bon *user*, il *dure* long temps. Ce meuble *durera* un siècle, cela *durera* jusqu'au bout.

**DURER**, avec la negative, signifie, Résister, souffrir quelque mal, quelque peine, quelque incommodité. On ne *sçait* *durer* avec cette femme-là, tant elle est criarde. Je ne puis plus *durer* avec cette colique. On ne *sçait* *durer* à la maison par ce beau temps-là. On ne *sçait* *durer* en ce poste, il est trop exposé à l'artillerie. On n'y *dure* point, on n'y peut tenir. **MOI.** Pensez-vous que je puisse *durer* avec toutes ses turpitudes? **ID.** On dit aussi ne pouvoir *durer* de chaud & de froid &c. pour dire, être extrêmement incommodé du chaud, du froid &c.

On dit proverbialement, Il faut faire vie qui *dure*, lorsqu'on parle de menage, & qu'on veut empêcher la dissipation. On dit d'un niais qui n'a point vu le monde, qu'il est bien neuf, qu'il *durera* long temps. On dit que le temps *dure* à quelcon; pour dire, qu'il lui ennuie, qu'il attend quelque chose avec grande impatience. On dit aussi, qu'un homme ne *sçait* *durer* en sa peau, qu'il ne peut *durer* en place; pour dire, qu'il est inquiet & inconstant.

**DURET**, **ETTE**, adj. diminutif de *dur*. L'oïseau étoit bon, mais il étoit un peu *duret*. Il est du fil-le bas.

**DURETÉ**, f. f. Solidité, qualité de ce qui est dur. C'est la résistance que font les corps à la division, & à la separation des parties dont ils sont composés. Le repos, la liaison, & la contiguité des parties qui se touchent immédiatement sans le mouvoir, fait la *dureté* des corps. **ROH.** On a trouvé l'invention de donner au plâtre la *dureté* du marbre. La *dureté* des diamans fait la meilleure partie de leur valeur. Les viandes gelées ont de la *dureté*.

En termes de Medecine on appelle *dureté*, certaines tumeurs ou callosités de corps & d'humeurs qui s'endurcissent. On sent des *duretés* dans les mains des hommes de travail. **ABLAN.** On dit aussi, une *dureté* de ventre, quand on est constipé; une *dureté* d'oreille, quand on est presque sourd.

**DURETÉ**, se dit figurément en choses spirituelles & morales, & signifie, Indocilité, insensibilité, cruauté. La *dureté* du cœur des Juifs obligea Moïse à leur permettre le divorce. **LA MAI.** Il a une *dureté* d'esprit qui fait qu'il ne peut rien comprendre; une *dureté* de cœur, qui fait qu'il n'aime personne. Nous joindrons nos forces pour attaquer la *dureté* de son humeur. **MOI.** Le cœur, & le temperament des Stoïciens ne s'accoutumoit pas toujours de la *dureté* philosophique dont ils faisoient profession. **OR. M.** Les opinions de Senèque ont trop de *dureté*. **ST. EV.** Un peu de *dureté* sied bien aux grandes âmes. **CORN.** Pensez-vous que je vous pardonne toutes les *duretés* que vous m'avez dites? Vous avez eu la *dureté* de me dire que la conversation de cette Dame vous avoit plu. **LE T. PORTUG.** La *dureté* des termes choque d'autant plus, qu'elle enferme quelque sorte d'indifférence, & de mépris. **NIC.**

Je renonce à la vanité  
De cette dureté farouche,  
Que l'on appelle fermeté. **QUIN.**

## D U R.

parceque sa chair est plus ferme que celle des autres pèches.

**DURANDAL**, f. m. est le nom de l'épée de Roland Chevalier Heros de l'Arioste. On s'en sert en cette phrase proverbiale : pour expliquer qu'une viande est fort dure, on dit que c'est *durandal*, l'épée de Roland.

**DURANT**, Preposition. Pendant, tandis qu'une chose subsistera. *Durant* qu'on est dans l'emploi il faut faire sa fortune. Il faut faire les provisions *durant* l'été. N'ai-je pas su que *durant* votre voyage, vous avez été de la plus belle humeur du monde? **LE T. PORTUG.** Si jamais la voye du Chretien est étroite, c'est *durant* les persecutions. **FL.** Ce mot se met quelquefois après le nom qu'il regit. J'y été malade six ans *durant*. On lui a assigné une pension fa vie *durant*.

**DURCIR**, v. act. & n. Rendre dur. On *durcit* le fer à force de le battre. Le soleil *durcit* l'ambre, *durcit* les perles. L'air *durcit* le corail. Un œuf trop cuit se *durcit*. La viande *durcit* pendant la gelée.

**DURCIR**, se dit aussi figurément de l'esprit, & signifie, Rendre ferme. Cela *durcit* l'esprit.

**DURCI**, se. part.

**DURE**, f. f. On ne le dit qu'en cette phrase : Coucher sur la *dure*; c'est-à-dire, sur la terre, ou sans matelas.

**DURE**, f. f. Perseverance des choses dans leur être; temps mesuré par la subsistance de quelque chose. Le temps est défini par les Philosophes, La *durée* d'un mouvement. Dieu a promis à ses élus une gloire d'éternelle *durée*. Cette fougue est trop violente, elle ne sera pas de *durée*. Nous ne jouissons de la vie qu'à mesure que nous la perdons : chaque moment en abrège la *durée*. **MORALE DE P.** On juge de la *durée* du temps selon la disposition où l'on se trouve : celui qui est accablé de tristesse s'ennuie de la *durée* du temps, parcequ'elle lui est pénible, & qu'il y fait plus d'attention. **MALB.** La *durée* des heures, au regard de l'ennui, & du chagrin, se fait plus sentir que celle des années. **BOU.** Les Dieux ne sont immortels que par la *durée* de leurs plaisirs. **DAC.** Les passions veulent être conduites avec art pour en étendre la *durée*, afin qu'elles ne s'épuisent pas trop tôt. **LE CH. D'H.** La *durée* de nos passions ne dépend pas plus de nous, que la *durée* de notre vie. **LA ROCH.** Je ne mesure pas ma vie par la *durée* du temps; mais par la *durée* de la gloire.

**BOU.** Les Dames pour l'ordinaire trouvent leurs maris de longue *durée*. **LE CH. DE M.** Cette femme s'est mis dans l'esprit d'égaliser la *durée* de son deuil à celle de sa vie, & a choisi cette triste, & fatigante voye pour acquiescer de la reputation. **M. ESP.** Il n'est rien ici bas d'éternelle *durée*. **MALB.** Cette tendre amitié par tant de fois jurée, Qui devoit surpasser les siècles en *durée*, A la fin s'est éteinte. **VOI.**

**DUREMENT**, adv. D'une manière dure. Il a été traité *durement* par son Maître, &c. Ces Religieux sont couchés bien *durement*. Luther s'est exprimé *durement* en parlant de la predestination. **CL.** Il ne faut pas dire *durement* les choses dures. **NIC.**

**DURE-MERE**, f. f. Terme d'Anatomie. C'est la membrane qui enveloppe le cerveau. Membrane du cerveau grossière & dure, qui est attachée à l'os du crâne.

**DURER**, v. n. Subsister pendant quelque espace de temps. Une femme se défait de son galant quand elle veut; mais il faut qu'elle garde son mari tant qu'il *dure*. **LE CH. DE M.** L'absence, pour peu qu'elle *dure*, nuit à l'amitié aussi bien qu'à l'amour. **ID.** Rien n'approche de l'ennui que donne une passion qui *dure* trop.

## D U R.

**ST. EV.** Un engagement qui doit *durer* jusqu'à la mort, ne se doit jamais faire qu'avec de grandes précautions. **MOI.** L'amour *durait* un monde au bon vieux temps. **MAR.** Le monde a déjà *duré* cinq à six mille ans. Ce que Malherbe écrit *dure* éternellement. **MALB.**

**DURER**, se dit aussi de ce qui est solide; qui subsiste long temps; qui est fort; qui s'use difficilement. Le drap d'Espagne est d'un bon *user*, il *dure* long temps. Ce meuble *durera* un siècle, cela *durera* jusqu'au bout.

**DURER**, avec la negative, signifie, Résister, souffrir quelque mal, quelque peine, quelque incommodité. On ne *sçait* *durer* avec cette femme-là, tant elle est criarde. Je ne puis plus *durer* avec cette colique. On ne *sçait* *durer* à la maison par ce beau temps-là. On ne *sçait* *durer* en ce poste, il est trop exposé à l'artillerie. On n'y *dure* point, on n'y peut tenir. **MOI.** Pensez-vous que je puisse *durer* avec toutes ses turpitudes? **ID.** On dit aussi ne pouvoir *durer* de chaud & de froid &c. pour dire, être extrêmement incommodé du chaud, du froid &c.

On dit proverbialement, Il faut faire vie qui *dure*, lorsqu'on parle de menage, & qu'on veut empêcher la dissipation. On dit d'un niais qui n'a point vu le monde, qu'il est bien neuf, qu'il *durera* long temps. On dit que le temps *dure* à quelcon; pour dire, qu'il lui ennuie, qu'il attend quelque chose avec grande impatience. On dit aussi, qu'un homme ne *sçait* *durer* en sa peau, qu'il ne peut *durer* en place; pour dire, qu'il est inquiet & inconstant.

**DURET**, **ETTE**, adj. diminutif de *dur*. L'oïseau étoit bon, mais il étoit un peu *duret*. Il est du fil-le bas.

**DURETÉ**, f. f. Solidité, qualité de ce qui est dur. C'est la résistance que font les corps à la division, & à la separation des parties dont ils sont composés. Le repos, la liaison, & la contiguité des parties qui se touchent immédiatement sans le mouvoir, fait la *dureté* des corps. **ROH.** On a trouvé l'invention de donner au plâtre la *dureté* du marbre. La *dureté* des diamans fait la meilleure partie de leur valeur. Les viandes gelées ont de la *dureté*.

En termes de Medecine on appelle *dureté*, certaines tumeurs ou callosités de corps & d'humeurs qui s'endurcissent. On sent des *duretés* dans les mains des hommes de travail. **ABLAN.** On dit aussi, une *dureté* de ventre, quand on est constipé; une *dureté* d'oreille, quand on est presque sourd.

**DURETÉ**, se dit figurément en choses spirituelles & morales, & signifie, Indocilité, insensibilité, cruauté. La *dureté* du cœur des Juifs obligea Moïse à leur permettre le divorce. **LA MAI.** Il a une *dureté* d'esprit qui fait qu'il ne peut rien comprendre; une *dureté* de cœur, qui fait qu'il n'aime personne. Nous joindrons nos forces pour attaquer la *dureté* de son humeur. **MOI.** Le cœur, & le temperament des Stoïciens ne s'accoutumoit pas toujours de la *dureté* philosophique dont ils faisoient profession. **OR. M.** Les opinions de Senèque ont trop de *dureté*. **ST. EV.** Un peu de *dureté* sied bien aux grandes âmes. **CORN.** Pensez-vous que je vous pardonne toutes les *duretés* que vous m'avez dites? Vous avez eu la *dureté* de me dire que la conversation de cette Dame vous avoit plu. **LE T. PORTUG.** La *dureté* des termes choque d'autant plus, qu'elle enferme quelque sorte d'indifférence, & de mépris. **NIC.**

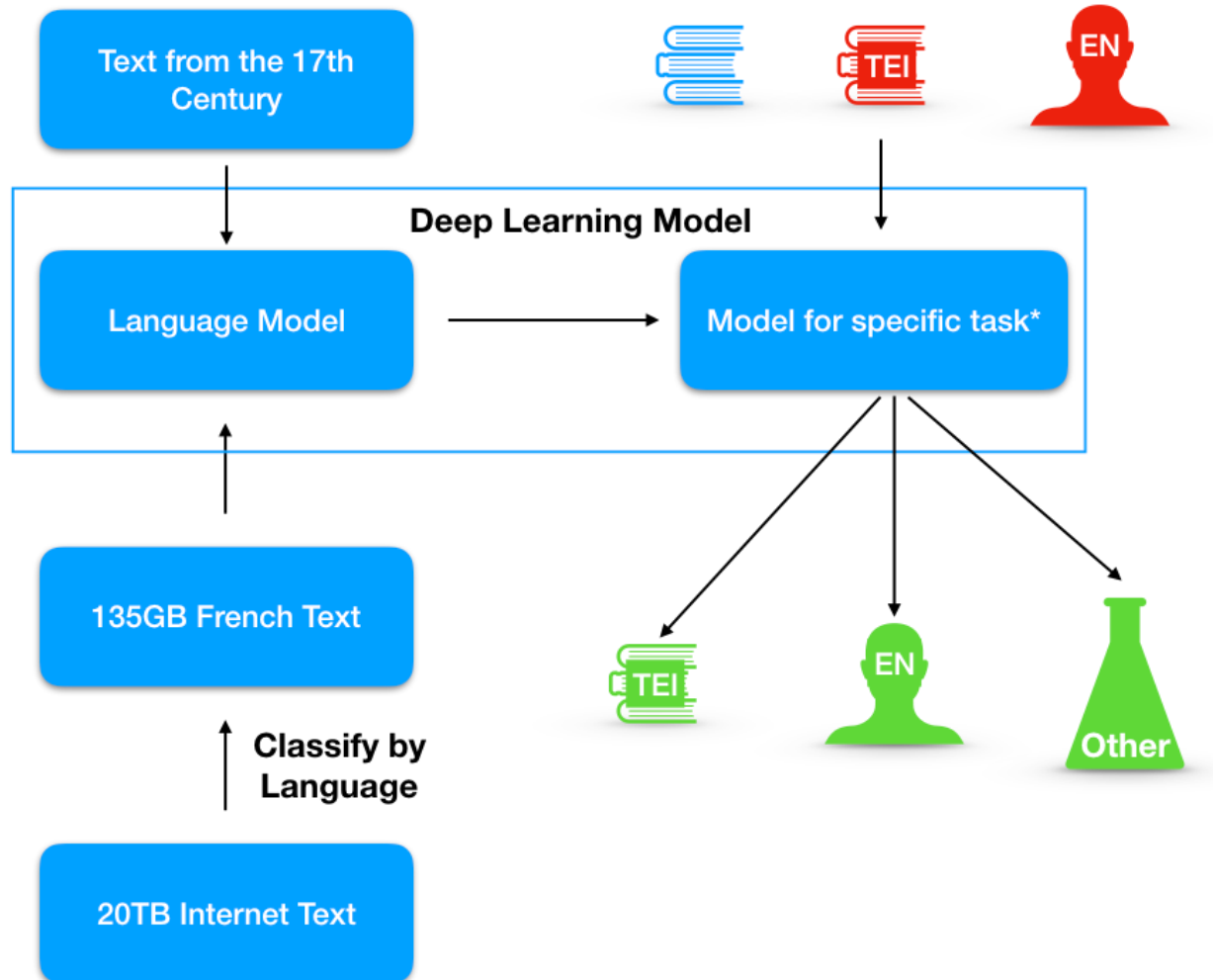
Je renonce à la vanité  
De cette dureté farouche,  
Que l'on appelle fermeté. **QUIN.**

# Experiments: Lexical Entry Model

	<i>Sample 1</i>			<i>Sample 2</i>		
<i>TEI element</i>	<b>Precision</b>	<b>Recall</b>	<b>F1</b>	<b>Precision</b>	<b>Recall</b>	<b>F1</b>
<b>&lt;etym&gt;</b>	87.5	60	71.19	73.68	71.79	72.73
<b>&lt;form&gt;</b>	94.44	92.73	93.58	92.24	96.4	94.27
<b>&lt;pc&gt;</b>	90.91	69.44	78.74	88.97	80.13	84.32
<b>&lt;re&gt;</b>	33.33	9.09	14.29	55.56	22.73	32.26
<b>&lt;sense&gt;</b>	67.65	59.28	63.19	77	76.65	76.84
<b>&lt;xr&gt;</b>	100	80	88.89	100	100	100

**Table 3: Field Level Evaluation of the Lexical Entry Model**

# GROBID: the next generation



\*Changes by task

Pedro Javier Ortiz Suárez, Laurent Romary, Benoît Sagot. Preparing the Dictionnaire Universel for Automatic Enrichment. *10th International Conference on Historical Lexicography and Lexicology (ICHLL)*, Jun 2019, Leeuwarden, Netherlands. [hal-02131598](#)

# Wrapping up

- Dictionaries are cool things
  - But we all share this...
- A great deal of standardisation work has already been done
  - A strong basis for improving interoperability
  - Further convergence work is needed
- Huge expectation around automatic annotation
  - Cf. Basnage: a drop in the legacy ocean (e.g. SIL)
    - But lack of generalisation across dictionaries
  - Future:
    - Families of dictionaries
    - Deep learning
- Towards a wealth of dictionary sources
  - Changing scale for more lexical knowledge
  - Towards a lexical time-space machine => bringing back knowledge to the lexicographic folk

Merci pour votre attention!