

Ceci n'est pas un dictionnaire

Adding and Extending Lexicographical Data of Medieval Romance Languages to and through a Multilingual Lexico-Ontological Project

REM. n°4: *Fiel de terre* traduit lt. *fel terrae* qui désigne la petite centaurée (*Centaureum erythraea*), cf. ThesLL 6¹,425,40.RIF1 8,49 enregistre norm. *fiel de terre* et lang. *fél dé terro*. Cp. aussi it. *fielle di terra* (dès 14^es., TLIO [FIELE](#)) et le calque all. *Erdgalle* (J.C.Adelung, Grammatisch-kritisches Wörterbuch der hochdeutschen Mundart, 1811, 1,1877). En mfr. et frm., *fiel de terre* désigne aussi la fumeterre, une plante annuelle, dicotylédone de la famille des Fumariacées (*Fumaria officinalis*), cf. FEW 3,445b. V. ci-dessous.]

☐ ☒

ALMA
(*fiel* PsOxfM 68,26; DébCorpsSamP V 137; ... M 68,24 (p. 1... recette... 65,7; EneasS² 8221; PhilomB 966; RouH II 1104; BenDucF 15272; RecMédJuteH 253; ... FougL 1019; etc. ... *fel* ... 10... sionA 279; BestGuillR 2901; GrossetChastM 1132; GuillJoiesRi 98; SAul...; LettrHippoT 104; Plattract... f°107v... 139 (f°102v; 2 att.); MirAGn2K 4,198; 25,6; JGarl HuntTeach 6; I...mplMedl... etc.etc., *fiel* RecMéd... H 458... Fevres f°9...v°10; f°165v°2; f°165v°3; etc., *feil* PhThSibS 317; AncrRiwlerT... cette H... 271... (1...), [*feil*... 9,153...], *fius* GautArrErR 3688 [c.r. sg.]

WISSENSNETZE IN DER MITTELALTERLICHEN ROMANIA

☐ ☒ ◆ 1^o t. de méd. “liquide verdâtre et amer qui est contenu dans la vésicule biliaire, bile” (dep. ca. 1160, EneasS² 8221 [*el cors*

Agenda

- Introduction: The Role of Dictionaries – The Situation of Dictionaries
- ALMA Project: Dictionary ...
 - ... Re-use
 - ... Extension
 - ... Further Processing
 - ... Dissemination
- Evaluation
- Conclusion

Introduction

- The Role of Dictionaries
 - At core of historical lexicology, text philology, historiography
 - Provide means for grounding research on stable knowledge
 - Both **are** and **facilitate** foundational research
 - Our context: dictionaries of Old / Middle French, Old Occitan, Old Gascon, medieval Italian, etc.
- The Situation of Dictionaries
 - Fundings are coming to an end => difficult juncture

Knowledge Networks in Medieval Romance Speaking Europe



- Inter-institutional project: Heidelberg / Bavarian / Mainz Academies of Sciences and Humanities;
 - six positions; 22 years
 - Maria Selig, Wolfgang Schweickard / Elton Prifti, Tittel
- Incorporates four dictionaries into its work:
 - *Dictionnaire onomasiologique de l'ancien gascon* DAG (HAdW, **funding ended**)
 - *Dictionnaire étymologique de l'ancien français* DEAF (HAdW, **funding ended**)
 - *Dictionnaire de l'occitan medieval* – DOM (BAdW, **funding ended**)
 - *Lessico etimologico italiano* – LEI (Mainz Academy, **ongoing**)




ALMA in a Nutshell

- **Aim:** investigate interaction between language, knowledge and scholarship in the Middle Ages
- **Field of observation:** Romance cultural sphere => emergence of new knowledge networks expressed in vernacular languages
- **Time period:** ca 1100 – ca 1500 AD
- **Languages:** Medieval French, Italian, Occitan, Gascon (Spanish, Catalan)
- **Question:** How are these languages developed into languages of knowledge / scholarship within new functional areas technically / conceptually complex?
- **Domains:** medicine and law / jurisdiction



© Sabine Tittel

Cultural Heritage

- Important part of the intellectual and cultural heritage of Europe
- Romance languages major carriers of a cultural exchange
 - in the Middle Ages
 - establish the European identity as a knowledge society

ALMA – Dictionaries

Manifold relations

1) Dictionary Re-use

- Implementation of quantitative machine-driven with competence-linguistic methods
- Controlled by dictionary re-use
 - FEW (*Französisches Etymologisches Wörterbuch*), DMF (*Dictionnaire du moyen français*)
 - Iberoromania and (medieval) Latin
 - DAG, DEAF, DOM, LEI (on the level of the databases / raw data)
- Cognitive step from
 - given lexeme => one or several concepts
 - analyzing the meaning(s) of the lexeme in constant confrontation with the language system documented only in the dictionaries

2) Dictionary Extension

- ALMA corpora will enlarge the material basis for lexicography in a significant way
- By digitizing existing text editions
 - *Chirurgie*, Henri de Mondeville [Old French 1314]: exploited by dictionaries but only partly and manually.
- By new text editions
 - *Chirurgia magna*, Gui de Chauliac (Lat. 1363): so far only first treatise of French translation exploited by lexicography

Lexical-Semantic Studies

- Lexical-semantic studies with features similar to dictionary articles
 - Lemmatized header (in each language); etymology; semantic tree structure; genus-differentia-definitions; apparatus of graphical realizations; corpus material
- Extend, advance, or replace entries of DAG, DEAF, DOM, LEI
- Put dictionary data into a multilingual, Pan-Romance context
 - Shedding new light on terminology, history of lexemes, etymology, history of concepts
 - Enhance comprehension of inter-relatedness of medieval vernacular languages hitherto scattered among individual dictionaries

fr. **MIRAC** m. – it. **MIRACH** m. – occ. **MIRAC** m. [– esp. **MIRAQUE** m.]

De l'ar. MARĀQQ pl. (مرق) “parties molles de l'abdomen” (Freytag 1,178; cp. Steingass 1207b). La signification de l'ar. *marāqq* témoigne, comme terme de médecine depuis Rāzī (av. 925) dans la littérature médicale arabe, d'un sens large comprenant (une région de) l'abdomen ou la paroi abdominale avec plusieurs parties membraneuses, fines, dont le péritoine: «“the thin, or delicate, and soft, or tender, parts of the belly; or the lower part thereof with what surrounds it, that is thin or delicate, or the lower part of the belly in the region of the [...] (peritoneum)” [...] and “abdominal wall, the skin and superficial fascia of the abdominal wall, the hypogastric region of the abdominal wall”», MedicalSynList 1,325s; Hyrtl 177ss. Par le biais des traductions des auteurs médiévaux, notamment ceux d'Avicenne et de Rhazès, le terme a trouvé son chemin au mlt. (le type en -r qui n'existe pas dans la tradition arabe résulte vraisemblablement d'une fausse vocalisation pendant le processus de la traduction⁽¹⁾): *mirach*, *mirath* m. “paroi abdominale antérieure” (dès fin 12^{s.}, AvicennaLiber₃ 24r; 318r; DC 5,563c; DMLBS 1801c MIRACH [att. fin 13^{s.}-ca. 1542]).⁽²⁾ V. Hyrtl 177ss pour une liste détaillée des différents emplois du mot dans les textes mlt. qui reflètent le sens large du terme arabe.

Dans les langues vernaculaires, le concept symbolisé par *mirac* (et var. congénères dans les différentes langues) est plus étroit qu'en arabe: le terme désigne une structure anatomique qu'il faut comprendre, avec l'anatomie moderne, comme la paroi abdominale antérieure. La paroi abdominale renferme l'ensemble des structures qui servent à délimiter les parties antérieures et latérales de la cavité abdominale, c'est-à-dire l'épiderme, le pannicule, le fascia superficiel de l'abdomen, les abdominaux, le fascia transversalis et le péritoine (cf. Fig. 1). Dans les textes médiévaux, le concept comprend la peau, le pannicule, la cavité abdominale, le péritoine, v. les contextes cités ci-dessous. Les deux att. du 16^{s.} (1546 Charles Etienne; av. 1584 Antoine Du Pinet, v. ci-dessous) expliquent *mirac* comme un synonyme ar. pour l'épigastre qui est la région supérieure et moyenne de l'abdomen, ce qui n'a rien de commun avec le concept moderne. — Cf. aussi Baader dans Keil / Assion, *Pachprosaforschung*, 1974, p. 104; AltieriLessico 97; Ineichen 2,292; Pellegrini 84; Els Sheikh 2,215; NasserEmprunts 271; CorrienteArabismos 391; ELAA 3,391 [*mirac* / *almirach*].

Le terme anatomique *mirac* (et var. congénères) est très intéressant tant du point de vue de contact linguistique et culturel, tant de celui de l'histoire de la médecine. Il apparaît, dans nombre de langues vernaculaires, sous des formes qui sont des variantes pour le péritoine qui remontent à l'ar. *sifāq* v. “péritoine” (Hyrtl 211; MedicalSynList 1,325s): mlt. *siphac* s. (DMLBS 3100c SIPHAC “(anat.) peritoneum”) et → (afr. SIFAC / it. SIFACH / aocc. SIFAC). Le péritoine (la membrane constituant l'intérieur de la cavité abdominale) adhère au fascia transversalis ce qui rend l'identification anatomique difficile (cf. BertoliniLeutert 2,248). Cependant, la spécification du *sifac* est nettement exprimée par

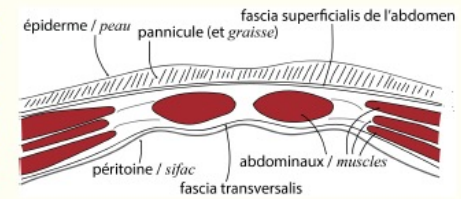


Figure 1: La paroi abdominale [illustration moderne avec termes modernes / anciens].

3) Dictionary Processing

- Semantic Web technologies: model project's results as Linked Open Data (LOD; text editions + studies)
- Lexico-semantic mapping
- Development of domain-specific, historicized ontologies for medicine and law
 - LOD modeling includes original dictionary articles with lexico-semantic mapping

LOD Modeling

- Main vocabulary:
OntoLex-Lemon
- Resource Description
Framework (RDF)
- Automatic processing
XML2RDF or
XML/TEI2RDF

```
1 @prefix dbr:      <https://dbpedia.org/resource/> .
2 @prefix dct:      <http://purl.org/dc/terms/> .
3 @prefix deaf:     <https://deaf.ub.uni-heidelberg.de/lemme/> .
4 @prefix decomp:   <http://www.w3.org/ns/lemon/decomp#> .
5 @prefix lexinfo:  <https://lexinfo.net/ontology/3.0/lexinfo#> .
6 @prefix olia:     <http://purl.org/olia/olia.owl#> .
7 @prefix ontolex:  <http://www.w3.org/ns/lemon/ontolex#> .
8 @prefix rdfs:     <http://www.w3.org/2000/01/rdf-schema#> .
9 @prefix skos:     <http://www.w3.org/2004/02/skos/core#> .
10 @prefix vartrans: <http://www.w3.org/ns/lemon/vartrans#> .
11
12 # --- lexical entry -----
13 deaf:fiel a ontolex:LexicalEntry , ontolex:Word ;
14   lexinfo:partOfSpeech "m."@fr ,
15   lexinfo:Noun ;
16   lexinfo:gender lexinfo:masculine ;
17   ontolex:canonicalForm deaf:fiel_form_fiel .
18 deaf:fiel_form_fiel a ontolex:Form ;
19   ontolex:writtenRep "fiel"@fro .
20
21 # graphical variant
22 deaf:fiel ontolex:otherForm deaf:fiel_form_fel .
23 deaf:fiel_form_fel a ontolex:Form ;
24   ontolex:writtenRep "fel"@fro .
25
26 # collocation "fiel de terre", sense sense1.h
27 deaf:fiel_de_terre a ontolex:LexicalEntry , ontolex:MultiwordExpression ;
28   decomp:subterm deaf:fiel ;
29   vartrans:lexicalRel lexinfo:collocation ;
30   rdfs:label "fiel de terre"@fro .
31
32 deaf:fiel_de_terre ontolex:sense deaf:fiel_sense1.h ;
33   ontolex:evokes deaf:fiel_sense1.h_lexConcept .
34
35 deaf:fiel_sense1.h a ontolex:LexicalSense ;
36   ontolex:isLexicalizedSenseOf deaf:fiel_sense1.h_lexConcept ;
37   ontolex:usage dbr:Metonymy ;
38   olia:hasRegister olia:TechnicalRegister ;
39   dct:subject dbr:Botany .
40
41 deaf:fiel_sense1.h_lexConcept a ontolex:LexicalConcept ;
42   skos:definition "plante herbacée [...], petite centaurée"@fr ;
43   ontolex:isConceptOf dbr:Centaurium_erythraea ;
44   ontolex:lexicalizedSense deaf:fiel_sense1.h .
```


More Than Old Wine in New Bottles

An Evaluation of Contributions: Lexicographical Data  ALMA

Contribution of Lex. Data ALMA

1. Reflections on corpus integration and why dictionaries help
 2. Lemmatization
 3. Anchoring the lexical-semantic studies
 4. Shedding Light on Cross-Domain Relations
-
- A. Enlargement of the Lexeme Inventory and Enhancement of Entries
 - B. Modeling as Linked Open Data

1. Reflections on corpus integration and why dictionaries help

- Limiting lexicological research to the material of a corpus = a problematic approach
 - Corpus borders generate absences
 - Results: relevant to subset of the language in the corpus
 - E.g., corpus with works of Chrétien de Troye

1. Reflections on corpus integration and why dictionaries help

- constitutive feature for corpus = discourse tradition (technical texts)
- **Problem**
 - terminology also in many other texts
E.g.: Old French nomenclature of ships and navigation in the *Vie seint Edmond le rei* (Anglo-Norman hagiographic poetry)
 - Internal differentiation of the corpus must be supported by comparison with other literary / technical texts
E.g.: words connected to sales in a *deed* versus in a *chanson de geste* => does not imply that they necessarily have to be diaphasically bound
- **Solution**: recontextualization within language as a whole by matching against broader lexicographical works

2. Lemmatization

- Tokenization / part-of-speech annotation of text (in a corpus), lemmatization
- Models established by the state-of-the-art dictionaries
 - E.g., for Old French, the lemma list of the DEAF

3. Anchoring the Lexical-Semantic Studies

- Lexicography examines / describes language as a system within all functional areas
- Enables anchoring the findings within framework of entire language
- Crucial for proper grasp of
 - Technical terms
 - Words of the textual context, whether terminological or general-language words

4. Shedding Light on Cross-Domain Relations

- Lexical-semantic studies identify connections between different knowledge domains
- Important for medicine, metabolic-pathological field: astronomy / astrology
 - Concepts of “healthy” and “sick”

=> lexicographie!

A. Enlargement of the Lexeme Inventory and Enhancement of Entries

Middle French *addicion* f. “protuberance of an osseous or cartilaginous structure” (GuiChaulMT 316; 318; 390; etc.)

- DEAFél (DEAFprés)
 - Valuable yet unverified material
 - Old French *addicion* f. “action d'ajouter qch., ce qu'on ajoute à qch; accroissement”
 - “our” sense not registered
- DAGél
 - no gascon cognate of the term *addicion* (no attestation at all!)

- DOM 1, 175b
 - Old Occitan *adicion* f. with several sense definitions (mainly in medical texts)
 - “our” sense not registered
- LEI 1,627
 - Italian *addizione* f. “addition”; “supplement”

A. Enlargement of the Lexeme Inventory and Enhancement of Entries

Middle French *addicion* f. “protuberance of an osseous or cartilaginous structure” (GuiChaulMT 316; 318; 390; etc.)

• DEAFól (DEAFpró)

The existing dictionary entries need reviewing or lack altogether

- no gascon cognate of the term *addicion* (no attestation at all!)

- DOM 1, 175b
 - Old Occitan *adicion* f. with several sense definitions (mainly in medical texts)
 - “our” sense not registered
- LEI 1,627
 - Italian *addizione* f. “addition”; “supplement”

A. Enlargement of the Lexeme Inventory and Enhancement of Entries

Middle French *addicion* f. “protuberance of an osseous or cartilaginous structure” (GuiChaulMT 316; 318; 390; etc.)

• DEAFól (DEAFpró)

The existing dictionary entries need reviewing or lack altogether

- no gascon cognate of the term *addicion* (no attestation at all!)

25/06/2023

• DOM 1 175b

Significant enhancement of all dictionaries in question will be achieved by ALMA's contributions, both with respect to improving existing entries and to filling gaps in their lexeme inventory.

“supplement”

B. Modeling as Linked Open Data

- Modeling of dictionary articles relevant for lexical-semantic studies
- Possible extension: more dictionary articles once workflows have been installed and tested
 - Including lexico-semantic mapping to ALMA's domain ontology
- ALMA's contribution beyond scope of own lexeme inventory
- Linguistic, historical, historic-cultural dictionary knowledge will be placed within new contexts / correlations / new knowledge circulation
 - Add Onomasiological-ontological component
 - Add benefits of Linked Open Data
 - Enhance visibility / re-usability / availability (F.A.I.R.)

Conclusion



Ceci n'est pas un dictionnaire

Ceci n'est pas un dictionnaire



- ALMA is *not a real dictionary* but it can be interpreted as a particular – *real* – representation of a dictionary
- less and more
 - Less: focuses only on a part of the language covered by the comprehensive dictionaries, i.e., on medical and juridical terminology;
 - More: conceptual entanglement benefits from the combination
 - re-using well-tried dictionary knowledge,
 - with the addition of new corpus material,
 - integrating machine-driven methods,
 - introducing a Pan-Romance perspective.

Dict. picture: © Mega Pixel, Shutterstock

Magritte, Pipe: © C. Herscovici / Artists Rights Society (ARS), New York

Ceci n'est pas un dictionnaire



- ALMA is *not a real dictionary* but it can be interpreted as a particular – *real* – representation of a dictionary
- less and more
 - Less: focuses only on a part of the language covered by the comprehensive dictionaries, i.e., on medical and juridical terminology;
 - More: conceptual entanglement benefits from the combination
 - re-using well-tried dictionary knowledge,
 - with the addition of new corpus material,
 - integrating machine-driven methods,
 - introducing a Pan-Romance perspective.

Linked-Data modeling +
Lexico-semantic
mapping

Thank you!