



HAL
open science

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 Klamer; Iztok Kosem, Nov 2019, Leiden, Netherlands. hal-02358218

HAL Id: hal-02358218

<https://inria.hal.science/hal-02358218v1>

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 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

ANR project BASNUM

Colleagues involved: Geoffrey Williams, Mohamed Khemakhem (Univ. de Grenoble), Ioana Galleron, Clarissa Stincone (Univ. Sorbone Nouvelle), Benoit Sagot, Laurent Romary, Pedro Ortiz (Inria)

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 *Billettes*; & les *Carmes Dechaussez*, qui ont été reformez des autres. Dans des Theses soutenuës 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 ses 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">Billettes</hi>; & les <seg type="terme_variant" rendition="#i">Carmes Dechaussez</seg>,

qui ont été reformez des autres. Dans des Theses soutenuës 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

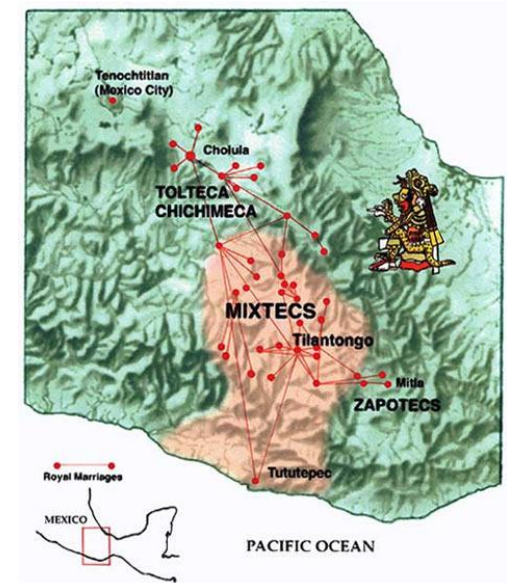
- What source texts? Dictionaries and other texts?
- What links between the DU and *Histoire des ouvrages des savants*
- What lexicographical model between prescription and description?
 - the place of 'good' usage and the 'best' authors
 - the dictionary as a language teaching tool – what users? what 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.

A 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. nahadzi. nahacadzí.

A. interiección. del que se queixa de alguna enfermedad. Hmij.

A. interiección desiderantis. ha.

A. interiección ridentis. ha ha ha.

A. interiección 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. yodadzí siña fiñu, dzutu ñohofina fiñu, dzutu yocuvui dzini, yocuvui nuu.

Abad. ser. yondadzini siñañuhu, f. cond. ñohondi siñañuhu. f. coho. yocuvuidzinindi. yocuvui nuundi.

Abadia dignidad. f. f. indadzi siñañuhu, f. f. siñohu siñañuhu, f. f. sicuvinuu.

Abadefa, prelada de monjas. fiñaha yondacañaha yuq dzehe. fiñahayocuvuidzini.

Abadejo efcaraajo. tenuyuq.

Abahar algo con el huelgo. yochidzo yoco yuhundi yotaa yoco yuhundi.

Abahar así cosa honda. yotniño yoco yuhundi.

Abahada cosa así. f. así cohondaa yoco. f. f. sinucoho yoco yuhu f. f. siná codzo yoco yuhu.

Abahar algo poniendolo al baho que fale dela olla. yodzaqnindi. yodzacuidzindi yoco.

Abahar algo dentro de alguna olla para que se cogerue con el baho y no se pudra. yodzacaanuuyocondi.

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. yodzaqhúdi yoco.

Abahar aposento. yodzacuidzindi huahi. yotniño yocondi. yondadzacuidzindi. yondadzainindi.

Abahado aposento. huahi ninucui ñe huidzi. huahi ninucoho yoco. huahi ninacuidzi. huahi nindu-vui ini.

Abalançarfe, hechándose por los suelos. yocoo cavuan dayédi, yofacavuandayendi. f. q. cavua.

Abalançarfe de arriba a baxo. yon diyoninondi.

Abalançarfe metiéndose entre otros. yofánudzavuadi. f. q. nanu. yofivanudzavuandi. f. q. nanu, yofivui cuinomañundi, yocai nana dzavuandi. f. quivui. yofivuidu utnahandi.

Aballar mouer có dificultad. dzuz

folio 1r column 1

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 quando se me ha olvidado algo. dzi. vt. nahadzi. nahacadzí.

A. interiección. del que se queixa de alguna enfermedad. Hmii.

A. interiección desiderantis. ha.

A. interiección ridentis. ha ha ha.

A. interiección 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. yodadzí siña fiñu, dzutu ñoho siña fiñu, dzutu yocuvui dzini, yocuvui nuu.

Abad. ser. yondadzini siñañuhu, f. cond. ñohondi siñañuhu. f. coho. yocuvuidzinindi. yocuvui nuundi.

Abadia dignidad. sasindadzi siñañuhu, f. f. sasíño siñañuhu, f. f. sasucuvuu.

Abadesa, prelada de monjas. fiñaha yondacañaha yuq dzehe. fiñahayocuvuidzini.

Abadejo escaraajo. tenuyuq.

Abahar algo con el huelgo. yochidzo yoco yuhundi yotaa yoco yuhundi.

Abahar así cosa honda. yotniño yoco yuhundi.

Abahada cosa así. f. así cohondaa yoco. f. f. sinucoho yoco yuhu f. f. siná 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 conserue con el baho y no se pudra. yodzacaanuuyocondi.

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. yodzaqhúdi yoco.

Abahar aposento. yodzacuidzindi huahi. yotniño yocondi. yondadzacuidzindi. yondadzainindi.

Abahado aposento. huahi ninucui ñe huidzi. huahi ninucoho yoco. huahi ninacuidzi. huahi nindu-vui ini.

Abalançarfe, hechándose por los suelos. yocoo cavuandayendi, yofacavuandayendi. f. q. cavua.

Abalançarfe de arriba abaxo. yon diyoninondi.

Abalançarfe metiéndose entre otros. yosanudzavuandi. f. q. nanu. yosanudzavuandi. f. q. nanu, yosivui cuinomañundi, yocai nana dzavuandi. f. quivui. yosivuidu utnahandi.

Aballar mouer con dificultad, dzuz

A h

folio 1r column 2

Abahar así cosa honda. yotniño yoco yuhundi.

Abahada cosa así. f. así cohondaa yoco. f. f. sinucoho yoco yuhu f. f. siná 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 conserue con el baho y no se pudra. yodzacaanuuyocondi.

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. yodzaqhúdi yoco.

Abahar aposento. yodzacuidzindi huahi. yotniño yocondi. yondadzacuidzindi. yondadzainindi.

Abahado aposento. huahi ninucui ñe huidzi. huahi ninucoho yoco. huahi ninacuidzi. huahi nindu-vui ini.

Abalançarfe, hechándose por los suelos. yocoo cavuandayendi, yofacavuandayendi. f. q. cavua.

Abalançarfe de arriba abaxo. yon diyoninondi.

Abalançarfe metiéndose entre otros. yosanudzavuandi. f. q. nanu. yosanudzavuandi. f. q. nanu, yosivui cuinomañundi, yocai nana dzavuandi. f. quivui. yosivuidu utnahandi.

Aballar mouer con dificultad, dzuz

A h

a a: a (del que halla a otro en maleficio)

a a a: a (*interiección 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 huéhica 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 huui: así, así (sonriéndose)

aa ndica huui: 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 q yoo: por ventura alguno

adzi yoo ee ñahando: por ventura alguno de vosotros

ahua: ay, quejándose la mujer

ama: así

A

ama: bien está (otorgando); sí

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

añiñe: palacio

anuhu: abismo; centro de la tierra; infierno, lugar de dañados

anuhu: profundo

anuhu maa: profundo

anuhu naa: infierno, lugar de dañados

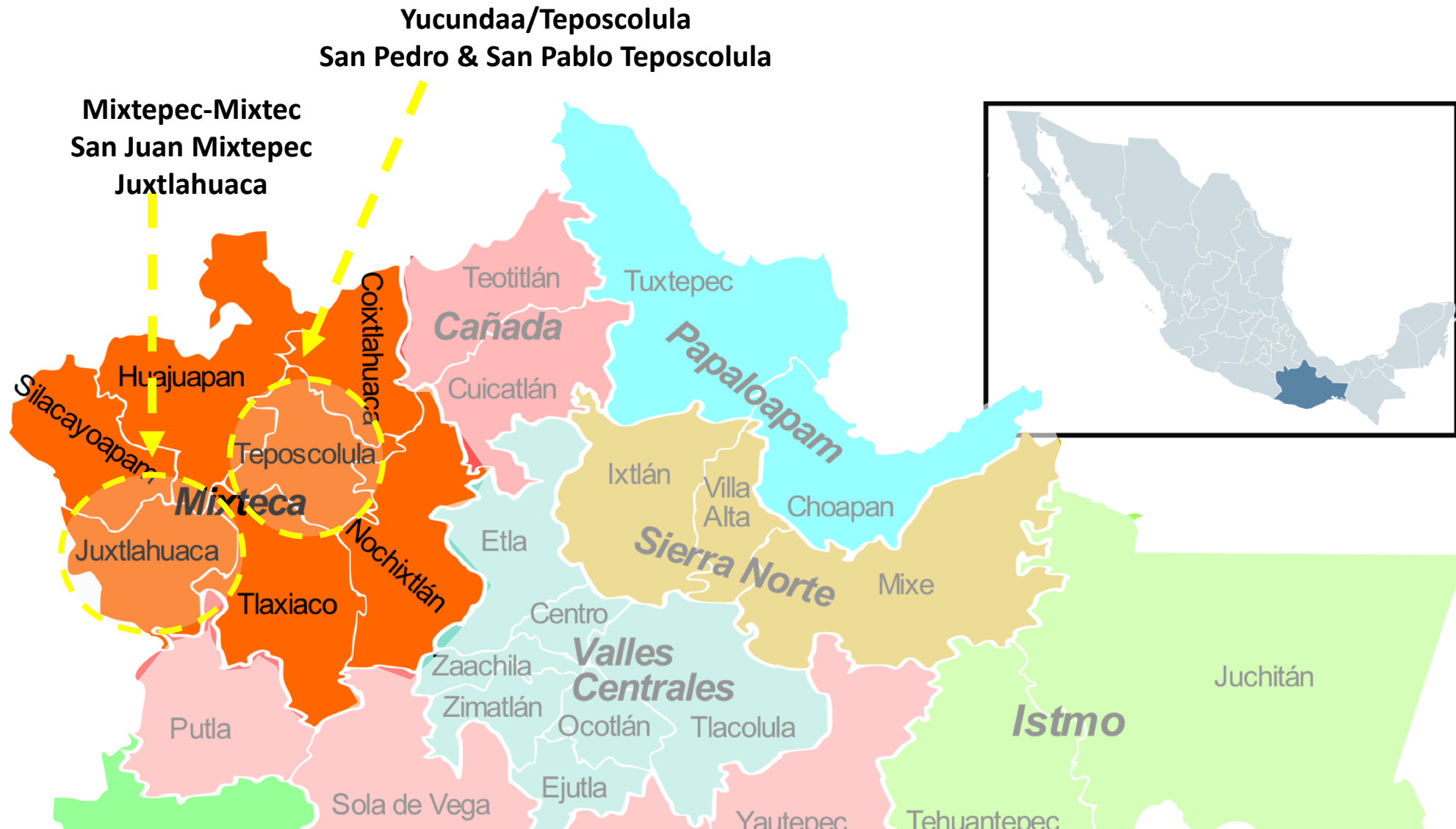
anuhu ndahui: infierno, lugar de dañados

atana: ojalá

atu: amarga cosa; áspero al gusto

aya: amarga cosa; áspero al gusto

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">nɔ̃ʒiká</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 [nɔ̃ʒiká] (*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

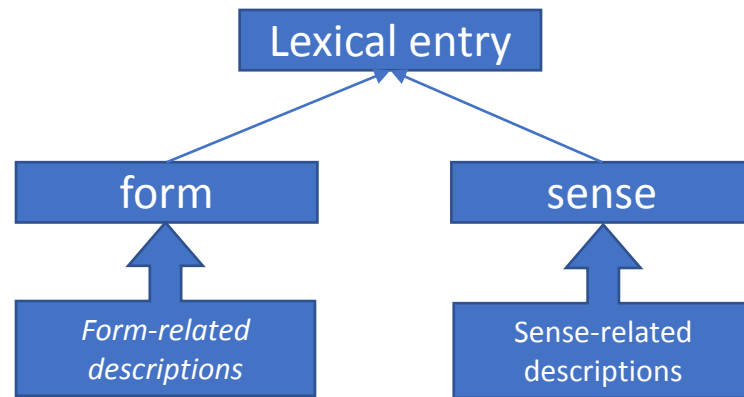
- Large coverage
- All parts of speech
- Build-in polysemy
 - Multiple senses for the same entry
- Referential synonymy

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

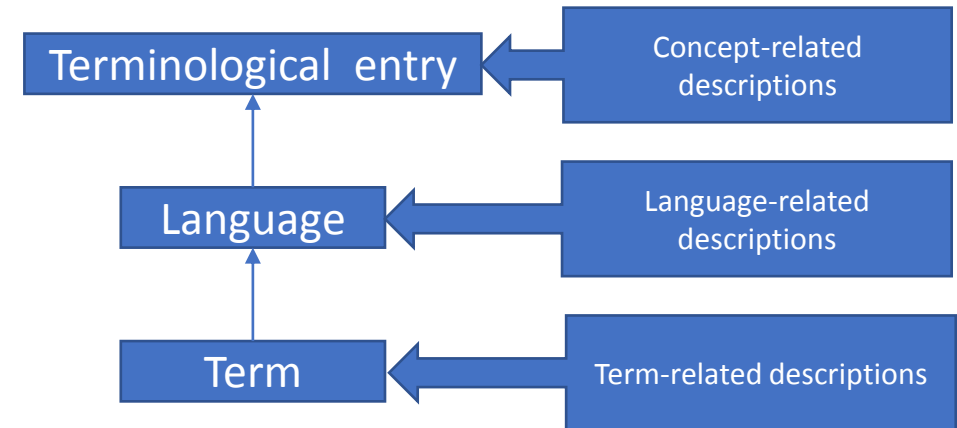
Semasiological models



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



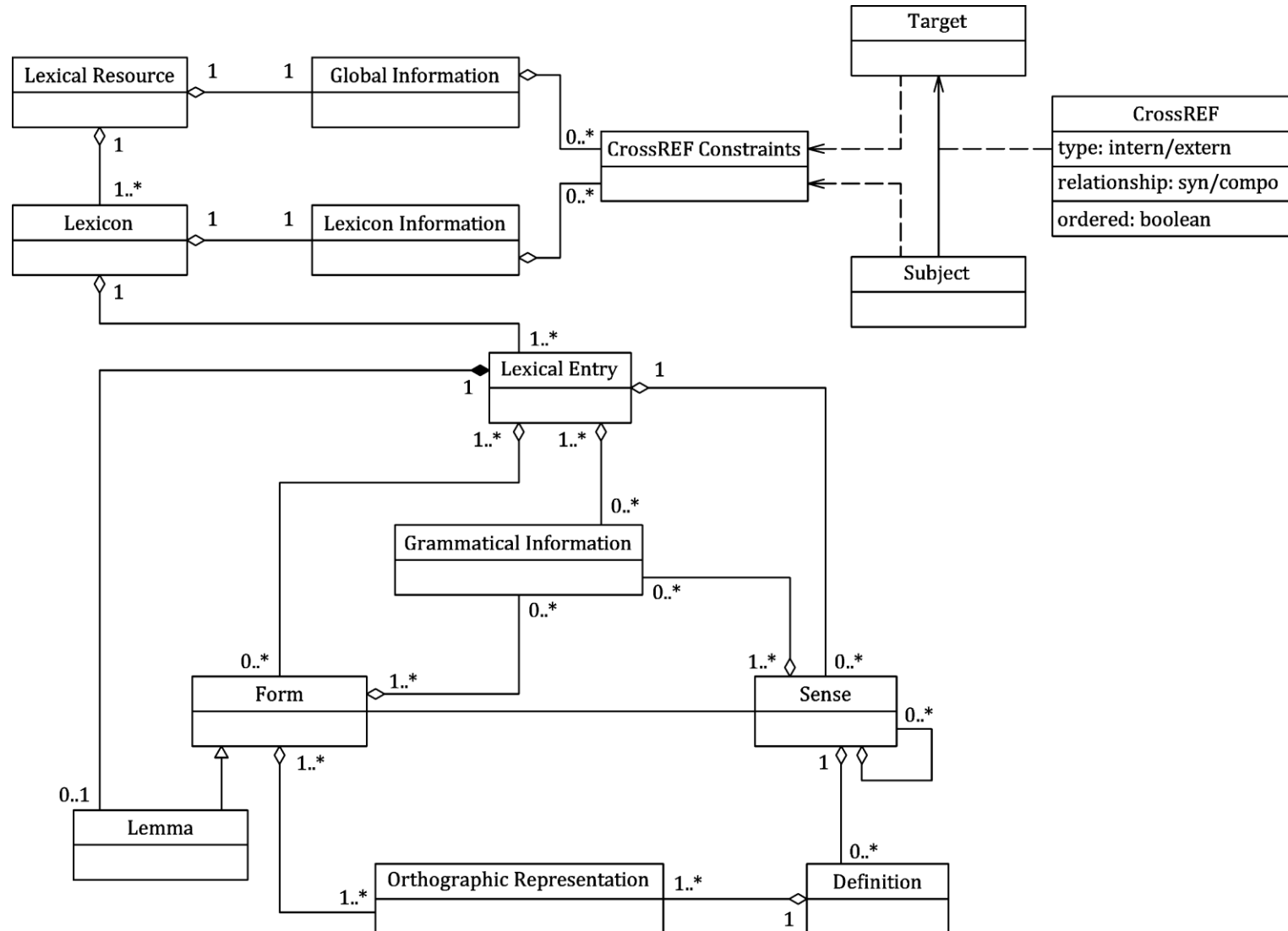
Laurent Romary, Mohamed Khemakhem, Fahad Khan, Jack Bowers, Nicoletta Calzolari, et al.. LMF Reloaded. *AsiaLex 2019: Past, Present and Future*, Jun 2019, Istanbul, Turkey. [hal-02118319](#)

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

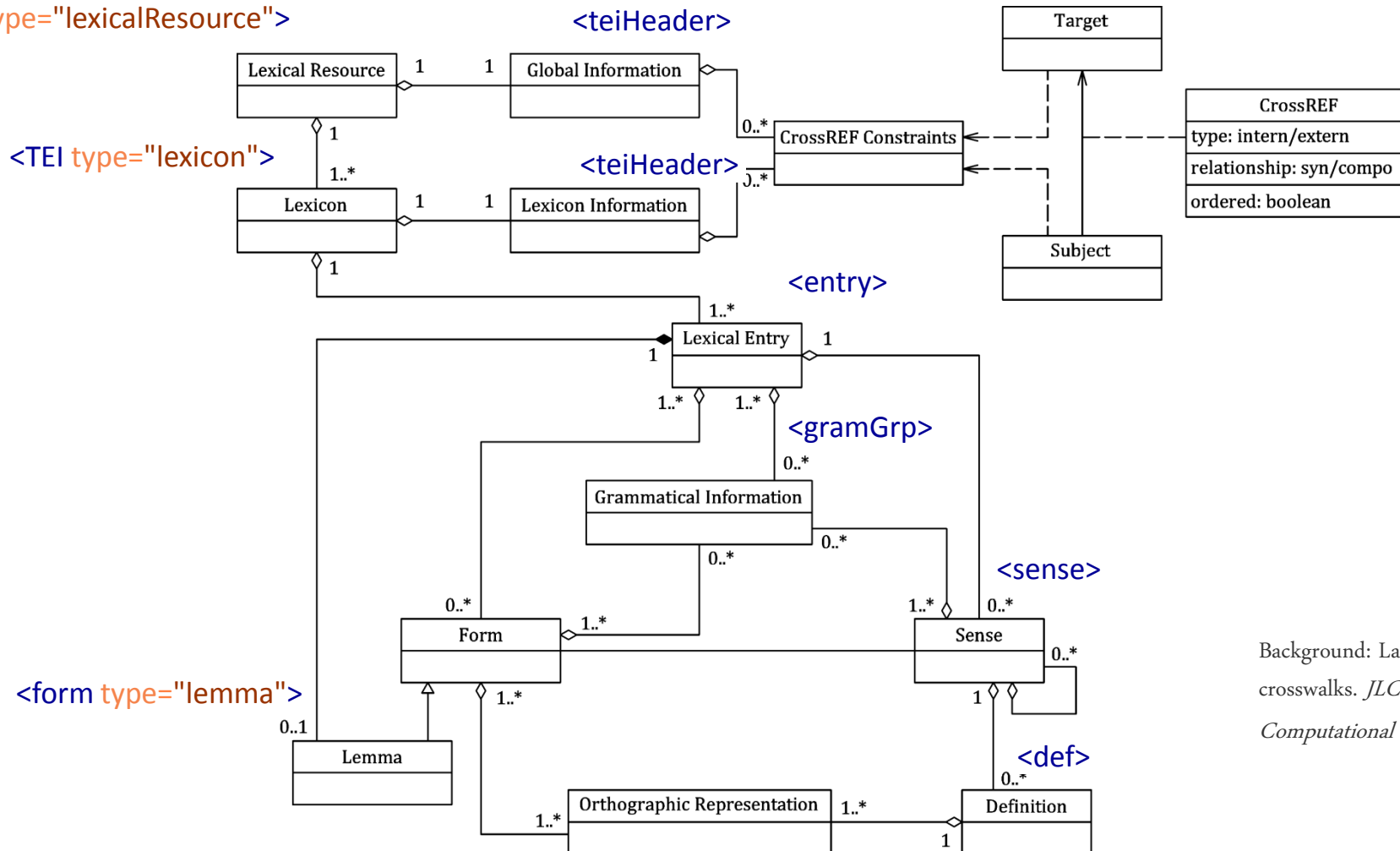
Language resource management — Lexical markup framework (LMF) — Part 1: Core model



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

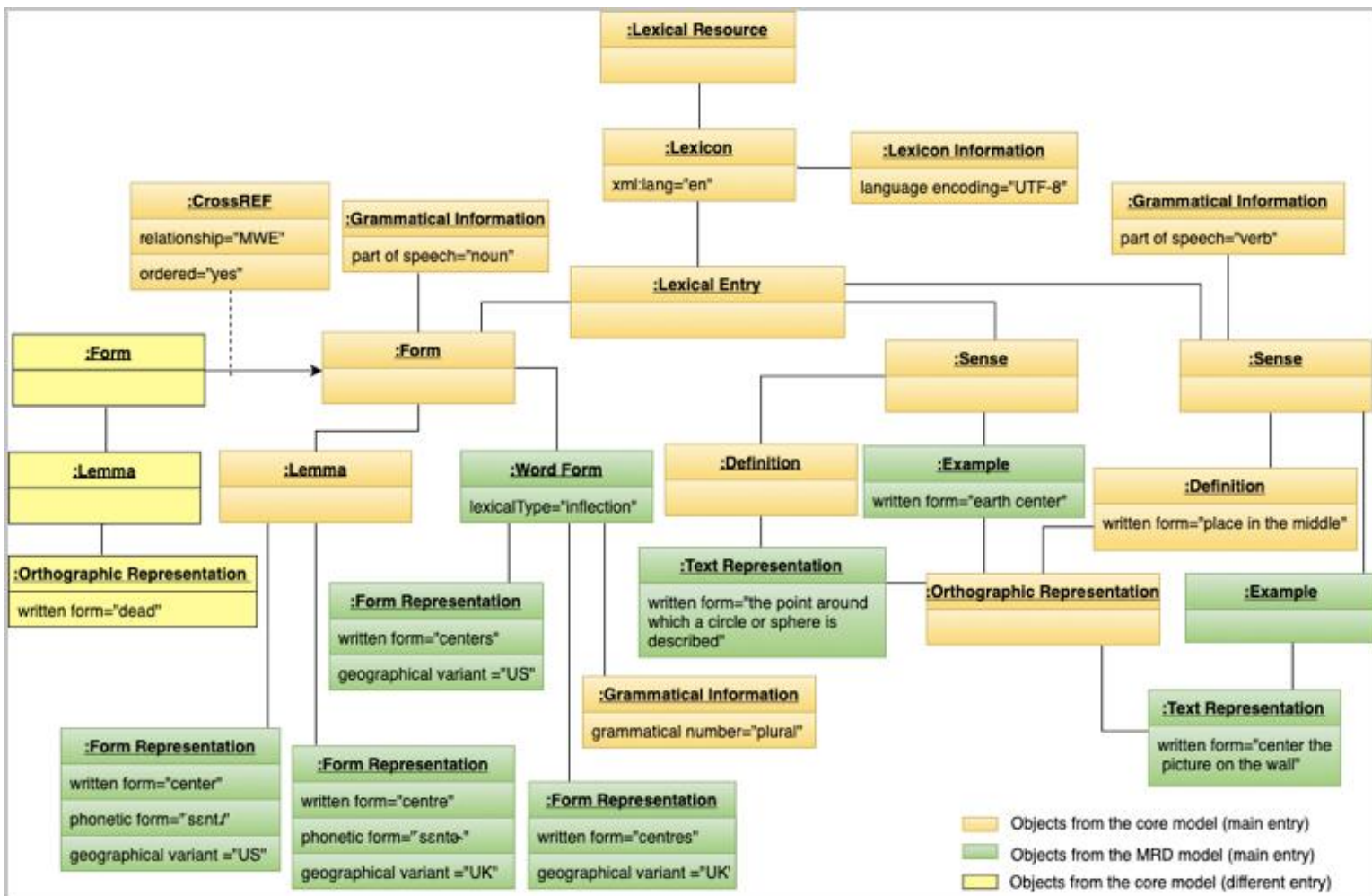
<teiCorpus type="lexicalResource">

<TEI type="lexicon">



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

Example: inflectional (full-form) lexicon



```

<entry>
  <form type="Lemma" xml:id="center_form">
    <orth>center</orth>
    <pron>'sɛntə</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>'sɛntə</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>
  
```

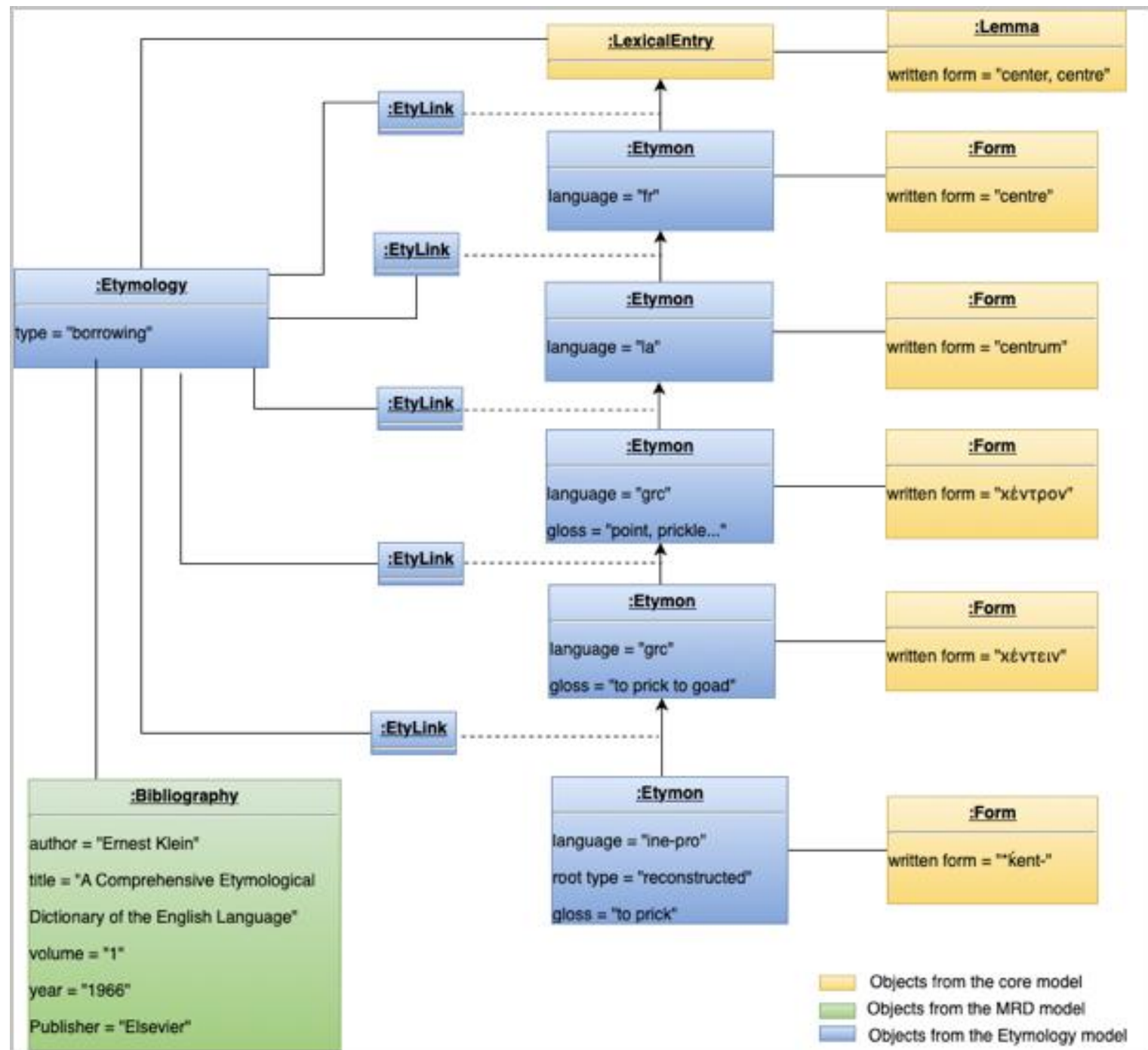
LMF reloaded - Etymology

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’, κέστρα, ‘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*



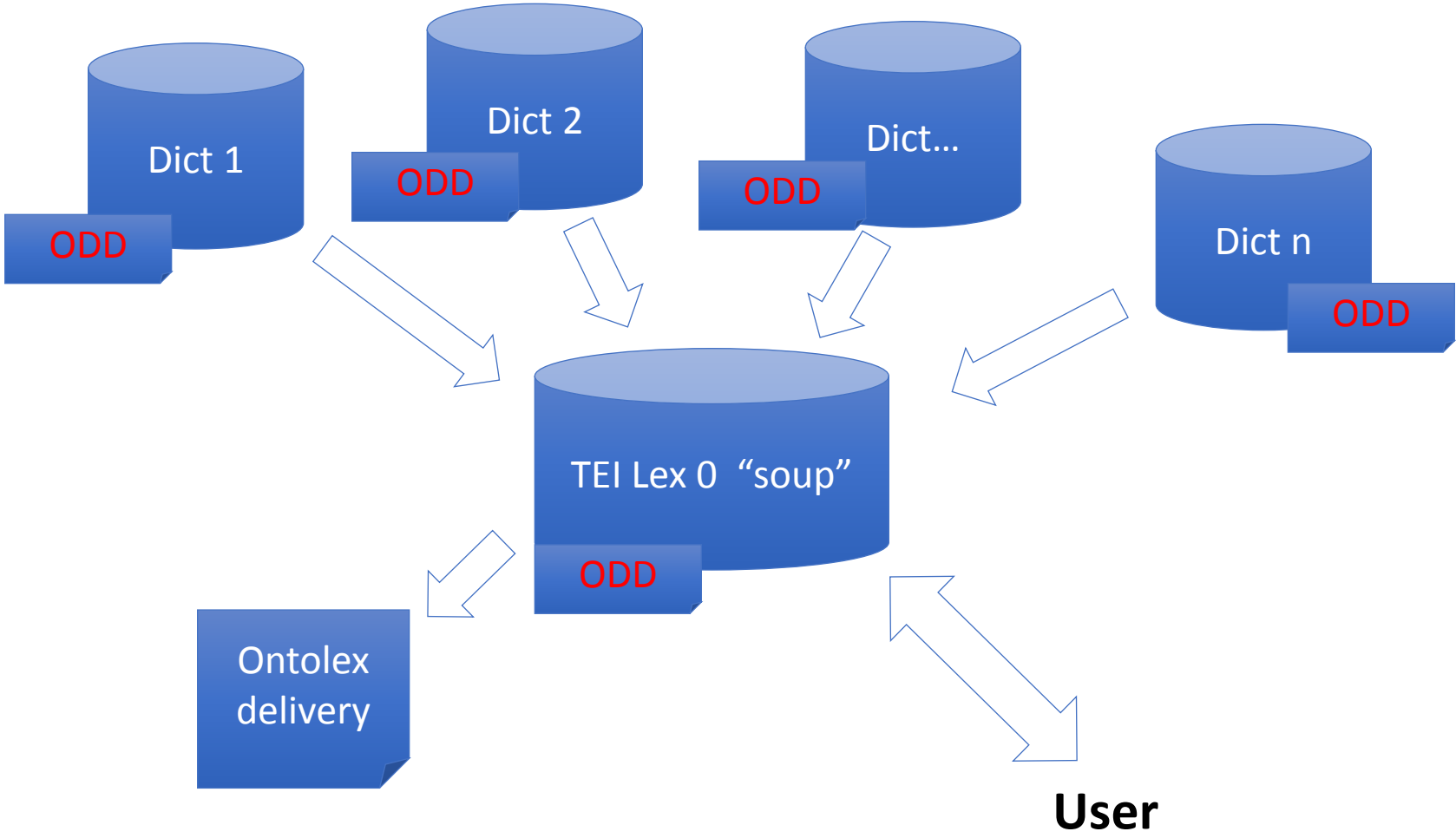
TEI lex 0: tightening the guidelines

- 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

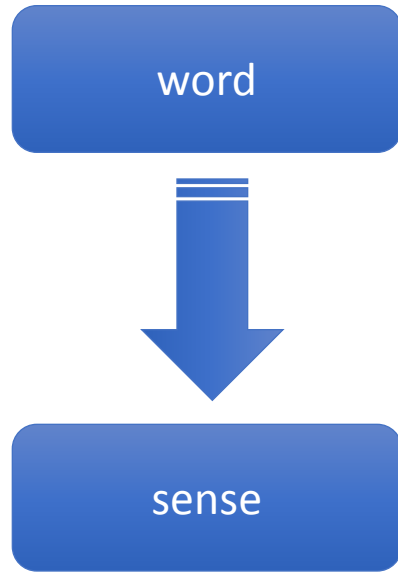
Application – the ELEXIS project

- 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>

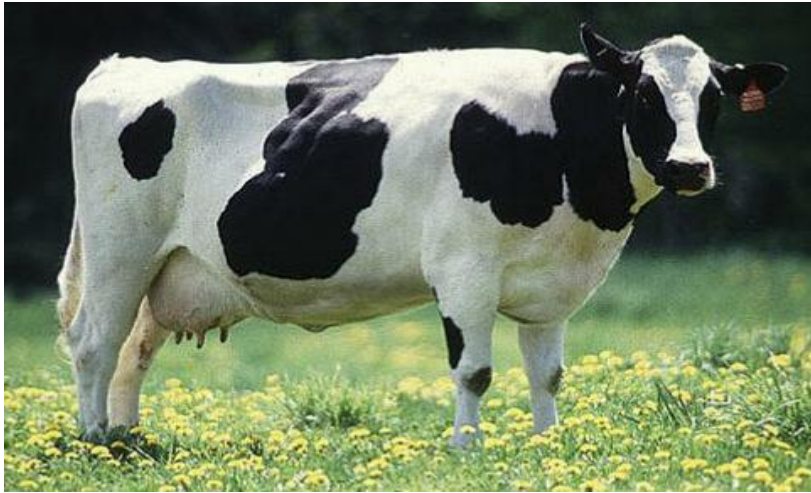
<!-- possibly some shared usg information -->
<entry type="wordForm">
  <form type="lemma">
    <orth expand="Hausaltar">-altar</orth>
    </pc>,</pc>
    <gramGrp>
      <gen value="masculine">der</gen>
    </gramGrp>
  </form>
  <sense>...</sense>
</entry>
<entry type="wordForm">
  <form type="lemma">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

Do we really have to encode all
this manually?

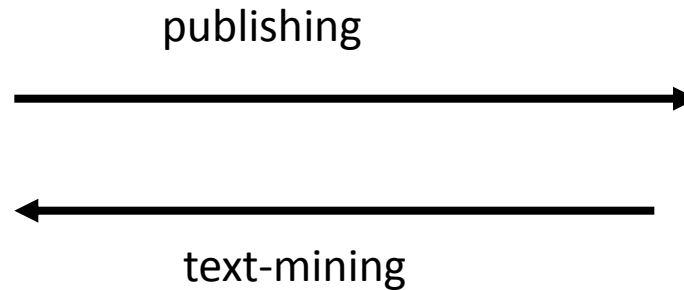
Considering machine learning techniques

- 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'Église catholique en face de la Réforme protestante], de Vatican I (1870) [où fut défini le dogme de l'infaillibilité pontificale], de Vatican II (1962-1965) [où fut définie l'attitude de l'Église 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 pape.

condenser [kɔ̃dɑ̃sɛ] 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. I. Donner son avis, ses conclusions : se prononcer : **on vous demande de conclure.** || Être proham : **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. Prétentions respectives de chacune des parties dans un procès. || Écrit exposant ces prétentions. || 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.

concordant, e adj. Qui s'accorde : **témoignages concordants.**
concordat [kɔ̃kɔrdɑ] 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^{er}. Le concordat entre

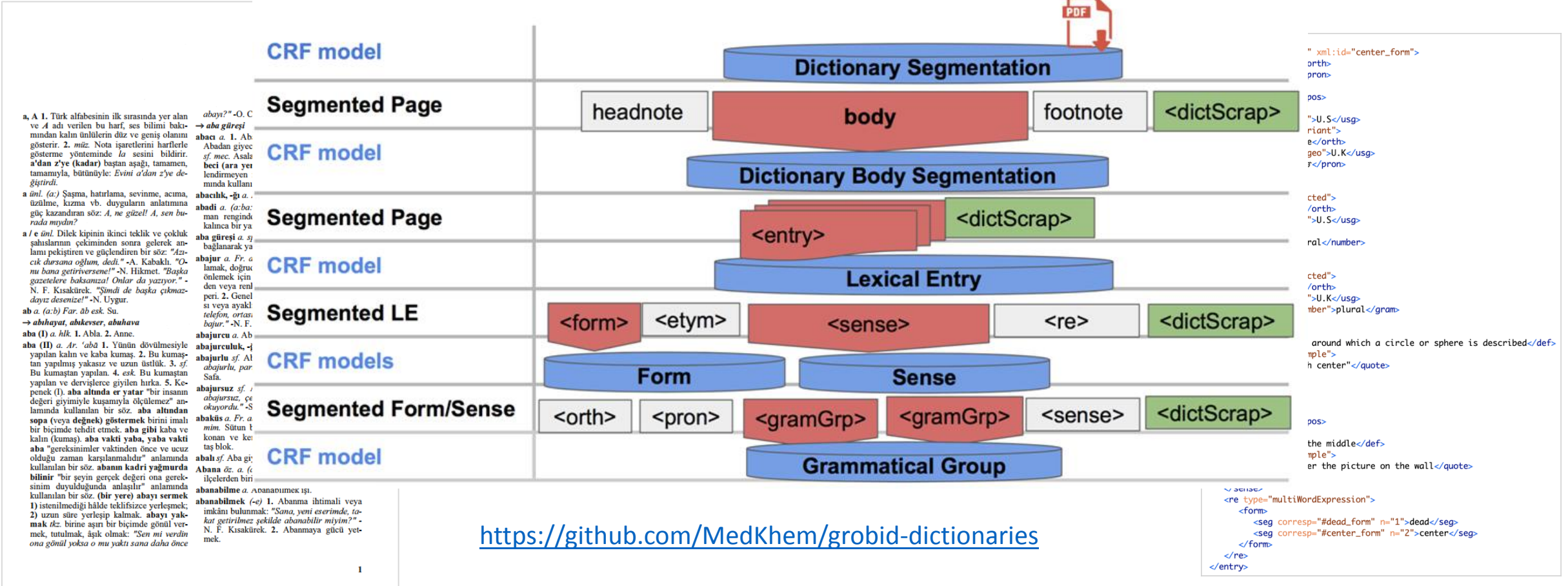
concupiscence n. f. (du lat. concupiscere, désirer). Penchant à jouir des biens terrestres, particulièrement des plaisirs sensuels.
concupiscent [kɔ̃kypisɑ̃] • l.-at] adj. Qui exprime la concupiscence : **regards concupiscent.** || Attaché aux plaisirs sensuels.
concurrer [kɔ̃kyʁɑ̃] v. t. Par un concours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

concours. || Lutte sportive : **concours hippique.** • **Concours général,** concours annuel entre les premiers élèves des classes supérieures des lycées, collèges et écoles normales.
concret [kɔ̃kʁɛ], **ète** [-ɛt] adj. (lat. concretus). Epais, condensé : **Auile concrète** (vieilli). || 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. (conç. 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. État d'un homme et d'une femme qui vivent ensemble sans être mariés. (On dit aussi UNION LIBRE.)



condensation.
condamnabile adj. Qui mérite d'être condamné : acte **condamnabile.**
condamnation [kɔ̃dɑ̃nasjɔ̃] n. f. (lat. condemnatio). Décision d'un tribunal important 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.**
condamnatore 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 [kɔ̃dɑ̃nɛ] 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.** || Barre, murer : **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 [kɔ̃dɑ̃sɛ] 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 :

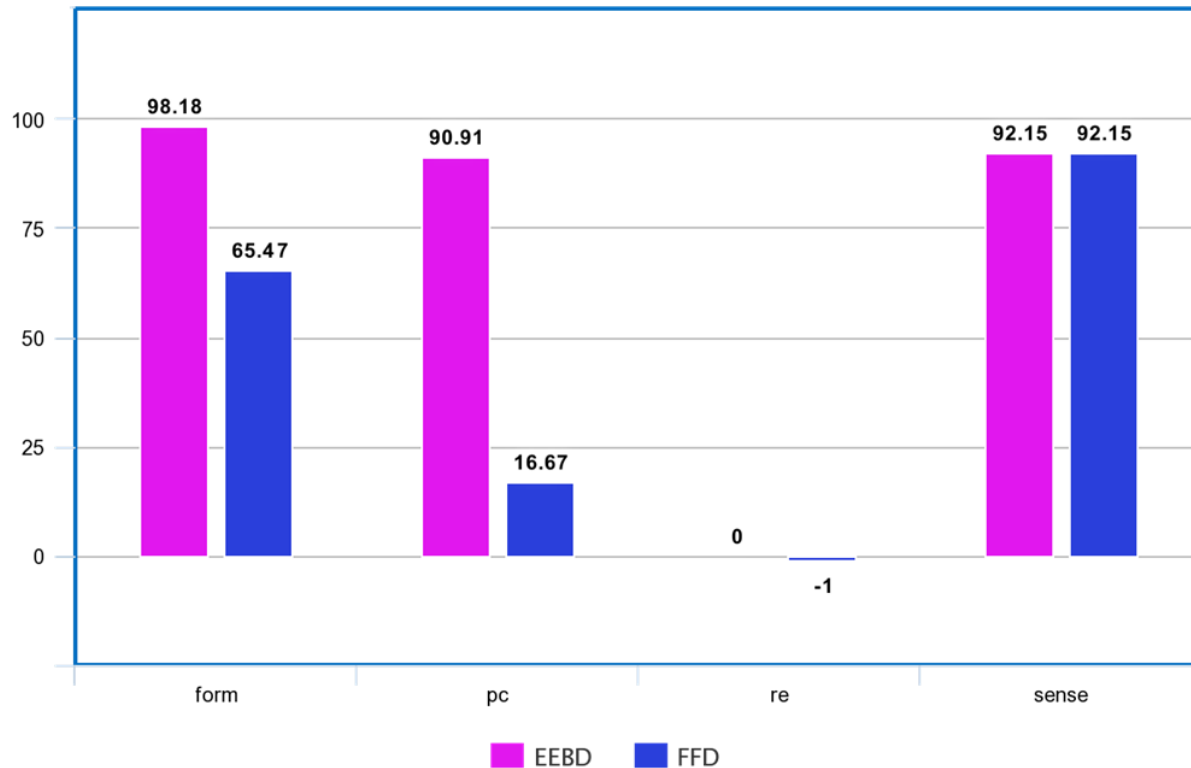
Cascading CRF models



Evaluation: Token Level - F1 Score

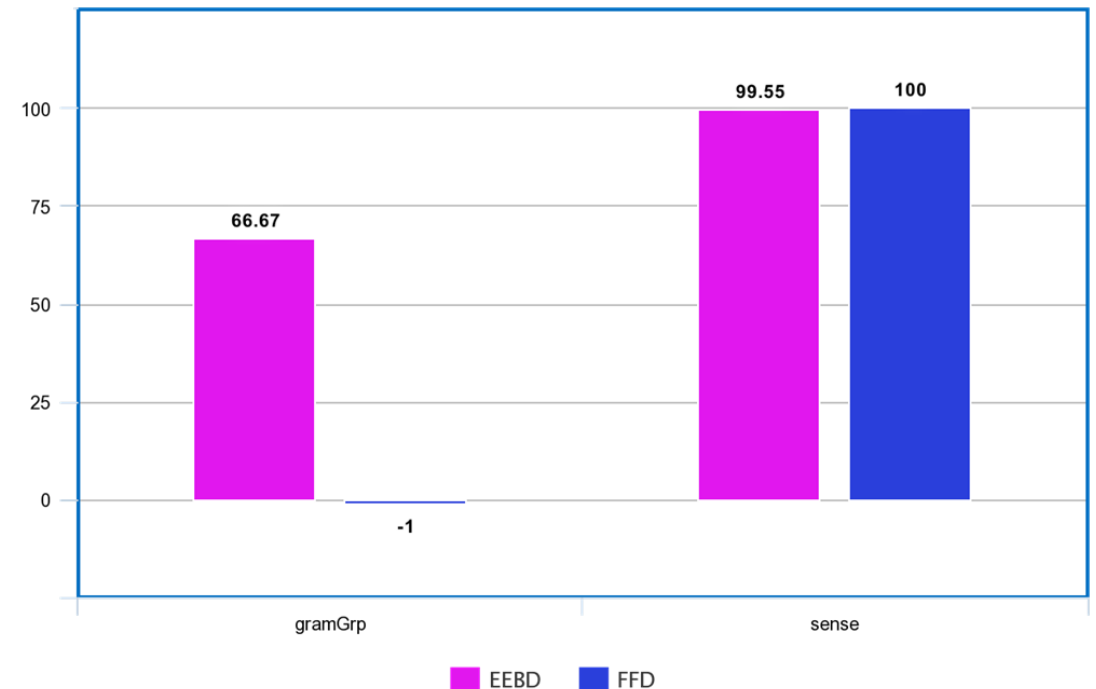
Lexical Entry

- 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-MEZ. f. m. Vieux mot qui signifioit autrefois un *maître*.
CACHER. v. act. Mettre quelque chose en un lieu secret, où il ne puisse être vu ni trouvé par d'autres agents dans la terre, afin que les soldats ne le puissent trouver. Menage après Guyot derive ce mot de *caeter*, qui signifie *chasser, pousser*. On dit en ce sens, que la nature nous a *caché* les trefors, les plus merveilleuses operations.
CACHER. signifie aussi, Voiler, déguiser, ne paroître pas à la vue. Cette fille est si modeste, qu'elle se *cache* le visage de ses coiffes, de son masque. Cet homme m'a *caché*, m'a déguisé son nom. Il se *cache* de moi; pour dire, Il fait cela à mon insçu. Dans cette éclipse la Lune *cache* la moitié du Soleil. Les arbres en se couchant se *cachent* dans l'onde. Voilà un bois qui nous *cache* la vue de ce chateau. On demandoit à quelcun qui portoit quelque chose *caché* sous son manteau, ce qu'il portoit; je le *cache*, dit-il, afin qu'on ne le sache point. M. M.
Heureux qui fait fait de son humble fortune,
Vit dans l'état obscur où les Dieux l'ont caché. R. A. C.
 On dit *cache* son jeu, & cette expression a trois divers sens. Elle signifie l'empêcher que quelcun ne voye son jeu. II. Diffimuler son dessein, en faisant semblant de ne sçavoir pas bien joier. III. *Cache* ses desseins, ensoit que personne ne les puisse découvrir. Il est tout-à-fait figuré en ce dernier sens.
CACHER. se dit figurément en choses morales. C'est un hypocrite qui sçait bien *cache* sa turpitude. Ce sont de beaux amis qui ne *cachent* rien l'un à l'autre. Il est avantageux souvent de se *cache*; de *cache*, de diffimuler la colère, son amour. Les Payens *cachent* beaucoup de secrets de la nature sous le voile de leurs fables. Dieu a *caché* ses mystères aux sages du siècle, & les a révélés aux simples. *Baron cache* une grande prudence sous une apparence de folie. Il ne vous fera point permis de *cache* plus long temps vos vices par vos dissimulations. O. M. La basselle de cet homme paroît d'autant plus qu'on la veut *cache*. On peut *cache* ses sentimens sous des fables ingénieuses. Dieu *cache* l'avenir sous d'épaisse ténèbres, & se rit de nos craintes injustes & déraisonnables. P. O. R. T. R. Il a de l'adresse à bien *cache* sa passion. A. B. I. *Cache* sa haine sous des fausses carences. R. A. C. On s'étudie plus dans le monde à *cache* ses passions, qu'à acquiescer la vertu. W. c. Rien n'est plus aimable que la simplicité d'une jeune Bergère, qui ne peut ni se montrer, ni se *cache* sans plaisir. F. O. N. T. Le soin de se *cache* vaut encore mieux que l'indolence de ceux qui ne se donnent pas la peine de déguiser leurs défauts. B. I. L. L.
 On dit absolument, Se *cache*, 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 asiles. Après avoir écrit un tel affront, il se fait *cache*, & ne plus paroître en public.
 On dit proverbialement, *Cache* ta vie: c'est un des préceptes d'Epictète, dont Plutarque a fait un beau Traité; pour dire, qu'il ne faut pas faire connoître à tous les hommes ce que l'on fait. Le péché que l'on *cache* est d'autant plus puni, & dissimulé. Une science *cache*, celle qui est abstraite, ou connue de peu de personnes,

C A C.
 comme l'Algebre, la Cabale, la Steganographie; L'écriture dit qu'il n'y a rien de si *cache* qui ne se revele, qui ne paroisse quelque jour. Dieu nous tient ses secrets *cache*, afin que nous ne cessions pas de prier. M. D. M.
Quand la vertu genit sous le pouvoir du vice,
Ainsi du Strigour les jugemens cachez.
 L' A. B. T. E. U.
CACHET. f. m. Petite sceau qui porte une gravure particulière de quelques Armes ou chiffres qu'on imprimé 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 *cache*, à cause qu'il sert à *cache* l'écriture. M. M.
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 *cache* imprimée sur un papier, avec lequel on pourra fermer quand on voudra une lettre qu'on donne ouverte.
LETTRÉ 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.
CACHETÉ R. v. act. Appliquer un *cachet* sur quelque chose qu'on veut envoyer fermé. *Cacheter* un paquet, une boîte, une bouteille.
CACHETÉ, é. s. 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 *peché* & 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*. P. A. T. R. U. On ne doit pas user de duel si on peut tuer son homme en *cachette*. P. A. S. C.
CACHOS. f. m. Plante qu'on ne trouve que dans les montagnes du Pérou. Elle croît comme un arbrisseau, & 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 agreable & sans acrimonie, contenant une femence fort menue. Les Indiens font beaucoup de cas de cette plante à cause de ses rares qualitez: car elle fait uriner & chasse le sable & la pierre hors des reins, & ce qui est plus admirable, c'est qu'on tient que par son usage elle buse la pierre dans la vessie, si elle est encore tendre, & qu'elle se puisse rompre par quelque quodiamment. En Latin *cachos*, ou *glanum pampiferaum joto venando tenui*.
CACHOT. f. m. Prison noire & obscure, qui est au dessous du rez de chauffée, & où on ne gite que sur la paille. On met dans les *cachots* les criminels condamnés, ou accusés de grands crimes, ou qui sont des rebelles dans la prison. Vous décririez-je ces *cachots*, ou plutôt ces sepulchres funestes, où l'on entere des hommes vivans, pour qui il semble que le soleil ait cessé de brüler, & que la nuit ait pris la place du jour? F. T.
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.
CACHOU. f. m. Petit grain qui se fait d'une composition de musc & d'ambre, qui sert à parfumer l'haléine. Sa base est une gomme qui se tire d'une décoction épaisse d'un certain arbre qui croît aux Indes. Ce *cachou* que les Auteurs appellent *kanis*, & qu'on dit aussi en nomme

ETYMOLOGISCHES WÖRTERBUCH DER PREUSSISCHEN SPRACHE.

[278]

A.

abbai 'beide' **abbans** acc.: lit. **abu** le. **abi** abg. **oba**: gr. ἄββαω 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. **abarā** '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. **ēglė** aus ***edlė**, poln. **jodła**.
aglo V. 470 'Regen' für **aglu**: gr. ἀχλύς 'Nebel, Wolke': **akh-** neben **ak-**? in **āklas** 'blind', lat. **aquilō** 'Nordwind'.
ayculo V. 470 'Nadel': gr. αἴχλοι · αἰ γωνίαι τοῦ βέλους Hesych.
ag. **igla** wäre nur zu vgl. wenn es für **aygulo** verschrieben.
ains 'ein', acc. **ainan**, gen. **ainassei**, dat. **ainesmu** fem. **ainā**: lit. **vėnas** abg. **inb** lat. **oinos** gr. οἶνή 'Eins auf dem Würfel' got. **ains**.
ainonts 'Jemand': acc. **ainontin**, gen. **ainontsi**, dat. **ainontsmu**: lit. **vėnuntas** 'einzig', **vėnunta** 'jemals'. **niainonts** acc. **niainonton** 'niemand'.
erains dat. **erainesmu** 'Jeder'. **er** 'bis': le. **ikwĩns** 'jeder', **ik** 'bis'.
niains, **niainan**; gen. **niainessa**; **niainā** f. 'kein'.
ainat 'allezeit': lit. **vėnāt** 'einzig, nur'. **terains** 'allein'.
aysmis V. 363 'Spieß': lit. **ėszmas** le. **ėšms** '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** 'Schat': ai. **etaša** 'bunt, Antilope', ved. **etā** 'schnelles Tier'.
aketes V. 255 'Eggen': lit. **akėtės**, **akėczios**, **ođiva** Hes; lat. **occa** cambr. **ocet** ahd. **egida**.

[279]

D U R.

patteque 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 la fortune. Il faut faire les provisions *durant* l'été. N'ai-je pas sçu 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'ay 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, IE, 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 abrégé 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'ennuye 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'épuient 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 acquérir 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 grosse & dure, qui est attachée à l'os du crâne.

DURER, v. n. Sublister pendant quelque espace de temps. Une femme se defait 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 precautions. **MO L.** 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, Resister, souffrir quelque mal, quelque peine, quelque incommodité. On ne *sçauroit durer* avec cette femme-là, tant elle est criarde. Je ne puis plus *durer* avec cette colique. On ne *sçauroit durer* à la maison par ce beau temps-là. On ne *sçauroit durer* en ce poste, il est trop exposé à l'artillerie. On n'y *dure* point, on n'y peut tenir. **MO L.** Pensez-vous que je puisse *durer* avec toutes ses turlupinades? **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 ennuye, qu'il attend quelque chose avec grande impatience. On dit aussi, qu'un homme ne *sçauroit 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.

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

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

DURÉTÉ, se dit figurément en choses spirituelles & morales, & signifie, Indocilité, insensibilité, cruauté. La *durété* du cœur des Juifs obligea Moïse à leur permettre le divorce. **LA MAI.** Il a une *durété* d'esprit qui fait qu'il ne peut rien comprendre; une *durété* de cœur, qui fait qu'il n'aime personne. Nous joindrons nos forces pour attaquer la *durété* de son humeur. **MO L.** Le cœur, & le temperament des Stoïciens ne s'accordoient pas toujours de la *durété* philosophique dont ils faisoient profession. **OR. M.** Les opinions de Senèque ont trop de *durété*. **ST. EV.** Un peu de *durété* sied bien aux grandes âmes. **CORN.** Pensez-vous que je vous pardonne toutes les *durétés* que vous m'avez dites? Vous avez eu la *durété* de me dire que la conversation de cette Dame vous avoit plu. **LE T. PORTUG.** La *durété* des termes choque d'autant plus, qu'elle enferme quelque sorte d'indifference, & de mepris. **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 la fortune. Il faut faire les provisions *durant* l'été. N'ai-je pas sçu 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'ay 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, IE, 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 abrégé 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'ennuye 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'épuient 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 acquérir 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 grosse & dure, qui est attachée à l'os du crâne.

DURER, v. n. Sublister pendant quelque espace de temps. Une femme se defait 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 precautions. **MO L.** 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, Resister, souffrir quelque mal, quelque peine, quelque incommodité. On ne *sçauroit durer* avec cette femme-là, tant elle est criarde. Je ne puis plus *durer* avec cette colique. On ne *sçauroit durer* à la maison par ce beau temps-là. On ne *sçauroit durer* en ce poste, il est trop exposé à l'artillerie. On n'y *dure* point, on n'y peut tenir. **MO L.** Pensez-vous que je puisse *durer* avec toutes ses turlupinades? **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 ennuye, qu'il attend quelque chose avec grande impatience. On dit aussi, qu'un homme ne *sçauroit 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.

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

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

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

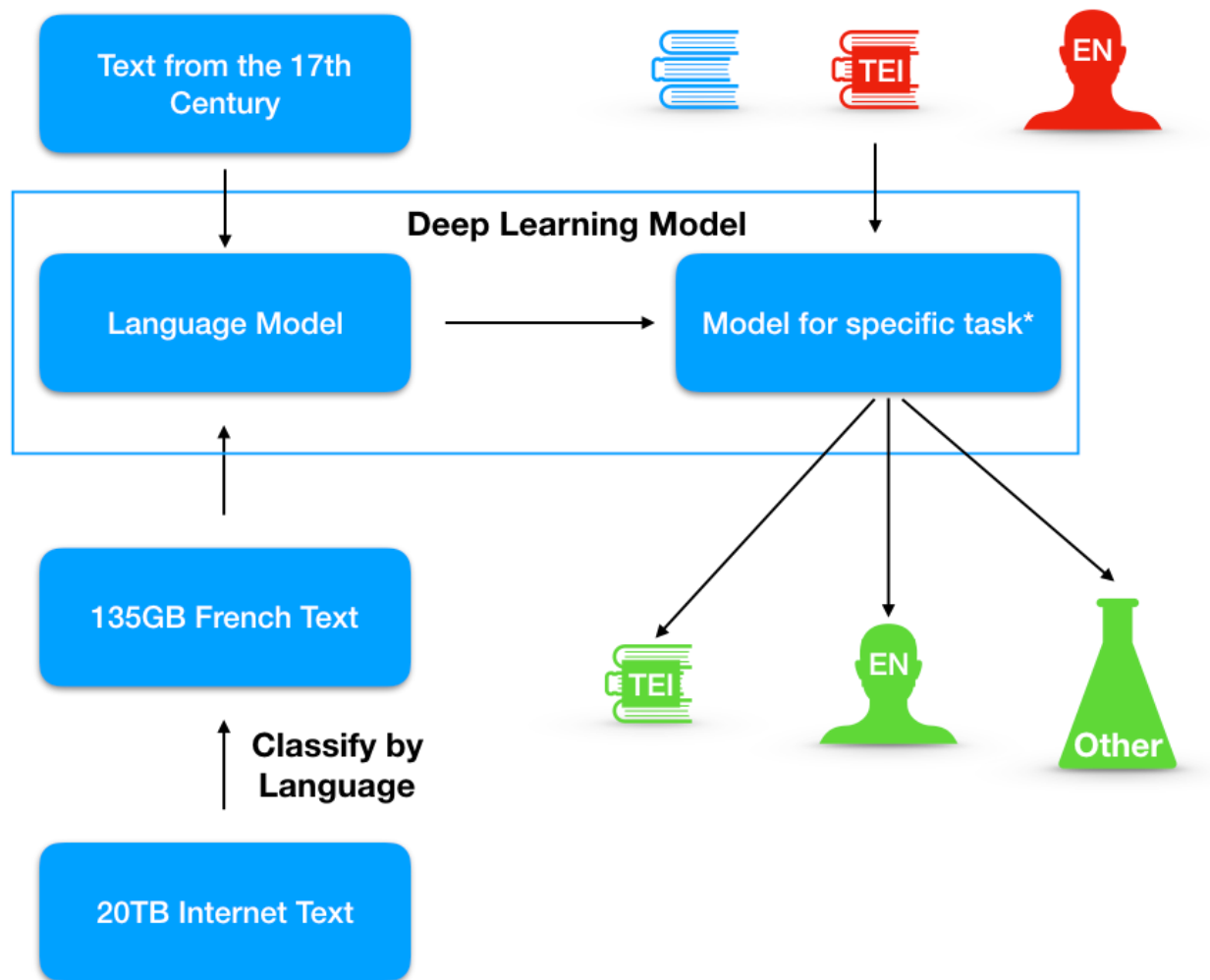
*Je renonce à la vanité
 De cette dureté farouche,
 Que l'on appelle fermeté.* **QUIN.**

Experiments: Lexical Entry Model

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

Table 3: Field Level Evaluation of the Lexical Entry Model

GROBID: the next generation



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](#)

*Changes by task

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!