

Tools and Instruments for Building and Querying Diachronic Computational Lexica

Modeling, Construction and Application of
Lexical Resources for the DH community

Fahad Khan*+
Andrea Bellandi *
Monica Monachini *

* ILC – CNR, Pisa
+ Ca' Foscari University, Venice

- Focussing on building **lexical linked open data** sets using an extension of the *lemon* model (Mc Crae et al. 2010), ***lemondDia***
- ***lemondDia*** (Khan et al. 2014 and 2016) allows the addition of **temporal information** and the representation of **meaning shift**, which are crucial for lexica for ancient/classical languages.
- The dataset is on a Old English lexicon (Díaz Vera et al. 2016)
 1. To demonstrate the power of *lemonDia* to formalize and query lexica with temporal info and meaning changes making use with of powerful SW technologies
 2. To present a web tool designed to help non expert users to build, query and use lexica constructed using lemonDia.

The *lemonDia* web interface

lemonDia
Old English Case Study Interface

Icon Info

nglish period created.
nglish_1 up to 950 created.
nglish_2 from 950 to 1050 created.
nglish_3 from 1050 to 1150 created.

on SHAME_OE.xlsx successfully imported.

al entries found: 8

antic shifts found: 8 See the diagram below.

Lexicon Model

```
<?xml version="1.0"?>
<rdf:RDF xmlns="http://www.semanticweb.org/lexicon#"
xml:base="http://www.semanticweb.org/lexicon"
xmlns:lemon="http://lemon-model.net/lemon#"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:owl="http://www.w3.org/2002/07/owl#"
xmlns:xmli="http://www.w3.org/XML/1998/namespace"
xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:lemonDia="http://www.languagelibrary.eu/lemonDia/lemonDia#"
xmlns:lexinfo="http://www.lexinfo.net/ontology/2.0/lexinfo#">
<owl:Ontology rdf:about="http://www.semanticweb.org/lexicon">
<owl:imports rdf:resource="http://www.languagelibrary.eu/lemonDia/semantics"/>
<owl:imports rdf:resource="http://www.languagelibrary.eu/lemonDia/lemonDia"/>
<owl:imports rdf:resource="http://www.lexinfo.net/ontology/2.0/lexinfo"/>
<owl:imports rdf:resource="http://www.languagelibrary.eu/lemonDia/anglosaxtimeline"/>
</owl:Ontology>
```

Antic Shift

Shift	Interval
love back (shame produces motion backwards)	METONYMY (left-open interval) [0, 1100]
cover (shame is a cover)	METAPHOR (left-open interval) [0, 950]
become red (shame produces redness in the face)	METONYMY [950, 1100]
careful (shame is caused by dishonour)	METONYMY (left-open interval) [0, 1100]
rotten (shame produces rottenness)	METONYMY (left-open interval) [0, 1100]
shame (shame is caused by nakedness)	METONYMY (left-open interval) [0, 950]
downward (shame produces motion downwards)	METONYMY (left-open interval) [0, 1100]

0 50 100 150 200 250 300 350 400 450 500 550 600 650 700 750 800 850 900 950 1000 1050 1100 1150 1200 1250 1300 1350 1400 1450

Aims and Context

- Lexical resources for DH
- Open data lexica based on *lemon* with temporal information, *lemonDia*
- The aims are:
 - Demonstrate the power of *lemonDia* to query lexica with temporal information with semantic web technologies based on OWL
 - Present a web tool to help users in creating lexica based on the *lemonDia* model

Why

- LOD ensures
 - Easy to publish and to link to other datasets through RDF
 - Exploiting RDF based technologies, OWL
- LOD enables querying the lexicon through
 - SWRL
 - SQWRL (enhanced queries)
- Temporal information and meaning shift are important for
 - Classical languages (language evolution)
 - Modern languages (language development)
 - Facilitating typological studies

The dataset

- Study of emotions
- Cognitive implications of meaning changes
- The time period is OE:
 - OE1 before 950
 - OE2 950-1050
 - OE3 1050-1150
- Encoded as *ProperInterval* from the time vocabulary

Lemon & LemonDia

- *Lemon*: born as a model for enriching ontologies with linguistic information
- De facto standard for publishing lexical resources as LOD
- The ontology represents the extension of senses
- Time: *usedSince* in *lemon* considered not flexible
- *LemonDia* allows addition of temporal info
- Time added to Lexical Entry--Lexical Sense
- Lexical Sense is treated as perdurant with inherent temporal extent
- *LexicalpSense* subclass of Lexical Sense is linked to *time:ProperInterval* via property *temporalExtent*

Areodian “redden with shame”

```
:AREODIAN_VB a lemon:LexicalEntry ;  
lemon:language "ang" ;  
lemon:sense :sense_Red_AREODIAN_VB,  
:sense_Shame_AREODIAN_VB ;  
lexinfo:partOfSpeech lexinfo:Verb.
```

```
:::sense_Shame_AREODIAN_VB a  
lemond:LexicalpSense ;  
lemon:reference dbpedia:Shame ;  
lemond:temporalExtent anglo:OE23.
```

...

```
:OE23 rdf:type owl:NamedIndividual ,  
<http://www.w3.org/2006/time#ProperInterval>  
> ;  
<http://www.w3.org/2006/time#intervalStarts  
> :OE2 ;  
<http://www.w3.org/2006/time#intervalFinishes> :OE3.
```


Temporal Info

- Time has often no specific start and end date
- Allen '83 defines time periods in relation to each other
- Batsakis '09 encodes Allen's relations in SWRL rules
- These rules combine the Allen's intervals to define new intervals: $O23 = \text{intervalStarts } OE2 \text{ and } \text{intervalFinishes } OE3$
- SWRL exploits logical axioms using a OWL reasoner and calculates knowledge that is implicit in the dataset

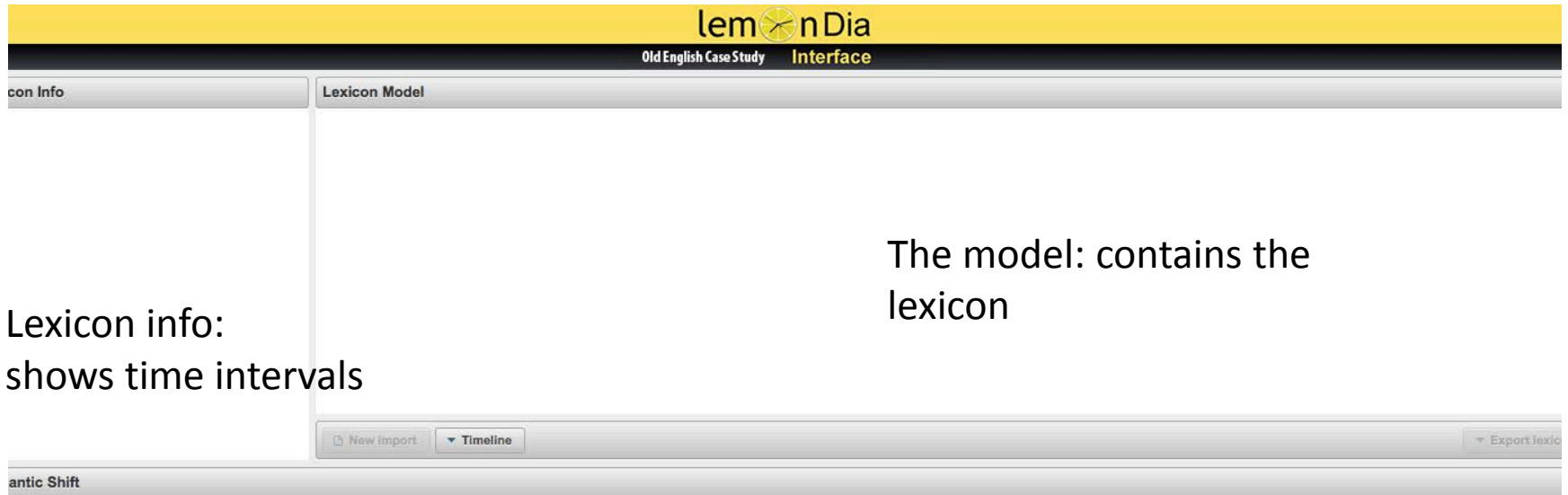
Tools for querying lexica: The literature

- LEXINFO Cimiano 2011: lmf and lemon
- LexGrid Johnson 2005: terminologies and ontologies integration
- LEXUS Ringersma 2007: LMF + ISO DatCat lexicon creation
- COLDIC Bel 2008: computational lexica
- LIR Ponsoda 2008: plug-in of NEON for localising ontologies
- Buitelaar 2009: another NEON plug-in
- OTR Tohuami 2011: manual enrichment of Ontological Terminological Resources
- Kenter 2012: corpus based lexica
- Platform for creating LLOD Ponsoda 2012: collaborative editing of resources by many users concurrently.

The *lemonDia* interface

- A user-friendly interface facilitates users in the building of diachronic lexica using the lemonDia model without involving them too deeply in the details of the model
- The interface is composed of 3 panels

The interface: 3 panels



Lexicon info:
shows time intervals

The model: contains the
lexicon

The semantic shift: contains
temporal evolution of the
sense in the lexicon

Icon Info

Lexicon Model

New Import

Timeline

Export lexicon

Command Shift

+ Create

Modify

The interface: time intervals

The screenshot displays the 'lem nDia' interface for an 'Old English Case Study'. A 'Timeline Creation' dialog box is open, allowing the user to define time intervals for a period named 'OldEnglish'. The dialog is titled 'Timeline Creation' and has a close button (X) in the top right corner.

Period name: OldEnglish

First interval between 0 and 950

closed left-open right-open

Second interval between 950 and 1050

closed left-open right-open

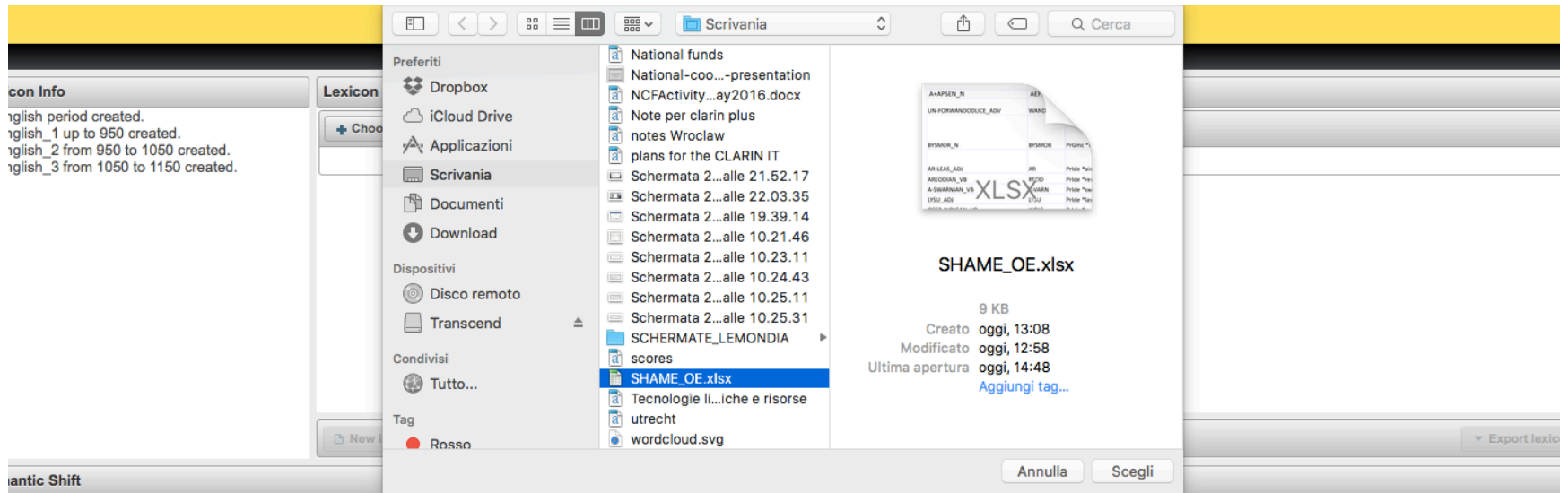
Third interval between 1050 and 1150

closed left-open right-open

Submit

The background interface shows a 'Lexicon Model' tab and a 'Timeline Creation' dialog box. The dialog box is titled 'Timeline Creation' and has a close button (X) in the top right corner. It contains a text input field for 'Period name' with the value 'OldEnglish'. Below this, there are three sections for defining time intervals. The first section is 'First interval between 0 and 950', with radio buttons for 'closed', 'left-open', and 'right-open', and a slider below it. The second section is 'Second interval between 950 and 1050', with radio buttons for 'closed', 'left-open', and 'right-open', and a slider below it. The third section is 'Third interval between 1050 and 1150', with radio buttons for 'closed', 'left-open', and 'right-open', and a slider below it. At the bottom of the dialog is a 'Submit' button.

The interface: import data



The interface: the semantic shift

Icon Info

Old English period created.
 Old English_1 up to 950 created.
 Old English_2 from 950 to 1050 created.
 Old English_3 from 1050 to 1150 created.

Old English SHAME_OE.xlsx successfully imported.

Old English entries found: 8

Old English semantic shifts found: 8 See the diagram below.

Lexicon Model

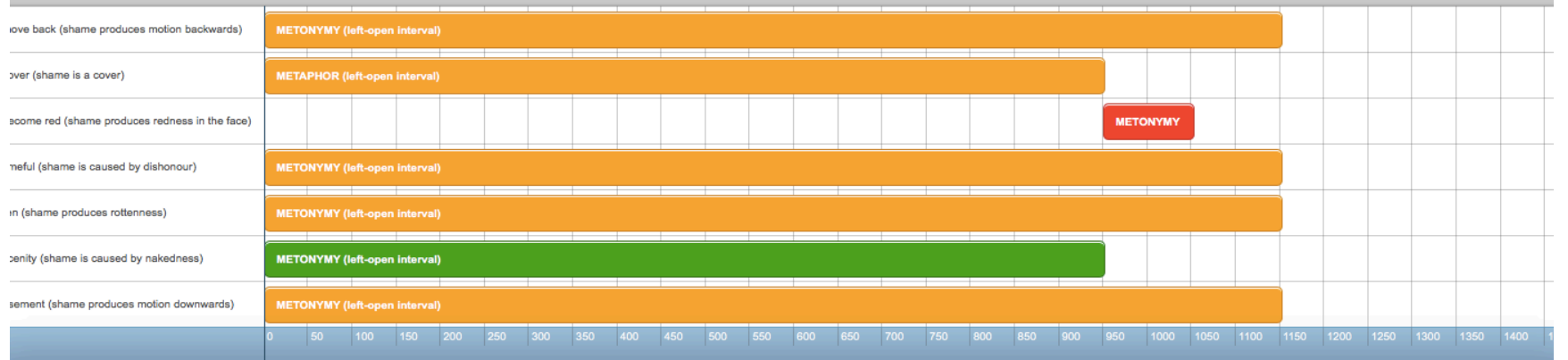
```
<?xml version="1.0"?>
<rdf:RDF xmlns="http://www.semanticweb.org/lexicon#"
xml:base="http://www.semanticweb.org/lexicon"
xmlns:lemon="http://lemon-model.net/lemon#"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:owl="http://www.w3.org/2002/07/owl#"
xmlns:xml="http://www.w3.org/XML/1998/namespace"
xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:lemonDia="http://www.languagelibrary.eu/lemonDia/lemonDia#"
xmlns:lexinfo="http://www.lexinfo.net/ontology/2.0/lexinfo#">
<owl:Ontology rdf:about="http://www.semanticweb.org/lexicon">
<owl:imports rdf:resource="http://www.languagelibrary.eu/lemonDia/semantics"/>
<owl:imports rdf:resource="http://www.languagelibrary.eu/lemonDia/lemonDia"/>
<owl:imports rdf:resource="http://www.lexinfo.net/ontology/2.0/lexinfo"/>
<owl:imports rdf:resource="http://www.languagelibrary.eu/lemonDia/anglosaxtimeline"/>
</owl:Ontology>
```

New import

Timeline

Export lexicon

Semantic Shift



Technical data

- Three tier architecture
- Apache Tomcat 7.0 webserver
- Java 2 SE implementation language
- OWLAPI model management
- Java Server Face JSF and Primefaces 5.1 for the presentation tier (cuncurrent access)